Development in Sub-Saharan Africa

New Micro-Level Evidence on Education, Geography and Trade

Dissertation
zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften
der Rechts- und Wirtschaftswissenschaftlichen Fakultät
der Universität Bayreuth

Vorgelegt von
Frederik Wild aus
BAB 8 Stuttgart-München bei km 10.7

Dekan:	Prof. Dr. André Mayer
Erstberichterstatter:	Prof. Dr. David Stadelmann
Zweitberichterstatter:	Prof. Dr. Mario Larch
Tag der mündlichen Prüfung:	18.04.2024

Für meine Familie

Acknowledgments

First and foremost, I am deeply indebted to my supervisor, David Stadelmann, for affording me the opportunity to pursue this endeavor. Without your trust and support, I could not imagine having developed the deep passion and satisfaction for academic work that I now enjoy. You gave me guidance when needed, yet also allowed me the freedom to explore the topics I am passionate about. I will remember the many insightful, often lively debates we had in the office, on our many travels, or simply between doors. Your intellectual acumen is truly inspiring.

Secondly, I would like to thank Mario Larch, who took on the role of second supervisor without hesitation and was always available for advice. I also won't forget Mario's summertime seminars, which were academically fruitful and, above all, encouraging environments. Not to mention the great barbecues! I am also very grateful to Benno Torgler for inviting me to QUT, where I spent an amazing two months during my research visit.

I am appreciative of the many colleagues I have met during my time in Bayreuth and of the faculty for providing a supportive and constructive environment. I would like to thank the team of VWL II, and especially Heidi, for always being there with an open ear and for a delightful morning talk. Not to forget Helge Wiechmann, with whom I have had the pleasure of sharing the office and who became a true friend during our dissertation chapters.

Importantly, I would like to thank Julia, who, in comparison to all of the above, could not escape the pleasure of listening to countless sermons on my projects. I don't take your move to Bayreuth for granted one bit. Thank you for your understanding, your limitless kindness, and for centering my mind. I have learned to count on your unwavering support. I also want to thank my best friend Dominik for always keeping me sane, be it with solid life advice or with some cheerful banter.

Last and certainly not least, my greatest gratitude is directed to my parents. I am deeply indebted to you for instilling in me the respect, devotion, and work ethic necessary for such an endeavor, all while not having had the chance to pursue an academic path yourselves. Your unshakeable support and your seemingly naïve trust in my doing is second to none and certainly to the surprise of my two brothers, Philipp and Maximilian, who have had the pleasure of watching the "Nesthäkchen" receive one or the other additional break throughout our lives. I thank you for always keeping a watchful eye.

Contents

Chapter 1

Introduction

Income differences across the world remain strikingly large, particularly along the North-South divide. For instance, as a region, sub-Saharan Africa's GDP per capita amounts to $3,767 compared to the European Union's $45,917 (World Bank 2022).[1] Importantly, these differences in aggregate economic performance are associated with deprivation across many vital dimensions of individual human development, such as health and education, along with broader individual capabilities and freedoms (Sen 1985). As such, the research on the causes and consequences of these persistent disparities remains significant. Particularly so in sub-Saharan Africa, home to the world's poorest countries.

This dissertation contributes to the understanding of development in sub-Saharan Africa by presenting three independent and self-contained scientific articles that address distinct developmental challenges of the continent, namely education, geography, and trade. More specifically, they provide new micro-level evidence on these broadly discussed aspects of development, which allows for a direct study of these factors in their contribution to individual economic welfare. Thus, while the chapters address distinct research questions, they are aligned in analyzing economic development from the individual- and household perspective.

Notes to the Reader. The individual chapters of this dissertation are structured as singular scientific papers and can be read independently from one another and in any sequence preferred by the reader. As such, there exists no general list of figures and tables nor a collated list of references. Each of these items is provided in the respective chapters of the dissertation. Modified versions of Chapter 2 and Chapter 3, on geography and education, have been published in peer-reviewed Journals, namely the *Review of Development Economics* and the *Journal of African Economies*, respectively. Chapter 4 is designed as a typical job market paper, which presents an empirical as well as a theoretical treatment of regional market integration in East Africa. As is common, this paper, in particular, was presented at numerous national and international conferences and workshops. In what follows, I provide a brief non-technical overview of each of the three chapters, individually.

[1] GDP figures are expressed in constant 2017 international (PPP) USD.

Chapter 2: Coastal Proximity and Individual Living Standards: Econometric Evidence from Geo-Referenced Household Surveys in Sub-Saharan Africa

This chapter re-visits the relevance of physical geography in economic development. In particular, we analyze coastal access as a distinct predictor of individual welfare, which has been suggested as influential for comparative development in as early as the "Wealth of Nations" (Smith 1776). Cross-country studies have confirmed these observations and find coastal access as a robust indicator for income differences across countries (e.g. Bloom et al. 1998; Gallup et al. 1999). However, the inability to isolate such features from other country-specific factors, such as institutions or culture, has prompted studies that analyze the link from a regional perspective (Gennaioli et al. 2013; Henderson et al. 2018).

This chapter systematically extends this latter line of research by drawing evidence from a granular large-scale, geo-referenced household-level survey covering 28 sub-Saharan African countries over 20 years. Analyzing individual-level data allows us to test whether the insights from cross-country and cross-regional contexts also apply at the individual level. Moreover, we can utilize the comprehensiveness of our dataset to explore a large set of indicators and potential channels of influence through which coastal access may matter for individual living standards.

The following main results emerge from the analysis. First, coastal proximity is confirmed as a relevant indicator of within country income disparities. Second, while coastal proximity remains robust for visually all controls entered, there are distinct factors that mediate the results. In particular, the inclusion of human capital, infrastructure, and urbanization seem to account for a large part of the identified welfare gap. And lastly, our study thereby also strengthens the notion that geography need not be "destiny", as policy may cater towards influencing these three factors.

Chapter 3: Heterogeneous Effects of Women's Schooling on Fertility, Literacy and Work: Evidence from Burundi's Free Primary Education Policy

Education has long been regarded as one of the primary instruments in fostering equitable economic development, as prominently outlined in the Millenium Development Goals (United Nations 2001). Particularly for women, schooling is widely forwarded as an effective means to tackle challenges connected to gender equality, and particularly for reducing teenage pregnancies and the negative consequences linked to them (Lloyd and Young, 2009; World Bank, 2017).

However, several empirical issues complicate studying the education-fertility nexus in a causal manner, as educational attainment is jointly influenced by factors such as family background, wealth, or selection effects by skill. As such, recent research has employed quasi-experimental methods to provide a more robust estimate that can be interpreted causally (e.g. Keats 2018).

This chapter contributes to this recent and now growing literature by exploiting Burundi's Free Primary Education Policy of 2005 as a natural experiment. To identify causal variation in education, we employ a Regression Discontinuity Design and analyze the outcomes affected by schooling through an instrumental variable (IV) approach. Contrary to the extant studies in this field, we are able to provide distinct and relevant evidence on the differential treatment effects of education. While poor women profit in terms of increases in literacy, remunerated employment opportunities as well as a reduction in teenage births, none of these effects of additional education are observed for women from the wealthier households of our sample. The evidence of such a marked heterogeneity helps to evaluate under which conditions the literature's findings may generalize.

Chapter 4: Regional Market Integration and Household Welfare: Spatial Evidence from the East African Community

The role of trade in economic development has been widely studied and, from a country-level perspective, usually points to positive effects on income and growth (e.g. Frankel and Romer 1999). However, donor agencies have emphasized the potentially inequality-enhancing impact of trade within countries (World Bank 2009), and there is a growing body of literature dedicated to examining these distributional effects (for an overview see Engel et al. 2021). One aspect that has garnered particular attention is whether trade increases regional inequalities within countries, i.e. across space (see for an overview Brülhart 2011). However, in Africa, the literature on such concerns is scant and mainly relies on the use of lights emitted by night as a proxy of economic development (e.g. Eberhard-Ruiz and Moradi 2019).

In this chapter, I provide novel evidence on the distributional effects of trade liberalization in Africa by combining the spatial considerations of market integration with a household-level analysis. I thereby treat the re-establishment of the East African Community (EAC) in 2001 as a regional policy intervention having differential effects on individual households governed by their geo-spatial location within the countries. I derive these predictions from an extension of a canonical New Economic Geography (NEG) model which accommodates the key features of the EAC's spatial layout. Contrary to the model's prediction, the reduced-form empirical results do not show evidence of increased household welfare in regions closer to the newly accessed markets. Rather, I observe strengthened agglomeration tendencies in the pre-existing economic hubs of Nairobi, Dar es Salaam and Kampala.

The empirical results are in contrast to the ones found on similar integration processes in other contexts and thereby provide relevant insights for policy makers. The findings may also contribute to the understanding of the causes and consequences of regional disparities in developing countries, particularly on the increasingly studied and peculiar urbanization tendencies.

Chapter 2

Coastal Proximity and Individual Living Standards: Econometric Evidence from Geo-Referenced Household Surveys in Sub-Saharan Africa[1]

Abstract. We investigate geo-referenced household-level data consisting of up to 128,609 individuals living in 11,261 localities across 17 coastal sub-Saharan African countries over 20 years. We analyze the relevance of coastal proximity, measured by geographic distance to harbors, as a predictor of individual economic living standards. Our setting allows us to account for country-time fixed effects as well as individual-specific controls. Results reveal that individuals living further away from the coast are significantly poorer measured along an array of welfare indicators. Our findings are robust to the inclusion of other geographic covariates of development such as climate (e.g. temperature, precipitation) or terrain conditions (e.g. ruggedness, land suitability). We also explore mechanisms through which coastal proximity may matter for individual welfare and decompose the estimated effect of coastal proximity via formal mediation analysis. Our results highlight the role of human capital, urbanization as well as infrastructural endowments in explaining within-country differences in individual economic welfare.

JEL Classification: O15, O18, R12, O55

Keywords: Geography, Coastal Proximity, Sub-Saharan Africa, Mediation Analysis

[1] A modified version has been published under the identical title in the *Review of Development Economics*, 2022, 26 (4), 1883-1901. DOI: 10.1111/rode.12901 (joint work with David Stadelmann). The version presented here systematically expands on the article to include additional evidence, robustness tests and discussions.
Acknowledgements. Funding from the DFG (project number 390713894) is gratefully acknowledged. We thank two anonymous referees and the handling editor Andy McKay. We also thank the participants at the European Public Choice Society Annual Meeting 2022 (Braga, Portugal) and at the Bayreuth Young Economics Research Seminar 2022 (Bayreuth, Germany) for fruitful comments and discussions. All remaining errors are my own.

2.1 Introduction

> "As by means of water-carriage a more extensive market is opened to every sort of industry than what land-carriage alone can afford it, so it is upon the sea-coast, and along the banks of navigable rivers, that industry of every kind naturally begins to subdivide and improve itself [...]. There are in Africa none of those great inlets, such as the Baltic and Adriatic seas in Europe, the Mediterranean [...] to carry maritime commerce into the interior parts of that great continent: [...]"
>
> Adam Smith (1776)[2]

Cross-country studies investigating the link between physical geography and economic development consistently provide evidence of a positive and statistically significant association between coastal access and national income (e.g. Bloom et al. 1998; Radelet and Sachs 1998; Gallup et al. 1999; Easterly and Levine 2003; Sala-I-Martin et al. 2004; UN-OHRLLS 2013). More recent literature analyzing subnational variation in economic activity also suggests coastal access as well as coastal proximity as relevant indicators of *within-country* differences in income and related developmental outcomes (e.g. Rappaport and Sachs 2003; Gennaioli et al. 2013; Motamed et al. 2014; Mitton 2016; Flückiger and Ludwig 2018; Henderson et al. 2018; Jetter et al. 2019).

To systematically complement the literature that focused on outcomes at the national or regional level, this paper analyzes the relevance of coastal proximity in predicting *individual* economic welfare. We employ a repeated cross-sectional dataset from the Afrobarometer spanning almost 20 years and consisting of up to 128,609 individuals living in 11,261 geo-referenced localities across 17 coastal sub-Saharan African countries. Particularly in Africa, countries and regions with coastal access have had higher levels of economic development compared to more remote areas, which has been attributed to factors such as lower costs of trade, the distribution of natural resources, as well as the amplifying forces of urbanization and agglomeration (e.g. Bloom et al. 1998a; Radelet and Sachs 1998; Gallup et al. 1999; Limão and Venables 2001; Atkin and Donaldson 2015; Storeygard 2016; Henderson et al. 2018). Spatial inequalities such as these have been shown to persist even when the initial advantages of (coastal) regions may have declined in relevance (Bleakley and Lin 2012; Jedwab et al. 2017).

Our results confirm coastal proximity as a robust indicator of *individual* economic welfare across African countries: living further away from the coast is associated with a significant and meaningful reduction in the likelihood of having cash employment (income), increases in the occurrence of cash-, food-, water- and medicinal droughts (deprivation), as well as lower overall household wealth (possessions). Our results are robust to the inclusion of relevant individual-

[2] As cited in (Smith and Campbell 2009; 19,21).

level covariates, country-time specific influences via fixed-effects, as well as an extensive set of further geographic variables related to development such as latitude, elevation, climatic factors (e.g. temperature, precipitation) and features of the terrain (e.g. ruggedness, land suitability).

We also explore potential mechanisms on how coastal proximity may matter for individual living standards and investigate several candidate factors shown to contribute to spatial disparities in the literature (for an overview see Breinlich et al. 2014). In particular, we analyze the relevance of human capital (e.g. Skoufias and Katayama 2011; Gennaioli et al. 2013; Flückiger and Ludwig 2018), urbanization-agglomeration (e.g. Young 2013; Motamed et al. 2014; Chauvin et al. 2017; Gollin et al. 2017; Henderson et al. 2018), institutions (e.g. Acemoglu et al. 2001; Nunn and Wantchekon 2011; Radeny and Bulte 2012; Michalopoulos and Papaioannou 2014; Mitton 2016), infrastructure (e.g. Calderón and Servén 2010; Dinkelman 2011; Jedwab and Moradi 2016; Bluhm et al. 2018; Donaldson 2018; Jetter et al. 2019) as well as market access and trade (e.g. Brülhart 2011; Bosker and Garretsen 2012; Hirte et al. 2020; Jedwab and Storeygard 2020). We consider these factors in turn and assess their power in mediating the relationship between coastal proximity and individual welfare through a formal mediation analysis. The results highlight human capital, urbanization as well as infrastructural endowments as the predominant channels via which the presented within-country differences in individual economic welfare may be explained.

Our findings at the individual level emphasize the relevance of coastal proximity as an indicator of economic development and lend further support to the previously discussed interrelation between first- and second-nature causes of development (see Rodrik et al. 2004; Breinlich et al. 2014; Lessmann and Seidel 2017).

The remainder of this paper is structured as follows: Section 2.2 presents the data, and Section 2.3 the estimation strategy. Our results are given in Section 2.4, where we also present the insights from the mediation analysis. Concluding remarks are offered in Section 2.5.

2.2 Data

We employ the complete set of the geo-referenced Afrobarometer survey rounds, spanning a timeframe of 20 years (from 1999 to 2018) across seven survey waves (Afrobarometer 2019).[3] Afrobarometer surveys are representative at the national level and respondents are adults of the sampled households. They carry individual- and household-level information on basic characteristics such as living conditions and household assets, and additionally, provide information on individuals' sentiments as well as opinions towards the economy, democracy, governance and society. Afrobarometer fits geo-coordinates (latitude and longitude) to respondents at the level of their respective enumeration area (EA) (BenYishay et al. 2017). The sampling procedure aims for eight individuals/households per EA. Our main (extended)

[3] Surveys were sampled in 1999-2001, 2002-2004, 2005-2006, 2008-2009, 2011-2013, 2014-2015 and 2016-2018, respectively.

sample of countries consists of 128,609 (212,037) individuals living in 11,261 (17,319) geo-referenced localities across 17 (28) coastal sub-Saharan African countries (see Figure 2.1). We chose to restrict the main sample to coastal countries, so as to separate the distance effect from a more general "landlockedness" effect which potentially confounds distance with other influences such as administrative dependencies on transit countries (see Faye et al. 2004; UN-OHRLLS 2013). We investigate the extended sample, including individuals living in landlocked countries in our robustness tests.

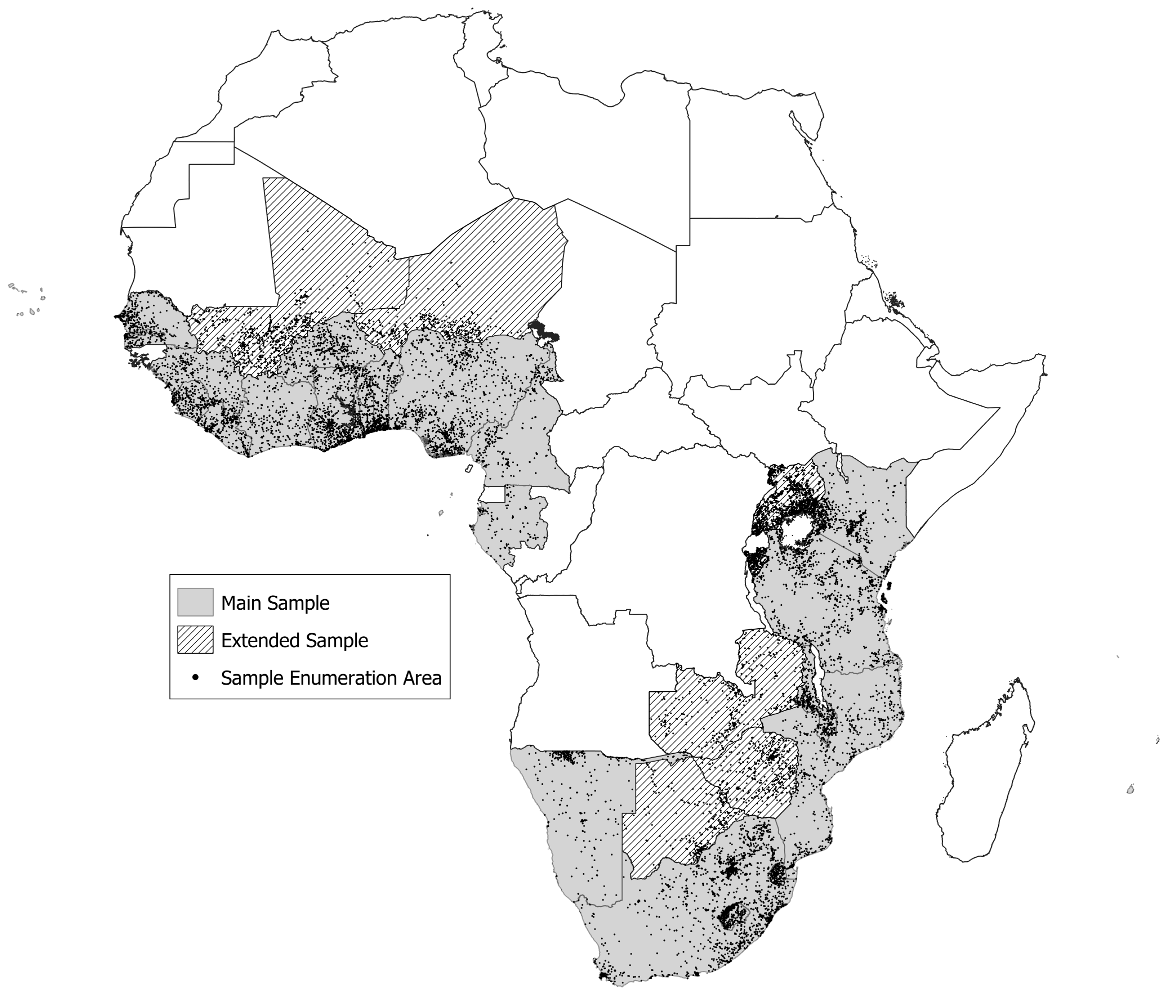


Figure 2.1: Sample Coverage

2.2.1. Dependent Variables and Channels of Influence

We employ three main dependent variables as indicators of individual economic welfare: 1) The dichotomous variable *Cash Employment* {0,1}, which indicates whether survey respondents currently have part- or full-time cash employment, i.e. serving as a measure of individuals' income. 2) The index *How often gone without enough: (Water / Food / Cash Income / Medical Care)* [0,4] which serves as a measure for individual- as well as household deprivation and is constructed by averaging individuals' responses in these four categories.[4] 3) The index *Possessions: (Radio / TV / Motor Vehicle)* [0,1] which serves as a measure for survey respondents' wealth.[5] To explore the potential channels through which coastal proximity may matter for individual living standards, we make use of Afrobarometer's opinion polling and first investigate individuals' sentiments regarding the most important issue of their respective country, *Most Important Issue: (Education/Institutions/Infrastructure)* {0,1}.[6] Thereafter, we directly investigate the lived realities around these concerns via *Education Level* {0,1}, two further composite indices, *Institutions Score* [1,4][7] and *Infrastructure Present in Enumeration Area: (Electricity Grid / Piped Water / Sewage / School / Paved Road / Health Clinic)* [0,1][8], as well as the dummy variable *Urban* {0,1}. We also analyze individuals' opinions towards supranational organizations aimed at increasing political as well as economic integration, *Helps your Country: (AU or ECOWAS/SADC/EAC/IGAD…)* {0,3} to directly relate coastal proximity with (regional) trade considerations.[9] To further explore a potential trade channel, we use information on individuals' occupation and test for a differential effect of distance using *Commercial Farmer* {0,1}, a dichotomous indicator for individuals working as farmers who grow their produce mainly for sale.[10]

2.2.2. Main Independent Variables

To construct our main explanatory variable of interest, *log(Distance to Harbor)*, we measure the shortest geodesic (ellipsoidal) within-country distance from each respondent's enumeration

[4] The four different questions read: "Over the past year, how often, if ever, have you or anyone in your family gone without: Enough clean water for home use" / "[…]: Enough food to eat" / "[…]: A cash income" / "[…]: Medicines or medical treatment?". Answers are: "Never", "Just once or twice", "Several times", "Many times", "Always", ranging from 0 through 4, respectively. These questions are consistently available in all Afrobarometer survey rounds. Using each question separately does not affect our main insights as shown in Table A.2 of the Appendix.

[5] The questions read "Which of these things do you personally own? Radio" / "[…]? Television" / "[…]? Motor Vehicle-Car-Motorcycle". Wealth possessions were surveyed from Round 3 and onwards. Using each question separately does not affect our main insights as shown Table A.2 of the Appendix.

[6] The question reads: "In your opinion, what are the most important problems facing this country that government should address?".

[7] To measure the quality of institutions, we construct an "Institutions Score" similar to Mitton (2016), which is based on an array of questions regarding local- authorities, processes and government. The score is constituted of 21 questions measuring individuals' trust in (local) courts, police and government, their experience with the procedures of local authorities, especially regarding bribery (corruption), the enforcement of crime, and the ease of handling administrative matters. Higher values indicate fewer negative experiences/better judgments of (local) institutions.

[8] Nunn and Wantchekon (2011) use an identical measure für the provision of public goods, excluding roads.

[9] Regional Economic Communities have the proclaimed aim to foster the movement of goods and people, and to improve living standards. The question reads: "*In your opinion, how much do each of the following do to help your country, or haven't you heard enough to say".*

[10] We thank an anonymous referee for pointing out this additional extension.

area to the respective country's major harbor(s).[11] Similar to Rappaport and Sachs (2003), we define all large and medium-sized ports listed in the World Port Index (WPI) as "major harbors" (NGA 2019). We also employ alternative conceptions of coastal proximity for robustness checks, namely shortest within-country distance to the coastline, *log(Distance to Coastline),* distance to major harbors using beelines (as the crow flies), *log(Beeline Distance to Harbor)*, as well as distance to the coastline using beelines, *log(Beeline Distance to Coast).*[12] Shapefile data for country administrative areas, the boundaries of which we use to calculate within country distances – and also from which we construct the coastline – come from the Center for Spatial Sciences at the University of California (GADM 2020).

Further Covariates. To isolate coastal proximity from other, potentially correlated, geographic influences of development, we closely follow Henderson et al. (2018) and add an extensive set of geographic covariates. We include *Elevation* (Farr et al. 2007), (*Abs.) Latitude*, *Ruggedness* (Nunn and Puga 2012) and *Malaria Ecology* (Sachs et al. 2004) as well as agricultural characteristics such as *Land Suitability* (Ramankutty et al. 2002)*, Growing Days* (FAO and IIASA 2019)*, Monthly Temperature* and *Monthly Rainfall* (Fick and Hijmans 2017). We also include seven dummy variables indicating the dominant natural vegetation of the area according to Olson et al. (2001).[13] We account for individuals' access to rivers or lakes by adding two dummy variables indicating whether individuals live within 25 kilometers of a navigable river or major lake, i.e. *Navigable River* {0,1} and *Major Lake* {0,1}, and thereby analyze an extended set of trade-related covariates together with our main explanatory variable, *log(Distance to Harbor)* (see Henderson et al. 2018).[14] We also add the individual-level covariates *Age*, *squared Age* and a dichotomous indicator of gender, *Female* {0,1}. The importance of urbanization-agglomeration aspects, argued to be particularly relevant in African contexts (see Young 2013; Motamed, Florax, and Masters 2014; Chauvin et al. 2017; Jedwab, Kerby, and Moradi 2017; Gollin, Kirchberger, and Lagakos 2017; Flückiger and Ludwig 2018; Henderson et al. 2018), is encapsulated by three distinct indicators of urbanization: *Primate City* {0,1}, a dummy indicating whether individuals live within 25 kilometers of a capital or primate city, *Population Density* (CIESIN 2017), a continuous measure of population density (per sq. kilometer), as well as *Urban* {0,1}, a dichotomous indicator included in the Afrobarometer survey.

Descriptive statistics for all variables are presented in Table 2.1 panels a) and b).

---

[11] We measure distances using the projection of coordinates along the earth's ellipsoid (using WGS 84, EPSG 7030). We add +1 (kilometer) to our distance measure prior to taking the logarithm.
[12] Beeline distances disregard country boundaries, i.e. cross country borders for shorter distances.
[13] Following Henderson et al. (2018) for the definition of those dummies leaves us with seven indicators relevant to our sample: *Mediterranean* {0,1}, *Desert* {0,1}, *Mangroves* {0,1}, *Tropical Forest* {0,1}, *Tropical Grassland* {0,1}, *Temperate Grassland* {0,1} and *Montane Grassland* {0,1}.
[14] The inclusion criteria for both rivers, i.e. "navigability" as well as lakes, i.e. "major", is defined as in Henderson et al. (2018): we select all natural rivers within size categories 1-5 (scale 1-7) as defined in Natural Earth (2019) and lakes with a surface area of over 5,000 sq. kilometers (Lehner and Döll 2004).

Table 2.1: Summary Statistics

Panel a)	Mean	St.Dev.	Min.	1st Quartile	Median	3rd Quartile	Max.	N
			Distribution across Sample					
Basic Characteristics								
Age	36	14	18	25	33	44	130	127,462
Female {0,1}	0.5	0.5	0.0	0.0	1.0	1.0	1.0	128,747
Educational Level {0,9}	3.4	2.2	0.0	2.0	4.0	5.0	9.0	128,211
Trade-related Covariates								
Distance to Harbor (in km)	345	289	0	82	282	564	1346	128,804
Distance to Coast (in km)	278	276	0	26	188	468	1328	128,804
Distance to Navigable River (in km)	296	229	0	127	234	422	1111	128,868
Distance to Major Lake (in km)	814	683	0	230	623	1309	2749	128,868
Urbanization Covariates								
Urban {0,1}	0.5	0.5	0.0	0.0	0.0	1.0	1.0	128,656
Primate City ≤ 25km {0,1}	0.2	0.4	0.0	0.0	0.0	0.0	1.0	128,868
Population Density (per sq. km)	2.0	4.8	0.0	0.1	0.2	1.3	125.4	128,848
Geographical Covariates								
Absolute Latitude	12	9	0	6	8	16	35	128,868
Elevation (in m)	543	595	0	48	276	1094	3914	128,860
Terrain Ruggedness (standardized)	0.0	1.0	-0.7	-0.6	-0.3	0.1	17.1	128,860
Land Suitability [0,1]	0.5	0.2	0.0	0.3	0.4	0.6	1.0	128,802
Average Monthly Temperature (in Celsius)	26	4	8	24	28	29	33	128,860
Average Monthly Rainfall (in mm)	108	63	1	68	94	127	384	128,860
Growing Days {0,365}	231	82	0	178	244	296	365	128,868
Malaria Ecology Index	12	10	0	0	13	23	33	128,802
Mediterranean {0,1}	0.0	0.1	0.0	0.0	0.0	0.0	1.0	128,868
Desert {0,1}	0.0	0.2	0.0	0.0	0.0	0.0	1.0	128,868
Mangroves {0,1}	0.0	0.2	0.0	0.0	0.0	0.0	1.0	128,868
Tropical Forest {0,1}	0.3	0.5	0.0	0.0	0.0	1.0	1.0	128,868
Tropical Grassland {0,1}	0.5	0.5	0.0	0.0	1.0	1.0	1.0	128,868
Temperate Grassland {0,1}	0.0	0.1	0.0	0.0	0.0	0.0	1.0	128,868
Montane Grassland {0,1}	0.1	0.3	0.0	0.0	0.0	0.0	1.0	128,868

Notes: The table depicts summary statistics corresponding to the main sample used in the estimations across the paper. Data encompasses individual-level responses from 17 coastal countries in sub-Saharan Africa and come from the 1999-2001, 2002-2004, 2005-2006, 2008-2009, 2011-2013, 2014-2015 and 2016-2018 (i.e. Round 1 through Round 7) geo-referenced Afrobarometer survey rounds. Variation in the number of observations size stem from differences in response rates of variables as well as changes in questions asked across surveys. Geographic covariates come from an array of sources described in the section on data (2.2).

	Distribution across Distance (within Quartiles)							
Panel b)	*Mean*	*St.Dev.*	*Min.*	*1st Q.*	*2nd Q.*	*3rd Q.*	*4th Q.*	*N*
Dependent Variables								
Cash Employment {0,1}	0.39	0.49	0	0.40	0.38	0.40	0.38	123,857
How often: Gone without (...) [0,4]	1.28	0.99	0	1.20	1.33	1.28	1.30	128,673
How often: Gone without Food {0,4}	0.97	1.16	0	0.94	1.03	0.93	0.99	128,420
How often: Gone without Water {0,4}	1.14	1.36	0	1.04	1.21	1.16	1.15	128,491
How often: Gone without Cash Income {0,4}	1.90	1.36	0	1.80	1.95	1.90	1.94	122,427
How often: Gone without Medical Care {0,4}	1.15	1.26	0	1.05	1.22	1.16	1.18	128,169
Possessions: (...) [0,1]	0.51	0.34	0	0.58	0.49	0.50	0.45	103,953†
Possessions: TV {0,1}	0.48	0.50	0	0.65	0.46	0.46	0.33	103,646†
Possessions: Radio {0,1}	0.75	0.43	0	0.77	0.74	0.74	0.73	103,902†
Possessions: Motor Vehicle {0,1}	0.30	0.46	0	0.32	0.28	0.31	0.29	103,452†
Pathways								
Most Important Issue: Education {0,1}	0.06	0.24	0	0.06	0.07	0.06	0.06	122,062
Most Important Issue: Institutions {0,1}	0.09	0.29	0	0.10	0.07	0.10	0.10	122,062
Most Important Issue: Infrastructure {0,1}	0.06	0.24	0	0.05	0.08	0.06	0.05	122,062
Present in EA: (...) [0,1]	0.56	0.30	0	0.67	0.54	0.54	0.47	117,554‡
Present in EA: Electricity Grid {0,1}	0.64	0.48	0	0.82	0.61	0.61	0.53	116,989‡
Present in EA: Piped Water {0,1}	0.55	0.50	0	0.74	0.50	0.55	0.40	116,304‡
Present in EA: Sewage {0,1}	0.29	0.45	0	0.41	0.26	0.30	0.18	115,159‡
Present in EA: Paved Road {0,1}	0.43	0.49	0	0.54	0.41	0.42	0.33	116,975‡
Present in EA: School {0,1}	0.86	0.35	0	0.88	0.88	0.83	0.84	116,668‡
Present in EA: Health Clinic {0,1}	0.57	0.49	0	0.63	0.58	0.54	0.53	115,179‡
Institutions Score [1,4]	2.81	0.53	1	2.72	2.75	2.85	2.91	128,347
Helps your Country: REC {0,4}	1.80	0.97	0	1.72	1.79	1.81	1.86	35,710'
Helps your Country: AU {0,4}	1.68	0.99	0	1.64	1.64	1.71	1.71	47,726*
Occupation: Commercial Farmer {0,1}	0.01	0.10	0	0.01	0.01	0.01	0.01	37,349**

Notes: The table depicts summary statistics corresponding to the main sample used in the estimations across the paper. Data encompasses individual-level responses from 17 coastal countries in sub-Saharan Africa and come from the 1999-2001, 2002-2004, 2005-2006, 2008-2009, 2011-2013, 2014-2015 and 2016-2018 (i.e. Round 1 through Round 7) geo-referenced Afrobarometer survey rounds. Variation in the number of observations size stem from differences in response rates of variables as well as changes in questions asked across surveys. Geographic covariates come from an array of sources described in the section on data (2.2).

† Not asked in survey rounds 1 and 2. ‡ Not asked in survey round 1. ' Only asked in survey rounds 2, 4 and 6. * Only asked in survey rounds 2, 4, 5 and 6. ** Only asked in survey rounds 1, 2 and 3.

2.3 Empirical Strategy

We employ the following regression control approach to analyze the link between coastal proximity and individual economic welfare:

$$Y_{i,c,t} = \alpha + \beta \log(\textit{Distance to Harbor})_i + \gamma \boldsymbol{X'}_i + \delta_{c,t} + \varepsilon_{i,c,t} \tag{2.1}$$

$Y_{i,c,t}$ represents the respective welfare indicator of individual i in country c, surveyed at survey-sampling period t. β captures the influence of the logged (within-country) distance to major harbors such that the link between distance and the respective welfare indicator can be interpreted as a semi-elasticity. Standard errors are clustered at the level of the survey enumeration area, i.e. at the survey cluster level. Binary dependent variables are estimated with a simple Linear Probability Model (LPM) specification.[15] $\boldsymbol{X}$ represents a vector of control variables which allows us to account for all influences potentially conflating the relationship between coastal distance and individual economic welfare. In contrast to the cross-country (cross-regional) literature, our setting allows us to account for country-time fixed effects $\delta_{c,t}$ such that we can explore a within-country estimate of distance to harbor on (individual) outcomes *net of* time-specific influences as well as country-specific influences at specific points in time, such as the Kenyan Post-Election Crisis of 2007-2008. $\varepsilon_{i,c,t}$ is an idiosyncratic error term.

We explore potential mechanisms and factors affecting the link between coastal proximity and individual living standards both via a "bad control" approach as well as a formal mediation analysis, after establishing the relevance of coastal proximity for individual living standards. Numerous robustness checks for the persistence of the observed links are offered (mostly relegated to the Appendix).

2.4 Results

Table 2.2 presents the main estimation results employing our three distinct individual welfare indicators as dependent variables. We report a parsimonious specification including country-time fixed effects in the odd-numbered columns. Even numbered columns include the full set of controls and represent our stringent setting.[16]

The results systematically indicate that distance to harbors is inversely related to individual economic welfare throughout all specifications.[17] To facilitate the interpretation of

[15] Results for binary dependent variables estimated via *Probit* yields qualitatively identical and quantitatively similar marginal effects. Results can be obtained from the authors.
[16] See Table A.1 for coefficient estimates of all (geographic) control variables.
[17] The variation in the number of observation stems from missing values. Holding the sample size constant by eliminating observations for which not all dependent/independent variables are available does not change our main insights as shown in Table A.8 of the Appendix.

Table 2.2: Coastal Proximity and Individual Living Standards

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.018***	-0.009***	0.073***	0.035***	-0.037***	-0.019***
	(0.001)	(0.002)	(0.004)	(0.006)	(0.001)	(0.002)
Discrete Change of Distance from Harbor to the 3rd Quartile (564km)	*-0.115*	*-0.055*	*0.460*	*0.223*	*-0.234*	*-0.120*
Sample Mean of Dep. Var.	[0.39]		[1.28]		[0.51]	
Basic Controls						
Age		0.031***		0.011***		0.012***
		(0.001)		(0.001)		(0.000)
Age^2		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
Female {0,1}		-0.106***		0.020***		-0.083***
		(0.003)		(0.004)		(0.002)
Urbanization Controls						
Urban {0,1}		0.055***		-0.284***		0.134***
		(0.004)		(0.010)		(0.003)
Primate City {0,1}		0.029***		-0.084***		0.029***
		(0.007)		(0.016)		(0.006)
Population Density		-0.001		0.004***		-0.001*
		(0.000)		(0.001)		(0.000)
Trade-related Controls						
Navigable River {0,1}		-0.013		-0.121***		0.029***
		(0.009)		(0.021)		(0.008)
Major Lake {0,1}		0.010		0.033		-0.023***
		(0.012)		(0.024)		(0.008)
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	123,793	122,238	128,609	126,982	103,889	102,990
R-Squared	0.09	0.14	0.17	0.20	0.15	0.22

Notes: Results in each column come from separate regressions and are estimated using the main sample of coastal sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The samples used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in these rounds. We also report estimated interquartile differences in the respective dependent variables between minimum- and third quartile harbor distances within the sample. Binary dependent variables are estimated through a simple linear probability model specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

the quantitative relevance of the main explanatory variable of interest, *log(Distance to Harbor)*, we report the predicted change of the respective dependent variable when moving from the minimum distance in the sample (i.e. effectively living by a major port) to living as far as 564 kilometers away from the harbor (3^{rd} quartile of sample) and compare the predicted change of each individual welfare indicator to the respective sample mean reported in brackets. The results show that distance to harbors is statistically significantly and negatively related to cash employment (columns 1 and 2) and positively, statistically significantly, related to deprivation (columns 3 and 4). Quantitatively, increasing individuals' distance to major ports to the 3^{rd} quartile in the sample translates into a 5.5 percentage-point decrease in the probability of having part- or full-time cash-employment (column 2) and can explain 17% of the occurrence of monetary-, medicinal- as well as food- and water-related shortages compared to the mean in the sample (column 4). Coastal remoteness is significantly related to having fewer (wealth) possessions: increasing individuals' distance to the 3^{rd} quartile corresponds to a 12 percentage-point decrease in the probability of owning a radio, a tv or a motor vehicle, accordingly, a 23% reduction compared to the mean in the sample (column 6). Importantly, the results for our indices of deprivation (columns 3 and 4) and possessions (columns 5 and 6) also hold when analyzing the variables that compose our indices separately (see Table A.2).

2.4.1. Robustness Checks and Extensions

We conduct a large array of robustness checks on our main results and summarize them in Table A.3 of the Appendix. All interpretations regarding the relevance of coastal proximity for individual living standards remain robust. (a) We re-estimate our main results by altering the distance specification to a simple "beeline" ("as the crow flies") measure. (b) We use a different conceptualization of coastal proximity by regressing our outcome variables on individuals' distance to the coastline instead of port locations, using *log(Distance to Coastline).* (c) Accordingly, we test beeline distances to the coastline with *log(Beeline Distance to Coastline).* We add dummies for living within 25 kilometers to a major harbor (d) or the coast (e), *Harbor* {0,1} and *Coast* {0,1} to separate the distance effect from a pure "coastal access" effect. We also investigate differing effects at distinct distance increments in Figure A.1. Distance to harbors is consistently, and importantly, increasingly related to lower living standards, negating a potential "binary effect" of merely living at these major harbors or not. (f) We keep observations constant across rows and columns. (g) We exclude distances larger than the 80^{th} percentile (629 kilometers) from the sample. (h) We exclude localities marked with a precision code of 2 and larger in the Afrobarometer survey (scale 1-8) from our sample. (i) We include survey sampling weights. (j) We employ clustering at the country-sample level. Moreover, we check the main coefficient's stability to potentially excluded controls via Oster tests (Oster 2019) in Table A.13. All robustness checks corroborate our general findings of a negative, independent, statistically significant relationship between coastal proximity and individual living standards. The results reiterate the relevance of coastal proximity, in varying conceptualizations, in predicting individual economic welfare.

Next to the above-mentioned robustness tests, we extend our analysis and (a) expand our main sample to include individuals living in landlocked countries (see "Extended Sample" in Figure 2.1) and also (b) analyze the persistence of our estimated effects over time. For (a), we include individuals living in landlocked countries, to explore a potential "placebo" group compared to individuals living in coastal countries. This allows us to compare the effect of sheer coastal distance within countries from a landlockedness-effect, the one often explored in the literature (UN-OHRLLS 2013). The idea is that differences in individual coastal proximity *within* landlocked countries should influence individual welfare to a lower degree given that national borders need to be crossed, creating other, potentially large restrictions unrelated to sheer distance (Faye et al. 2004).[18] As expected, Table A.14 suggests that the relevance of individual distance to harbors tends to be less pronounced for individuals living in landlocked countries. (b) The relative importance of trade-related factors of geography might be expected to change along a country's developmental path (see Henderson et al. 2018). Hence, we estimate differential effects using an interaction effect constituted of *log(Distance to Harbor)* and *Young {0,1}*, which indicates respondents below the median age in the sample (33). The results in Table A.15 of the Appendix show a clear pattern. The negative effect of distance becomes less stark for younger generations, potentially hinting at a reduction in the relevance of trade-related aspects over time (see Henderson et al. 2018).[19]

2.4.2. Mechanisms Explaining the Relevance of Coastal Proximity

Table 2.3 explores potential mechanisms through which coastal proximity may influence individual economic welfare. Following the literature, we focus on the link between coastal proximity and human capital, urbanization, institutional quality, infrastructural development and the perceived relevance of trade to investigate potential (indirect) channels that explain the spatial economic disparity given by individual geographic distance to harbors (see Breinlich et al. 2014 for an overview).

Individual educational attainment has been linked to economic welfare at the cross-regional level (e.g. Skoufias and Katayama 2011; Gennaioli et al. 2013; Chauvin et al. 2017; Flückiger and Ludwig 2018). Individuals' opinions regarding education, as shown in column (1), do not mirror these findings, as respondents living in more remote locations do not report education as the most important issue (facing the country/government) more often than individuals living closer to the coast. However, we do find that individuals' actual educational attainment decreases substantially along coastal distance (column 2): moving from the

[18] Empirically, we add an interaction term constituted of log(*Distance to Harbor*) and a binary variable indicating whether the country is *Landlocked*. The sum of the coefficients log(*Distance to Harbor)* and the interaction term represents the total effect of distance to coast for individuals living in landlocked countries.
[19] We thank an anonymous referee for pointing out this additional extension.

Table 2.3: Exploring Potential Mechanisms of Influence of Coastal Proximity

	Dependent Variable								
	Most important Issue: Education {0,1}	Education: Educational Level {0,9}	Type of Residence: Urban {0,1}	Most important Issue: Institutions {0,1}	Institutions Score [1,4]	Most important Issue: Infrastructure {0,1}	Infrastr. in EA: (Elec.Grid/Pip.Water/Sewage/School/Pave.Road/Clinic) [0,1]	Helps your Country: REC {0,3}	Helps your Country: AU {0,3}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(1+Distance to Harbor)	0.001 (0.001)	-0.106*** (0.012)	-0.071*** (0.007)	0.002 (0.001)	0.025*** (0.003)	0.005*** (0.001)	-0.009*** (0.003)	0.034*** (0.009)	0.046*** (0.008)
Discrete Change of Distance from Harbor to the 3rd Quartile (564km)	*0.007*	*-0.670*	*-0.449*	*0.010*	*0.158*	*0.030*	*-0.056*	*0.214*	*0.293*
Sample Mean of Dep. Var.	[0.06]	[3.40]	[0.46]	[0.09]	[2.81]	[0.06]	[0.56]	[1.80]	[1.68]
Basic Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Urbanization Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Trade-related Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Full Geographic Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country-Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	120,461	126,555	127,075	120,461	126,683	120,461	116,174	35,409	47,298
R-Squared	0.03	0.29	0.30	0.06	0.21	0.04	0.43	0.11	0.07

Notes: Results in each column come from separate regressions and are estimated using the main sample of coastal sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in column (7) includes individuals from rounds 2 through 7 of the Afrobarometer, columns (8) and [9] includes data from rounds 2, 4, [5], and 6. We also report estimated interquartile differences in the respective dependent variables between minimum- and third quartile harbor distances within the sample. Binary dependent variables are estimated through a simple linear probability model specification. The standard errors reported are clustered at the survey enumeration area level. AU, African Union; REC, Regional Economic Community. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

minimum distance to the 3rd quartile within the sample reduces the level of education by .670, which corresponds to about 20% of the sample mean.

As shown by recent literature (Motamed et al. 2014; Henderson et al. 2018), levels of urbanization are negatively correlated with increased distance to coast in Africa, and can negatively impact individual economic welfare directly, or indirectly, via agglomeration economies (e.g. Skoufias and Katayama 2011; Bosker and Garretsen 2012; Young 2013; Gollin, Kirchberger, and Lagakos 2017; Flückiger and Ludwig 2018; Henderson et al. 2018). Consistent with this literature, we also find a negative, statistically significant, and quantitatively large relationship between (coastal) remoteness and living in urban environments (column 3). Given the strong interconnection between coastal proximity and urbanization, we also explore the differential effects of distance for individuals living in urban environments in Table A.16 separately, by estimating the interaction term *log(Distance to Harbor)* $\times$ *Urban* {0,1}. The results show that, while less pronounced, the distance penalty remains for two of our three outcomes for respondents from urban settings.

Regarding institutions, we proceed similarly to Mitton (2016; 107) and construct an index, *Institutions Score* [1,4], which combines responses concerning individuals' experiences with- and opinions on local authorities, offices and government. The results suggest that individuals living further away from the coast do not report institutions to be at the top of their concerns (column 4), nor is the institutional score negatively affected when living further way from major harbors (column 5). Recent literature has also suggested a weak link between institutions and differences in subnational development within countries (Radeny and Bulte 2012; Michalopoulos and Papaioannou 2014; Mitton 2016). In fact, individuals living in interior regions even seem to evaluate institutions more positively compared to coastal areas in our findings (column 5), a result which mirrors the one in Radeny and Bulte (2012) as well as Nunn and Wantchekon (2011), whereby distance to coast positively influenced levels of trust via a lower intensity of slave trade.

Infrastructure has been highlighted as a relevant factor for regional development (e.g. Calderón and Servén 2010; Dinkelman 2011; Jedwab and Moradi 2016; Storeygard 2016; Bluhm et al. 2018; Donaldson 2018; Jetter, Mösle, and Stadelmann 2019). Consistent with this literature, our results at the individual level show that coastal proximity is negatively associated with respondents' sentiments that infrastructure needs are issues of concern (see column 6 and 7). The actual access to basic infrastructure (as measured by our composite infrastructure index), is also negatively associated with distance to major ports.

Increased trade costs and reduced market access have been shown to be an inherent issue of remote areas in Africa (e.g. Bosker and Garretsen 2012; Atkin and Donaldson 2015; Henderson et al. 2018; Jedwab and Storeygard 2020). While trade volumes are necessarily an aggregate phenomenon, we find that survey respondents further away from the coast exhibit a higher tendency to report their respective Regional Economic Communities (RECs) or the

African Union (AU) as helpful to their country, which is consistent with them wishing to improve trade opportunities. Moreover, Table A.17 shows that the distance penalty is significantly increased for commercial farmers, i.e. farmers who mainly grow their produce for sale. Commercial farming is likely to depend on access to markets and trade opportunities, leaving commercial farmers more vulnerable to a distance penalty.

2.4.3. Bad Controls and Mediation Analysis

All results highlight coastal proximity as a statistically as well as an economically meaningful indicator of individual living standards and as a relevant predictor for diverse mechanisms that systematically relate and contribute to economic development and spatial inequalities. As coastal remoteness need not be destiny (Motamed et al. 2014), we aim to gauge the empirical importance of our controls as well as the potential mechanisms on our main explanatory variable by investigating the relevance of a bad controls problem and by performing a formal mediation analysis.

Bad Controls. We add in all of our baseline covariates and the explored mechanisms in step-wise fashion and report the corresponding changes to our main coefficient, *log(Distance to Harbor)*, as well as changes in the residual variance. Results are presented in Table 2.4. Row (a) shows the coefficient of *log(Distance to Harbor)* in a regression including country-time fixed effects only. Row (b) proceeds to add in our basic controls, i.e. *Age*, *Age squared* and *Female* {0,1}, as is done in Table 2.2. Row (c) adds our three urbanization controls to the specification, and so on.[21] The results show that, while the coefficient size of *log(Distance to Harbor)* diminishes, as is expected, coastal proximity remains a statistically relevant predictor of individual living standards throughout all rows and columns. The covariates contributing most to the specifications, as seen by changes in the coefficient (odd column numbers) as well as changes in the R-squared (even column numbers), are *Urbanization Controls, Educational Level* and *Infrastructure*, which are the ones we will explore as potential mediators next.

[21] We do not add sentiments of RECs or the AU to the list of covariates as their availability across survey rounds is sparse, observations size would drop by ~50%.

Table 2.4: Bad Controls, Relevance of included Covariates

	Dependent Variable					
	Cash Employment {0,1}		How often gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	log(1+Distance to Harbor)	Δ R-Squared	log(1+Distance to Harbor)	Δ R-Squared	log(1+Distance to Harbor)	Δ R-Squared
	(1)	(2)	(3)	(4)	(5)	(6)
(a) No Controls	-0.020*** (0.001)	-	0.076*** (0.004)	-	-0.037*** (0.001)	-
(b) = (a) + Basic Controls	-0.020*** (0.001)	[0.054]	0.075*** (0.004)	[0.003]	-0.037*** (0.001)	[0.026]
(c) = (b) + Urbanization Controls	-0.009*** (0.002)	[0.004]	0.024*** (0.004)	[0.021]	-0.014*** (0.001)	[0.036]
(d) = (c) + Trade-related Controls	-0.009*** (0.002)	[0.000]	0.024*** (0.004)	[0.000]	-0.014*** (0.001)	[0.000]
(e) = (d) + Geographic Controls	-0.009*** (0.002)	[0.001]	0.024*** (0.004)	[0.000]	-0.019*** (0.002)	[0.003]
(f) = (e) + Educational Level	-0.006** (0.002)	[0.017]	0.029*** (0.006)	[0.035]	-0.015*** (0.002)	[0.056]
(g) = (f) + Institutions Score	-0.006** (0.002)	[0.000]	0.036*** (0.006)	[0.016]	-0.015*** (0.002)	[0.000]
(h) = (g) + Infrastructure	-0.005** (0.002)	[0.000]	0.032*** (0.006)	[0.011]	-0.014*** (0.002)	[0.010]
Country-Time FE	YES		YES		YES	
Observations	114,857		115,307		102,287	
R-Squared	0.18		0.25		0.28	

Notes: Odd columns present the coefficient (changes) of our main explanatory variable log(1+Distance to Harbor) when subsequently adding seven distinct (sets of) control variables to a parsimonious baseline regression, constituted of our main regressor and country-time fixed effects. Even columns report the corresponding changes in the total R-squared values compared to the previous specification. The results in each row come from separate regressions, and observations are held constant across rows. Inclusion of mediating factors in (h), variables on infrastructure, limits the sample to rounds 2 to 7 of coastal sub-Saharan African countries included in the Afrobarometer. The remaining changes in the number of observations across columns stem from differences in the response rates of dependent variables (see notes in Table 2.2). Binary dependent variables are estimated through a simple linear probability model specification. Test statistics at the bottom of the table are produced from the full regression, that is, specification (h). The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Mediation Analysis. To further evaluate the link between coastal proximity and individual economic welfare, as well as its potential channels of influence, we conduct a formal mediation analysis. We empirically decompose the total effect of coastal proximity and individual welfare into *indirect effects*, i.e. effects which run through the proposed mediating factors, and *direct effects*, i.e. effects of coastal proximity that are unrelated to the proposed channels.

$$Y_{i,c,t} = \alpha_1 + \beta_1 \log(Distance\ to\ Harbor)_i + \theta M_i + \gamma_1 \boldsymbol{X'}_i + \delta_{c,t} + \varepsilon_{i,c,t} \tag{2.2}$$

$$M_{i,c,t} = \alpha_2 + \beta_2 \log(Distance\ to\ Harbor)_i + \gamma_2 \boldsymbol{X'}_i + \delta_{c,t} + \mu_{i,c,t} \tag{2.3}$$

β_1 measures the *direct effect* of coastal proximity on our different welfare indicators Y, and β_2 measures the effect of distance to harbor on the respective mediator M (e.g. education, urbanity, infrastructure). θ represents the direct effect of the mediator M on the outcome variable such that the *indirect effect* is retrieved by multiplying $\beta_2 \times \theta$ (Alwin and Hauser 1975; MacKinnon et al. 2007). The total effect is then given by a summation of the *direct* (β_1) and *indirect effects* ($\beta_2 \times \theta$).[22] Figure A.2 provides a visual representation of the mediation analysis. As before, $\boldsymbol{X}$ is a vector including all of our usual controls. We keep country-time fixed effects $\delta_{c,t}$ to evaluate a stringent setting.

Table 2.5 reports the coefficients of the total, direct and indirect effects of coastal proximity on individual economic welfare. Estimations are performed via structural equation modelling (SEM). To save space, we present the mechanisms on which distance to harbor had the largest impact in Table 2.3 and Table 2.4, *Education Level*, *Urbanization Controls* and *Infrastructure* and estimate their mediating effect on our three main outcome variables (results for our proxies of *Institutions* and *Trade* are relegated to Table A.18 in the Appendix).

The results in Table 2.5 suggest that a substantial part of the total effect of distance to harbors is mediated by educational attainment. Including respondents' level of schooling in the main specification (equation (2.2)) reduces the coefficient size of the direct effect of coastal proximity by 28% (see *proportion mediated* at the bottom of the table) on average, i.e. across outcome variables. In other words, coastal proximity matters for educational outcomes, and through education, it matters for individuals' living standards, subsequently. The direct effect of education on living standards is quantitively large and statistically significant throughout all estimations, indicating a relevant effect of education on economic welfare on its own. These results are in line with cross-country and subnational evidence, identifying educational differences as an important factor for explaining disparities in economic development (see Skoufias and Katayama 2011; Gennaioli et al. 2013; Chauvin et al. 2017; Flückiger and Ludwig 2018).

[22] Slight deviations in coefficients between the total effect in Table 2.2 and Table 2.5 arise because of missing values of the respective mediator variables introduced.

Table 2.5: Mediation Analysis: Direct and Indirect Links of Distance to Harbor

	Dependent Variable								
	Cash Employment {0,1}			How often gone without: (Water / Food / Cash Inc. / Med.) [0,4]			Possessions: (Radio / TV / Motor Vehicle) [0,1]		
	Mediator: Education {0,9}	Mediator: Urban {0,1}	Mediator: Infrastructure [0,1]	Mediator: Education {0,9}	Mediator: Urban {0,1}	Mediator: Infrastructure [0,1]	Mediator: Education {0,9}	Mediator: Urban {0,1}	Mediator: Infrastructure [0,1]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Baseline (Total) Effect:									
log(1+Distance to Harbor)	-0.009***	-0.012***	-0.009***	0.035***	0.055***	0.037***	-0.019***	-0.028***	-0.019***
	(0.002)	(0.002)	(0.002)	(0.006)	(0.006)	(0.006)	(0.002)	(0.002)	(0.002)
Direct Effect:									
log(1+Distance to Harbor)	-0.005**	-0.009***	-0.008***	0.027***	0.035***	0.032***	-0.015***	-0.019***	-0.017***
	(0.002)	(0.002)	(0.002)	(0.006)	(0.006)	(0.006)	(0.002)	(0.002)	(0.002)
Indirect Effect:									
Distance to Harbor via Mediator	-0.003***	-0.004***	-0.001***	0.009***	0.020***	0.005***	-0.004***	-0.009***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Direct Effect of Mediator:									
(Education / Urban / Infrastructure)	0.033***	0.055***	0.089***	-0.082***	-0.284***	-0.582***	0.044***	0.134***	0.210***
	(0.001)	(0.004)	(0.008)	(0.002)	(0.010)	(0.019)	(0.001)	(0.003)	(0.006)
Proportion Mediated	[0.39]	[0.30]	[0.09]	[0.25]	[0.36]	[0.14]	[0.21]	[0.33]	[0.10]
Basic Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Urbanization Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Trade-related Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Full Geographic Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country-Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	121,823	122,238	115,480	126,401	126,982	116,084	102,514	102,990	102,990

Notes: This table presents results from a formal mediation analysis testing the influence of three potential mediators of log(1+Distance to Harbor) on our three main outcome variables, respectively. Row 1 presents the baseline effect; its counterpart is shown in Table 2.2. Row 2 shows the direct effect of our main explanatory variable, that is, the effect of distance not attributable to the mediating factor, while row 3 depicts the part of the effect that runs precisely via its influence on the mediator. Row 4 shows the direct effect of the mediator on the respective outcome variable. Row 5 provides the proportion mediated, which is given by dividing the indirect effect by the total effect. Results in each column come from a separate SEM regression. The sample used is comprised of coastal sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Columns 3, 6, [7], [8], and [9] do not include individuals surveyed in rounds 1 [and 2] of the Afrobarometer, as questions on infrastructure (household items) were not asked in this round. Binary dependent variables are estimated through a simple linear probability model specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Similar insights arise for the role of urbanity in explaining relevant parts of distance's effect on living standards, mediating 33% of the effect on average. The mediator Urban {0,1} is therefore picking up a substantial part of the total effect of coastal distance, in similar magnitude as do educational differences. It is important to note that while both education and urbanization absorb variation in explaining individual living standards on their own (Table 2.4), as well as through their mediation of coastal proximity (Table 2.5), empirically, we cannot fully separate them. Indeed, existing literature has provided evidence suggesting that they are interrelated and mutually reenforced (e.g. Skoufias and Katayama 2011; Chauvin et al. 2017; Flückiger and Ludwig 2018).

Infrastructure, proxied by our composite measure Infrastructure Present in Enumeration Area: (Electricity Grid / Piped Water / Sewage / School / Paved Road / Health Clinic) [0,1], while relevant, does not show for an influence in similar magnitudes as do education or urbanization, mediating only an average of 11% of the effect. Table A.18 explores the role of institutions as a mediator in explaining the distance penalty. Contrary to human capital and urbanity, the pronounced gap in individual living standards across distances does not seem to be associated with perceived differences in (local) institutional quality when controlling for country-time fixed effects. Also, our evidence for a positive, direct effect of institutional quality on individual economic welfare is mixed (see row 4), consistent with other findings from subnational (regional) contexts (see Michalopoulos and Papaioannou 2014; Gennaioli et al. 2013; Mitton 2016).

The relevance of coastal proximity on economic development has often been ascribed to trade-related factors, especially among "late developers" (see Henderson et al. 2018). Table A.18 explores this link, estimating the direct and indirect effect of regional and supra-regional institutions fostering trade, as measured by respondents' evaluation of the African Union (AU) and their "corresponding" Regional Economic Community (REC), respectively. The results show that more positively perceived trade organizations correlate positively with individual living standards (row 4), which emphasizes a potential need for trade facilitation independent of individuals' distance to harbors, i.e. (global) markets.

2.5 Conclusion

We systematically investigate the role of coastal proximity in explaining intra-national differences in *individual* living standards across sub-Saharan Africa economies using an extensive dataset covering up to 128,609 observations distributed across 11,261 localities over 20 years. We employ geo-referenced individual-level data to complement the existing literature that focuses on outcomes at the national or regional level. Analyzing individuals' distance to harbors and their corresponding living standards allows us to test whether the insights of the cross-country and cross-regional contexts also apply at the individual level. Moreover, we can utilize the comprehensiveness of our dataset to explore a large set of indicators and potential channels of influence to gauge the relevance of coastal proximity and to investigate the mechanism through which it may matter for individual living standards.

Our results show that coastal proximity, as measured by geographical distance to harbors, predicts a relevant part of individual living standards and remains a strong predictor of individual economic welfare controlling for individual-level covariates, country-time specific influences via fixed-effects, as well as an extensive set of other established geographical influences of development.

Exploring potential channels, we find that human capital, urbanization, as well as access to infrastructure mediate relevant parts of the link between coastal proximity and economic development. This highlights that even though coastal proximity is a relevant indicator for individual living standards across Africa, coastal proximity need not be “destiny”. Fostering education as well as infrastructural outlays might help in mitigating problems associated with coastal remoteness. Nevertheless, the systematic robustness of coastal proximity as a predictor for individual living standards, even in stringent settings, suggests that there are relevant development costs of remoteness alone that need to be addressed (see also UN-OHRLLS 2013).

2.6 References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. 'The Colonial Origins of Comparative Development: An Empirical Investigation'. *American Economic Review* 91 (5): 1369–1401.

Afrobarometer, [Kenya, Tanzania, Uganda] [Round 1, 2, 3, 4, 5, 6, 7]. 2019. 'Afrobarometer Data'.

Alwin, Duane F., and Robert M. Hauser. 1975. 'The Decomposition of Effects in Path Analysis'. *American Sociological Review* 40 (1): 37.

Atkin, David, and Dave Donaldson. 2015. 'Who's Getting Globalized? The Size and Implications of Intra-National Trade Costs'. w21439. Cambridge, MA: National Bureau of Economic Research.

BenYishay, Ariel, Renee Rotberg, Jessica Wells, Zhonghui Lv, Seth Goodman, Lidia Kovacevic, and Dan Runfola. 2017. 'Geocoding Afrobarometer Rounds 1-6: Methodology & Data Quality'. Williamsburg, VA: AidData at William & Mary.

Bleakley, Hoyt, and Jeffrey Lin. 2012. 'Portage and Path Dependence *'. *The Quarterly Journal of Economics* 127 (2): 587–644.

Bloom, David E., Jeffrey D. Sachs, Paul Collier, and Christopher Udry. 1998. 'Geography, Demography, and Economic Growth in Africa'. *Brookings Papers on Economic Activity* 1998 (2): 207.

Bluhm, Richard, Axel Dreher, Andreas Fuchs, Bradley Parks, and Austin Strange. 2018. 'Connective Financing: Chinese Infrastructure Projects and the Diffusion of Economic Activity in Developing Countries'. In *AidData Working Paper #64*. Williamsburg, VA: Data at William & Mary.

Bosker, Maarten, and Harry Garretsen. 2012. 'Economic Geography and Economic Development in Sub-Saharan Africa'. *The World Bank Economic Review* 26 (3): 443–85.

Breinlich, Holger, Gianmarco I.P. Ottaviano, and Jonathan R.W. Temple. 2014. 'Regional Growth and Regional Decline'. In *Handbook of Economic Growth*, 2:683–779. Amsterdam.

Brülhart, Marius. 2011. 'The Spatial Effects of Trade Openness: A Survey'. *Review of World Economics* 147 (1): 59–83.

Calderón, César, and Luis Servén. 2010. 'Infrastructure and Economic Development in Sub-Saharan Africa'. *Journal of African Economies* 19 (suppl_1): i13–87.

Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. 2017. 'What Is Different about Urbanization in Rich and Poor Countries? Cities in Brazil, China, India and the United States'. *Journal of Urban Economics* 98: 17–49.

CIESIN. 2017. 'Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11'. Center For International Earth Science Information Network-CIESIN-Columbia University. Socioeconomic Data and Applications Center (SEDAC).

Dinkelman, Taryn. 2011. 'The Effects of Rural Electrification on Employment: New Evidence from South Africa'. *American Economic Review* 101 (7): 3078–3108.

Donaldson, Dave. 2018. 'Railroads of the Raj: Estimating the Impact of Transportation Infrastructure'. *American Economic Review* 108 (4–5): 899–934.

Easterly, William, and Ross Levine. 2003. 'Tropics, Germs, and Crops: How Endowments Influence Economic Development'. *Journal of Monetary Economics* 50 (1): 3–39.

Eberhard-Ruiz, Andreas, and Alexander Moradi. 2019. 'Regional Market Integration in East Africa: Local but No Regional Effects?' *Journal of Development Economics* 140 (September): 255–68.

Engel, Jakob, Deeksha Kokas, Gladys López Acevedo, and Maryla Maliszewska. 2021. The Distributional Impacts of Trade: Empirical Innovations, Analytical Tools, and Policy Responses. Trade and Development. Washington, DC: World Bank.

FAO, and IIASA. 2019. 'Global Agro Ecological Zones Version 4 (GAEZ V4)'.

Faye, Michael L., John W. McArthur, Jeffrey D. Sachs, and Thomas Snow. 2004. 'The Challenges Facing Landlocked Developing Countries'. *Journal of Human Development* 5 (1): 31–68.

Fick, Stephen E., and Robert J. Hijmans. 2017. 'WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas'. *International Journal of Climatology* 37 (12): 4302–15.

Flückiger, Matthias, and Markus Ludwig. 2018. 'Geography, Human Capital and Urbanization: A Regional Analysis'. *Economics Letters* 168: 10–14.

Frankel, Jeffrey A, and David Romer. 1999. 'Does Trade Cause Growth?' *American Economic Review* 89 (3): 379–99.

GADM. 2020. 'Global Administrative Areas [Digital Geospatial Data]'. Center for Spatial Sciences at the University of California, Davis.

Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger. 1999. 'Geography and Economic Development'. *International Regional Science Review* 22 (2): 179–232.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2013. 'Human Capital and Regional Development'. *The Quarterly Journal of Economics* 128 (1): 105–64.

Gollin, Douglas, Martina Kirchberger, and David Lagakos. 2017. 'In Search of a Spatial Equilibrium in the Developing World'. w23916. Cambridge, MA: National Bureau of Economic Research.

Henderson, Vernon J., Tim Squires, Adam Storeygard, and David Weil. 2018. 'The Global Distribution of Economic Activity: Nature, History, and the Role of Trade'. *The Quarterly Journal of Economics* 133 (1): 357–406.

Hirte, Georg, Christian Lessmann, and André Seidel. 2020. 'International Trade, Geographic Heterogeneity and Interregional Inequality'. *European Economic Review* 127 (August): 103427.

Jedwab, Remi, Edward Kerby, and Alexander Moradi. 2017. 'History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities In Kenya'. *The Economic Journal* 127 (603): 1467–94.

Jedwab, Remi, and Alexander Moradi. 2016. 'The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa'. *Review of Economics and Statistics* 98 (2): 268–84.

Jedwab, Remi, and Adam Storeygard. 2020. 'The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960-2010'. w27670. Cambridge, MA: National Bureau of Economic Research.

Jetter, Michael, Saskia Mösle, and David Stadelmann. 2019. 'Cursed by No Coast: How Regional Landlockedness Affects Income within Countries'. *Economics Letters* 181: 70–73.

Keats, Anthony. 2018. 'Women's Schooling, Fertility, and Child Health Outcomes: Evidence from Uganda's Free Primary Education Program'. *Journal of Development Economics* 135: 142–59.

Lehner, Bernhard, and Petra Döll. 2004. 'Development and Validation of a Global Database of Lakes, Reservoirs and Wetlands'. *Journal of Hydrology* 296 (1–4): 1–22.

Lessmann, Christian, and André Seidel. 2017. 'Regional Inequality, Convergence, and Its Determinants – A View from Outer Space'. *European Economic Review* 92 (February): 110–32.

Limão, Nuno, and Anthony J. Venables. 2001. 'Infrastructure, Geographical Disadvantage, Transport Costs, and Trade', 29.

MacKinnon, David P., Amanda J. Fairchild, and Matthew S. Fritz. 2007. 'Mediation Analysis'. *Annual Review of Psychology* 58 (1): 593–614.

Michalopoulos, Stelios, and Elias Papaioannou. 2014. 'National Institutions and Subnational Development in Africa'. *The Quarterly Journal of Economics* 129 (1): 151–213.

Mitton, Todd. 2016. 'The Wealth of Subnations: Geography, Institutions, and within-Country Development'. *Journal of Development Economics* 118: 88–111.

Motamed, Mesbah J., Raymond J. G. M. Florax, and William A. Masters. 2014. 'Agriculture, Transportation and the Timing of Urbanization: Global Analysis at the Grid Cell Level'. *Journal of Economic Growth* 19 (3): 339–68.

Natural Earth. 2019. 'Natural Earth. Free Vector and Raster Map Data.'

NGA, Natioal Geospace-Intelligence Agency. 2019. 'World Port Index 27th Edition'. United States Government.

Nunn, Nathan, and Diego Puga. 2012. 'Ruggedness: The Blessing of Bad Geography in Africa'. *Review of Economics and Statistics* 94 (1): 20–36.

Nunn, Nathan, and Leonard Wantchekon. 2011. 'The Slave Trade and the Origins of Mistrust in Africa'. *American Economic Review* 101 (7): 3221–52.

Olson, David M., Eric Dinerstein, Eric D. Wikramanayake, Neil D. Burgess, George V. N. Powell, Emma C. Underwood, Jennifer A. D'amico, et al. 2001. 'Terrestrial Ecoregions of the World: A New Map of Life on Earth'. *BioScience* 51 (11): 933.

Oster, Emily. 2019. 'Unobservable Selection and Coefficient Stability: Theory and Evidence'. *Journal of Business & Economic Statistics* 37 (2): 187–204.

Radelet, Steven, and Jeffrey D. Sachs. 1998. 'Shipping Costs, Manufactured Exports, and Economic Growth'. *National Bureau of Economic Research* Working Paper 19933.

Radeny, Maren, and Erwin Bulte. 2012. 'Determinants of Rural Income: The Role of Geography and Institutions in Kenya'. *Journal of African Economies* 21 (2): 307–41.

Ramankutty, Navin, Jonathan A. Foley, John Norman, and Kevin McSweeney. 2002. 'The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change: *Cropland Sensitivity to Climate Change*'. *Global Ecology and Biogeography* 11 (5): 377–92.

Rappaport, Jordan, and Jeffrey D. Sachs. 2003. 'The United States as a Coastal Nation'. *Journal of Economic Growth* 8: 5–46.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2004. 'Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development'. *Journal of Economic Growth* 9 (2): 131–65.

Sachs, Sonia Ehrlich, Andrew Mellinger, Pia Malaney, Jeffrey Sachs, Anthony Kiszewski, and Andrew Spielman. 2004. 'A Global Index Representing the Stability of Malaria Transmission'. *The American Journal of Tropical Medicine and Hygiene* 70 (5): 486–98.

Sala-I-Martin, Xavier, Gernot Doppelhofer, and Ronald I. Miller. 2004. 'Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach'. *American Economic Review* 94 (4): 813–35.

Sen, Amartya. 1985. *Commodities and Capabilities*. Lectures in Economics 7. Amsterdam New York Oxford: North-Holland.

Skoufias, E., and R. S. Katayama. 2011. 'Sources of Welfare Disparities between and within Regions of Brazil: Evidence from the 2002-2003 Household Budget Survey (POF)'. *Journal of Economic Geography* 11 (5): 897–918.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan and T. Cadell.

Smith, Adam, and R. H. Campbell. 2009. The Glasgow Edition of the Works and Correspondence of Adam Smith. 2 1: An Inquiry into the Nature and Causes of the Wealth of Nations / Adam Smith. General Editors R. H. Campbell. 11. print. Indianapolis, IN: Liberty Fund.

Storeygard, Adam. 2016. 'Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa'. *The Review of Economic Studies* 83 (3): 1263–95.

United Nations. 2001. 'Road Map towards the Implementation of the United Nations Millennium Declaration: Report of the Secretary-General.' New York: United Nations Department of Public Information.

UN-OHRLLS, United Nations Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and Small Island Developing States. 2013. 'The Development Economics of Landlockedness: Understanding the Development Costs of Being Landlocked'. New York, NY: United Nations - OHRLLS.

Wild, Frederik, and David Stadelmann. 2022. 'Coastal Proximity and Individual Living Standards: Econometric Evidence from Georeferenced Household Surveys in sub-Saharan Africa'. *Review of Development Economics* 26 (4): 1883–1901.

World Bank. 2009. *World Development Report 2009: Reshaping Economic Geography*. 2009. Washington, D.C.: World Bank.

———. 2022. 'Data.Worldbank.Org'. 2022.

Young, Alwyn. 2013. 'Inequality, the Urban-Rural Gap, and Migration'. *The Quarterly Journal of Economics* 128 (4): 1727–85.

Appendix A

Coastal Proximity and Individual Living Standards: Econometric Evidence from Geo-Referenced Household Surveys in Sub-Saharan Africa

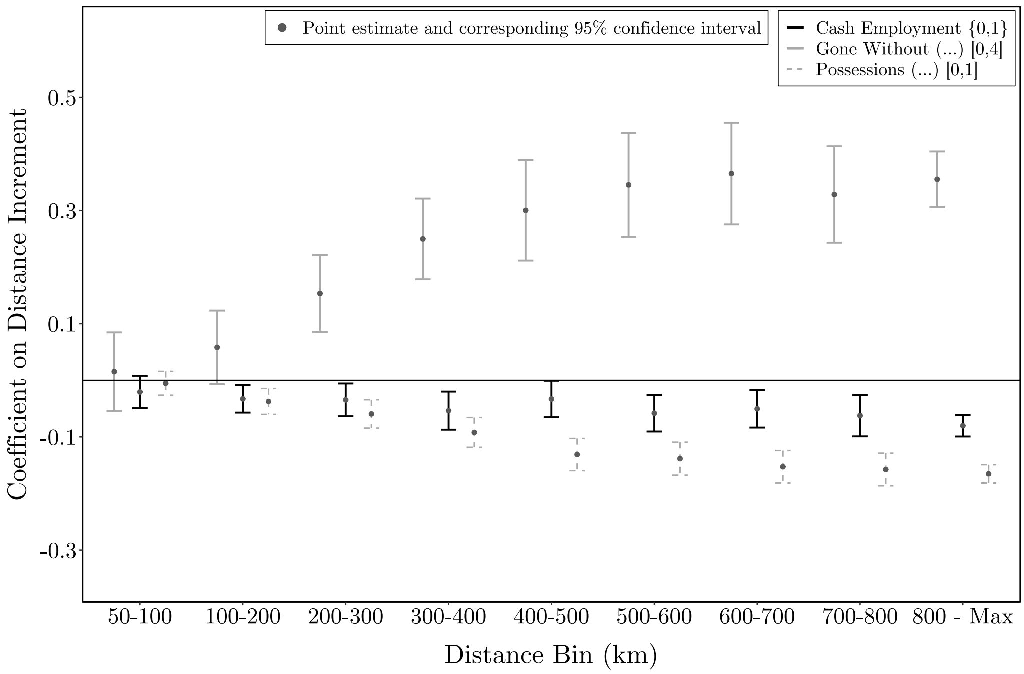


Figure A.1: Distance Bins

Notes: The plot depicts point estimates as well as their corresponding 95% confidence intervals produced from three separate regressions of the three main outcome variables - distinguished by black, grey and dashed-grey figures - on all harbor distance increments shown on the X-axis (the omitted category is 0-50 km), and including country-sample fixed effects. Therefore, coefficients are interpreted as the average change in the outcome variable for individuals living within the distance increments to living within 50km to the harbor. Results are produced using the main sample of coastal, sub-Saharan African countries from round 1 through round 7 of the geo-coded Afrobarometer surveys, except for the dashed estimates (Possessions), given that this question is available only from round 3 and onwards. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. Standard errors are clustered at the survey-enumeration level.

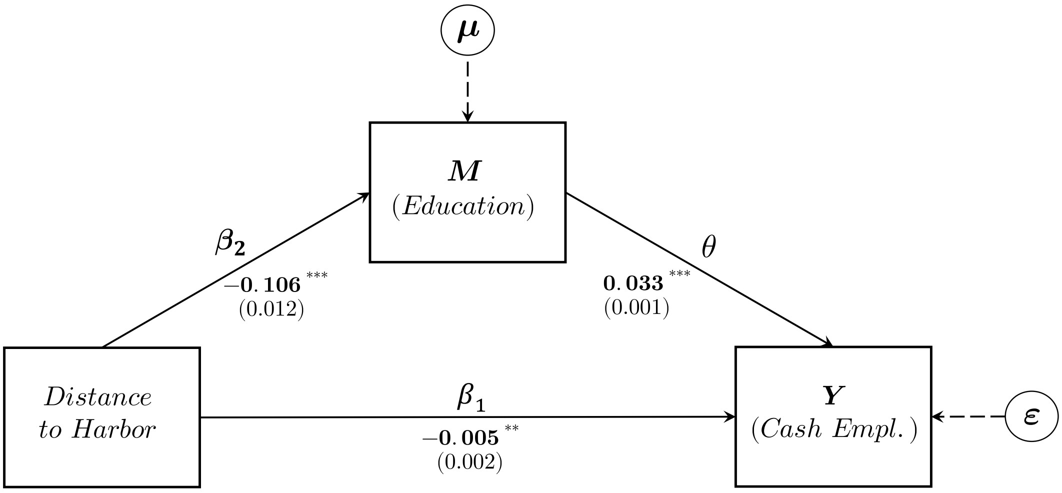


Figure A.2: Structure of Mediation Analysis

Notes: The figure illustrates the envisioned mediation structure discussed in the manuscript. The estimate depicted represents the results of column (1) of Table 2.5. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table A.1: Coastal Proximity and Individual Living Standards: Full Table

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.018***	-0.009***	0.073***	0.035***	-0.037***	-0.019***
	(0.001)	(0.002)	(0.004)	(0.006)	(0.001)	(0.002)
Discrete Change of Distance from Harbor to the 3rd Quartile (564km)	*-0.115*	*-0.055*	*0.460*	*0.223*	*-0.234*	*-0.120*
Sample Mean of Dep. Var.	[0.39]		[1.28]		[0.51]	
Basic Controls						
Age		0.031***		0.011***		0.012***
		(0.001)		(0.001)		(0.000)
Age^2		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
Female {0,1}		-0.106***		0.020***		-0.083***
		(0.003)		(0.004)		(0.002)
Urbanization Controls						
Urban {0,1}		0.055***		-0.284***		0.134***
		(0.004)		(0.010)		(0.003)
Primate City {0,1}		0.029***		-0.084***		0.029***
		(0.007)		(0.016)		(0.006)
Population Density		-0.001		0.004***		-0.001*
		(0.000)		(0.001)		(0.000)
Trade-related Controls						
Navigable River {0,1}		-0.013		-0.121***		0.029***
		(0.009)		(0.021)		(0.008)
Major Lake {0,1}		0.010		0.033		-0.023***
		(0.012)		(0.024)		(0.008)
Full Geographic Controls						
Abs. Latitude		-0.002		-0.015***		0.001
		(0.001)		(0.002)		(0.001)
Elevation (km)		-0.016		-0.115***		0.038***
		(0.013)		(0.028)		(0.012)
Ruggedness (Standardized)		-0.004		0.015*		-0.006**
		(0.004)		(0.008)		(0.003)
Land Suitability [0,1]		0.021*		0.189***		0.011
		(0.012)		(0.027)		(0.009)
Monthly Temperature (Celsius)		-0.009***		0.020***		-0.001
		(0.002)		(0.005)		(0.002)
Monthly Rainfall (Standardized)		-0.003		0.003		-0.001
		(0.004)		(0.009)		(0.003)
Growing Days {0,365}		0.000		-0.001***		0.000**
		(0.000)		(0.000)		(0.000)
Malaria Index		0.000		0.004***		-0.001***
		(0.000)		(0.001)		(0.000)
Mediterraenean {0,1}		0.014		0.038		0.038
		(0.037)		(0.084)		(0.031)
Desert {0,1}		0.028		0.170**		0.024
		(0.035)		(0.077)		(0.030)

Notes: Table continued on next page.

	(1)	(2)	(3)	(4)	(5)	(6)
Mangroves {0,1}		0.017		0.129*		-0.005
		(0.034)		(0.074)		(0.028)
Tropical Forest {0,1}		0.041		0.037		0.025
		(0.032)		(0.072)		(0.027)
Tropical Grassland {0,1}		0.029		0.070		0.017
		(0.032)		(0.071)		(0.027)
Temperate Grassland {0,1}		0.046		0.146*		0.016
		(0.039)		(0.089)		(0.032)
Montane Grassland {0,1}		0.032		0.229***		-0.007
		(0.034)		(0.076)		(0.029)
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	123,793	122,238	128,609	126,982	103,889	102,990
R-Squared	0.09	0.14	0.17	0.20	0.15	0.22

Notes: Results in each column come from separate regressions and are estimated using the main sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in these rounds. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.2: De-Composed Indices

	Dependent Variable													
	How often gone without: (Cash Income) {0,4}		How often gone without: (Water) {0,4}		How often gone without: (Food) {0,4}		How often gone without: (Medical Care) {0,4}		Possessions: (Radio) {0,1}		Possessions: (TV) {0,1}		Possessions: (Motor Vehicle) {0,1}	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
log(1+Distance to Harbor)	0.095***	0.051***	0.069***	0.015*	0.036***	0.023***	0.096***	0.053***	-0.019***	-0.017***	-0.085***	-0.039***	-0.008***	0.000
	(0.004)	(0.007)	(0.005)	(0.009)	(0.004)	(0.007)	(0.004)	(0.007)	(0.001)	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)
Discrete Change of Distance from Harbor to the 3rd Quartile (564km)	*0.599*	*0.322*	*0.433*	*0.093*	*0.230*	*0.143*	*0.604*	*0.334*	*-0.121*	*-0.109*	*-0.535*	*-0.248*	*-0.050*	*-0.003*
Sample Mean of Dep. Var.	[1.90]		[1.14]		[0.97]		[1.15]		[0.75]		[0.48]		[0.30]	
Basic Controls														
Age		0.007***		0.010***		0.010***		0.015***		0.014***		0.011***		0.010***
		(0.001)		(0.001)		(0.001)		(0.001)		(0.001)		(0.001)		(0.000)
Age^2		0.000***		0.000***		0.000***		0.000***		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
Female {0,1}		0.043***		-0.002		0.041***		0.002		-0.119***		-0.038***		-0.091***
		(0.006)		(0.006)		(0.006)		(0.006)		(0.003)		(0.002)		(0.003)
Urbanization Controls														
Urban {0,1}		-0.287***		-0.298***		-0.214***		-0.350***		0.062***		0.259***		0.083***
		(0.012)		(0.015)		(0.011)		(0.012)		(0.004)		(0.006)		(0.004)
Primate City {0,1}		-0.144***		-0.069***		-0.045***		-0.079***		0.004		0.051***		0.033***
		(0.021)		(0.023)		(0.016)		(0.018)		(0.006)		(0.009)		(0.008)
Population Density		0.007***		0.002*		0.004***		0.004***		-0.001*		0.002***		-0.003***
		(0.001)		(0.001)		(0.001)		(0.001)		(0.000)		(0.001)		(0.001)
Trade-related Controls														
Navigable River {0,1}		-0.125***		-0.137***		-0.152***		-0.075***		0.021**		0.044***		0.022***
		(0.023)		(0.030)		(0.024)		(0.025)		(0.009)		(0.011)		(0.008)
Major Lake {0,1}		0.060*		-0.001		0.053*		0.031		-0.008		-0.040***		-0.017**
		(0.032)		(0.039)		(0.027)		(0.029)		(0.010)		(0.013)		(0.009)
Full Geographic Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	122,363	120,792	128,427	126,808	128,356	126,761	128,105	126,508	103,838	102,940	103,582	102,687	103,388	102,493
R-Squared	0.17	0.19	0.09	0.12	0.07	0.09	0.11	0.14	0.04	0.08	0.23	0.30	0.23	0.25

Notes: This table reports results for all items of our two main composite living-standards indices tested in Table 2.2. Results in each column come from separate regressions and are estimated using the main sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (9) through (14) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in these rounds. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table A.3: Summary of Robustness Tests

	Coefficient on log(1+Distance to Harbor) - if not indicated otherwise					
	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficient	-0.018***	-0.009***	0.073***	0.035***	-0.037***	-0.019***
	(0.001)	(0.002)	(0.004)	(0.006)	(0.001)	(0.002)
(a) *log(1+Beeline Distance to Harbor)*	-0.019***	-0.011***	0.071***	0.033***	-0.037***	-0.019***
(see Table A.4)	(0.001)	(0.003)	(0.004)	(0.006)	(0.001)	(0.002)
(b) log(1+*Distance to Coastline*)	-0.014***	-0.008***	0.050***	0.031***	-0.026***	-0.010***
(see Table A.5)	(0.001)	(0.002)	(0.003)	(0.005)	(0.001)	(0.002)
(c) *log(1+Beeline Distance to Coastline)*	-0.014***	-0.011***	0.048***	0.027***	-0.026***	-0.011***
(see Table A.6)	(0.001)	(0.002)	(0.003)	(0.005)	(0.001)	(0.002)
(d) Including Harbor ≤ 25km Dummy	-0.028***	-0.021***	0.038***	0.037***	-0.028***	-0.023***
(see Table A.7)	(0.002)	(0.003)	(0.006)	(0.007)	(0.002)	(0.002)
(e) Including Coast ≤ 25km Dummy	-0.019***	-0.010***	0.060***	0.031***	-0.033***	-0.018***
(see Table A.7)	(0.002)	(0.003)	(0.005)	(0.006)	(0.002)	(0.002)
(f) Keeping Observations Constant	-0.019***	-0.008***	0.074***	0.032***	-0.037***	-0.019***
(see Table A.8)	(0.001)	(0.003)	(0.004)	(0.006)	(0.001)	(0.002)
(g) Excluding Distances > 629 km	-0.015***	-0.005**	0.074***	0.039***	-0.035***	-0.016***
(see Table A.9)	(0.002)	(0.003)	(0.004)	(0.006)	(0.001)	(0.002)
(h) Excluding Low-Precision Localities	-0.018***	-0.011***	0.062***	0.029***	-0.034***	-0.018***
(see Table A.10)	(0.002)	(0.003)	(0.005)	(0.008)	(0.002)	(0.003)
(i) Including Survey Weights	-0.018***	-0.006**	0.070***	0.029***	-0.037***	-0.017***
(see Table A.11)	(0.002)	(0.003)	(0.004)	(0.006)	(0.001)	(0.002)
(j) Country-Sample Clustering	-0.018***	-0.009*	0.073***	0.035***	-0.037***	-0.019***
(see Table A.12)	(0.003)	(0.005)	(0.008)	(0.009)	(0.003)	(0.003)
Basic Controls	NO	YES	NO	YES	NO	YES
Urbanization Controls	NO	YES	NO	YES	NO	YES
Trade-related Controls	NO	YES	NO	YES	NO	YES
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES

Notes: This table summarizes the robustness checks on our main findings. Results in each row and column come from separate regressions. a) Replaces the main explanatory variable of (logged) within-country distance used in Table 2.2 with a simple beeline distance to the nearest harbor. b) Replaces the main explanatory variable with a within-country distance to the country's coastline. (c) Replaces the main explanatory variable with a beeline distance to the coastline. (d) Adds a dummy indicator of living within 25 kilometers to a major harbor to our baseline specification. (e) Includes a dummy indicator of living within 25 kilometers to the coastline to our baseline specification. (f) Estimates the baseline specification but holds the number of observations constant across all columns, using the maximum number of individuals for which all in- and dependent variables are available. Note that this drops rounds 1 and 2 of the Afrobarometer entirely, given that questions on ownership of household items were not asked. (g) Estimates the baseline specification but excludes distances larger than the 80th percentile. (h) Estimates the baseline specification but excludes localities marked with a precision code of 2 or larger (scale 1-8). (i) Estimates the baseline specification but uses the included Afrobarometer survey weights. (j) Estimates the baseline specification but implements standard error clustering at the country-sample level instead of the survey enumeration area level. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level, if not indicated otherwise indicated. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.4: Robustness Test: "*beeline*" Distance to Harbor

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Beeline Distance to Harbor)	-0.019***	-0.011***	0.071***	0.033***	-0.037***	-0.019***
	(0.001)	(0.003)	(0.004)	(0.006)	(0.001)	(0.002)
Discrete Change of Distance from Harbor to the 3rd Quartile (506km)	*-0.117*	*-0.067*	*0.440*	*0.204*	*-0.229*	*-0.115*
Sample Mean of Dep. Var.	[0.39]		[1.28]		[0.51]	
Basic Controls						
Age		0.031***		0.011***		0.012***
		(0.001)		(0.001)		(0.000)
Age^2		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
Female {0,1}		-0.106***		0.020***		-0.083***
		(0.003)		(0.004)		(0.002)
Urbanization Controls						
Urban {0,1}		0.055***		-0.285***		0.135***
		(0.004)		(0.010)		(0.003)
Primate City {0,1}		0.026***		-0.087***		0.029***
		(0.007)		(0.016)		(0.006)
Population Density		-0.001		0.004***		-0.001
		(0.000)		(0.001)		(0.000)
Trade-related Controls						
Navigable River {0,1}		-0.012		-0.125***		0.031***
		(0.009)		(0.020)		(0.008)
Major Lake {0,1}		0.011		0.033		-0.023***
		(0.012)		(0.024)		(0.008)
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	123,793	122,238	128,609	126,982	103,889	102,990
R-Squared	0.10	0.14	0.17	0.20	0.15	0.22

Notes: This table is equivalent to Table 2.2 but exchanges the within-country distance to harbors used in Table 1 with a beeline (as the crow flies) distance to the nearest harbor. Results in each column come from separate regressions and are estimated using the main sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in these rounds. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.5: Robustness Test: Distance to Coastline

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Coastline)	-0.014***	-0.008***	0.050***	0.031***	-0.026***	-0.010***
	(0.001)	(0.002)	(0.003)	(0.005)	(0.001)	(0.002)
Discrete Change of Distance from the Coast to the 3rd Quartile (468km)	*-0.085*	*-0.047*	*0.303*	*0.188*	*-0.161*	*-0.063*
Sample Mean of Dep. Var.	[0.39]		[1.28]		[0.51]	
Basic Controls						
Age		0.031***		0.011***		0.012***
		(0.001)		(0.001)		(0.000)
Age^2		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
Female {0,1}		-0.106***		0.020***		-0.082***
		(0.003)		(0.004)		(0.002)
Urbanization Controls						
Urban {0,1}		0.055***		-0.285***		0.136***
		(0.004)		(0.010)		(0.003)
Primate City {0,1}		0.033***		-0.102***		0.043***
		(0.007)		(0.014)		(0.005)
Population Density		-0.001		0.004***		0.000
		(0.000)		(0.001)		(0.000)
Trade-related Controls						
Navigable River {0,1}		-0.011		-0.130***		0.032***
		(0.009)		(0.021)		(0.008)
Major Lake {0,1}		0.012		0.026		-0.023***
		(0.012)		(0.024)		(0.008)
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	123,793	122,238	128,609	126,982	103,889	102,990
R-Squared	0.09	0.14	0.16	0.20	0.15	0.22

Notes: This table is equivalent to Table 2.2, but exchanges the within-country distance to harbors used in Table 1 with within-country distance to the coastline. Results in each column come from separate regressions and are estimated using the main sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in these rounds. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.6: Robustness Test: "*beeline*" Distance to Coastline

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Beeline Distance to Coastline)	-0.014***	-0.011***	0.048***	0.027***	-0.026***	-0.011***
	(0.001)	(0.002)	(0.003)	(0.005)	(0.001)	(0.002)
Discrete Change of Distance from the Coast to the 3rd Quartile (426km)	*-0.087*	*-0.063*	*0.288*	*0.162*	*-0.158*	*-0.063*
Sample Mean of Dep. Var.	[0.39]		[1.28]		[0.51]	
Basic Controls						
Age		0.031***		0.011***		0.012***
		(0.001)		(0.001)		(0.000)
Age^2		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
Female {0,1}		-0.106***		0.020***		-0.082***
		(0.003)		(0.004)		(0.002)
Urbanization Controls						
Urban {0,1}		0.055***		-0.286***		0.136***
		(0.004)		(0.010)		(0.003)
Primate City {0,1}		0.031***		-0.105***		0.043***
		(0.007)		(0.015)		(0.005)
Population Density		-0.001		0.004***		0.000
		(0.000)		(0.001)		(0.000)
Trade-related Controls						
Navigable River {0,1}		-0.010		-0.130***		0.033***
		(0.009)		(0.021)		(0.008)
Major Lake {0,1}		0.013		0.028		-0.023***
		(0.012)		(0.024)		(0.008)
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	123,793	122,238	128,609	126,982	103,889	102,990
R-Squared	0.09	0.14	0.16	0.20	0.15	0.22

Notes: This table is equivalent to Table 2.2, but exchanges the within-country distance to harbors used in Table 1 with a beeline (as the crow flies) distance to the coastline. Results in each column come from separate regressions and are estimated using the main sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in these rounds. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.7: Robustness Test: Adding Harbor/Coastal Dummies

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.021***	-0.010***	0.037***	0.031***	-0.023***	-0.018***
	(0.003)	(0.003)	(0.007)	(0.006)	(0.002)	(0.002)
Discrete Change of Distance from Harbor to the 3rd Quartile (564km)	*-0.133*	*-0.064*	*0.232*	*0.195*	*-0.147*	*-0.116*
Sample Mean of Dependent Var.	[0.39]		[1.28]		[0.51]	
Trade-related Controls						
Harbor {0,1}	-0.071***		0.008		-0.025***	
	(0.010)		(0.025)		(0.008)	
Coast {0,1}		-0.013*		-0.039**		0.006
		(0.008)		(0.018)		(0.006)
Navigable River {0,1}	-0.013	-0.013	-0.121***	-0.122***	0.029***	0.029***
	(0.009)	(0.009)	(0.021)	(0.020)	(0.008)	(0.008)
Major Lake {0,1}	0.016	0.010	0.032	0.032	-0.021***	-0.023***
	(0.012)	(0.012)	(0.024)	(0.024)	(0.008)	(0.008)
Urbanization Controls						
Urban {0,1}	0.056***	0.055***	-0.285***	-0.283***	0.135***	0.134***
	(0.004)	0.0043	(0.010)	0.0099	(0.003)	0.0035
Primate City {0,1}	0.044***	0.031***	-0.085***	-0.078***	0.034***	0.028***
	(0.008)	(0.007)	(0.016)	(0.016)	(0.006)	(0.006)
Population Density	0.000	-0.001	0.004***	0.004***	-0.001*	-0.001*
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Basic Controls						
Age	0.031***	0.031***	0.011***	0.011***	0.012***	0.012***
	(0.001)	0.0006	(0.001)	0.0009	(0.000)	0.0004
Age^2	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female {0,1}	-0.106***	-0.106***	0.020***	0.020***	-0.083***	-0.083***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)
Full Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	122,238	122,238	126,982	126,982	102,990	102,990
R-Squared	0.15	0.14	0.20	0.20	0.22	0.22

Notes: This table is equivalent to Table 2.2, but adds dummies indicating individuals living within 25km of the harbor or coast, in turn. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.8 Robustness Test: Constant Observations

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.019***	-0.008***	0.074***	0.032***	-0.037***	-0.019***
	(0.001)	(0.003)	(0.004)	(0.006)	(0.001)	(0.002)
Discrete Change of Distance from Harbor to the 3rd Quartile (547km)	*-0.118*	*-0.049*	*0.463*	*0.201*	*-0.232*	*-0.118*
Sample Mean of Dep. Var.	[0.39]		[1.33]		[0.51]	
Basic Controls						
Age		0.035***		0.011***		0.012***
		(0.001)		(0.001)		(0.000)
Age^2		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
Female {0,1}		-0.114***		0.022***		-0.083***
		(0.003)		(0.005)		(0.002)
Urbanization Controls						
Urban {0,1}		0.060***		-0.288***		0.135***
		(0.005)		(0.011)		(0.003)
Primate City {0,1}		0.030***		-0.076***		0.029***
		(0.007)		(0.018)		(0.006)
Population Density		-0.001**		0.003***		-0.001*
		(0.000)		(0.001)		(0.000)
Trade-related Controls						
Navigable River {0,1}		-0.018*		-0.143***		0.029***
		(0.010)		(0.022)		(0.008)
Major Lake {0,1}		-0.001		0.030***		-0.023***
		(0.012)		(0.003)		(0.008)
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	102,460	102,460	102,460	102,460	102,460	102,460
R-Squared	0.11	0.17	0.17	0.20	0.15	0.22

Notes: This table is equivalent to Table 2.2, but holds the number of observations constant across all columns, using the maximum number of individuals for which all independent and dependent variables are available. Note that this drops rounds 1 and 2 of the Afrobarometer entirely, given that questions on ownership of household items were not asked in these rounds. Therefore, the sample used is comprised of coastal, sub-Saharan African countries included in survey rounds 3 through 7 of the Afrobarometer. Results in each column come from separate regressions. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the country-sample level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.9: Robustness Test: Excluding Large Distances

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.015***	-0.005**	0.074***	0.039***	-0.035***	-0.016***
	(0.002)	(0.003)	(0.004)	(0.006)	(0.001)	(0.002)
Discrete Change of Distance from Harbor to the 3rd Quartile (410km)	*-0.093*	*-0.032*	*0.444*	*0.232*	*-0.207*	*-0.097*
Sample Mean of Dep. Var.	[0.39]		[1.27]		[0.53]	
Basic Controls						
Age		0.032***		0.011***		0.012***
		(0.001)		(0.001)		(0.000)
Age^2		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
Female {0,1}		-0.101***		0.027***		-0.087***
		(0.003)		(0.005)		(0.002)
Urbanization Controls						
Urban {0,1}		0.050***		-0.296***		0.131***
		(0.005)		(0.011)		(0.004)
Primate City {0,1}		0.033***		-0.076***		0.033***
		(0.007)		(0.016)		(0.006)
Population Density		-0.001**		0.004***		-0.001**
		(0.000)		(0.001)		(0.000)
Trade-related Controls						
Navigable River {0,1}		-0.008		-0.108***		0.027***
		(0.010)		(0.024)		(0.008)
Major Lake {0,1}		0.042**		0.015***		-0.019
		(0.019)		(0.003)		(0.016)
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	99,302	98,269	103,029	101,945	84,589	83,929
R-Squared	0.10	0.15	0.18	0.22	0.15	0.21

Notes: This table is equivalent to Table 2.2, but excludes distances larger than the 80th percentile. Results in each column come from separate regressions and are estimated using the main sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in these rounds. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the country-sample level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.10: Robustness Test: Excluding Low-Precision Geo-clusters

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.018***	-0.011***	0.062***	0.029***	-0.034***	-0.018***
	(0.002)	(0.003)	(0.005)	(0.008)	(0.002)	(0.003)
Discrete Change of Distance from Harbor to the 3rd Quartile (542km)	*-0.111*	*-0.068*	*0.374*	*0.173*	*-0.204*	*-0.108*
Sample Mean of Dep. Var.	[0.40]		[1.21]		[0.51]	
Basic Controls						
Age		0.034***		0.011***		0.017***
		(0.001)		(0.001)		(0.001)
Age^2		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
Female {0,1}		-0.105***		0.022***		-0.088***
		(0.004)		(0.006)		(0.003)
Urbanization Controls						
Urban {0,1}		0.046***		-0.291***		0.124***
		(0.006)		(0.014)		(0.005)
Primate City {0,1}		0.035***		-0.117***		0.044***
		(0.010)		(0.021)		(0.007)
Population Density		-0.001		0.006***		-0.001**
		(0.001)		(0.001)		(0.001)
Trade-related Controls						
Navigable River {0,1}		-0.012		-0.110***		0.031***
		(0.012)		(0.027)		(0.010)
Major Lake {0,1}		-0.002		0.002		-0.025**
		(0.018)		(0.003)		(0.012)
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	61,731	60,869	64,023	63,122	49,067	48,594
R-Squared	0.08	0.14	0.18	0.21	0.15	0.21

Notes: This table is equivalent to Table 2.2, but excludes localities marked with a precision code of 2 or larger (scale 1-8). Results in each column come from separate regressions and are estimated using the main sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in these rounds. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the country-sample level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.11: Robustness Test: Including Survey Weights

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.018***	-0.006**	0.070***	0.029***	-0.037***	-0.017***
	(0.002)	(0.003)	(0.004)	(0.006)	(0.001)	(0.002)
Discrete Change of Distance from Harbor to the 3rd Quartile (425km)	*-0.112*	*-0.039*	*0.446*	*0.186*	*-0.232*	*-0.106*
Sample Mean of Dep. Var.	[0.39]		[1.28]		[0.51]	
Basic Controls						
Age		0.031***		0.011***		0.012***
		(0.001)		(0.001)		(0.000)
Age^2		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
Female {0,1}		-0.105***		0.021***		-0.083***
		(0.003)		(0.005)		(0.002)
Urbanization Controls						
Urban {0,1}		0.055***		-0.281***		0.135***
		(0.005)		(0.011)		(0.004)
Primate City {0,1}		0.030***		-0.090***		0.030***
		(0.008)		(0.017)		(0.006)
Population Density		0.000		0.004***		-0.001
		(0.000)		(0.001)		(0.000)
Trade-related Controls						
Navigable River {0,1}		-0.009		-0.111***		0.027***
		(0.010)		(0.022)		(0.008)
Major Lake {0,1}		0.019		0.023		-0.020**
		(0.013)		(0.025)		(0.009)
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	123,793	122,238	128,608	126,982	103,889	102,990
R-Squared	0.10	0.15	0.16	0.20	0.15	0.22

Notes: This table is equivalent to Table 2.2, but uses the included Afrobarometer survey weights to produce the estimates. Results in each column come from separate regressions and are estimated using the main sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in these rounds. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.12: Robustness Test: Country-Sample Clustering

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.018***	-0.009*	0.073***	0.035***	-0.037***	-0.019***
	(0.003)	(0.005)	(0.008)	(0.009)	(0.003)	(0.003)
Discrete Change of Distance from Harbor to the 3rd Quartile (564km)	*-0.115*	*-0.055*	*0.460*	*0.223*	*-0.234*	*-0.120*
Sample Mean of Dep. Var.	[0.39]		[1.28]		[0.51]	
Basic Controls						
Age		0.031***		0.011***		0.012***
		(0.002)		(0.001)		(0.001)
Age^2		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
Female {0,1}		-0.106***		0.020***		-0.083***
		(0.007)		(0.005)		(0.006)
Urbanization Controls						
Urban {0,1}		0.055***		-0.284***		0.134***
		(0.007)		(0.018)		(0.007)
Primate City {0,1}		0.029**		-0.084***		0.029***
		(0.011)		(0.027)		(0.009)
Population Density		-0.001		0.004***		-0.001
		(0.001)		(0.001)		(0.001)
Trade-related Controls						
Navigable River {0,1}		-0.013		-0.121***		0.029***
		(0.012)		(0.026)		(0.010)
Major Lake {0,1}		0.010		0.033		-0.023**
		(0.020)		(0.036)		(0.010)
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	123,793	122,238	128,609	126,982	103,889	102,990
R-Squared	0.09	0.14	0.17	0.20	0.15	0.22

Notes: This table is equivalent to Table 2.2, but implements standard error clustering at the country-sample level instead of the survey enumeration area level. Results in each column come from separate regressions and are estimated using the main sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in these rounds. We also report estimated interquartile differences in the respective dependent variables between minimum-, and 3rd quartile harbor distances within the sample. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the country-sample level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.2.

Table A.13: Robustness Test: Oster (2019) Tests

	Explanatory Variable: log(1+Distance to Harbor), Specification: δ=0.5, Rmax=1				
Independent Variable	β° Baseline Effect	β' Controlled Effect	$R^{\circ 2}$ Baseline	R'^{2} Controlled	$[\beta', \beta^{*}]$ Identified Set
	(1)	(2)	(3)	(4)	(5)
Cash Employment {0,1}	-0.003 (0.002)	-0.009*** (0.002)	[0.00]	[0.14]	[-0.027 , -0.009]
How often gone without: (Water / Food / Cash Inc. / Med.) [0,4]	0.025*** (0.006)	0.035*** (0.006)	[0.00]	[0.20]	[0.035 , 0.055]
Possessions (Radio / TV / Motor Vehicle) [0,1]	-0.030*** (0.002)	-0.019*** (0.002)	[0.02]	[0.22]	[-0.019 , 0.002]
Basic Controls	NO	YES	NO	YES	YES
Urbanization Controls	NO	YES	NO	YES	YES
Trade-related Controls	NO	YES	NO	YES	YES
Full Geographic Controls	NO	YES	NO	YES	YES
Country-Time FE	NO	YES	NO	YES	YES

Notes: This table presents result from a formal analysis of coefficient stability and influence of unobservables according to Oster (2019), analyzing changes in the estimate of our main explanatory variable "log(1+Distance to Harbor)" when adding the full set of controls as well as fixed-effects, using our three main outcome variables. Columns (1) and (2) present the uncontrolled β°, as well as the controlled β' and columns (3) and (4) depict their respective regression's R-Squared. Column (5) shows the lower- and upper bound estimate of the identified set assuming Rmax = 1 and δ = 0.5. The bias-adjusted upper bound is calculated using $\beta^{*} = \beta' - \delta((\beta^{\circ} - \beta')*(Rmax - R'2))/(R'2 - R^{\circ}2))$, the lower bound is given by β'. Results are estimated using the main sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. The sample used row 3 do not include individuals surveyed in Rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in this round. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey-enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table A.14: Extensions: "Landlockedness"

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.018***	0.003	0.073***	0.017***	-0.037***	-0.009***
	(0.001)	(0.002)	(0.004)	(0.005)	(0.001)	(0.002)
Interaction:						
log(1+Distance to Harbor) × Landlocked {0,1}	0.030**	0.019	0.026	0.057*	0.025*	0.005
	(0.013)	(0.013)	(0.039)	(0.033)	(0.014)	(0.010)
Isolated Effect of the Distance to Harbor in Landlocked Countries						
Combined Effect:						
log(1+Distance to Harbor) + *Interaction*	0.011	0.022*	0.099**	0.074**	-0.012	-0.005
	[0.39]	[0.08]	[0.01]	[0.03]	[0.40]	[0.65]
Basic Controls	NO	YES	NO	YES	NO	YES
Urbanization Controls	NO	YES	NO	YES	NO	YES
Trade-related Controls	NO	YES	NO	YES	NO	YES
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	204,717	200,128	212,037	207,211	169,590	166,242
R-Squared	0.11	0.15	0.15	0.19	0.16	0.24

Notes: We analyze the differential effect of our main explanatory variable for individuals living in landlocked countries. Row one shows the uninteracted effect of log(1+Distance to Harbor), i.e. the distance effect of individuals in coastal countries, row two shows the differential effect for being landlocked. Row three depicts the combined effect of the two constituent terms, i.e. the effect of log(1+Distance to Harbor) for individuals living in landlocked countries together with the corresponding p-value in brackets. Results in each column come from separate regressions and are estimated using the sample of coastal and landlocked sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in this round. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table A.15: Extension: Generations (Young) Interaction

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.023***	-0.014***	0.069***	0.031***	-0.039***	-0.021***
	(0.002)	(0.003)	(0.004)	(0.006)	(0.001)	(0.002)
Interaction:						
log(1+Distance to Harbor) × Young {0,1}	0.008***	0.010***	0.007*	0.008**	0.003*	0.004***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
Isolated Effect of the Distance to Harbor for Individuals below median Age (33)						
Combined Effect:						
log(1+Distance to Harbor) + *Interaction*	-0.015***	-0.004	0.075***	0.039***	-0.036***	-0.017***
	[0.00]	[0.11]	[0.00]	[0.00]	[0.00]	[0.00]
Basic Controls	NO	YES	NO	YES	NO	YES
Urbanization Controls	NO	YES	NO	YES	NO	YES
Trade-related Controls	NO	YES	NO	YES	NO	YES
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	122,555	122,238	127,305	126,982	103,083	102,990
R-Squared	0.10	0.15	0.17	0.20	0.16	0.22

Notes: We analyze the differential effect of our main explanatory variable for individuals above and below the median age within the sample. Row one shows the uninteracted effect of log(1+Distance to Harbor), i.e. the distance effect of individuals at and above the median age (33), row two shows the differential effect for being in the younger strata. Row three depicts the combined effect of the two constituent terms, i.e. the effect of log(1+Distance to Harbor) for individuals younger than the median age (33) together with the corresponding p-value in brackets. Results in each column come from separate regressions and are estimated using the sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in this round. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table A.16: Extension: "Remote Urbanities"

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.023***	-0.025***	0.026***	0.039***	-0.027***	-0.029***
	(0.003)	(0.003)	(0.007)	(0.008)	(0.002)	(0.003)
Interaction:						
log(1+Distance to Harbor) × Urban {0,1}	0.016***	0.022***	0.010	-0.004	0.010***	0.014***
	(0.003)	(0.003)	(0.007)	(0.008)	(0.002)	(0.002)
Isolated Effect of the Distance to Harbor for Individuals living in Urban Environments						
Combined Effect:						
log(1+Distance to Harbor) + *Interaction*	-0.007***	-0.003	0.036***	0.034***	-0.016***	-0.015***
	[0.00]	[0.23]	[0.00]	[0.00]	[0.00]	[0.00]
Basic Controls	NO	YES	NO	YES	NO	YES
Urbanization Controls	NO	YES	NO	YES	NO	YES
Trade-related Controls	NO	YES	NO	YES	NO	YES
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	123,584	122,238	128,397	126,982	103,889	102,990
R-Squared	0.10	0.15	0.19	0.20	0.19	0.22

Notes: We analyze the differential effect of our main explanatory variable for individuals living in urban environments. Row one shows the uninterested effect of log(1+Distance to Harbor), i.e. the distance effect of individuals in rural settings, row two shows the differential effect for urban sample participants. Row three depicts the combined effect of the two constituent terms, i.e. the effect of log(1+Distance to Harbor) for individuals living in urban environments together with the corresponding p-value in brackets. Results in each column come from separate regressions and are estimated using the sample of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in this round. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table A.17: Extension: Commercial Farmers

	Dependent Variable					
	Cash Employment {0,1}		Freq. gone without: (Water / Food / Cash Inc. / Med.) [0,4]		Possessions: (Radio / TV / Motor Vehicle) [0,1]	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+Distance to Harbor)	-0.015***	-0.011**	0.083***	0.045***	-0.032***	-0.023***
	(0.004)	(0.006)	(0.008)	(0.012)	(0.003)	(0.006)
Interaction:						
log(1+Distance to Harbor) × Commercial Farmer {0,1}	-0.047**	-0.047**	0.063	0.088**	0.005	-0.001
	(0.022)	(0.022)	(0.041)	(0.040)	(0.018)	(0.018)
Isolated Effect of the Distance to Harbor for Individuals working as Commercial Farmers						
Combined Effect:						
log(1+Distance to Harbor) + *Interaction*	-0.062***	-0.058***	0.146***	0.134***	-0.026***	-0.025***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Basic Controls	NO	YES	NO	YES	NO	YES
Urbanization Controls	NO	YES	NO	YES	NO	YES
Trade-related Controls	NO	YES	NO	YES	NO	YES
Full Geographic Controls	NO	YES	NO	YES	NO	YES
Country-Time FE	YES	YES	YES	YES	YES	YES
Observations	33,084	32,296	37,481	36,637	13,239	13,080
R-Squared	0.06	0.10	0.11	0.15	0.12	0.22

Notes: We analyze the differential effect of our main explanatory variable for individuals working as commercial farmers, i.e. farmers who produce mainly for sale. Row one shows the uninteracted effect of log(1+Distance to Harbor), i.e. the distance effect of individuals not working as commercial farmers, row two shows the differential effect for commercial farmers. Row three depicts the combined effect of the two constituent terms, i.e. the effect of log(1+Distance to Harbor) for individuals working as commercial farmers together with the corresponding p-value in brackets. Results in each column come from separate regressions and are estimated using the sample of coastal, sub-Saharan African countries included in survey rounds 1, 2, and 3 of the Afrobarometer for which this detailed occupational data is available. Remaining changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used in columns (5) and (6) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items were not asked in this round. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table A.18: Mediation Analysis: Direct and Indirect Links of Distance to Harbor

	Dependent Variable								
	Cash Employment {0,1}			How often gone without: (Water / Food / Cash Inc. / Med.) [0,4]			Possessions: (Radio / TV / Motor Vehicle) [0,1]		
	Mediator: Institutions {1,4}	Mediator: REC {0,3}	Mediator: AU {0,3}	Mediator: Institutions {1,4}	Mediator: REC {0,3}	Mediator: AU {0,3}	Mediator: Institutions {1,4}	Mediator: REC {0,3}	Mediator: AU {0,3}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Baseline (Total) Effect:									
log(1+Distance toHarbor)	-0.009***	-0.008**	-0.016***	0.035***	0.037***	0.038***	-0.019***	-0.020***	-0.020***
	(0.002)	(0.004)	(0.004)	(0.006)	(0.010)	(0.009)	(0.002)	(0.004)	(0.003)
Direct Effect:									
log(1+Distance toHarbor)	-0.009***	-0.009**	-0.016***	0.041***	0.038***	0.040***	-0.019***	-0.021***	-0.021***
	(0.002)	(0.004)	(0.004)	(0.006)	(0.010)	(0.009)	(0.002)	(0.004)	(0.003)
Indirect Effect:									
Distance to Harbor via Mediator	0.000	0.000***	0.000***	-0.005***	-0.001***	-0.002***	0.000**	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Direct Effect of Mediator:									
(Institutions / REC / AU)	0.001	0.011***	0.010***	-0.216***	-0.037***	-0.039***	-0.007**	0.016***	0.014***
	(0.004)	(0.003)	(0.002)	(0.008)	(0.006)	(0.005)	(0.003)	(0.002)	(0.002)
Proportion Mediated	[0.00]	[0.05]	[0.03]	[0.15]	[0.04]	[0.05]	[0.01]	[0.04]	[0.04]
Basic Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Urbanization Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Trade-related Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Full Geographic Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country-Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	121,897	35,216	47,063	126,602	35,380	47,265	102,677	27,769	39,882

Notes: This table presents results from a formal mediation analysis testing the influence of three potential mediators of log(1+Distance to Harbor) on our three main outcome variables, respectively. Row one presents the baseline effect, its counterpart is depictable in Table 1. Row two shows the direct effect of our main explanatory variable, i.e. the effect of distance not attributable to the mediating factor, while row 3 depicts the part of the effect which runs precisely via its influence on the mediator. Row four shows the direct effect of the mediator on the respective outcome variable. Results in each column come from a separate SEM regression. Changes in the number of observations across columns stem from differences in the response rates of dependent/independent variables. The sample used is comprised of coastal, sub-Saharan African countries included in survey rounds 1 through 7 of the Afrobarometer. The sample used in columns (7), (8) and (9) do not include individuals surveyed in rounds 1 and 2 of the Afrobarometer, as questions on ownership of household items (infrastructure) were not asked in this round. Similarly, the sample in columns (2), (5), (8) only include individuals surveyed in rounds 2, 4, and 6 and the sample in columns (3), (6), (9) only include individuals surveyed in rounds 2, 4, 5 and 6. Binary dependent variables are estimated through a simple LPM (Linear Probability Model) specification. The standard errors reported are clustered at the survey-enumeration area level. * p < .1, ** p < .05, *** p < .01.

Chapter 3

Heterogeneous Effects of Women's Schooling on Fertility, Literacy and Work: Evidence from Burundi's Free Primary Schooling Policy[1]

Abstract. This article investigates the effect of women's schooling on fertility as well as on associated mechanisms by leveraging Burundi's free primary education policy (FPE) of 2005 as a natural experiment. Exogenous variation in schooling is identified through a fuzzy regression discontinuity design (RDD). Our results show that educational attainment was positively influenced by Burundi's FPE for women situated at all wealth levels. However, the relevant downstream effects of schooling – measured by fertility, literacy, and work outcomes – reveal heterogeneous treatment effects which are moderated by women's household wealth. While poor women profit in terms of increases in literacy (6.7 percentage-point increase for each year of policy-induced schooling), remunerated employment opportunities (5.7 percentage-point increase), as well as a reduction in desired and actual fertility outcomes (6.9 percentage-point reduction in teenage childbirth), none of these effects of additional education are observed for women from the wealthier households of our sample. The evidence of such a marked heterogeneity contributes to the growing literature examining the nexus between education and fertility in developing countries and helps to evaluate under which conditions the literature's findings may generalize.

JEL Classification: I25, I26, J13, O15

Keywords: Female Education, Fertility, sub-Saharan Africa, Regression Discontinuity Design

---

[1] This Chapter is joint work with David Stadelmann. A modified version has been published under the identical title in the *Journal of African Economies*, 2024, 33 (1), 67-91. DOI: 10.1093/jae/ejad002. The version presented here expands on the article to include further evidence on extensions and robustness.
Acknowledgements. Funding from the DFG (project number 390713894) is gratefully acknowledged. We thank two anonymous referees and the handling editor Brian Dillon for their helpful remarks. I also thank the participants at the Centre for the Study of African Economies Conference 2023 (Oxford, UK), at the 2020 German Economic Association (VFS) Annual Meeting (Cologne, Germany), and at the Biannual Graduate Seminar of the faculty of economics at the University of Bayreuth in 2020 for their insightful comments and remarks at various stages of the project. All remaining errors are my own.

3.1 Introduction

> "The increase in the education of women and girls contributes to greater empowerment of women, to a postponement of the age of marriage and to a reduction in the size of families."
>
> (United Nations 1995; 76)

Teenage childbearing is a prevailing problem in many developing countries. Particularly in sub-Saharan Africa, where 10% of women between the ages of 15 and 19 have already given birth to at least one child (United Nations 2022). Some of the negative consequences associated with teenage childbearing are adverse infant and child health, increased maternal mortality, as well as lower human capital, and less favorable labor market outcomes (see Levine et al. 2009). Educating women is regarded as an effective policy for tackling many of the developmental challenges connected to gender equality at large, as prominently outlined in the Millennium Development Goals (United Nations 2001), and particularly for reducing women's teenage pregnancies and the negative consequences linked to them (Lloyd and Young 2009; World Bank 2017).

Indeed, early empirical literature has consistently shown a negative association between schooling and fertility (e.g. Cochrane 1979; Martin 1995; Ainsworth et al. 1996). However, given that early childbirth or continued schooling is often jointly determined (e.g. Martínez and Odhiambo 2018), causal interpretations of these findings are challenging. A growing body of research therefore explores the education-fertility nexus by employing quasi-experimental methods (for an overview see Psaki et al. 2019; Evans and Mendez Acosta 2021). In sub-Saharan Africa, this literature regularly exploits the enactment of *free primary education* policies (hereafter FPE) as credibly exogenous shocks to schooling, such as in Ethiopia (Behrman 2015b; Moussa and Omoeva 2020; Chicoine 2021), Ghana (Adu-Boahen and Yamauchi 2018), Malawi (Behrman 2015a, 2015b; Grant 2015; Makate and Makate 2016; Moussa and Omoeva 2020; Zenebe Gebre 2020), Nigeria (Osili and Long 2008), and Uganda (Behrman 2015a, 2015b; Makate and Makate 2018; Masuda and Yamauchi 2020; Moussa and Omoeva 2020). These more recent findings identify a negative causal effect of education on teenage fertility in sub-Saharan African countries. A small set of these studies also started to explore potential differential effects of FPE on schooling and, importantly, the outcomes influenced by schooling. Where tested, the results provide evidence that FPE policies almost exclusively influenced the educational outcomes of the poorer groups in the population (Grépin and Bharadwaj 2015; Makate and Makate 2016; Keats 2018). However, the absence of statistically significant increases in FPE-induced education for the wealthy precludes a comparison of econometrically reliable estimates on downstream effects such as fertility, literacy, or work between poor and wealthy subgroups.[2] Still, decisions on schooling and fertility are dependent on women's socio-economic background and, correspondingly, their time-use, particularly in developing contexts (Ilahi 2000; Levine et al. 2009). Thus, identifying and investigating potential heterogeneities along women's household wealth, especially in the downstream effects of female education, represents a relevant research gap.

We contribute to the literature in two ways. *First*, we add on to the existing evidence by presenting a statistically robust negative effect of education on women's teenage fertility and by providing support for commonly associated mechanisms. *Second*, we demonstrate that the estimated downstream effects of policy-induced schooling are distinctly heterogeneous. That is, there is a marked differential treatment effect governed by women's household wealth. We thereby extend previous findings and provide novel (credibly causal) evidence on heterogeneities materializing in the downstream factors influenced by education. Methodologically, we proceed similarly to previous studies and consider the implementation of Burundi's free primary education (FPE) policy in 2005 as a quasi-experiment, which to our knowledge, has not been utilized to explore the education-fertility nexus before. We employ a fuzzy regression discontinuity design (RDD) to identify exogenous increases in schooling and analyze the downstream effects of schooling through an instrumental variable (IV) approach. To assess women's full teenage fertility, we follow the literature and measure education as well as outcomes influenced by education several years after the policy was implemented. As such, we use data from Burundi's Demographic and Health Surveys (DHS) conducted in 2010/2011 and 2016/2017.

Our results show that Burundi's FPE policy triggered large and statistically significant increases in education, both for poor and for wealthy women. Being young enough to benefit from FPE (i.e. being 13 years and younger at the time of implementation) increased girls' education by 1.22 years compared to women too old to benefit from the policy (14 years and older). This corresponds to a 34% increase compared to the mean years of schooling for untreated cohorts. Given that education was strongly increased for both poor and wealthy women, we are able to investigate this particularly interesting source of heterogeneity in the outcomes influenced by education, i.e. employ instrumental variables (IV) regressions for poor and wealthy subgroups separately. These IV-estimates provide evidence that policy-induced education mainly affects the poor: One additional year of schooling reduces their desired number of children by 6.4% relative to untreated cohorts and lowers the likelihood of having their first child during teenage years by 6.9 percentage-points. Estimates of these outcomes for women from wealthy households are considerably lower, at 2.2% and 1.2 percentage-points, respectively, and importantly, statistically indistinguishable from zero.

[2] In settings where increases in schooling were apparent for both poor and wealthy subgroups (see Adu-Boahen and Yamauchi 2018; Makate and Makate 2018), results are not indicative of a heterogeneity in the downstream factors influenced by schooling.

Regarding the mechanisms through which education is suggested to influence fertility, literature has largely proposed two main pathways, often categorized by *income* and *learning* (Michael 1973; McCrary and Royer 2011). The *income* mechanism rests on the seminal work of Gary Becker (1960, 1965) and Jacob Mincer (1962, 1963) and emphasizes the increasing opportunity costs of childbirth and rearing in response to an increase in women's schooling. Given the time intensity of these activities, comparative statics have suggested substitution away from a higher quantity of children towards fewer, more qualitative children, i.e. children that acquire more human capital (see Becker and Lewis 1973; Willis 1973; Lam and Duryea 1999). Concerning the second mechanism *learning*, education may influence fertility through the dissemination of new information as well as the overall expansion of the capacity to access and absorb information. As such, education may directly exert an effect on women's reproductive behavior through curricula (Bankole et al. 2007), both by raising the overall stock of health knowledge and by facilitating its effective usage (Grossman 1972; Rosenzweig and Schultz 1989). Education may also influence fertility decisions and health knowledge by providing women with the skills necessary to attain knowledge about family planning outside the premises of the school, for instance, through the usage of mass media outlets (Thomas et al. 1991; Glewwe 1999). Moreover, schools are regarded as an important venue for the socialization of students, which reinforces aspects of the *learning* channel and may affect both channels alike through "assortative mating" (Caldwell 1976, 1980; Mare 1991; Bongaarts and Watkins 1996; Behrman and Rosenzweig 2002). Besides these two distinct mechanisms, it has been argued that schooling may reduce fertility simply by keeping women in school and thereby shortening the available time for sexual activity, a so-called "incarceration effect" (Black et al. 2008; Berthelon and Kruger 2011). Much of the more recent literature explores these suggested pathways by testing mediating factors such as women's employment and remuneration on the one hand, as well as literacy, access and usage of mass media outlets, knowledge, and implementation of family planning, sexual activity and indicators of assortative mating (partner's characteristics) on the other. So far, the evidence is mixed regarding the empirical support for each of these mechanisms (see Psaki et al. 2019).

The findings we present support both *learning* and *income* pathways alike: one additional year of schooling raises women's literacy rates by 6.7 percentage-points and increases the likelihood of working outside of the household for cash by 5.7 percentage-points. Given that both of these effects are apparent only for the poor women in our sample, these results agree with the differential fertility responses we observe, substantiating the potential significance of these two channels.

We conduct an array of robustness checks to validate our results and the heterogeneous nature of them. We thereby test for potentially confounding differences between treated and untreated groups going back the time of intervention, assess the influence of differing conflict intensities of Burundi's Civil War, and check the robustness of our results to placebo cut-offs, differing bandwidths and functional forms (see details in Section 3.5). Given that our employed wealth score measures household wealth at the time of the survey, we cannot preclude that FPE may have induced women to increase household wealth levels via added schooling which represents a potential endogeneity issue concerning our differential treatment effects. When directly exploring whether women's wealth score as well as other direct measures of income and living standards are affected by education in our IV setting, we find no evidence for this type of endogeneity. Given that over half of the women in our treated sample are still considered dependents, it is not surprising that significant wealth changes have not (yet) materialized through FPE-induced schooling. We also investigate data from the Multiple Indicator Cluster Surveys (MICS) to test differences in outcomes of poor as well as wealthy treated and untreated groups directly before and after Burundi's free primary education policy in a difference-in-differences framework. All of our results and interpretations remain robust, particularly their heterogeneous nature.

The remainder of this paper is structured as follows: We introduce the background to our identification and the empirical strategy in Section 3.2. The data employed is introduced in 3.3. Section 3.4 presents the empirical results and a discussion thereof. Section 3.5 provides our robustness and validity checks. Section 3.6 offers concluding remarks.

3.2 Institutional Background and Identification Strategy

3.2.1. Institutional Background

The World Declaration On Education For All in 1990 and the Dakar Framework For Action in 2000 manifested the promise of developing countries to achieve universal primary education (UPE) by the year 2015. Prompted by these commitments, many countries in sub-Saharan Africa enacted policy changes toward free primary education at the turn of the century, mainly by focusing on the elimination of user fees (Kattan 2006). User fees have been identified as a crucial obstacle to school attendance in developing regions, particularly for the poor (Deininger 2003; Grogan 2008; Bold et al. 2013; Travaglianti 2017; Burlando and Bbaale 2022). Subsequently, African countries have made tremendous strides in increasing educational enrolment due to these interventions, improving the net enrolment rate (NER) from 53% to 79% between 1990 and 2018 (UIS 2022). However, many of them are still facing obstacles in moving towards universal primary education: Large within-country differences in educational attainment prevail and are outlined by factors such as individuals' wealth, place of residence, ethnicity, and gender (Levine et al. 2009; World Bank 2017).

Burundi is one of the poorest countries in the world, with an income per capita of $793 and over 65% of the population living below the national poverty line (World Bank 2021). The country is densely populated, and most of its 11.2 million inhabitants (87%) live in rural areas (Dunlop and King 2019). Burundi's annual fertility rate and population growth stand at 5%

and 2.7%, respectively, well above the sub-Saharan African average of 4.5% and 2.5% (United Nations 2022). Since its independence in 1962, Burundi has undergone several periods of violence. The most noteworthy conflict was the genocidal massacre of 1993 which triggered the subsequent civil war lasting, at least formally, until the year 2000, in which the Arusha Peace and Reconciliation Agreement (APRA) was signed (Bundervoet et al. 2009; Travaglianti 2017). Access to education was part of the struggle citizens fought for in the war and a core element of APRA as well as the constitution adopted in 2005, when the direct fees of primary schooling were removed (Travaglianti 2017; Dunlop and King 2019).[3] The waiver of school fees was accompanied by a surge in the construction of schools and an overall increase in budgetary spending on education, especially at the primary level, which rose from 37% in 2001 to 54% in 2008 of the total educational budget (Travaglianti 2017). As a result, primary school enrolment increased from 0.97 million in 2004 to 1.32 million in 2006, expanding the gross enrolment ratio (GER) from 76% to 103% in just two years (UIS 2022). School participation continuously increased thereafter, arriving at a maximum GER of 141% in 2011. Note that a shift of children from wealthier families into private schools was not observed. At around 1.5% in the years before the policy, the proportion of primary school children in private institutions has historically been low and actually decreased after the introduction of FPE (UIS 2022).

3.2.2. Identification Strategy

To establish credibly causal effects of education, we draw upon the recent body of research and consider Burundi's free primary education policy (FPE) of 2005 as a natural experiment. We identify exogenous (policy-induced) variation in schooling by employing a fuzzy regression discontinuity design (RDD), exploiting the fact that FPE exogenously sorts women into treatment or control, depending on women's age at the time of implementation. To the best of our knowledge, Burundi's FPE of 2005 has not been used to study the education-fertility nexus before. In Burundi, the full primary school cycle entails six years of schooling to be attained during the official primary school ages of 7 to 12 (Ministere e la Presidence charge de la Bonne Gouvernance et du Plan - MPBGP et al. 2017). However, overage enrolment is common. Data from the 2005 Multiple Indicator Cluster Survey (MICS) shows a mean age of 8.2 for girls entering first grade at the time of policy implementation (UNICEF 2005), which differs somewhat between girls living in wealthier households (7.78 years) and girls from poorer ones (8.46 years). Following the literature, we account for this de facto age of entry into primary school in our estimations and consider a primary school leaving age of 13 rather than 12 (see also Adu-Boahen and Yamauchi 2018; Keats 2018). As such, given that fees were removed from the school year starting in September of 2005, women who were born in 1992 or later (aged 13

[3] While the direct fees of schooling were removed, other costs of schooling such as uniforms, materials and contributions to the school's budget continued to persist, as was the case in many other countries implementing FPE. Still, removing the direct schooling fees had a "psychological effect" (Travaglianti 2017; 112), deeming school overall more affordable than before.

or younger at policy implementation), are considered young enough to benefit from free schooling and are therefore sorted into the treatment group. Accordingly, women born in 1991 or earlier (aged 14 and older) were too old to benefit from the policy and sorted into the control group. Hence, the RDD design in our setting compares women who were "just-treated" to women who were just too old to benefit from FPE (see e.g. Behrman 2015a, 2015b; Makate and Makate 2018; Keats 2018; Moussa and Omoeva 2020). Next to a late entry into school, factors such as absenteeism and grade repetition are common in sub-Saharan Africa. In 2006, there were approximately 24% overaged children enrolled in Burundi's primary school grades (UIS 2022). It is thereby plausible that women older than the de facto primary school age were exposed to the policy change, by enrolling in lower primary school grades at later ages or by continuing with their remaining primary school years (Behrman 2015b). On the other hand, some girls younger than the de facto primary school ages have already left schooling behind permanently, be it for marriage, for work, or similar duties, and will thereby not benefit from the removal of tuition fees. Given such noncompliance, the estimation will be made through a fuzzy regression discontinuity design.

Note that the effect estimated in a fuzzy RDD resembles an instrumental variable estimate: It is defined as the ratio of the reduced form estimate of the discontinuity in the outcome variable Y_i to the estimate of the discontinuity in the treatment, $Schooling_i$, both determined by the assignment variable B_i (year of birth) and evaluated at the cut-off c (Hahn et al. 2001; Imbens and Lemieux 2008; Lee and Lemieux 2010). The estimation of our downstream effects of schooling therefore consists of two stages: The first-stage, regressing the treatment $Schooling_i$ on the identifying instrument Z_i, whereby $Schooling_i$ is a continuous measure for years of schooling and Z_i is a binary indicator $Z_i = 1[Year\ of\ Birth \geq c(1992)]$. The second-stage, regressing the outcome variable Y_i on the instrumented endogenous variable, $\widehat{Schooling_i}$, estimated in the first stage.

The regression discontinuity literature favors a non-parametric, or local, estimation with lower-order polynomials over a global specification with higher-order polynomials (Gelman and Imbens 2019). Thus, an appropriate window of data (hereafter bandwidth) has to be selected for both sides of the cut-off. We choose the bandwidth that minimizes the modified "leave-one-out" cross-validation function (CV) introduced by Ludwig and Miller (2005) and countercheck these results with the Imbens and Kalyanaraman (2012) proposed bandwidth selection procedure. The value of the CV function (MSE) stabilizes at a bandwidth size of five birth year cohorts on either side of the cut-off.[4] Thus, all our estimates are based on a bandwidth size of five, effectively comparing women born in 1987-1991 to women born in 1992-1996 (aged 9-13 and 14-18 at the time of the policy). Both our first and second-stage results are robust to varying bandwidth sizes (see Section 3.5).

[4] The CV confirms the bandwidth size of five within the subsamples of poor and wealthy women as well.

To fit an appropriate functional form to the bandwidth selected, we test six different specifications of the regression equation. As such, we estimate linear, quadratic, and cubic specifications, each with and without an interaction term between the treatment indicator Z_i and the assignment variable B_i (allowing the slope to vary before and after the cut-off). We select a linear functional form with a varying slope before and after the cut-off as our baseline specification, given that it minimizes the Akaike information criterion (AIC) among the tested specifications. Both our first- and second-stage results are robust to the choice of the functional form (see Section 3.5). The main results presented in this paper are estimated through the following two-stage regression equation:

$$Schooling_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 (B_i - 1992) + \alpha_3 Z_i \times (B_i - 1992) + \alpha_k X_i + e_i \tag{3.1}$$

$$Y_i = \beta_0 + \beta_{1RD} \widehat{Schooling_i} + \beta_2 (B_i - 1992) + \beta_3 Z_i \times (B_i - 1992) + \beta_k X_i + u_i \tag{3.2}$$

All variables are defined as above. X_k is an additional vector of control variables and e_i, u_i are the idiosyncratic error terms. β_{1RD} is the causal effect of schooling and represents the local average treatment effect (LATE) on compliers (Angrist et al. 1996; Lee and Lemieux 2010). As is common in the literature employing RD designs, the assignment variable B_i is centered at the cut-off point $c = 1992$ such that the constituent terms of the interaction, namely α_1 and β_{1RD} , can be directly interpreted as the effect at the discontinuity. Binary outcome variables are estimated through a linear-probability model (e.g. Adu-Boahen and Yamauchi 2018; Keats 2018; Ozier 2018).[5]

3.3 Data and Descriptive Statistics

The data we use come from the individual recodes of the 2010/2011 and the 2016/2017 Burundi Demographic and Health Surveys (DHS). DHS are cross-sectional, household-based surveys collecting a broad array of information on topics such as demographics, education, employment and occupation, as well as fertility and family planning (Croft et al. 2018). The main respondents are women of reproductive age (15-49).

As a result of the empirical specification (bandwidth), the sample is restricted to women born five years before and after the cut-off year 1992, resulting in our main sample of 7,714 women aged 20-30 at the time of the survey (born between 1987-1996). We exclude women younger than 20 years of age to assess women's full teenage fertility and to analyze outcomes for women who have mostly completed schooling (see Ferré 2009; Keats 2018).[6] We thereby measure outcomes of women several years after the policy was conducted. Following the

[5] Our estimates on binary outcome variables are robust to the use of a *Probit* model.
[6] The DHS Household recodes show that only 24% of 20-year-old women are still attending school. Of them, 88% are enrolled in secondary school.

recommendation by the DHS, we weight all analyses with the included survey weights, making the outputs representative at the national level, and account for the pooling of datasets by normalizing the sum of weights across surveys. Table 3.1 reports summary statistics for the main sample as well as for the poor and wealthy subgroups separately.

To explore heterogeneities, we employ the DHS wealth index score, which is constructed by assessing households' dwelling and living arrangements, i.e. construction materials, sanitation facilities, as well as household durables and assets such as radio and TV, bicycles, cars, etc., and by placing them on a relative scale of wealth within the sample (Rutstein and Johnson 2004). We consider respondents from households scoring on and below the median wealth score as "poor", and correspondingly, respondents with a denoted household wealth level above the median wealth score as "wealthy". Summary Statistics (Table 3.1) reveal that the poor subgroup is significantly deprived in terms of access to basic utilities and services: Less than 1% of poor households have access to electricity or a flush toilet, only 17% have drinking water piped into their dwelling, and almost all of the households in the poorer subgroup (99%) are dependent on wood as a fuel source. The increased effort in the provision of many of these basic goods can figure largely in household members' time-use, particularly in females' on whom these tasks are often levied (Ilahi 2000; Levine et al. 2009). Wealthy households are endowed much more favorably in these basic goods and utilities, which is why an investigation into the differential effects of women's schooling along household wealth can be highly informative. Note that the DHS wealth index captures respondents' wealth at the time of the survey. We discuss and address potential caveats pertaining to this wealth grouping, alternatives, and robustness checks in Section 3.5.

3.3.1. Dependent Variables

To assess adolescent fertility outcomes, we employ women's reported age at first birth and construct a dichotomous indicator of having had a first birth before the age of 20. Similarly, we investigate the incidence of teenage marriage by estimating the probability of having been married before the age of 20.

To explore the pathways through which education is suggested to influence fertility, we first investigate respondents' levels of literacy as well as their desired number of children. These variables serve as proxies for investigating a potential *learning* pathway. Note that additional years of education must not necessarily lead to improved learning outcomes, especially given that the abolition of school fees in Africa has been associated with the overcrowding of schools and compromised learning environments for children (Deininger 2003; Lucas and Mbiti 2012; Bold et al. 2013). The measure on women's literacy thereby also serves as a direct assessment of the quality and effectiveness of FPE-induced schooling. To investigate a potential *income* pathway, we test whether women have work other than domestic housework, e.g. are working in the family business/farm, are selling items, are (self)-employed, etc., currently or anytime

Table 3.1: Summary Statistics

	Sample					
	Full Sample		Poor		Wealthy	
	Control	Treatment	Control	Treatment	Control	Treatment
Number of Observations	4476	3238	2302	1556	2174	1682
Age at Survey	24.19	21.96	24.16	21.93	24.23	21.99
	[3.05]	[1.46]	[3.03]	[1.47]	[3.08]	[1.45]
Education (completed schooling years)	3.55	5.60	1.99	3.98	5.56	7.45
	[4.18]	[4.30]	[2.79]	[3.60]	[4.77]	[4.29]
Never received schooling {0,1}	0.44	0.23	0.57	0.32	0.26	0.13
Residence: Rural {0,1}	0.87	0.85	0.99	0.99	0.73	0.69
Number of Siblings	6.20	5.91	6.18	5.95	6.22	5.86
	[2.47]	[2.42]	[2.45]	[2.33]	[2.51]	[2.53]
Religion						
Catholic {0,1}	0.59	0.55	0.62	0.61	0.54	0.49
Protestant {0,1}	0.34	0.36	0.31	0.31	0.38	0.42
Muslim {0,1}	0.03	0.04	0.01	0.02	0.05	0.05
Wealth Quintile (1-5)	3.01	3.09	1.94	1.97	4.38	4.38
Dwelling						
Electricity {0,1}	0.10	0.12	0.00	0.00	0.22	0.26
Wood Fuel {0,1}	0.85	0.81	0.99	0.98	0.66	0.61
Piped Water {0,1}	0.32	0.37	0.17	0.17	0.53	0.60
Flush Toilet {0,1}	0.04	0.06	0.00	0.00	0.10	0.13
Fertility						
Age at First Birth	19.67	18.98	19.59	18.99	19.80	18.97
	[2.55]	[2.06]	[2.43]	[1.98]	[2.72]	[2.20]
Has given Birth before Age 20 {0,1}	0.34	0.32	0.37	0.39	0.31	0.25
Reproductive Behavior						
Age First Marriage	18.78	18.17	18.60	18.03	19.07	18.42
	[2.72]	[2.19]	[2.52]	[2.09]	[2.98]	[2.34]
Married before Age 20 {0,1}	0.48	0.39	0.54	0.49	0.41	0.28
Age at First Sex	13.63	11.55	14.28	12.88	12.85	10.04
	[8.62]	[8.74]	[8.08]	[8.21]	[9.16]	[9.07]
Has had Sex before Age 20 {0,1}	0.55	0.51	0.59	0.58	0.50	0.42
Pathway: Learning						
Literacy: Able to read sentence {0,1}	0.56	0.71	0.43	0.60	0.73	0.84
Desired Number of Children	4.01	3.67	4.05	3.63	3.96	3.72
	[1.42]	[1.33]	[1.43]	[1.30]	[1.41]	[1.37]
Pathway: Income						
Worked Last Year {0,1}	0.86	0.81	0.92	0.90	0.78	0.71
Works outside of the HH. for Cash/Money {0,1}	0.08	0.09	0.08	0.10	0.07	0.08

Notes: The sample consists of the 2010/2011 and 2016/2017 Burundi DHS Female Surveys, using a bandwidth size of five birth year cohorts on either side of the cut-off. Respondents are restricted to age 20 and older at the time of the survey. The "Control" group is comprised of individuals born between 1987-1991, the "Treatment" group is comprised of individuals born between 1992-1996. The "poor" and "wealthy" subsamples include women whose households score on and below as well as above the median wealth level, respectively. The statistics are weighted using the DHS sample weights and account for the pooling of survey rounds.

within the last year. Additionally, we specifically test whether respondents' work is employed work and if it is remunerated, i.e. whether they are working outside of the household for cash/money. We investigate additional factors associated with the proposed mechanisms. This includes respondents' reproductive behavior, measured by women's age at first sexual intercourse as well as their knowledge and usage of contraceptive methods (condoms). We also investigate women's engagement with mass media outlets such as newspapers-, radio- and tv-usage and explore further aspects of assortative mating, i.e. test husbands' characteristics such as their age, education or fertility preferences.

3.3.2. Independent Variables

Our main explanatory variable of interest is women's years of schooling. DHS surveys report the number of completed schooling years, which is a continuous measurement of educational attainment. The variable is constructed by asking respondents: "What is the highest level of school you attended?" and subsequently, "what is the highest (standard/form/year) you completed at that level?" (The DHS Program 2015; 2).[7]

We include a set of control variables that were plausibly fixed before the FPE policy implementation. We thereby add dummy variables capturing women's religious affiliation, as it may influence women's behavior through doctrines and values stipulating certain roles of women within the family (Makate and Makate 2018). We also add a control variable indicating the respondent's number of siblings. A larger number of siblings could negatively influence girls in their ability to receive schooling (Ewemooje et al. 2020). It could also have a non-negligible impact on girls' own desired family size. To account for regional discrepancies in schooling as well as fertility patterns, we include province fixed effects for Burundi's eighteen provinces[8] and also add a dummy variable indicating whether the respondent resides in rural or urban dwellings. Lastly, we add survey fixed effects to account for differences in women's educational attainment or fertility outcomes, which might vary systematically between the two periods of the survey rounds.

Given the usual difficulty in choosing between a sufficient number of controls and overfitting, we also investigate the robustness of our first- and second-stage results to the Post-Double-Selection (PDS) algorithm developed in Belloni et al. (2014). We thereby include further potential control variables from the DHS as well as other sources and allow for higher order terms as well as the interaction among them. In total, we consider 896 controls. Our results are robust to PDS' selected controls, in fact, our coefficients are more precisely estimated, if anything. Results are provided in the Appendix, Table B.1 through Table B.4.

[7] For example, if the respondent's highest level of schooling is secondary education and she completed 2 years at that level, the variable is constructed by adding all primary school grades (6 in Burundi) + 2 years of secondary school = 8 years of total schooling.
[8] We use the geo-coded sample clusters of the 2010/2011 survey round to allocate households to Burundi's 18 provinces to "match" the province information provided in the 2016/2017 survey round.

3.4 Results

3.4.1. The Effect of FPE on Schooling: First-Stage RDD Evidence

Figure 3.1 and Figure 3.2 provide visual evidence that Burundi's FPE affected women's educational attainment. They depict long-term trends in education for women by birth year using the pooled sample (Figure 3.1) and the sub-samples split by household wealth (Figure 3.2).

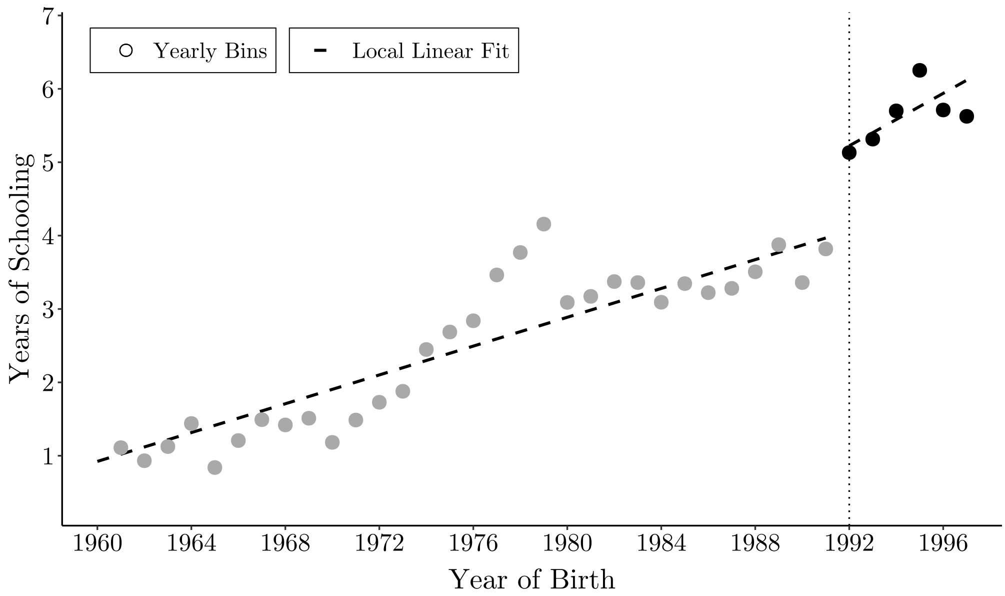


Figure 3.1: Women's Schooling by Birth Year Cohort

Notes: The plot depicts yearly averages of highest schooling grades achieved by individuals born within the respective birth cohort indicated on the x-axis. The "Local Linear Fit" fits the slope estimated from a regression of y on x using the (local) data window indicated by the different color hue. Data come from the 2010/2011 and the 2016/2017 Burundi DHS Female Surveys. Respondents are restricted to age 20 and older. All estimates are weighted using the DHS sample survey weights.

Figure 3.1 illustrates that women who benefited from free primary schooling, i.e. women born in or after 1992, show for an immediate and substantial increase in educational outcomes compared to the cohorts born prior. And given that girls born in later years had increasingly more (primary) schooling years left to attend without having to pay fees, educational attainment is increasing in birth year cohorts after the cut-off.

Figure 3.2 suggests that both poor and wealthy women were able to benefit from the introduction of FPE. However, the observation of increasing educational attainment for women born after the cut-off is dissimilar and can only be observed for the poor subgroup. Taking account of the very low levels of schooling of poor women born before the cut-off, it seems likely that mainly girls from currently poor households were constrained by the direct cost of schooling prior to FPE. A reason for the larger initial discontinuity, but non-increasing trend line, for the wealthy women after the introduction of the policy might be that FPE precisely induced the women from wealthy households who had achieved some primary education but were constrained to complete the full six-year cycle, to finish primary schooling. Note that we also observe a steep incline in educational attainment for cohorts born two decades earlier (in the 1970s), followed by a sharp drop-off for women born in and after 1980. The increase, as

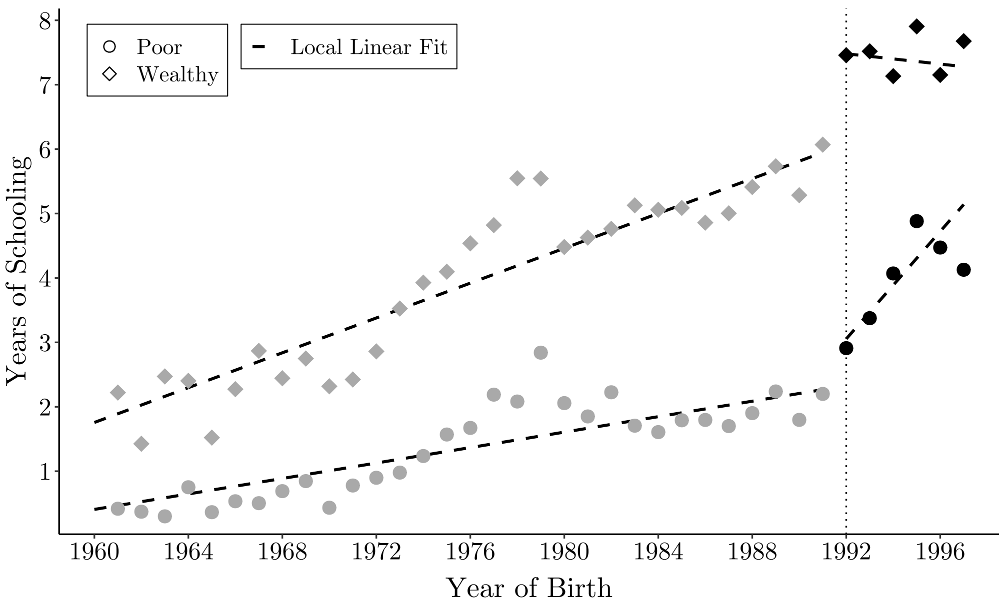


Figure 3.2: Women's Schooling by Birth Year Cohort, separated by Household Wealth

Notes: The plot depicts yearly averages of highest schooling grades achieved by individuals born within the respective birth cohort indicated on the x-axis. The "Local Linear Fit" fits the slope estimated from a regression of y on x using the (local) data window indicated by the different color hue. Estimates for the "poor" and "wealthy" are generated from subsamples split at the median household wealth. Data come from the 2010/2011 and the 2016/2017 Burundi DHS Female Surveys. Respondents are restricted to age 20 and older. All estimates are weighted using the DHS sample survey weights.

well as the decline in schooling for these cohorts, has been attributed to the expansion of the primary education sector in the 1980s, the successes of which were quickly eradicated for cohorts subject to the violent conflict during Burundi's Civil War (Obura 2008; Verwimp and Van Bavel 2014). Subsequent birth cohorts' educational attainment slowly recovered to pre-war levels (Travaglianti 2017) and ultimately surpassed them in one instance when tuition fees were removed in 2005. Given our identification strategy, these pre-policy developments do not affect our results as we draw on data from birth cohorts born in 1987 to 1992 for our estimation (see Subsection 3.2.2, and in particular, Section 3.5 for robustness and validity tests).

Table 3.2 presents first-stage discontinuity as specified in regression equation (3.1). The results are presented for the full sample (columns 1 and 2), as well as the poor (columns 3 and 4) and wealthy subsamples (columns 5 and 6), separately.[9] Odd columns show results of a parsimonious specification, only including survey fixed effects. For the discussion of results, we refer to the more stringent specification, including all controls and fixed effects presented in the even columns. The discontinuous increase in schooling generated by Burundi's FPE policy is presented in row one. The coefficient estimate represents the increase in the years of education

Table 3.2: First-Stage Estimates, Years of Schooling

	Regression Sample					
	Full Sample		Poor		Wealthy	
	(1)	(2)	(3)	(4)	(5)	(6)
1[Year of Birth ≥ 1992]	1.344***	1.223***	0.849***	0.908***	1.356***	1.280***
	(0.255)	(0.223)	(0.228)	(0.220)	(0.397)	(0.367)
Mean of Control Group	[3.55]		[1.99]		[5.56]	
Interaction: 1 × Year of Birth	0.135	0.164**	0.378***	0.339***	-0.179	-0.099
	(0.090)	(0.077)	(0.081)	(0.078)	(0.141)	(0.130)
Year of Birth	0.052	0.038	0.059	0.064	0.116	0.070
	(0.054)	(0.047)	(0.046)	(0.044)	(0.090)	(0.085)
Effect on the second cohort treated by FPE						
1[YoB ≥ 1992] + 1 × YoB + YoB		1.426***		1.312***		1.250***
		(0.197)		(0.187)		(0.338)
Observations	7,713	7,709	3,857	3,856	3,856	3,853
Adjusted R-squared	0.054	0.263	0.099	0.183	0.041	0.187
Controls	NO	YES	NO	YES	NO	YES
Province Fixed Effects	NO	YES	NO	YES	NO	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: The results in each column are produced by a separate regression. The sample used for the estimations consists of the 2010/2011 and 2016/2017 Burundi's DHS Female Surveys, using a bandwidth size of five birth year cohorts on either side of the cutoff. Respondents are restricted to age 20 and older at the time of the survey. Columns (1) and (2) are estimated from our full sample. Columns (3) and (4) are estimated using the subsample of poor women (below and including the median wealth score), and columns (5) and (6) are estimated using the subsample of wealthy women (scoring above the median wealth score). Odd columns include survey fixed effects. Even columns include region (province) fixed effects as well as controls on religious affiliation, number of siblings and urban/rural status. The variable "Year of Birth" has been re-centered such that the coefficient in row one can be directly interpreted as the discontinuous difference at the cutoff year 1992. "Mean of Control Group" indicates the mean of the dependent variable for women in untreated birth year cohorts (the control group). Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

[9] The regression results may be visually inspected in Figure B.1 through Figure B.3 in the Appendix (without controls and fixed effects).

for women who were 13 years old (born in 1992) at the time of the policy compared to women who were 14 years old (born in 1991). The estimate is economically large and statistically significant at the 1% level throughout all columns. Being just young enough to benefit from free schooling increased women's educational attainment between 0.908 (specification 4) and 1.280 years (specification 6). In comparison to related findings, girls "just-treated" by tuition-free primary schooling gained about 0.61 years of schooling in Ethiopia (Behrman 2015b) between 0.31 to 0.46 years in Malawi (Behrman 2015a, 2015b) and between 0.64 to 1.24 years in Uganda (Keats 2018; Makate and Makate 2018). Our slightly larger estimate may be explained by the low levels of education prior to the FPE. Interestingly, Burundian women from the wealthy subgroup had a larger increase in schooling at the cut-off than the poor. However, as Figure 3.2 shows, poor women born after 1992 continuously increased their educational attainment, which is captured by the large positive effect of the interaction term "1[Year of Birth ≥ 1992] × Year of Birth". Hence, when comparing the effect of the second treated cohort after the cut-off (born in 1993), the poor already show for larger increases in education compared to the wealthy, (0.908+0.339+0.064=) 1.312 years to (1.280-0.099+0.070=) 1.250 years, respectively. Note that the mean schooling years of the control cohorts differ largely between the wealthy and the poor. Poor women not exposed to the policy had an average of 1.99 years of schooling, compared to 5.56 years for the wealthy. As such, while FPE added about the same number of schooling years for poor and rich women, it more than doubled the amount of schooling for the poor in relative terms. See Subsection 3.4.3 for a more thorough discussion of this observation. The coefficient estimates are robust to the inclusion of all of our controls and fixed effects, as well as controls selected by the PDS algorithm (see Table B.1 in the Appendix).

3.4.2. The Effects of Schooling: Second-Stage Evidence

The first-stage estimates provide evidence for substantial policy-induced increases in schooling. Thus, they present strong instruments for estimating the effect of additional education on our dependent variables in an IV setting. Given that both poor and wealthy women exhibit large and significant discontinuities in schooling, we have the opportunity to additionally test for differential treatment effects of added education for poor and wealthy women separately.

Literacy and Desired Fertility. Table 3.3 panel A investigates the effect of increased schooling due to the FPE policy on women's literacy. The first-stage discontinuity delivers a strong instrument as shown by the F-Statistic (29.9). Results show that one additional year of schooling increased women's literacy by 5.0 percentage-points, which marks an increase of 8.8% compared to the mean of the control group. This is in line with findings from related studies which consistently show positive impacts of education on women's literacy, e.g. Keats (2018) or Behrman (2015b). However, there is relevant heterogeneity in our estimated effect of schooling as shown in columns (3) through (6). Poor women increased their rate of literacy through policy-induced schooling by 15.6% compared to (a statistically insignificant) increase of 3.7% by wealthy women. Importantly, note that untreated cohorts of wealthy women had higher levels of literacy overall, 73% compared to 43% for the poor.

These results suggest that for the poor, free access to schooling was successful in providing them with basic skills, presumably acquired at lower levels of schooling. It also hints towards the effectiveness of FPE-induced schooling in general and goes against some of the suggestions of depreciating school quality due to issues such as overcrowding (e.g. Deininger 2003; Lucas and Mbiti 2012; Bold, Kimenyi, and Sandefur 2013). The results are also consistent with the aggregate evidence that switching into private schools was not observed in Burundi in the years following the policy implementation (UIS 2022).

Table 3.3: Second-Stage Results, Literacy and Desired Fertility

	Regression Sample					
	Full Sample		Poor		Wealthy	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Literacy {0,1}						
Schooling	0.054***	0.050***	0.067**	0.067**	0.033	0.027
Instrumented by 1[YoB ≥ 1992]	(0.015)	(0.017)	(0.033)	(0.030)	(0.022)	(0.024)
Mean of Control Group	[0.56]		[0.43]		[0.73]	
F-Statistic	27.7	29.9	13.8	17.1	11.4	11.9
Observations	7,708	7,704	3,856	3,855	3,852	3,849
Panel B: Desired Number of Children						
Schooling	-0.144**	-0.151**	-0.307**	-0.261**	-0.093	-0.088
Instrumented by 1[YoB ≥ 1992]	(0.060)	(0.063)	(0.136)	(0.118)	(0.086)	(0.086)
Mean of Control Group	[4.01]		[4.05]		[3.96]	
F-Statistic	25.9	28.3	12.5	15.8	11.4	11.9
Observations	7,520	7,516	3,764	3,763	3,756	3,753
Controls	NO	YES	NO	YES	NO	YES
Province Fixed Effects	NO	YES	NO	YES	NO	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: The results in each panel and column are produced by a separate regression. The IV estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Burundi has high rates of fertility, with an average of 5 births per woman (United Nations 2022). Panel B of Table 3.3 shows that a large portion of these high rates of fertility is reflected in the number of children a woman desires to bear throughout her lifetime, with an average of

4 children per woman in our full sample. Our second-stage estimates suggest that education negatively influences women's desired lifetime fertility, as shown in panel B of Table 3.3. Our results are comparable to recent evidence from other countries: Behrman (2015b), Keats (2018), and Masuda and Yamauchi (Masuda and Yamauchi 2020) report decreases in the number of desired children between -0.11 and -0.34, which are well in line with our estimated coefficient of -0.151. However, different from the extant literature, our results show that the effect is driven mainly by women from the lower socio-economic strata of the population (columns 3 to 6). Compared to women in the control group, treated women decreased their fertility preference by 3.7%, the poorer subgroup by 6.4%, and the wealthy by (a statistically insignificant) 2.2%. An explanation for this is that education exerts its negative impact on women's desired fertility through increased levels of literacy, which were raised only for the poor. This finding is supportive evidence of the pathway *learning* since literacy acts as a proxy for girls' learning (capacity). Overall, these results suggest an education-driven convergence of fertility preference between the poor and the wealthy.

Fertility and Reproductive Behavior. Panel A and B of Table 3.4 investigate the effect of education on women's fertility outcomes. A one-year increase in schooling postpones women's age at first birth by almost half a year (panel A, column 2) and reduces the likelihood of teenage childbearing by 3.4 percentage-points (panel B, column 2).[10] Once more, the effect of added education seems to be driven by women from the poorer half of the population, which is shown when comparing the estimates of column (4) with those in column (6) both in panel A and in panel B, exploring women's age at first birth and the probability of having had a first birth during teenage years.[11] One additional year of education for the poor decreases the probability of having had a first birth before the age of 20 by 6.9 percentage-points, compared to a statistically insignificant 1.2 percentage-points for women from wealthier households. Marriage and cohabitation are arguably one of the most direct channels through which teenage childbearing is promoted (Presler-Marshall and Jones 2012). Our results show that one year of added schooling decreases the likelihood of being married before the age of 20 by 6.1 percentage-points, which corresponds to a 12.7% decrease in comparison to the control group (panel C, column 2). As such, the evidence from Burundi corresponds to that of studies on other developing countries assessing the link between education and fertility, which consistently report significant reductions in the likelihood of teenage marriage alongside decreased teenage childbearing (e.g. Keats 2018; Masuda and Yamauchi 2020).

[10] We further test the probability of having had a first birth at additional age increments (from 16 to 23) in Table B.5 in the Appendix and show that the effect of decreased teenage childbirth was apparent at the ages 19 and 20. This might be interpreted as evidence against an "incarceration effect", given that women were most likely out of school at these age ranges.

[11] Note that the IV-estimate for wealthy women in panel A of Table 3.4 suffers from weak instrument bias, with an F-Statistic of 2.1 (Staiger and Stock 1997). This is due to a sample reduction given that not all women in our sample, especially the wealthy, have born children at the time of the survey. Estimates from panel B are arguably a more reliable indicator of altered fertility behavior via increased schooling.

Table 3.4: Second-Stage Results, Fertility and Reproductive Behavior

	Regression Sample					
	Full		Poor		Wealthy	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Age at first Birth						
Schooling	0.440**	0.468**	0.435	0.449*	0.640*	0.757
Instrumented by 1[YoB ≥ 1992]	(0.179)	(0.204)	(0.274)	(0.263)	(0.381)	(0.564)
Mean Ctrl. Group	[19.67]		[19.59]		[19.80]	
F-Statistic	14.9	11.9	9.1	10.3	3.9	2.1
Observations	4,918	4,917	2,830	2,830	2,088	2,087
Panel B: First Birth before Age 20 {0,1}						
Schooling	-0.037**	-0.034*	-0.056	-0.069*	-0.026	-0.012
Instrumented by 1[YoB ≥ 1992]	(0.018)	(0.019)	(0.041)	(0.039)	(0.024)	(0.025)
Mean of Control Group	[0.34]		[0.37]		[0.31]	
F-Statistic	27.9	30.0	13.8	17.0	11.7	12.2
Observations	7,713	7,709	3,857	3,856	3,856	3,853
Panel C: Married before Age 20 {0,1}						
Schooling	-0.063***	-0.061***	-0.078**	-0.090**	-0.059**	-0.045*
Instrumented by 1[YoB ≥ 1992]	(0.018)	(0.018)	(0.039)	(0.037)	(0.025)	(0.025)
Mean of Control Group	[0.48]		[0.54]		[0.41]	
F-Statistic	27.9	30.0	13.8	17.0	11.7	12.2
Observations	7,713	7,709	3,857	3,856	3,856	3,853
Panel D: Sex before Age 20 {0,1}						
Schooling	-0.044**	-0.042**	-0.047	-0.064*	-0.046*	-0.031
Instrumented by 1[YoB ≥ 1992]	(0.018)	(0.019)	(0.037)	(0.035)	(0.028)	(0.029)
Mean of Control Group	[0.55]		[0.59]		[0.50]	
F-Statistic	28.5	29.9	14.8	17.9	11.2	11.4
Observations	7,623	7,619	3,807	3,806	3,816	3,813
Controls	NO	YES	NO	YES	NO	YES
Province Fixed Effects	NO	YES	NO	YES	NO	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: The results in each panel and column are produced by a separate regression. The IV estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB)≥1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

As before, we explore the effects at further age increments in Table B.6 showing that the negative effect of schooling on marriage is only significant at one age increment. Importantly, the occurrence of teenage marriage was decreased for both the poor (panel C, column 4) and wealthy women (panel C, column 6), by 16.7% and 11%, respectively. Given that wealthy women delayed their first marriage but not their first birth, our results suggest that a reduction in teenage marriage may not automatically lead to lower levels of (desired) fertility.

To further explore the effect of education on reproductive behavior, we investigate women's age at first sexual intercourse. Panel D shows that an additional year of education makes women 7.6% less likely to report having had their first reported sexual intercourse before the age of 20. Exploring estimates for all age increments shows that sexual exposure was reduced also before the age of 18 (Table B.7). Given that all of our fertility-indicating variables (including sexual activity, birth, and marriage) were significantly reduced in later teenage years, i.e. years in which women were most likely already out of school, reduced fertility through "incarceration" seems unlikely. Once more, the effects of increased education are heterogeneous and only statistically significant for poorer women. Unaltered sexual exposure for the wealthy could explain the gap between delayed marriage but unreduced fertility described above.

Income. Recent research reports inconclusive evidence concerning the support for an *income* mechanism (Ali and Gurmu 2018; Makate and Makate 2018; Masuda and Yamauchi 2020; Moussa and Omoeva 2020; Zenebe Gebre 2020). Only Chicoine (2021), Grépin and Bharadwaj (2015), and Keats (2018) and find increases in (employed) work opportunities, a shift to higher-skilled employment (contrary to agricultural work or similar), and higher remuneration.

Table 3.5 shows that policy-induced education did not increase the likelihood of having any type of work or occupation different than domestic housework (panel A). However, added education did seem to help women specifically into remunerated employment opportunities, that is, significantly more women are now working outside of the household for cash/money (panel B). The now common heterogeneity regarding poor and rich women emerges again: each year of policy-induced schooling increased poor women's remunerated employment opportunities by 5.7 percentage-points, compared to a statistically insignificant 0.8 percentage-point change for the wealthy. These findings lend support to an *income* pathway: More educated (poor) women show for increased opportunity costs of childbearing, given by an increase in remunerated employment opportunities. Higher opportunity costs are reflected in (poor) women's updated fertility preferences, which are adjusted accordingly (downwards).

Learning. The reported increase in women's literacy and the accompanying decrease in their desired number of children emphasize the potential significance of women's *learning* on subsequent fertility outcomes. This link is especially relevant for poorer women. To further explore the workings of this channel, we first test women's knowledge and usage of family planning (condoms) in Table B.8 panel A, and panel B in the Appendix. While we do not find an effect of increased awareness of modern contraceptives (panel A)[12], education increases the likelihood of having used a condom with the last sexual partner (panel B). Again, this effect is

[12] General knowledge on condoms as a contraceptive method is already high, 94% for untreated cohorts, such as any potential effect must be small.

Table 3.5: Second-Stage Results, Income

	Regression Sample					
	Full Sample		Poor		Wealthy	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Worked last Year {0,1}						
Schooling	-0.003	-0.002	0.017	0.005	0.000	-0.002
Instrumented by 1[YoB ≥ 1992]	(0.015)	(0.013)	(0.024)	(0.020)	(0.025)	(0.022)
Mean of Control Group	[0.86]		[0.92]		[0.78]	
F-Statistic	27.9	30.0	13.8	17.0	11.7	12.2
Observations	7,713	7,709	3,857	3,856	3,856	3,853
Panel B: Works outside of the Household for Cash/Money {0,1}						
Schooling	0.023**	0.026**	0.058*	0.057**	0.007	0.008
Instrumented by 1[YoB ≥ 1992]	(0.011)	(0.013)	(0.031)	(0.027)	(0.014)	(0.015)
Mean of Control Group	[0.08]		[0.08]		[0.07]	
F-Statistic	27.9	30.0	13.8	17.0	11.7	12.2
Observations	7,713	7,709	3,857	3,856	3,856	3,853
Controls	NO	YES	NO	YES	NO	YES
Province Fixed Effects	NO	YES	NO	YES	NO	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: The results in each panel and column are produced by a separate regression. The IV estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

likely driven by poor women. To assess the potential channels through which knowledge on modern family norms or contraceptive methods is dispersed, we test for women's engagement with mass media outlets such as newspapers, radio, and television in Table B.9. We do not find any robust indication of that more education also increased exposure to information via these sources. Inasmuch as assortative mating, and thereby altered husband-wife relationships mediate fertility outcomes, we test for the effect of education on husband's age, education as well as the difference in husbands' and wives' desired fertility. We find no statistically significant effects for such augmentation to take place (Table B.10).

3.4.3. Further Evidence on the Heterogeneity

A salient feature distinguishing poor and wealthy subgroups within our main sample is their educational exposure altogether and, correspondingly, the levels of schooling at which FPE-induced educational gains were realized. This subsection explores this difference more profoundly. Specifically, we investigate the grade levels at which the gains of education were realized through FPE for our full sample, as well as for the poor and wealthy subgroups, separately. Given that our IV-estimates resemble local average treatment effects (LATE), this exercise informs us about the sub-population of "compliers" driving our results (see Acemoglu and Angrist 2001; Angrist and Pischke 2009). Figure 3.3 depicts the shifts in grade levels achieved due to FPE visually, plotting the difference in the conditional probability (on the y-axis) of having completed at least a given school grade (on the x-axis) for women that were just young enough to benefit from the policy change (13 years and younger), compared to women who too old to benefit from the removal of fees (14 years and older). The differences

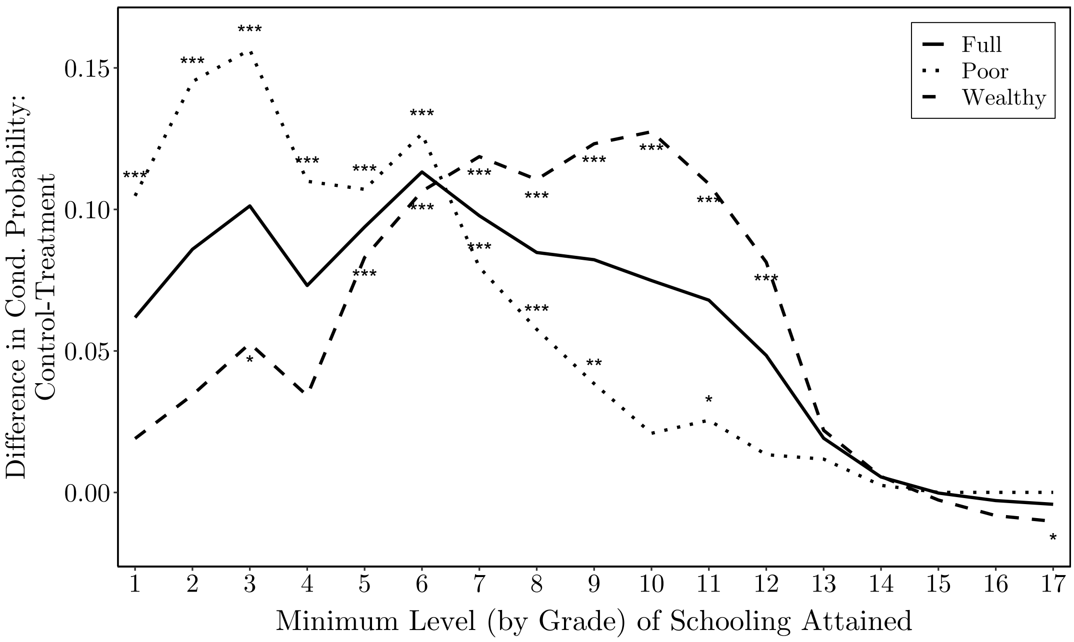


Figure 3.3: FPE-Induced Shifts in Grade Levels

Notes: The plot depicts differences in the conditional probability (on the y-axis) of having completed at least a given school grade (on the x-axis) for women that were just young enough to benefit from the policy change (13 years and younger), compared to women who too old to benefit from the removal of fees (14 years and older). The differences depicted are conditional CDF changes, estimated from regressions of having at least the respective school level on our treatment indicator and our usual covariates. Note that we do not report significance levels for the full sample, as they are significant at the 1% level (10% level) up until year 12 (year 13). Data come from the 2010/2011 and the 2016/2017 Burundi DHS Female Surveys. Respondents are restricted to age 20 and older. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds.

depicted are conditional CDF changes, estimated from regressions including our usual covariates and including survey weights.[13] Note that we do not report significance levels for the full sample, as they are significant at the 1% level (10% level) up until year 12 (year 13).

[13] Note that the changes of the three different samples are normalized by their respective first-stages as outlined in Angrist and Pischke (2009), giving us the contribution (weight) of the respective schooling level change towards the average causal response over all educational levels.

The figure shows that for the poor, schooling increases induced by FPE were highest at the (early) primary school level, increases for the wealthy were highest (and significant) at later primary- to secondary schooling years: Being born in or after 1992 increased poor women's probability of having attained at least some primary school years (between 1-3 years) between 10.8 and 16.3 percentage-points, contrary to (statistically insignificant) increases between 2 and 5.3 percentage-points for the wealthy for these earlier school years. Increases for wealthier women catch up to the level of poor women's increases while moving rightwards along the x-axis and exceed them at secondary school grades. This indicates that removing primary school fees induced wealthier women to also transition into secondary school (e.g. see also Keats 2018).

Overall, the exercise shows that treatment groups across the poor and wealthy subsamples differ systematically in their levels of added schooling induced by their exposure to FPE. Burundian women from the wealthier strata attained higher grade levels independent of the policy. For poor women, however, FPE mainly increased the probability to obtain lower grade levels of schooling, which proves sufficient to induce behavioral changes.[14] Reconciling these insights with our main findings suggests that next to women's household wealth, schooling attained at lower grade levels may matter more in influencing women's outcomes regarding literacy, fertility and work than later schooling years, at least judging by our specific setting for women exposed to FPE in Burundi.[15]

3.5 Robustness Checks and Validity Tests

The assumptions behind our IV-estimates require that women born (just) before, and women born (just) after the cut-off do not systematically vary in characteristics influencing their outcomes concerning fertility, literacy and work other than by their differential exposure to schooling. We included control variables as well as fixed effects to address this issue in the first instance and check these findings against a control selection algorithm (see PDS results in Table B.1 through Table B.4 of the Appendix). However, given that we do not have information on women at the time of policy implementation, there is a possibility of (further) unobserved differences affecting our results.

We investigate the robustness and validity of our results in four distinct ways. (a) We assimilate women's circumstances at the time of the intervention using Burundi's MICS survey from 2005 and investigate the continuous nature of several influencing characteristics of treatment and control cohorts, thereby measuring the potential influence of confounding factors. (b) We re-assess our baseline results, specifically their heterogeneous nature of them,

[14] This goes against findings prior to the year 2000 in which a handful of developing countries (including Burundi) show a curvilinear relationship between schooling and fertility, meaning that the early years of education actually increased women's levels of childbearing . Note, however, that these findings were attributed to a country's very (early) status along the fertility transition for one (Martin 1995) and secondly, to education's differential impact on mediating factors such as contraceptive usage and breastfeeding (Jain 1981).
[15] Our IV-estimates represent the local average treatment effect (LATE), which is specific to the setting and sample at hand (Angrist et al. 1996).

again by moving closer to the intervention period, using the MICS surveys of 2000 as well as 2005 in a difference-in-differences framework. (c) We further validate our results by assessing the impact of Burundi's Civil War as a potential source of bias. (d) We analyze the robustness of our estimates to varying bandwidths as well as functional forms and check our results against placebo cut-offs.

Accounting for potential Confounders. DHS surveys do not include retrospective information on respondents. To nonetheless explore the potentially discontinuous nature of important covariates at the time of policy implementation, we utilize the Burundi Multiple Indicator Cluster Survey (MICS) conducted in 2005 and identify household members born in our treatment and control cohorts 1992-1996 and 1987-1991, respectively (UNICEF 2005).[16] We test for the assumption of unconfoundedness by estimating equation (3.1), replacing women's schooling with various (household) characteristics of these earlier sampled individuals as dependent variables. Although not identical to the respondents in the main sample, individuals born within these years and surveyed in 2005 represent plausible proxy-respondents corresponding to our treatment and control cohorts situated at the time of policy implementation. To assess the potential influence of parents' socio-economic status and their child investments, we also test for discontinuities in characteristics of "potential mothers" (see Keats 2018) by identifying women in the 2005 MICS who have had births between the years 1987-1996. All of the results are presented in Table B.11 in the Appendix.

The estimates presented show statistically insignificant differences such that there are no apparent discontinuities between proxy treatment and proxy control cohorts at the time of policy implementation. It, therefore, seems unlikely that there were systematic differences influencing the eventual outcomes of sample respondents today, which are conflated with their differential access to schooling. Our results hold when dropping observations to include only female household members. They also hold when testing for a differential discontinuous effect among poor and wealthy subgroups separately.

Wealth Grouping. A potential concern regarding the grouping of women by wealth in our analysis is that women increased their (household) wealth due to policy-induced schooling. To directly test for this possibility, we replace our outcome variable in equation (3.2) with women's DHS wealth score, as well as several other indicators of wealth such as household's access to electricity, piped water, a flush toilet and several consumer durables such as a tv, radio and bicycle. The results are presented in Table B.12 and show no robust evidence for a direct (causal) effect of FPE-induced schooling on women's household wealth. Moreover, more than half of the women in our treated sample are still considered dependents, i.e. are neither the head of the household (8%) nor the wife to the head (40%). Thus, household wealth is unlikely to be affected by them and may be considered as given. Also note that if women's additional

[16] There are no DHS surveys available in Burundi between 1987 and 2010.

schooling did quickly allow them to move into the wealthier subgroup of the population, our estimated coefficients would be biased downwards, representing conservative estimates of the true effect of schooling on the poor.

We additionally test for potentially differing exposure as well as treatment between poor and wealthy subgroups, again by moving closer to the time of policy intervention. We thereby pool the cross-sectional Multiple Indicator Cluster Surveys (MICS) of 2000 and 2005 together with the 2010/2011 DHS Survey and conduct a difference-in-difference-in-differences (triple-difference) analysis effectively analyzing the difference of being wealthy or poor between treated (policy-eligible) and untreated women before and after the policy (adopted from Keats 2018). Specifically, we compare outcomes for women born in 1992-1996 and 1987-1991, i.e. aged 15-19 and 20-24 in the 2010/2011 Burundi DHS Female Survey (first difference). The difference in outcomes between these two groups is then adjusted by a secular time-trend, given by the difference between these two age cohorts before policy implementation, i.e. in 2000 and 2005 using the MICS Female Surveys (second difference). Comparing these differences both for women from poor and from wealthy households (third difference) leads us to estimate the following triple-difference specification:

$$
\begin{aligned}
Y_{i,r,t} = {} & \beta_0 + \beta_1(\mathit{Treated}_i \times \mathit{Post}_t) + \beta_2(\mathit{Treated}_i \times \mathit{Post}_t \times \mathit{Poor}_i) + \beta_3(\mathit{Treated}_i \times \mathit{Poor}_i) \\
& + \beta_4(\mathit{Post}_t \times \mathit{Poor}_i) + \beta_5 \mathit{Treated}_i + \beta_6 \mathrm{Poor}_i + \beta_k X_i + \gamma_t + \delta_r + u_{i,r,t}
\end{aligned}
\tag{3.3}
$$

$\mathit{Treated}_i$ is a dummy for individuals aged 15-19, Post_t indicates respondents sampled in the survey period 2010/2011 and Poor_i switches on for women living in households below the median wealth level. γ_t and δ_r represent survey and region fixed effects, respectively. We also include controls for women's age as well as the type of residence (urban/rural) and cluster standard errors both at the survey-cluster and birth year level. Results for educational attainment, literacy, fertility choices (concerning their last-born child) as well as household wealth are depicted in Figure 3.4 which plots coefficient β_1, as well as the linear combination of $\beta_1 + \beta_2$ together with the corresponding 95% confidence intervals, allowing for a direct difference-in-differences interpretation and comparison for both wealthy and poor subgroups.

The results show that educational attainment (measured in schooling levels) was positively influenced for both wealthy and poor women, the poor showing higher absolute increases in schooling, making it broadly consistent with our main results. The rates of literacy were influenced analogously, and women report wanting to have their last-born child later in life, the result only significant for the poor. Importantly, household wealth outcomes of treated individuals were not significantly influenced by FPE. Figure B.4 in the Appendix provides further evidence. In line with our main results, we do not find evidence of altered marriage or fertility outcomes at these younger ages (15-19). Moreover, we perform a "placebo" analysis, where we estimate the same triple-difference analysis, albeit measuring these changes from 2000 (pre) to 2005 (post), a non-intervention period.[17] These results show statistically indistinguishable differences for all tested outcomes between the two groups over time, both for wealthy and poor women, as should be expected for a "placebo" test.

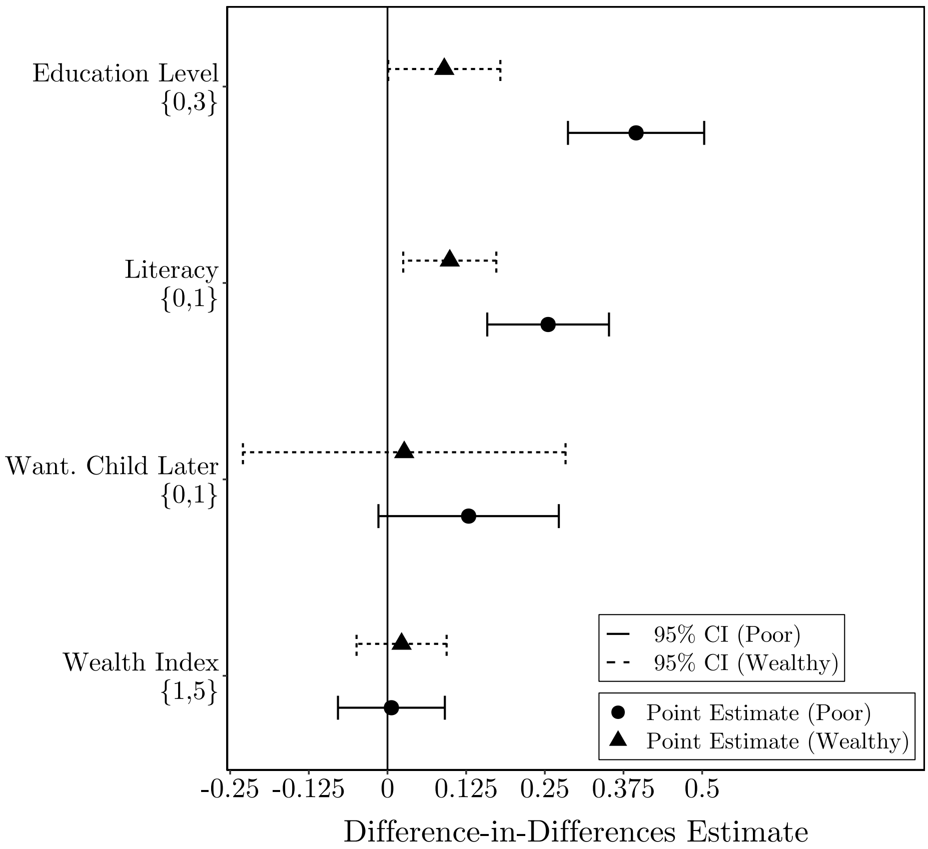


Figure 3.4: Difference-in-Difference-in-Differences

Notes: The "Point Estimates" represent difference-in-differences coefficients, including a 95% confidence interval, for the respective dependent variable indicated on the y-axis. Data come from the 2000 and 2005 Multiple Cluster Indicator Surveys (MICS) as well as the 2010/2011 Burundi DHS survey, using the women's recodes. The coefficients are estimated through "triple-difference" estimation (3.3), comparing individuals aged 15-19 to individuals aged 20-24 in 2000, 2005 and 2010 for both "wealthy" and for "poor" households (splitting at the median household wealth level). The estimates are weighted using MICS/DHS sample weights and account for the pooling of sample rounds. Standard errors are clustered at the birth-year and enumeration-area level.

As a further test for the robustness of our heterogeneous effects, in particular, their sensitivity to differing wealth groupings, we re-categorize individuals in our main analysis "poor" when living in households where the main flooring consists of sand, dung or dirt (see Keats 2018). Results are given in Table B.13, and again support our interpretations. Note that poor women tend to reside in rural areas (see Table 3.1), which is why we control for women's

[17] A question on fertility preference was not asked in the 2000 MICS, which is why there is no placebo-estimate on this outcome.

urban/rural status in all estimations. Importantly, our results also hold when conducting the (wealthy vs. poor) subsample analysis among rural regions only.[18]

Burundian Civil War. The massacres of 1993 and the subsequent civil war was the most violent period in Burundi's history, with an estimated total of 300,000 deaths (UNFPA 2002). Conflicts were most intense from 1993 to 1998, the years thereafter outlined by significantly lower conflict intensities (Mercier et al. 2020).

The massacres and the civil war in Burundi have been linked to negative effects on households' wealth, children's health, and children's likelihood of completing primary school (Bundervoet et al. 2009; Verwimp and Van Bavel 2014; Mercier et al. 2020).[19] Depending on the investigated outcome, these studies define individuals affected by the war born between 1981-1998, which includes both our treatment and our control cohorts born between 1987 and 1996 such that there is no differential effect to be expected of the war itself between individuals sorted into treatment and control. And although conflict intensities have varied, given the timeline and the spatial dispersal of conflict, effects are unlikely to systematically change for women born just after our chosen cut-off year in 1992. In other words, being 13 in 2005 should carry no special property for women but our identification in their differential access to schooling given by the timing of Burundi's FPE. Even if other peace-building policies by the newly formed government coincide with FPE's implementation year of 2005, it is unlikely for these measures to exert their effect in a sharp, discontinuous fashion starting with girls 13 years or younger only, i.e. such peace-building policies would affect all age groups from this point in time onward.

If there were conflict-induced changes affecting women's eventual fertility, literacy or work outcomes differently across treatment and control, it is sensible to think that the discontinuous nature of these outcomes should be apparent at "cut-off" years other than 1992, too. To investigate this possibility, we re-estimate equation (3.1), replacing women's schooling with all of our main outcome variables, and move our specified window size of five years around the corresponding placebo cut-off years starting from 1981 up until 1995. The results are plotted in Figure B.5 through Figure B.12. Across all outcome variables tested, there are no discontinuities apparent in other years but at our specified cut-off date, as we would expect if all effects were induced by FPE. As has also been shown in panel a) in Table B.11, there are no significant discontinuities in children's circumstances at the time of the policy implementation for our defined treatment cohort. Specifically, treated cohorts do not show a

[18] We are unable to conduct (poor vs. wealthy) subsample estimations in urban regions, as there are too few (poor) individuals living in urban environments altogether.

[19] Note that the negative effects on schooling were significantly smaller for women (Verwimp and Van Bavel 2014). Research on the impacts of similar civil wars, for example, in neighboring Rwanda has confirmed the negative impact on schooling (and the lesser negative impact on women) at least in the short- to medium term (Akresh and de Walque 2008; Bundervoet and Fransen 2018; La Mattina 2018) as well as negative effects on fertility outcomes (Kraehnert et al. 2019).

higher likelihood of having a living mother or a living father, which acts as a proxy for the war's intensity (Bundervoet 2009), nor having better maternal care or improved health inputs as a child.

Our findings are also robust to a varying bandwidth (see below), specifically to a smaller one. If conflict affected our treatment and control cohorts differently, estimates calculated from women born only a couple of birth years apart would likely cause biased effects to vanish. However, all of our effects remain significant at low bandwidth sizes, especially in the full sample. Hence, for our analysis of the impact of schooling induced by FPE, it is plausible to assume that treatment and control groups were similarly affected by the civil war, or other influencing factors following the war, like peace-building policies, such that we are confident in consistently identifying the isolated effect of education with our estimation strategy.

To nonetheless test for the potential influence of differing war exposure, we remove provinces with above median conflict intensities from our sample and re-estimate our main outcome variables (Table B.14). We use data from Bundervoet (2009), who calculates conflict intensity as the estimated percentage of a parent (mother or father) being killed in the massacres of 1993 using historical records from Chrétien and Mukuri (2002).[20] We subsequently drop provinces from our sample with above median conflict intensities, i.e. where above 6.6% of mothers or fathers are estimated to have been killed in the massacres. To compare, the mean conflict intensity across provinces is 9%, the upper quartile of provinces had over 16% of parental losses, and individuals in the highest affected province had an occurrence of over 22% parental deaths. Our interpretations remain largely unchanged and robust when removing these conflict-intense provinces.[21]

Varying Bandwidth and Functional Form. To investigate the sensibility of the estimated first-stage discontinuity as well as our second-stage results to a varying bandwidth, we re-estimate our main results using window sizes ranging from 2 to 10 birth year cohorts. Figure B.13 and Figure B.14 plot the estimated first-stage discontinuities for each bandwidth size together with the 95% confidence interval, each for the full sample, and separated along household wealth. We find significant increases in schooling from a bandwidth size of as little as two (three for the subgroup estimates) birth year cohorts on each side of the cut-off. The estimates stabilize from a bandwidth size of 5 onwards. Concerning the second-stage estimates, Figure B.15 through Figure B.22 provide evidence for a robust effect of education on fertility and associated factors regarding bandwidth size, and especially, concerning the heterogeneous nature of our main results. Estimates for the poorer subgroup are generally larger and more

[20] This may not fully account for provincial intensities of subsequent civil war years. Thus, we also explored the results by leaving out the most affected provinces of the first years of the civil war as defined by Bundervoet, Verwimp, and Akresh (2009). The results hold when excluding these regions.
[21] We do not report separate effects for poor and wealthy subsamples, given that the instrument strength for both subgroups decreased significantly caused by the stark drop of the sample size.

statistically significant at all data window sizes. This is especially true at smaller bandwidth sizes, which presumably, present less potential bias in our estimates given that longer-lived time trends are less apparent when comparing women born only a couple of years apart.

Next to the bandwidth choice, we test for sensitivity due to changing functional form as a further investigation into the robustness of our results. We adjust the functional form of equations (3.1) and (3.2) by adding higher-order polynomials of the running variable (year of birth), alternating the interaction of them with the treatment indicator (1[YOB≥1992]). We stick to our selected bandwidth size of five birth year cohorts on each side of the cut-off. The results are presented in Table B.15 and Table B.16 of the Appendix for first- and second-stage estimates, respectively. The estimate of the discontinuity remains significant throughout all common RDD specifications, apart from specification (f). The size of the significant, discontinuous effect varies from 1.170 to 1.498 years of education for the full sample. These first-stage estimates are carried over to the second-stage in Table B.16, providing an interesting insight into the heterogeneous nature of our results. Throughout the specifications, the main estimates remain broadly robust in size and in significance. With increasing terms added to the equation, the IV-estimate tends to drop in precision (note also the dwindling F-Statistic as can be inferred from the estimates in Table B.16), losing statistical significance for most estimates from specification (d) and onwards. It is noteworthy that estimates generated from the sample of poor women are more robust than the ones estimated from the wealthier half of the sample. Up to the quadratic specification (c), five of the seven usually significant coefficients remain (three of them at the 5% level) for the poor subgroup of women, whereas none of the estimates of the wealthy subgroup are significant.

3.6 Conclusion

This article investigates the effect of female education on fertility by exploiting Burundi's free primary education policy (FPE) of 2005 as a natural experiment. We identify exogenous increases in schooling through a fuzzy regression discontinuity design (RDD). Subsequently, we conduct instrumental variable estimations examining the causal effect of schooling on adolescent fertility outcomes as well as on commonly associated mechanisms.

Our results show that Burundi's FPE policy increased educational outcomes for women regardless of their socio-economic standing. Being young enough to benefit from free primary schooling (age 13 or younger) increased women's educational attainment by a substantial 1.22 years. Employing policy-induced variation in individual education as an instrument, we find that one year of additional education delayed women's first birth by half a year, and overall, decreased their probability of having their first childbirth before the age of 20. Investigating the mechanisms through which education is theorized to reduce women's fertility, we find supportive evidence for both *learning* and *income* pathways. Additional schooling led to higher literacy levels, decreased desired family sizes, and improved paid labor market opportunities.

Numerous robustness and sensitivity tests support our findings and interpretations.

Our estimates thereby confirm and complement the established negative causal effect of schooling on teenage fertility documented in the literature and provide further support for closely related mechanisms. What specifically distinguishes our results from the ones found in recent literature is the marked heterogeneity in the downstream effects of policy-induced schooling. While both poor and wealthy women gain additional years of schooling due to Burundi's FPE, the outcomes affected via policy-induced education are mainly driven by behavioral changes of the poor. Effects of additional schooling may therefore be expected to materialize differently depending on the socio-economic background of women, which complements the existing knowledge base. This insight holds relevant lessons when envisaging future educational interventions.

3.7 References

Acemoglu, Daron, and Joshua D. Angrist. 2001. ‘How Large Are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws’. In *NBER Macroeconomics Annual 2000*, edited by Ben Bernanke, Kenneth Rogoff, and National Bureau of Economic Research. Cambridge, Mass.: MIT Press.

Adu-Boahen, Emmanuel, and Chikako Yamauchi. 2018. ‘The Effect of Female Education on Adolescent Fertility and Early Marriage: Evidence from Free Compulsory Universal Basic Education in Ghana’. *Journal of African Economies* 27 (2): 227–48.

Ainsworth, Martha, Kathleen Beegle, and Andrew Nyamete. 1996. ‘The Impact of Women’s Schooling on Fertility and Contraceptive Use: A Study of Fourteen Sub-Saharan African Countries’. *The World Bank Economic Review* 10 (1): 85–122.

Akresh, Richard, and Damien de Walque. 2008. ‘Armed Conflict and Schooling: Evidence from the 1994 Rwandan Genocide’. *World Bank Policy Research Working Paper 4606*, 38.

Ali, Fatma Romeh M., and Shiferaw Gurmu. 2018. ‘The Impact of Female Education on Fertility: A Natural Experiment from Egypt’. *Review of Economics of the Household* 16 (3): 681–712.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. ‘Identification of Causal Effects Using Instrumental Variables’. *Journal of the American Statistical Association* 91 (434): 444–55.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press.

Bankole, Akinrinola, Fatima H. Ahmed, Stella Neema, Christine Ouedraogo, and Sidon Konyani. 2007. ‘Knowledge of Correct Condom Use and Consistency of Use among Adolescents in Four Countries in Sub-Saharan Africa’. *African Journal of Reproductive Health* 11 (3): 197–220.

Becker, Gary S. 1960. ‘An Economic Analysis of Fertility’. In *Demographic and Economic Change in Developed Countries*, edited by Universities-National Bureau, 209–40. New York: National Bureau of Economic Research, distributed by Columbia University Press.

———. 1965. ‘A Theory of the Allocation of Time’. *The Economic Journal* 75 (299): 493–517.

Becker, Gary S., and H. Gregg Lewis. 1973. ‘On the Interaction between the Quantity and Quality of Children’. *Journal of Political Economy* 81 (2, Part 2): S279–88.

Behrman, Jere R., and Mark R. Rosenzweig. 2002. ‘Does Increasing Women’s Schooling Raise the Schooling of the Next Generation?’ *American Economic Review* 92 (1): 323–34.

Behrman, Julia A. 2015a. ‘The Effect of Increased Primary Schooling on Adult Women’s HIV Status in Malawi and Uganda: Universal Primary Education as a Natural Experiment’. *Social Science & Medicine* 127 (February): 108–15.

———. 2015b. ‘Does Schooling Affect Women’s Desired Fertility? Evidence From Malawi, Uganda, and Ethiopia’. *Demography* 52 (3): 787–809.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. ‘Inference on Treatment Effects after Selection among High-Dimensional Controls’. *The Review of Economic Studies* 81 (2): 608–50.

Berthelon, Matias E., and Diana I. Kruger. 2011. ‘Risky Behavior among Youth: Incapacitation Effects of School on Adolescent Motherhood and Crime in Chile’. *Journal of Public Economics* 95 (1–2): 41–53.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2008. ‘Staying in the Classroom and out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births*’. *The Economic Journal* 118 (530): 1025–54.

Bold, Tessa, Mwangi S. Kimenyi, and Justin Sandefur. 2013. 'Public and Private Provision of Education in Kenya'. *Journal of African Economies* 22 (suppl 2): ii39–56.

Bongaarts, John, and Susan Cotts Watkins. 1996. 'Social Interactions and Contemporary Fertility Transitions'. *Population and Development Review* 22 (4): 639–82.

Bundervoet, Tom. 2009. 'Livestock, Land and Political Power: The 1993 Killings in Burundi'. *Journal of Peace Research* 46 (3): 357–76.

Bundervoet, Tom, and Sonja Fransen. 2018. 'The Educational Impact of Shocks in Utero: Evidence from Rwanda'. *Economics & Human Biology* 29: 88–101.

Bundervoet, Tom, Philip Verwimp, and Richard Akresh. 2009. 'Health and Civil War in Rural Burundi'. *Journal of Human Resources* 44 (2): 536–63.

Burlando, Alfredo, and Edward Bbaale. 2022. 'Fertility Responses to Schooling Costs: Evidence from Uganda's Universal Primary Education Policy'. *Economic Development and Cultural Change* 70 (3): 1017–39.

Caldwell, John C. 1976. 'Toward A Restatement of Demographic Transition Theory'. *Population and Development Review* 2 (3/4): 321–66.

———. 1980. 'Mass Education as a Determinant of the Timing of Fertility Decline'. *Population and Development Review* 6 (2): 225–55.

Chicoine, Luke. 2021. 'Free Primary Education, Fertility, and Women's Access to the Labor Market: Evidence from Ethiopia'. *The World Bank Economic Review* 35 (2): 480–98.

Chrétien, Jean-Pierre, and Melchior Mukuri, eds. 2002. *Burundi, La Fracture Identitaire: Logiques de Violence et Certitudes Ethniques, 1993-1996*. Hommes et Sociétés. Paris: Karthala.

Cochrane, Susan Hill. 1979. *Fertility and Education: What Do We Really Know?* World Bank Staff Occasional Papers 26. Baltimore: Published for the World Bank [by] Johns Hopkins University Press.

Croft, Trevor N., Aileen M.J. Marshall, and Courtney K. Allen. 2018. 'Guide to DHS Statistics'. Rockville, Maryland, USA: ICF International.

Deininger, Klaus. 2003. 'Does Cost of Schooling Affect Enrollment by the Poor? Universal Primary Education in Uganda'. *Economics of Education Review* 22 (3): 291–305.

Dunlop, Emily, and Elisabeth King. 2019. 'Education at the Intersection of Conflict and Peace: The Inclusion and Framing of Education Provisions in African Peace Agreements from 1975-2017'. *Compare: A Journal of Comparative and International Education* 0 (0): 1–21.

Evans, David K., and Amina Mendez Acosta. 2021. 'Education in Africa: What Are We Learning?' *Journal of African Economies* 30 (1): 13–54.

Ewemooje, Olusegun S., Elizabeth Biney, and Acheampong Yaw Amoateng. 2020. 'Determinants of Fertility Intentions among Women of Reproductive Age in South Africa: Evidence from the 2016 Demographic and Health Survey'. *Journal of Population Research* 37 (June): 265–89.

Ferré, Céline. 2009. 'Age At First Child: Does Education Delay Fertility Timing? The Case Of Kenya'. In *Policy Research Working Paper*. WPS 4833. The World Bank.

Gelman, Andrew, and Guido Imbens. 2019. 'Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs'. *Journal of Business & Economic Statistics* 37 (3): 447–56.

Glewwe, Paul. 1999. 'Why Does Mother's Schooling Raise Child Health in Developing Countries? Evidence from Morocco'. *The Journal of Human Resources* 34 (1): 124–59.

Grant, Monica J. 2015. 'The Demographic Promise of Expanded Female Education: Trends in the Age at First Birth in Malawi'. *Population and Development Review* 41 (3): 409–38.

Grépin, Karen A., and Prashant Bharadwaj. 2015. 'Maternal Education and Child Mortality in Zimbabwe'. *Journal of Health Economics* 44 (December): 97–117.

Grogan, Louise. 2008. ‘Universal Primary Education and School Entry in Uganda’. *Journal of African Economies* 18 (2): 183–211.
Grossman, Michael. 1972. ‘On the Concept of Health Capital and the Demand for Health’. *Journal of Political Economy* 80 (2): 223–55.
Hahn, Jinyong, Petra Todd, and Wilbert Klaauw. 2001. ‘Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design’. *Econometrica* 69 (1): 201–9.
Ilahi, Nadeem. 2000. ‘The Intra-Household Allocation of Time and Tasks: What Have We Learnt from the Empirical Literature?’ In *Policy Research Report on Gender and Development*, 48. Working Paper Series No. 13. Washington, D.C.: The World Bank.
Imbens, Guido W., and Karthik Kalyanaraman. 2012. ‘Optimal Bandwidth Choice for the Regression Discontinuity Estimator’. *The Review of Economic Studies* 79 (3): 933–59.
Imbens, Guido W., and Thomas Lemieux. 2008. ‘Regression Discontinuity Designs: A Guide to Practice’. *Journal of Econometrics* 142 (2): 615–35.
Jain, Anrudh K. 1981. ‘The Effect of Female Education on Fertility: A Simple Explanation’. *Demography* 18 (4): 577–95.
Kattan, Raja Bentaouet. 2006. ‘Implementation of Free Basic Education Policy’. 47186. The World Bank.
Keats, Anthony. 2018. ‘Women’s Schooling, Fertility, and Child Health Outcomes: Evidence from Uganda’s Free Primary Education Program’. *Journal of Development Economics* 135: 142–59.
Kraehnert, Kati, Tilman Brück, Michele Di Maio, and Roberto Nisticò. 2019. ‘The Effects of Conflict on Fertility: Evidence From the Genocide in Rwanda’. *Demography* 56 (3): 935–68.
La Mattina, Giulia. 2018. ‘How Persistent Is the Effect of Conflict on Primary Education? Long-Run Evidence from the Rwandan Genocide’. *Economics Letters* 163 (February): 32–35.
Lam, David, and Suzanne Duryea. 1999. ‘Effects of Schooling on Fertility, Labor Supply, and Investments in Children, with Evidence from Brazil’. *The Journal of Human Resources* 34 (1): 160.
Lee, David S., and Thomas Lemieux. 2010. ‘Regression Discontinuity Designs in Economics’. *Journal of Economic Literature* 48 (2): 281–355.
Levine, Ruth, Cynthia, Lloyd, Margaret Green, and Caren Grown. 2009. *Girls Count: A Global Investment & Action Agenda*. Washington, D.C.: Center for Global Development.
Lloyd, Cynthia B, and Juliet Young. 2009. ‘New Lessons: The Power of Educating Adolescent Girls’. Girls Count. New York: Population Council: Coalition for Adolescent Girls.
Lucas, Adrienne M., and Isaac M. Mbiti. 2012. ‘Access, Sorting, and Achievement: The Short-Run Effects of Free Primary Education in Kenya’. *American Economic Journal: Applied Economics* 4 (4): 226–53.
Ludwig, Jens, and Douglas L. Miller. 2005. ‘Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design’. In *NBER Working Paper Series 11702*. Cambridge, MA: National Bureau of Economic Research.
Makate, Marshall, and Clifton Makate. 2016. ‘The Causal Effect of Increased Primary Schooling on Child Mortality in Malawi: Universal Primary Education as a Natural Experiment’. *Social Science & Medicine* 168 (November): 72–83.
———. 2018. ‘Education and Teenage Childbirth in Uganda: Understanding the Links from the Uganda Demographic and Health Survey’. *International Journal of Social Economics* 45 (5): 746–64.
Mare, Robert D. 1991. ‘Five Decades of Educational Assortative Mating’. *American Sociological Review* 56 (1): 15–32.

Martin, Teresa Castro. 1995. 'Women's Education and Fertility: Results from 26 Demographic and Health Surveys'. *Studies in Family Planning* 26 (4): 187–202.

Martínez, Elin, and Agnes Odhiambo. 2018. *Leave No Girl behind in Africa - Discrimination in Education against Pregnant Girls and Adolescent Mothers*. New York, NY: Human Rights Watch.

Masuda, Kazuya, and Chikako Yamauchi. 2020. 'How Does Female Education Reduce Adolescent Pregnancy and Improve Child Health?: Evidence from Uganda's Universal Primary Education for Fully Treated Cohorts'. *The Journal of Development Studies* 56 (1): 63–86.

McCrary, Justin, and Heather Royer. 2011. 'The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth'. *American Economic Review* 101 (1): 158–95.

Mercier, Marion, Rama L. Ngenzebuke, and Philip Verwimp. 2020. 'Violence Exposure and Poverty: Evidence from the Burundi Civil War'. *Journal of Comparative Economics* 48 (4): 822–40.

Michael, Robert T. 1973. 'Education and the Derived Demand for Children'. *Journal of Political Economy* 81 (2): 128–64.

Mincer, Jacob. 1962. 'Labor Force Participation of Married Women: A Study of Labor Supply', 45.

———. 1963. 'Market Prices, Opportunity Costs, and Income Effects'. In *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfield*, 67–82. Stanford, CA: Carl F. Christ, Stanford University Press.

Ministere e la Presidence charge de la Bonne Gouvernance et du Plan - MPBGP, Ministere de la Sante Publique et de la Lutte contre le Sida - MSPLS, Institut de Statistiques et d'etudes economiques du Burundi - ISTEEBU, and ICF. 2017. *Burundi Troisieme Enquete Demographique et de Sante 2016-2017*. Bujumbura, Burundi: MPBGP, MSPLS, ISTEEBU, and ICF.

Moussa, Wael, and Carina Omoeva. 2020. 'The Long-Term Effects of Universal Primary Education: Evidence from Ethiopia, Malawi, and Uganda'. *Comparative Education Review* 64 (2): 179–206.

Obura, Anna P. 2008. *Staying Power: Struggling to Reconstruct Education in Burundi since 1993*. Paris: International Institute for Educational Planning.

Osili, Una Okonkwo, and Bridget Terry Long. 2008. 'Does Female Schooling Reduce Fertility? Evidence from Nigeria'. *Journal of Development Economics* 87 (1): 57–75.

Ozier, Owen. 2018. 'The Impact of Secondary Schooling in Kenya: A Regression Discontinuity Analysis'. *Journal of Human Resources* 53 (1): 157–88.

Presler-Marshall, Elizabeth, and Nicola Jones. 2012. 'Charting the Future: Empowering Girls to Prevent Early Pregnancy'. London: Overseas Development Institute.

Psaki, Stephanie R., Erica K. Chuang, Andrea J. Melnikas, David B. Wilson, and Barbara S. Mensch. 2019. 'Causal Effects of Education on Sexual and Reproductive Health in Low and Middle-Income Countries: A Systematic Review and Meta-Analysis'. *SSM - Population Health* 8.

Rosenzweig, Mark R., and T. Paul Schultz. 1989. 'Schooling, Information and Nonmarket Productivity: Contraceptive Use and Its Effectiveness'. *International Economic Review* 30 (2): 457–77.

Rutstein, Shea O., and Kiersten Johnson. 2004. 'The DHS Wealth Index'. In *DHS Comparative Reports No. 6*. Calverton, Maryland, USA: Macro International.

Staiger, Douglas, and James H. Stock. 1997. 'Instrumental Variables Regression with Weak Instruments'. *Econometrica* 65 (3): 557–86.

The DHS Program. 2015. 'Demographic and Health Surveys Model Questionnaire - Woman'.

Thomas, Duncan, John Strauss, and Maria-Helena Henriques. 1991. ‘How Does Mother’s Education Affect Child Height?’ *The Journal of Human Resources* 26 (2): 183–211.

Travaglianti, Manuela. 2017. ‘How Abolishing School Fees Increased Support for the Incumbent in Burundi’. *African Affairs* 116 (462): 101–24.

UIS. 2022. ‘UIS Statistics’. UNESCO Institute for Statistics. 2022.

UNFPA, United Nations Population Fund. 2002. ‘Situation De´mographique et Sociale Au Burundi. Re´sultats de l’Enqueˆte Sociode´mographique et de Sante´ de La Reproduction’. Burundi: De´partement de la Population.

UNICEF. 2005. ‘Multiple Indicator Cluster Surveys (MICS)’.

United Nations, ed. 1995. *Report of the International Conference on Population and Development: Cairo, 5-13 September 1994*. New York: United Nations.

———. 2001. ‘Road Map towards the Implementation of the United Nations Millennium Declaration: Report of the Secretary-General.’ New York: United Nations Department of Public Information.

———. 2022. ‘World Population Prospects: The 2022 Revision’. United Nations, Department of Economic and Social Affairs, Population Division.

Verwimp, Philip, and Jan Van Bavel. 2014. ‘Schooling, Violent Conflict, and Gender in Burundi’. *The World Bank Economic Review* 28 (2): 384–411.

Wild, Frederik, and David Stadelmann. 2024. ‘Heterogeneous Effects of Women’s Schooling on Fertility, Literacy and Work: Evidence from Burundi’s Free Primary Education Policy’. *Journal of African Economies* 33 (1): 67–91.

Willis, Robert J. 1973. ‘A New Approach to the Economic Theory of Fertility Behavior’. *Journal of Political Economy* 81 (2, Part 2): S14–64.

World Bank. 2017. *World Development Report 2018: Learning to Realize Education’s Promise*. The World Bank.

———. 2021. *World Development Indicators*. Washington, D.C.: The World Bank.

Zenebe Gebre, Tihtina. 2020. ‘Free Primary Education, Timing of Fertility, and Total Fertility’. *The World Bank Economic Review* 34 (3): 730–48.

Appendix B

Heterogeneous Effects of Women's Schooling on Fertility, Literacy and Work: Evidence from Burundi's Free Primary Education Policy

Table B.1: RD Estimates, Years of Schooling (Controls following Post-Double-Selection)

	Regression Sample								
	Full			Poor			Wealthy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1[Year of Birth ≥ 1992]	1.344***	1.223***	1.266***	0.849***	0.908***	0.893***	1.356***	1.280***	1.357***
	(0.255)	(0.223)	(0.211)	(0.228)	(0.220)	(0.210)	(0.397)	(0.367)	(0.362)
Mean of Control Group		[3.55]			[1.99]			[5.56]	
Interaction: 1 × Year of Birth	0.135	0.164**	0.081	0.378***	0.339***	0.196***	-0.179	-0.099	-0.123
	(0.090)	(0.077)	(0.073)	(0.081)	(0.078)	(0.073)	(0.141)	(0.130)	(0.121)
Year of Birth	0.052	0.038	-0.530**	0.059	0.064	0.098	0.116	0.070	-0.882**
	(0.054)	(0.047)	(0.226)	(0.046)	(0.044)	(0.228)	(0.090)	(0.085)	(0.360)
Effect on the second cohort treated by FPE									
1[YoB ≥ 1992] + 1 × YoB + YoB		1.426***			1.312***			1.250***	
		(0.197)			(0.187)			(0.338)	
Observations	7,713	7,709	7,263	3,857	3,856	3,783	3,856	3,853	3,480
Adj. R-squared	0.054	0.263	-	0.099	0.183	-	0.041	0.187	-
Controls (Post Double Selection)	NO	YES	(28)	NO	YES	(25)	NO	YES	(23)
Province Fixed Effects	NO	YES	PDS	NO	YES	PDS	NO	YES	PDS
Survey Fixed Effects	YES	YES	PDS	YES	YES	PDS	YES	YES	PDS

Notes: The results in each column are produced by a separate regression. Columns (3), (6) and (9) report results using the Post Double Selection (PDS) algorithm developed in Belloni et al. (2014), considering a total of 896 potential controls. The included number of PDS' selected controls are given in parenthesis at the bottom of the table. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.2: Second-Stage Results, Literacy and Desired Fertility (Post-Double-Selection)

	Regression Sample								
	Full			Poor			Wealthy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Literacy {0,1}									
Schooling	0.054***	0.050***	0.059***	0.067**	0.067**	0.072**	0.033	0.027	0.037
Instrumented by 1[YoB ≥ 1992]	(0.015)	(0.017)	(0.016)	(0.033)	(0.030)	(0.029)	(0.022)	(0.024)	(0.023)
Mean of Control Group		[0.56]			[0.43]			[0.73]	
F-Statistic	27.7	29.9	35.5	13.8	17.1	18.7	11.4	11.9	13.9
Observations	7,708	7,704	7,258	3,856	3,855	3,782	3,852	3,849	3,476
Panel B: Desired Number of Children									
Schooling	-0.144**	-0.151**	-0.153**	-0.307**	-0.261**	-0.281**	-0.093	-0.088	-0.079
Instrumented by 1[YoB ≥ 1992]	(0.060)	(0.063)	(0.061)	(0.136)	(0.118)	(0.119)	(0.086)	(0.086)	(0.079)
Mean of Control Group		[4.01]			[4.05]			[3.96]	
F-Statistic	25.9	28.3	34.9	12.5	15.8	17.5	11.4	11.9	14.7
Observations	7,520	7,516	7,087	3,764	3,763	3,691	3,756	3,753	3,396
Controls (Post Double Selection)	NO	YES	(34,36)	NO	YES	(27,27)	NO	YES	(23,26)
Province Fixed Effects	NO	YES	PDS	NO	YES	PDS	NO	YES	PDS
Survey Fixed Effects	YES	YES	PDS	YES	YES	PDS	YES	YES	PDS

Notes: The results in each panel and column are produced by a separate regression. Columns (3), (6) and (9) report results using the Post Double Selection (PDS) algorithm developed in Belloni et al. (2014), considering a total of 896 potential controls. The included number of PDS' selected controls are given in parenthesis at the bottom of the table. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.3: Second-Stage Results, Fertility and Reproductive Behavior (Post-Double-Selection)

	Regression Sample								
	Full			Poor			Wealthy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Age at First Birth									
Schooling	0.440**	0.468**	0.511***	0.435	0.449*	0.516**	0.640*	0.757	0.686
Instrumented by 1[YoB ≥ 1992]	(0.179)	(0.204)	(0.198)	(0.274)	(0.263)	(0.245)	(0.381)	(0.564)	(0.465)
Mean of Control Group		[19.67]			[19.59]			[19.80]	
F-Statistic	14.9	11.9	14.1	9.1	10.3	12.6	3.9	2.1	2.0
Observations	4,918	4,917	4,769	2,830	2,830	2,800	2,088	2,087	2,087
Panel B: First Birth before Age 20 {0,1}									
Schooling	-0.037**	-0.034*	-0.049***	-0.056	-0.069*	-0.086**	-0.026	-0.012	-0.020
Instrumented by 1[YoB ≥ 1992]	(0.018)	(0.019)	(0.019)	(0.041)	(0.039)	(0.040)	(0.024)	(0.025)	(0.024)
Mean of Control Group		[0.34]			[0.37]			[0.31]	
F-Statistic	27.9	30.0	36.7	13.8	17.0	18.5	11.7	12.2	14.1
Observations	7,713	7,709	7,263	3,857	3,856	3,783	3,856	3,853	3,480
Panel C: Married before Age 20 {0,1}									
Schooling	-0.063***	-0.061***	-0.081***	-0.078**	-0.090**	-0.113***	-0.059**	-0.045*	-0.064**
Instrumented by 1[YoB ≥ 1992]	(0.018)	(0.018)	(0.020)	(0.039)	(0.037)	(0.040)	(0.025)	(0.025)	(0.025)
Mean of Control Group		[0.48]			[0.54]			[0.41]	
F-Statistic	27.9	30.0	35.5	13.8	17.0	18.4	11.7	12.2	13.7
Observations	7,713	7,709	7,263	3,857	3,856	3,783	3,856	3,853	3,480
Panel D: Sex before Age 20 {0,1}									
Schooling	-0.044**	-0.042**	-0.063***	-0.047	-0.064*	-0.082**	-0.046*	-0.031	-0.052**
Instrumented by 1[YoB ≥ 1992]	(0.018)	(0.019)	(0.019)	(0.037)	(0.035)	(0.037)	(0.028)	(0.029)	(0.026)
Mean of Control Group		[0.55]			[0.59]			[0.50]	
F-Statistic	28.5	29.9	36.1	14.8	17.9	19.1	11.2	11.4	13.6
Observations	7,623	7,619	7,180	3,807	3,806	3,733	3,816	3,813	3,447
Controls (Post Double Selection)	NO	YES	(24,47,43,46)	NO	YES	(19,33,35,37)	NO	YES	(47,28,28,23)
Province Fixed Effects	NO	YES	PDS	NO	YES	PDS	NO	YES	PDS
Survey Fixed Effects	YES	YES	PDS	YES	YES	PDS	YES	YES	PDS

Notes: The results in each panel and column are produced by a separate regression. Columns (3), (6) and (9) report results using the Post Double Selection (PDS) algorithm developed in Belloni et al. (2014), considering a total of 896 potential controls. The included number of PDS' selected controls are given in parenthesis at the bottom of the table. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.4: Second-Stage Results, Income (Controls following Post-Double-Selection)

	Regression Sample								
	Full			Poor			Wealthy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Worked Last Year {0,1}									
Schooling	-0.003	-0.002	0.003	0.017	0.005	0.002	0.000	-0.002	0.004
Instrumented by 1[YoB ≥ 1992]	(0.015)	(0.013)	(0.014)	(0.024)	(0.020)	(0.020)	(0.025)	(0.022)	(0.021)
Mean of Control Group		[0.86]			[0.92]			[0.78]	
F-Statistic	27.9	30.0	34.5	13.8	17.0	18.8	11.7	12.2	13.5
Observations	7,713	7,709	7,263	3,857	3,856	3,783	3,856	3,853	3,480
Panel B: Works outside of the Household for Cash/Money {0,1}									
Schooling	0.023**	0.026**	0.024**	0.058*	0.057**	0.052**	0.007	0.008	0.003
Instrumented by 1[YoB ≥ 1992]	(0.011)	(0.013)	(0.012)	(0.031)	(0.027)	(0.025)	(0.014)	(0.015)	(0.014)
Mean of Control Group		[0.08]			[0.08]			[0.07]	
F-Statistic	27.9	30.0	35.1	13.8	17.0	18.7	11.7	12.2	13.3
Observations	7,713	7,709	7,263	3,857	3,856	3,783	3,856	3,853	3,480
Controls	NO	YES	(44,39)	NO	YES	(33,29)	NO	YES	(28,29)
Province Fixed Effects	NO	YES	PDS	NO	YES	PDS	NO	YES	PDS
Survey Fixed Effects	YES	YES	PDS	YES	YES	PDS	YES	YES	PDS

Notes: The results in each panel and column are produced by a separate regression. Columns (3), (6) and (9) report results using the Post Double Selection (PDS) algorithm developed in Belloni et al. (2014), considering a total of 896 potential controls. The included number of PDS' selected controls are given in parenthesis at the bottom of the table. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.5: Second-Stage Results, Birth before Age {0,1}

	First Birth before Age ... {0,1}							
	Age 16	Age 17	Age 18	Age 19	Age 20	Age 21	Age 22	Age 23
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Full								
Mean of Control Group	[0.03]	[0.07]	[0.12]	[0.22]	[0.34]	[0.48]	[0.60]	[0.68]
Schooling	0.002	0.012	0.009	-0.031*	-0.034*	-0.033	-0.033	-0.018
Instrumented by 1[YoB ≥ 1992]	(0.007)	(0.010)	(0.014)	(0.017)	(0.019)	(0.029)	(0.027)	(0.029)
F-Statistic	30.0	30.0	30.0	30.0	30.0	12.6	13.8	11.3
Observations	7,709	7,709	7,709	7,709	7,709	6,412	5,553	4,576
Panel B: Poor								
Mean of Control Group	[0.03]	[0.07]	[0.13]	[0.23]	[0.37]	[0.52]	[0.65]	[0.74]
Schooling	-0.003	0.019	0.016	-0.054	-0.069*	0.001	-0.041	-0.025
Instrumented by 1[YoB ≥ 1992]	(0.013)	(0.021)	(0.029)	(0.034)	(0.039)	(0.066)	(0.045)	(0.042)
F-Statistic	17.0	17.0	17.0	17.0	17.0	4.9	9.8	8.7
Observations	3,856	3,856	3,856	3,856	3,856	3,217	2,792	2,325
Panel C: Wealthy								
Mean of Control Group	[0.03]	[0.07]	[0.11]	[0.19]	[0.31]	[0.44]	[0.52]	[0.60]
Schooling	0.008	0.010	0.006	-0.021	-0.012	-0.062	-0.037	-0.008
Instrumented by 1[YoB ≥ 1992]	(0.009)	(0.014)	(0.018)	(0.023)	(0.025)	(0.039)	(0.043)	(0.058)
F-Statistic	12.2	12.2	12.2	12.2	12.2	5.9	4.5	3.1
Observations	3,853	3,853	3,853	3,853	3,853	3,195	2,761	2,251
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Province Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The results in each panel and column panel are produced by a separate regression. The IV-estimate at each age increment tests the outcome that women had a first birth before the indicated age while restricting the sample to respondents aged at least that of the investigated age. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.6: Second-Stage Results, Married before Age {0,1}

	Married before Age ... {0,1}							
	Age 16	Age 17	Age 18	Age 19	Age 20	Age 21	Age 22	Age 23
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Full								
Mean of Control Group	[0.07]	[0.13]	[0.22]	[0.34]	[0.48]	[0.58]	[0.66]	[0.71]
Schooling	-0.012	-0.007	-0.007	-0.032	-0.061***	-0.044	-0.032	-0.017
Instrumented by 1[YoB ≥ 1992]	(0.011)	(0.013)	(0.017)	(0.019)	(0.018)	(0.027)	(0.025)	(0.028)
F-Statistic	30.0	30.0	30.0	30.0	30.0	12.6	13.8	11.3
Observations	7,709	7,709	7,709	7,709	7,709	6,412	5,553	4,576
Panel B: Poor								
Mean of Control Group	[0.08]	[0.14]	[0.24]	[0.38]	[0.54]	[0.65]	[0.74]	[0.79]
Schooling	-0.021	0.009	-0.008	-0.059	-0.090**	-0.020	-0.024	-0.009
Instrumented by 1[YoB ≥ 1992]	(0.022)	(0.028)	(0.033)	(0.037)	(0.037)	(0.064)	(0.039)	(0.041)
F-Statistic	17.0	17.0	17.0	17.0	17.0	4.9	9.8	8.7
Observations	3,856	3,856	3,856	3,856	3,856	3,217	2,792	2,325
Panel C: Wealthy								
Mean of Control Group	[0.06]	[0.11]	[0.19]	[0.30]	[0.41]	[0.49]	[0.55]	[0.61]
Schooling	-0.006	-0.023	-0.004	-0.011	-0.045*	-0.066*	-0.042	-0.020
Instrumented by 1[YoB ≥ 1992]	(0.013)	(0.017)	(0.022)	(0.027)	(0.025)	(0.038)	(0.043)	(0.050)
F-Statistic	12.2	12.2	12.2	12.2	12.2	5.9	4.5	3.1
Observations	3,853	3,853	3,853	3,853	3,853	3,195	2,761	2,251
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Province Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The results in each panel and column panel are produced by a separate regression. The IV-estimate at each age increment tests the outcome that women had a first birth before the indicated age while restricting the sample to respondents aged at least that of the investigated age. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.7: Second-Stage Results, Sex before Age {0,1}

	First Sex before Age ... {0,1}							
	Age 16	Age 17	Age 18	Age 19	Age 20	Age 21	Age 22	Age 23
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Full								
Mean of Control Group	[0.09]	[0.15]	[0.26]	[0.42]	[0.55]	[0.66]	[0.73]	[0.78]
Schooling	-0.006	-0.002	-0.039**	0.000	-0.042**	-0.031	-0.016	-0.019
Instrumented by 1[YoB ≥ 1992]	(0.011)	(0.015)	(0.018)	(0.020)	(0.019)	(0.027)	(0.024)	(0.025)
F-Statistic	29.9	29.9	29.9	29.9	29.9	12.8	13.7	11.4
Observations	7,619	7,619	7,619	7,619	7,619	6,340	5,489	4,528
Panel B: Poor								
Mean of Control Group	[0.09]	[0.16]	[0.27]	[0.43]	[0.59]	[0.71]	[0.78]	[0.83]
Schooling	-0.017	0.002	-0.076**	-0.018	-0.064*	-0.010	0.001	-0.004
Instrumented by 1[YoB ≥ 1992]	(0.022)	(0.028)	(0.035)	(0.036)	(0.035)	(0.054)	(0.037)	(0.035)
F-Statistic	17.9	17.9	17.9	17.9	17.9	5.4	10.1	8.8
Observations	3,806	3,806	3,806	3,806	3,806	3,176	2,756	2,298
Panel C: Wealthy								
Mean of Control Group	[0.08]	[0.14]	[0.25]	[0.39]	[0.50]	[0.59]	[0.66]	[0.71]
Schooling	0.002	-0.006	-0.015	0.018	-0.031	-0.050	-0.039	-0.041
Instrumented by 1[YoB ≥ 1992]	(0.016)	(0.021)	(0.026)	(0.032)	(0.029)	(0.040)	(0.042)	(0.049)
F-Statistic	11.4	11.4	11.4	11.4	11.4	5.4	4.1	3.0
Observations	3,813	3,813	3,813	3,813	3,813	3,164	2,733	2,230
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Province Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The results in each panel and column panel are produced by a separate regression. The IV-estimate at each age increment tests the outcome that women had a first birth before the indicated age while restricting the sample to respondents aged at least that of the investigated age. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.8: Second-Stage Results, Knowledge and Use of Condoms

	Regression Sample					
	Full		Poor		Wealthy	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Knows Condom as contr. Method {0,1}						
Schooling	0.010	0.010	0.011	0.012	0.010	0.011
Instrumented by 1[YoB ≥ 1992]	(0.009)	(0.009)	(0.020)	(0.019)	(0.011)	(0.011)
Mean of Control Group	[0.94]		[0.92]		[0.96]	
F-Statistic	27.9	30.0	13.8	17.0	11.7	12.2
Observations	7,713	7,709	3,857	3,856	3,856	3,853
Panel B: Used Condom with last sex. Partner {0,1}						
Schooling	0.040**	0.042*	0.033*	0.031*	0.070	0.090
Instrumented by 1[YoB ≥ 1992]	(0.020)	(0.025)	(0.019)	(0.018)	(0.071)	(0.126)
Mean of Control Group	[0.02]		[0.01]		[0.04]	
F-Statistic	10.7	8.5	8.4	9.8	1.6	0.7
Observations	4,801	4,799	2,703	2,702	2,098	2,097
Controls	NO	YES	NO	YES	NO	YES
Province Fixed Effects	NO	YES	NO	YES	NO	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: The results in each panel and column are produced by a separate regression. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.9: Second-Stage Results, Engagement with Mass Media

	Regression Sample					
	Full		Poor		Wealthy	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Reads Newspaper at least once per Week {0,1}						
Schooling	-0.011	-0.013	-0.003	-0.003	-0.025	-0.026
Instrumented by 1[YoB ≥ 1992]	(0.008)	(0.008)	(0.011)	(0.011)	(0.015)	(0.016)
Mean of Control Group	[0.06]		[0.03]		[0.10]	
F-Statistic	27.9	30.0	13.9	17.1	11.7	12.2
Observations	7,712	7,708	3,856	3,855	3,856	3,853
Panel B: Listens to Radio at least once per Week {0,1}						
Schooling	0.029*	0.025	0.015	0.014	0.023	0.019
Instrumented by 1[YoB ≥ 1992]	(0.017)	(0.019)	(0.034)	(0.032)	(0.026)	(0.027)
Mean of Control Group	[0.42]		[0.30]		[0.58]	
F-Statistic	27.8	29.8	13.8	17.0	11.6	12.1
Observations	7,712	7,708	3,857	3,856	3,855	3,852
Panel C: Watches TV at least once per Week {0,1}						
Schooling	0.014	0.010	0.007*	0.007*	0.013	0.012
Instrumented by 1[YoB ≥ 1992]	(0.009)	(0.008)	(0.004)	(0.004)	(0.019)	(0.017)
Mean of Control Group	[0.06]		[0.00]		[0.14]	
F-Statistic	28.0	30.2	13.9	17.1	11.7	12.2
Observations	7,710	7,706	3,856	3,855	3,854	3,851
Controls	NO	YES	NO	YES	NO	YES
Province Fixed Effects	NO	YES	NO	YES	NO	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: The results in each panel and column are produced by a separate regression. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.10: Second-Stage Results, Husband's Characteristics

	Regression Sample					
	Full		Poor		Wealthy	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Husband's Age						
Schooling	-1.183	-1.411	-1.402	-1.177	-2.093	-4.269
Instrumented by 1[YoB ≥ 1992]	(0.854)	(0.994)	(1.148)	(0.990)	(3.407)	(10.813)
Mean of Control Group	[29.78]		[29.38]		[30.39]	
F-Statistic	6.3	5.2	5.7	7.3	0.6	0.2
Observations	4,430	4,430	2,576	2,576	1,854	1,854
Panel B: Husband's Education						
Schooling	0.593	0.591	0.115	0.164	1.575	2.322
Instrumented by 1[YoB ≥ 1992]	(0.392)	(0.441)	(0.451)	(0.399)	(2.225)	(4.940)
Mean of Control Group	[3.59]		[2.49]		[5.30]	
F-Statistic	5.9	5.1	5.7	7.2	0.4	0.2
Observations	4,485	4,485	2,606	2,606	1,879	1,879
Panel C: Husband's Desired Number of Children: Same? {0,1}						
Schooling	0.065	0.078	0.078	0.071	0.102	0.134
Instrumented by 1[YoB ≥ 1992]	(0.054)	(0.065)	(0.080)	(0.073)	(0.156)	(0.262)
Mean of Control Group	[0.63]		[0.62]		[0.63]	
F-Statistic	9.7	7.6	7.5	8.7	1.2	0.5
Observations	3,899	3,899	2,240	2,240	1,659	1,659
Controls	NO	YES	NO	YES	NO	YES
Province Fixed Effects	NO	YES	NO	YES	NO	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: The results in each panel and column are produced by a separate regression. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.11: Smoothness Test of Covariates, Multiple Indicator Cluster Survey (MICS)

	Multiple Indicator Cluster Survey (MICS)						
Panel A: Household Characteristics							
	Attended School {0,1}	Wealth Index {1,5}	Urban {0,1}	Household Size {1,14}	Mother Alive {0,1}	Father Alive {0,1}	Religion (Head) {1,6}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1[Year of Birth ≥ 1992]	-0.032	0.018	-0.016	-0.298	0.003	-0.068	0.038
	(0.046)	(0.144)	(0.014)	(0.204)	(0.041)	(0.055)	(0.093)
Mean of Control Group	[0.79]	[3.01]	[0.07]	[5.61]	[0.82]	[0.65]	[1.40]
Observations	2,518	2,523	2,523	2,523	2,205	2,196	2,523
Adj. R-squared	(0.001)	0.002	0.001	0.012	0.023	0.036	0.00

Panel B: Mothers' & Children's Characteristics					**Last Born Child**		
	Education Level {1,3}	Literacy {0,1}	Ever Married {0,1}	Children Born {1,13}	Height at Birth {1,5}	Was Breastfed {0,1}	Received Vitamin A {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1[Children's Year of Birth ≥ 1992]	-0.048	-0.031	0.014	-0.351*	0.064	0.032	0.047
	(0.088)	(0.064)	(0.044)	(0.185)	(0.159)	(0.029)	(0.088)
Mean of Control Group	[0.42]	[0.29]	[0.77]	[5.92]	[3.24]	[0.97]	[0.39]
Observations	1,124	1,792	1,913	1,913	857	875	874
Adj. R-squared	0.043	0.027	0.022	0.232	0.002	0.002	(0.00)

Notes: The results in each panel and column are produced by a separate regression, using the Burundi Multiple Indicator Cluster Survey (MICS) of 2005. The discontinuities estimated in panel A reflect changes in the characteristics of households whose members are born in or after 1992, compared to households of members born in or before 1991. The discontinuities estimated in panel B reflect changes in characteristics of "potential mothers", i.e. of women who have had their first birth in or after 1992. Estimates are weighted using the MICS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.12: Education and Women's Household Wealth

	Dependent Variable													
	DHS Wealth Score		*Dwelling:* Electricity {0,1}		*Dwelling:* Piped Water {0,1}		*Dwelling:* Flush Toilet {0,1}		*Own:* Radio {0,1}		*Own:* TV {0,1}		*Own:* Bicycle {0,1}	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Schooling	5755.257	3142.718	0.016	0.007	0.018	0.016	0.007	0.002	0.050***	0.051**	0.006	0.000	0.020	0.024
Instrumented by 1[YoB $\geq$ 1992]	(3913.7)	(2463.7)	(0.013)	(0.011)	(0.022)	(0.019)	(0.011)	(0.008)	(0.019)	(0.021)	(0.010)	(0.009)	(0.019)	(0.020)
Mean of Control Group	-[4444.59]		[0.10]		[0.32]		[0.04]		[0.46]		[0.06]		[0.22]	
F-Statistic	27.9	30.0	25.8	27.5	25.2	27.2	25.7	27.5	25.7	27.5	25.7	27.5	25.7	27.5
Observations	7,713	7,709	7,612	7,609	7,568	7,565	7,614	7,611	7,614	7,611	7,614	7,611	7,610	7,607
Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Province Fixed Effects	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The results in each column are produced by a separate regression. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) $\geq$ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.13: Second-Stage Results, Redefined Poor and Wealthy Subgroups

	Dependent Variable							
Panel A: Literacy, Preferences & Fertility	Literacy {0,1}		Desired Number of Childern		Age at First Birth		First Birth before Age 20 {0,1}	
	Poor	Wealthy	Poor	Wealthy	Poor	Wealthy	Poor	Wealthy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Schooling	0.055**	0.036	-0.215**	-0.036	0.604**	0.232	-0.051*	-0.014
Instrumented by 1[YoB ≥ 1992]	(0.024)	(0.023)	(0.091)	(0.104)	(0.276)	(0.305)	(0.028)	(0.029)
Mean of Control Group	[0.56]	[0.86]	[4.01]	[3.72]	[19.67]	[20.49]	[0.34]	[0.22]
F-Statistic	20.9	5.9	20.6	5.5	8.6	2.3	21.0	5.8
Observations	5,801	1,718	5,651	1,682	4,040	775	5,804	1,719

Panel B: Marriage, Sex & Labour Market Outcomes	Married before Age 20 {0,1}		Sex before Age 20 {0,1}		Worked last Year {0,1}		Works outside HH. For Cash/Money {0,1}	
	Poor	Wealthy	Poor	Wealthy	Poor	Wealthy	Poor	Wealthy
	(1)	(2)	(1)	(2)	(3)	(4)	(5)	(6)
Schooling	-0.088***	-0.003	-0.064**	0.003	0.004	-0.035	0.045**	-0.005
Instrumented by 1[YoB ≥ 1992]	(0.028)	(0.031)	(0.028)	(0.034)	(0.018)	(0.029)	(0.018)	(0.020)
Mean of Control Group	[0.48]	[0.25]	[0.55]	[0.41]	[0.86]	[0.66]	[0.08]	[0.12]
F-Statistic	21.0	5.8	21.0	5.6	21.0	5.8	21.0	5.8
Observations	5,804	1,719	5,737	1,701	5,804	1,719	5,804	1,719
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Province Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The results in each column are produced by a separate regression. The sample is comprised of provinces with below median conflict intensities. The IV-estimate and key explanatory variable "Schooling" is instrumented by the binary indicator of being born in or after the cutoff, "1[Year of Birth (YoB) ≥ 1992]". The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.14: Second-Stage Results, Dropping Provinces with above-median Conflict Intensity

	Dependent Variable							
	Literacy {0,1}	Desired Number of Childern	Age at first Birth	First Birth before Age 20 {0,1}	Married before Age 20 {0,1}	Sex before Age 20 {0,1}	Worked last Year {0,1}	Works outside HH. for Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Schooling	0.078***	-0.166**	0.448**	-0.043**	-0.062***	-0.038*	-0.016	0.019
Instrumented by 1[YoB ≥ 1992]	(0.017)	(0.076)	(0.206)	(0.020)	(0.022)	(0.021)	(0.016)	(0.013)
Mean of Control Group	[0.54]	[4.27]	[19.46]	[0.37]	[0.51]	[0.56]	[0.85]	[0.07]
F-Statistic	22.6	20.8	9.8	22.6	22.6	22.8	22.6	22.6
Observations	3,823	3,692	2,471	3,824	3,824	3,772	3,824	3,824
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Province Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The results in each column are produced by a separate regression. The estimate in row one "1[Year of Birth ≥ 1992]" represents the discontinuous increase in schooling by women "just-treated" by the policy (13 years old), compared to women just too old to benefit from free schooling (14 years old). Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.15: First-stage Discontinuity in Schooling, Varying Functional Form

	Functional Form																	
	(a) Linear			(b) Linear Interaction			(c) Quadratic			(d) Quadratic Interaction			(e) Cubic			(f) Cubic Interaction		
	Full	Poor	Wealthy	Full	Poor	Wealthy	Full	Poor	Wealthy	Full	Poor	Wealthy	Full	Poor	Wealthy	Full	Poor	Wealthy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
1[Year of Birth ≥ 1992]	1.234***	0.956***	1.284***	1.223***	0.908***	1.280***	1.170***	0.772***	1.348***	1.498***	0.843**	1.787***	1.228***	0.683**	1.526***	0.413	-0.510	0.569
	(0.223)	(0.220)	(0.366)	(0.223)	(0.220)	(0.367)	(0.228)	(0.230)	(0.365)	(0.365)	(0.355)	(0.619)	(0.296)	(0.296)	(0.477)	(0.731)	(0.717)	(1.283)
Interaction: 1 × Year of Birth				0.164**	0.339***	-0.099				0.428**	0.193	0.647**				-1.385	-1.832*	-1.698
				(0.077)	(0.078)	(0.130)				(0.169)	(0.141)	(0.286)				(0.959)	(1.000)	(1.596)
Interaction: 1 × (Year of Birth)^2										0.965***	0.922***	0.706				-1.385	-0.294	-0.182
										(0.302)	(0.328)	(0.484)				(0.959)	(0.449)	(0.696)
Interaction: 1 × (Year of Birth)^3																-0.116	-0.170***	-0.147*
																(0.408)	(0.063)	(0.085)
Year of Birth	0.105***	0.200***	0.028	0.038	0.064	0.070	0.127***	0.257***	0.004	-0.370	-0.081	-0.466	0.103	0.295***	-0.069	1.252	1.919**	1.301
	(0.037)	(0.038)	(0.058)	(0.047)	(0.044)	(0.085)	(0.040)	(0.043)	(0.060)	(0.241)	(0.236)	(0.415)	(0.095)	(0.097)	(0.148)	(0.874)	(0.881)	(1.523)
Year of Birth^2							0.011	0.028***	-0.013	-0.066*	-0.024	-0.086	0.013	0.026***	-0.008	0.535*	0.718**	0.568
							(0.007)	(0.007)	(0.013)	(0.039)	(0.038)	(0.066)	(0.009)	(0.010)	(0.015)	(0.315)	(0.322)	(0.549)
Year of Birth^3													0.001	-0.002	0.003	0.066*	0.081**	0.072
													(0.004)	(0.004)	(0.006)	(0.034)	(0.036)	(0.060)
Intercept	6.497***	3.238***	7.135***	6.288***	2.804***	7.258***	6.440***	3.099***	7.197***	5.753***	2.601***	6.559***	-3.873***	3.172***	7.051***	7.023***	4.172***	7.933***
	(0.436)	(0.788)	(0.596)	(0.442)	(0.779)	(0.622)	(0.436)	(0.780)	(0.601)	(0.518)	(0.854)	(0.812)	(0.284)	(0.811)	(0.659)	(0.830)	(0.986)	(1.403)
Observations	7709	3856	3853	7709	3856	3853	7709	3856	3853	7709	3856	3853	7709	3856	3853	7709	3856	3853
Adj. R-squaed	0.262	0.179	0.187	0.263	0.183	0.187	0.263	0.182	0.187	0.264	0.184	0.187	0.263	0.182	0.187	0.264	0.185	0.187
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Province Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The results in each column are produced by a separate regression. The estimate in row one "1[Year of Birth ≥ 1992]" represents the discontinuous increase in schooling by women "just-treated" by the policy (13 years old), compared to women just too old to benefit from free schooling (14 years old). The standard errors reported are clustered at the survey-cluster level. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

Table B.16: Second-Stage Results: Varying Functional Form

	Functional Form																	
	(a) Linear			(b) Linear Interaction			(c) Quadratic			(d) Quadratic Interaction			(e) Cubic			(f) Cubic Interaction		
	Full	Poor	Wealthy	Full	Poor	Wealthy	Full	Poor	Wealthy	Full	Poor	Wealthy	Full	Poor	Wealthy	Full	Poor	Wealthy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Literacy {0,1}																		
Schooling	0.051***	0.070**	0.027	0.050***	0.067**	0.027	0.043**	0.057	0.025	0.066***	0.081	0.041	0.052**	0.052	0.039	0.043	0.143	0.043
Instr. by 1[YoB $\geq$ 1992]	(0.017)	(0.028)	(0.024)	(0.017)	(0.030)	(0.024)	(0.018)	(0.035)	(0.022)	(0.023)	(0.055)	(0.030)	(0.023)	(0.058)	(0.027)	(0.203)	(0.204)	(0.216)
F-Statistic	[30.4]	[19.0]	[12.0]	[29.9]	[17.1]	[11.9]	[26.2]	[11.3]	[13.4]	[16.7]	[5.6]	[8.0]	[17.0]	[5.3]	[10.0]	[0.3]	[0.5]	[0.1]
Desired Number of Children																		
Schooling	-0.150**	-0.248**	-0.088	-0.151**	-0.261**	-0.088	-0.153**	-0.306**	-0.076	-0.035	-0.091	-0.027	-0.063	-0.177	-0.025	-1.233	0.759	-0.657
Instr. by 1[YoB $\geq$ 1992]	(0.063)	(0.112)	(0.086)	(0.063)	(0.118)	(0.086)	(0.066)	(0.148)	(0.079)	(0.079)	(0.205)	(0.093)	(0.075)	(0.207)	(0.084)	(2.850)	(1.049)	(1.416)
F-Statistic	[28.9]	[17.5]	[11.9]	[28.3]	[15.8]	[11.9]	[24.4]	[10.5]	[12.8]	[17.4]	[5.7]	[9.1]	[17.8]	[5.2]	[11.2]	[0.2]	[0.8]	[0.2]
Age at First Birth																		
Schooling	0.401**	0.367	0.710	0.468**	0.449*	0.757	0.632**	0.676**	0.837	0.225	0.433	0.115	0.324	0.566	0.238	1.942	8.224	11.278
Instr. by 1[YoB $\geq$ 1992]	(0.191)	(0.245)	(0.554)	(0.204)	(0.263)	(0.564)	(0.256)	(0.342)	(0.607)	(0.215)	(0.421)	(0.319)	(0.205)	(0.415)	(0.273)	(3.569)	(139.187)	(160.028)
F-Statistic	[12.9]	[11.6]	[2.1]	[11.9]	[10.3]	[2.1]	[9.5]	[7.6]	[2.0]	[10.3]	[5.0]	[3.0]	[11.0]	[5.2]	[4.0]	[0.3]	[0.0]	[0.0]
First Birth before Age 20 {0,1}																		
Schooling	-0.033*	-0.066*	-0.013	-0.034*	-0.069*	-0.012	-0.038**	-0.083*	-0.016	-0.038	-0.122	0.003	-0.037	-0.125	0.003	-0.251	0.167	-0.176
Instr. by 1[YoB $\geq$ 1992]	(0.019)	(0.037)	(0.025)	(0.019)	(0.039)	(0.025)	(0.020)	(0.047)	(0.023)	(0.029)	(0.079)	(0.035)	(0.027)	(0.080)	(0.032)	(0.454)	(0.370)	(0.419)
F-Statistic	[30.5]	[19.0]	[12.3]	[30.0]	[17.0]	[12.2]	[26.3]	[11.3]	[13.6]	[16.9]	[5.6]	[8.3]	[17.2]	[5.3]	[10.2]	[0.3]	[0.5]	[0.2]
Married before Age 20 {0,1}																		
Schooling	-0.062***	-0.088**	-0.044*	-0.061***	-0.090**	-0.045*	-0.059***	-0.099**	-0.037	-0.082***	-0.180**	-0.039	-0.074***	-0.179**	-0.027	-0.226	-0.030	-0.364
Instr. by 1[YoB $\geq$ 1992]	(0.018)	(0.035)	(0.025)	(0.018)	(0.037)	(0.025)	(0.019)	(0.044)	(0.023)	(0.030)	(0.090)	(0.034)	(0.027)	(0.087)	(0.031)	(0.399)	(0.239)	(0.763)
F-Statistic	[30.5]	[19.0]	[12.3]	[30.0]	[17.0]	[12.2]	[26.3]	[11.3]	[13.6]	[16.9]	[5.6]	[8.3]	[17.2]	[5.3]	[10.2]	[0.3]	[0.5]	[0.2]
Sex before Age 20 {0,1}																		
Schooling	-0.042**	-0.063*	-0.031	-0.042**	-0.064*	-0.031	-0.039**	-0.067	-0.025	-0.042	-0.110	-0.008	-0.043	-0.112	-0.011	-0.115	0.083	-0.037
Instr. by 1[YoB $\geq$ 1992]	(0.019)	(0.033)	(0.028)	(0.019)	(0.035)	(0.029)	(0.020)	(0.041)	(0.027)	(0.029)	(0.077)	(0.037)	(0.028)	(0.076)	(0.035)	(0.250)	(0.276)	(0.223)
F-Statistic	[30.3]	[19.8]	[11.6]	[29.9]	[17.9]	[11.4]	[26.5]	[12.0]	[13.0]	[17.3]	[5.8]	[8.2]	[17.4]	[5.5]	[9.8]	[0.4]	[0.5]	[0.3]
Worked Last Year {0,1}																		
Schooling	-0.002	0.003	-0.002	-0.002	0.005	-0.002	0.000	0.015	-0.004	0.003	0.005	0.005	0.009	0.010	0.013	-0.183	0.080	-0.209
Instr. by 1[YoB $\geq$ 1992]	(0.013)	(0.019)	(0.022)	(0.013)	(0.020)	(0.022)	(0.015)	(0.026)	(0.021)	(0.021)	(0.039)	(0.034)	(0.021)	(0.040)	(0.031)	(0.320)	(0.181)	(0.444)
F-Statistic	[30.5]	[19.0]	[12.3]	[30.0]	[17.0]	[12.2]	[26.3]	[11.3]	[13.6]	[16.9]	[5.6]	[8.3]	[17.2]	[5.3]	[10.2]	[0.3]	[0.5]	[0.2]

Notes: Table continued on next page.

Works outside of the Household for Cash/Money {0,1}

Schooling	0.026**	0.055**	0.008	0.026**	0.057**	0.008	0.024*	0.061*	0.006	0.058***	0.124*	0.038	0.063***	0.140*	0.036	0.054	-0.021	0.087
Instr. by 1[YoB ≥ 1992]	(0.012)	(0.026)	(0.015)	(0.013)	(0.027)	(0.015)	(0.013)	(0.033)	(0.015)	(0.023)	(0.073)	(0.024)	(0.023)	(0.078)	(0.024)	(0.166)	(0.145)	(0.245)
F-Statistic	[30.5]	[19.0]	[12.3]	[30.0]	[17.0]	[12.2]	[26.3]	[11.3]	[13.6]	[16.9]	[5.6]	[8.3]	[17.2]	[5.3]	[10.2]	[0.3]	[0.5]	[0.2]
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Province Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Survey Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The results in each panel and in each column are produced by a separate regression. The IV-estimate and key explanatory variable “Schooling” is instrumented by the binary indicator of being born in or after the cutoff, “1[Year of Birth (YoB) ≥ 1992]”. The strength of the excluded instrument is given by the reported Kleibergen-Paap rk Wald F-Statistic. Estimates are weighted using the DHS sample weights and account for the pooling of survey rounds. The standard errors reported are clustered at the survey-cluster level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.2.

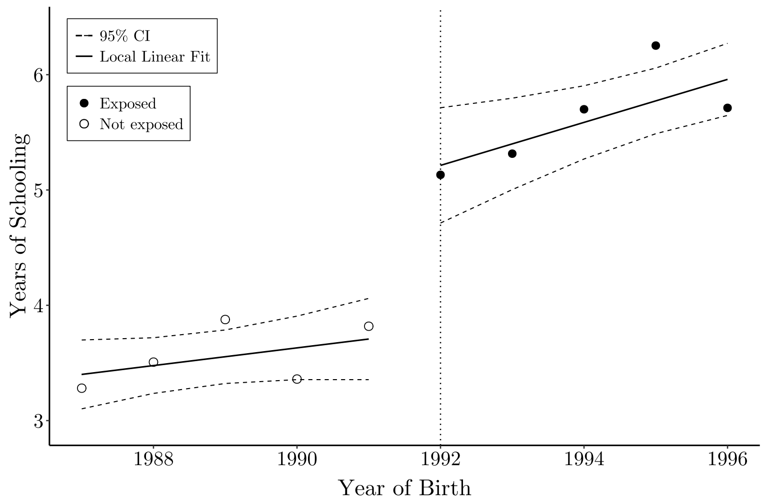


Figure B.1 Regression Discontinuity Estimate, Full Sample

Notes: The figure visually depicts the Regression Discontinuity estimate α_2 (slope before 1992) and the combined effect $\alpha_2 + \alpha_3$ (slope after 1992) produced from regression (3.1). See full notes in Table 3.2.

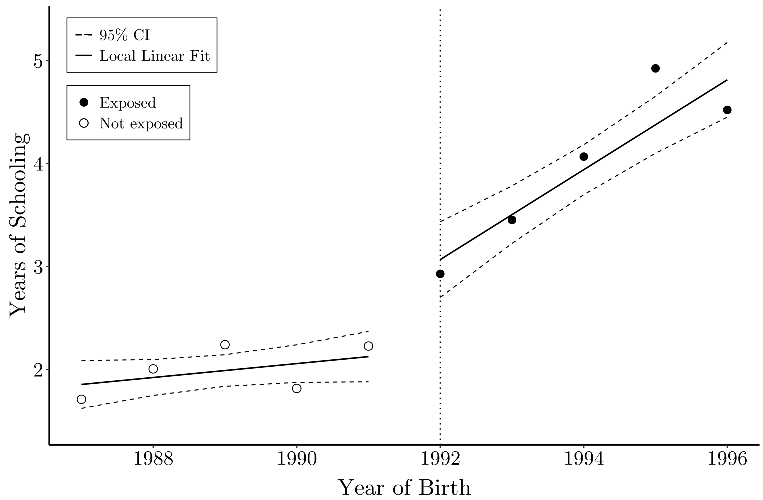


Figure B.2: Regression Discontinuity Estimate, Poor Subgroup

Notes: Identically produced as Figure B.1 using the subsample of "poor" women.

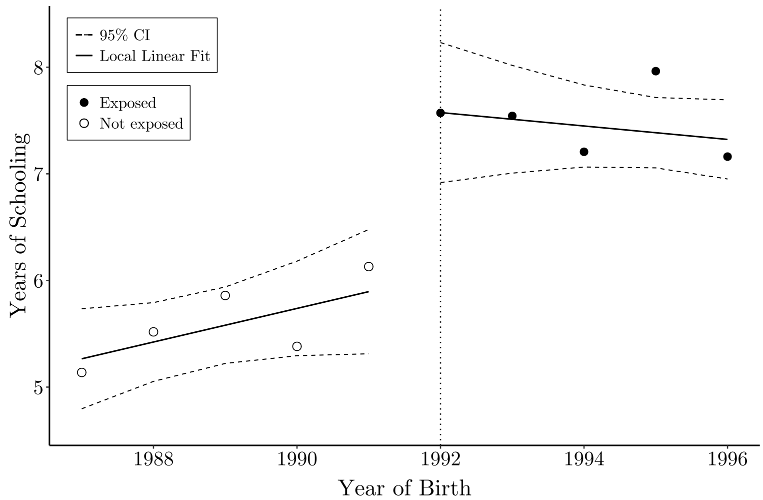


Figure B.3: Regression Discontinuity Estimate, Wealthy Subgroup

Notes: Identically produced as Figure B.1 using the subsample of "wealthy" women.

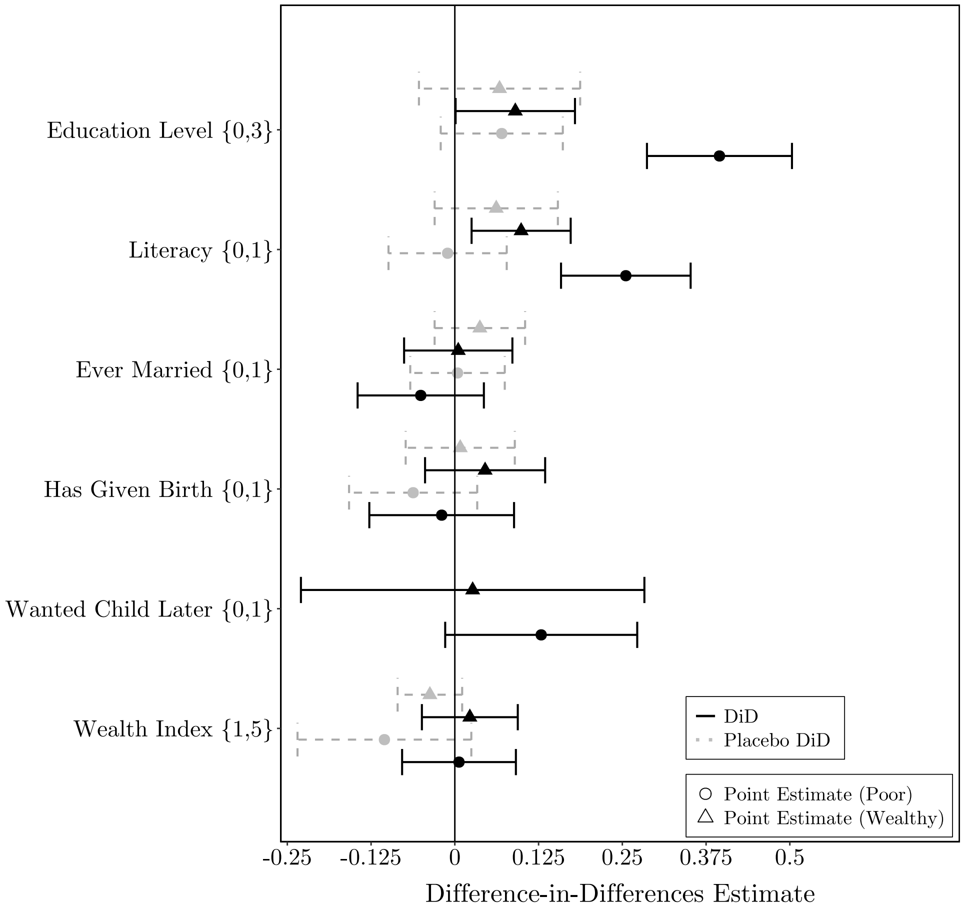


Figure B.4 Difference-in-Difference-in-Differences

Notes: The "Point Estimates" represent difference-in-differences coefficients, including a 95% confidence interval, for the respective dependent variable indicated on the y-axis. Data come from the 2000 and 2005 Multiple Cluster Indicator Surveys (MICS) as well as the 2010/2011 Burundi DHS survey, using the women's recodes. The coefficients are estimated through "triple-difference" estimation (3.3), comparing individuals aged 15-19 to individuals aged 20-24 in 2000, 2005 and 2010 for both "wealthy" and for "poor" households (splitting at the median household wealth level). The estimates are weighted using MICS/DHS sample weights and account for the pooling of sample rounds. Standard errors are clustered at the birth-year and enumeration-area level. See subsection 3.5 for a detailed discussion.

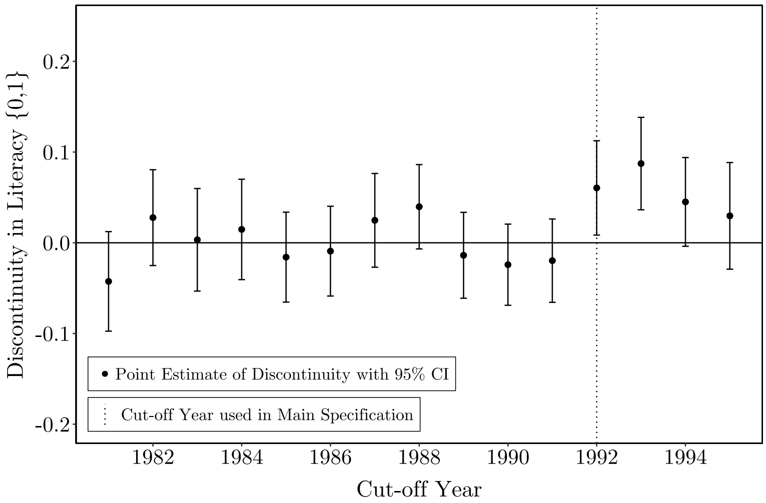


Figure B.5: Placebo Discontinuity, Literacy {0,1}

Notes: The figure visually depicts separate reduced form Regression Discontinuity estimate α_1 produced from regression (3.1). for varying (placebo) intervention years. See Section 3.5 for a detailed discussion.

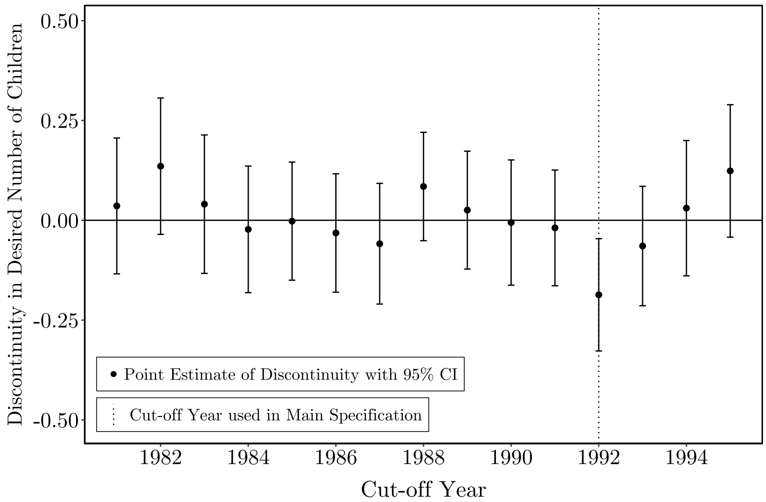


Figure B.6: Placebo Discontinuity, Desired Number of Children

Notes: Identically produced as Figure B.5 but replacing the outcome variable.

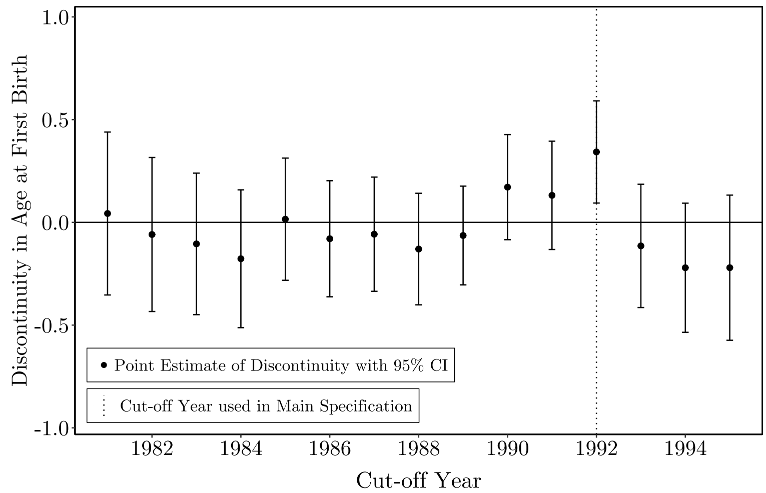


Figure B.7: Placebo Discontinuity, Age at First Birth

Notes: Identically produced as Figure B.5 but replacing the outcome variable.

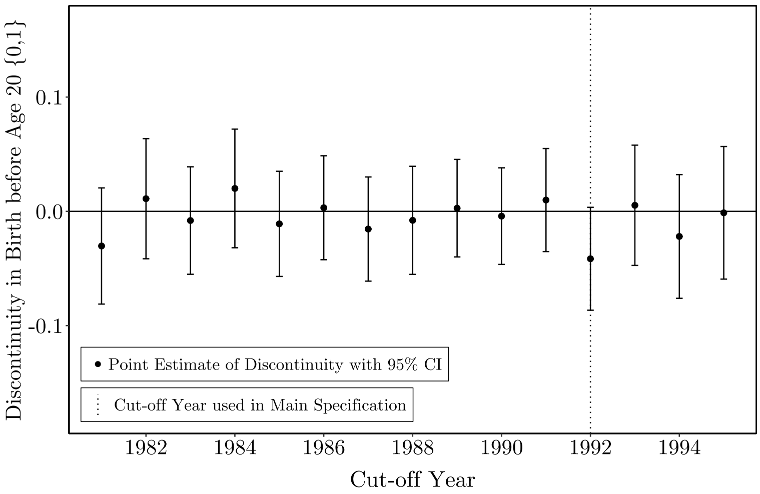


Figure B.8: Placebo Discontinuity, First Birth before Age 20 {0,1}

Notes: Identically produced as Figure B.5 but replacing the outcome variable.

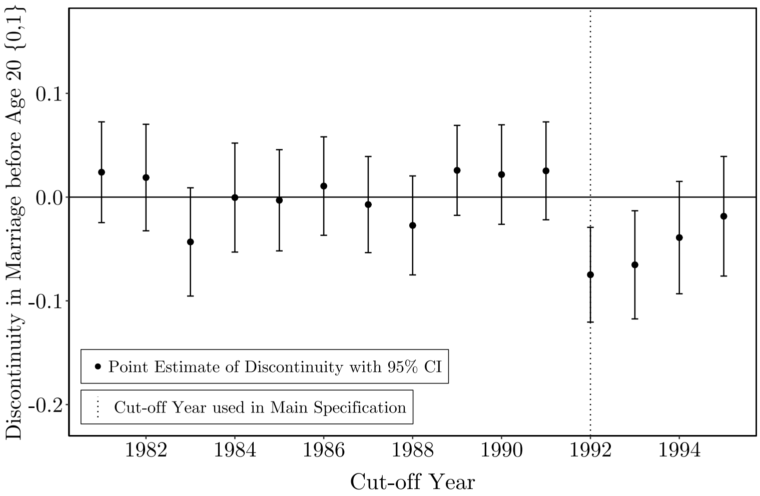


Figure B.9: Placebo Discontinuity, Married before Age 20 {0,1}

Notes: Identically produced as Figure B.5 but replacing the outcome variable.

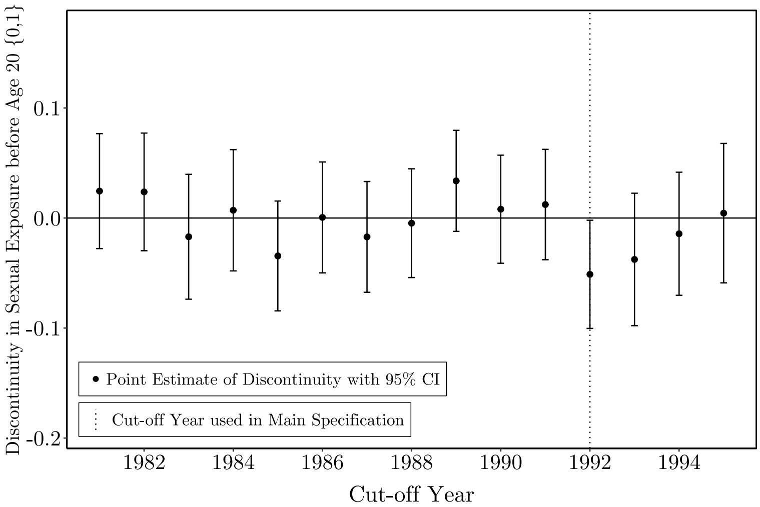


Figure B.10: Placebo Discontinuity, Sex before Age 20 {0,1}

Notes: Identically produced as Figure B.5 but replacing the outcome variable.

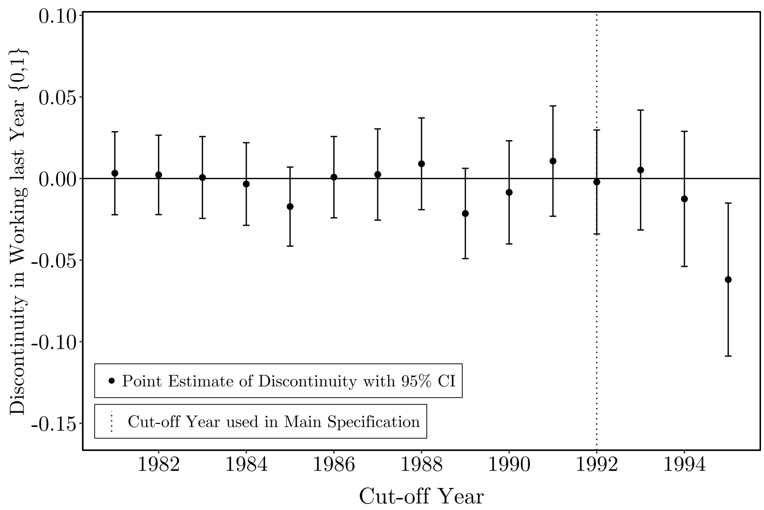


Figure B.11: Placebo Discontinuity, Worked the Last Year {0,1}

Notes: Identically produced as Figure B.5 but replacing the outcome variable.

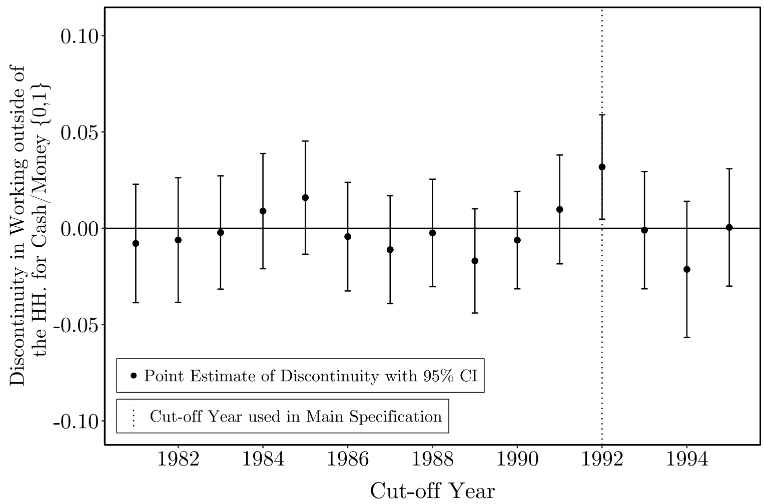


Figure B.12: Placebo Discontinuity, Works outside of the Household for Cash/Money {0,1}

Notes: Identically produced as Figure B.5 but replacing the outcome variable.

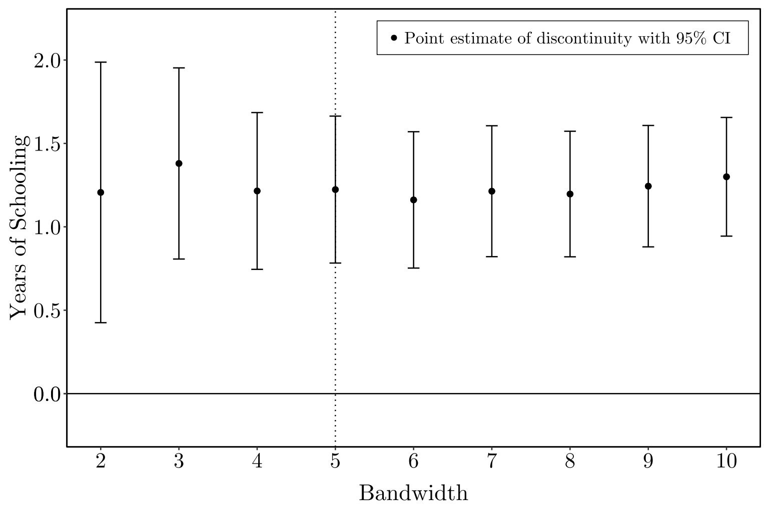


Figure B.13: First-Stage Discontinuity by Bandwidth

Notes: The figure plots separate first-stage regression estimates produced via equation (3.1) for various sample bandwidths. See Section 3.5 for a more detailed discussion.

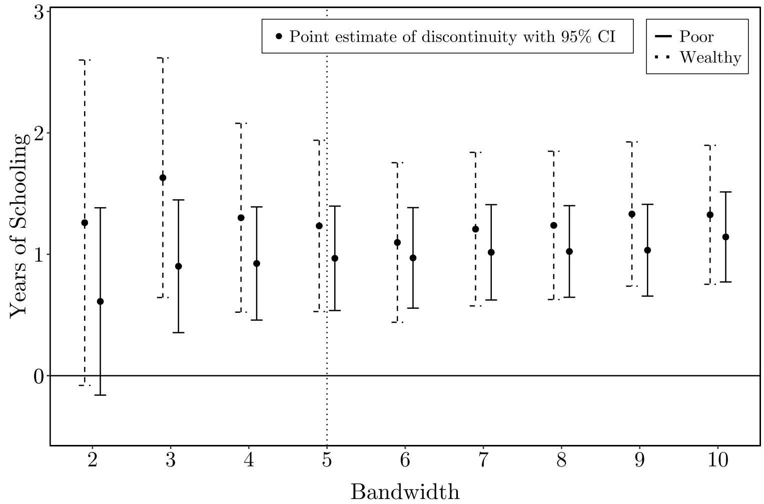


Figure B.14: First-Stage Discontinuity by Bandwidth, Separated by Wealth

Notes: Identically produced as Figure B.13 but using the "poor" and "wealthy" subsamples, respectively.

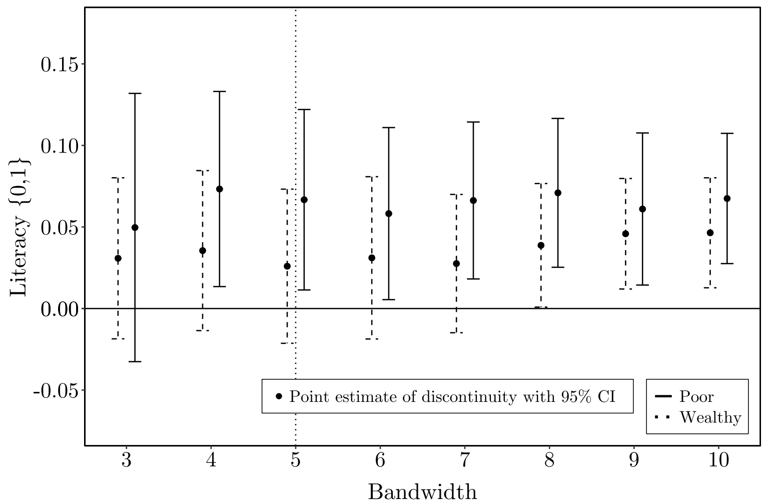


Figure B.15: Second-Stage Estimate by Bandwidth, Literacy {0,1}

Notes: The figure plots separate second-stage regression estimates produced via equation (3.2) for various sample bandwidths for both "Poor" and "Wealthy" subsamples. See Section 3.5 for a more detailed discussion.

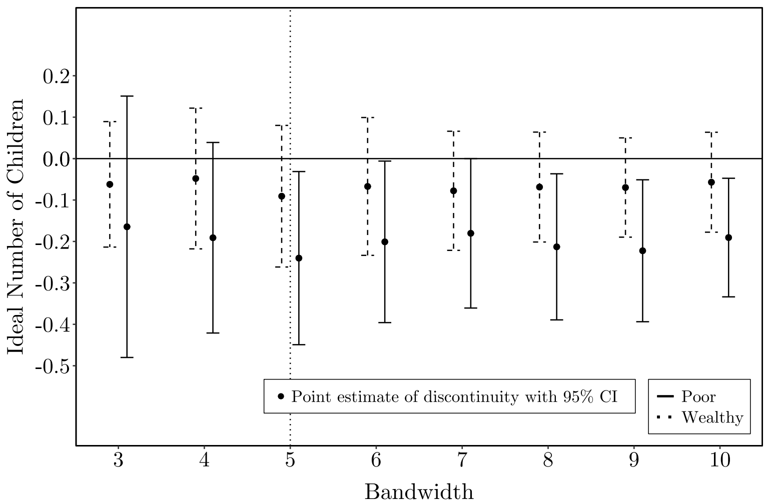


Figure B.16: Second-Stage Estimate by Bandwidth, Ideal Number of Children

Notes: Identically produced as Figure B.15 but replacing the outcome variable.

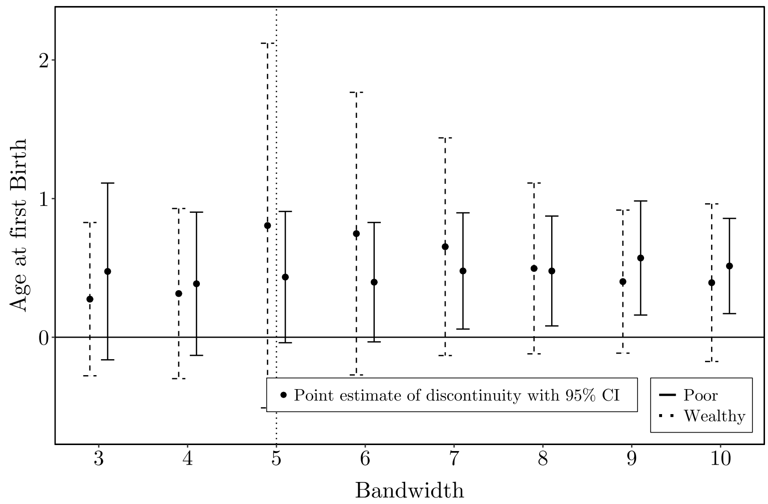


Figure B.17: Second-Stage Estimate by Bandwidth, Age at first Birth

Notes: Identically produced as Figure B.15 but replacing the outcome variable.

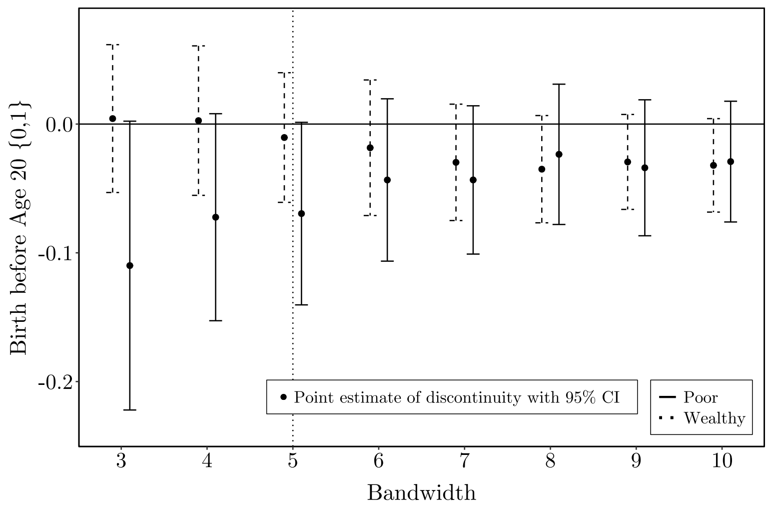


Figure B.18: Second-Stage Estimate by Bandwidth, Birth before Age 20 {0,1}

Notes: Identically produced as Figure B.15 but replacing the outcome variable.

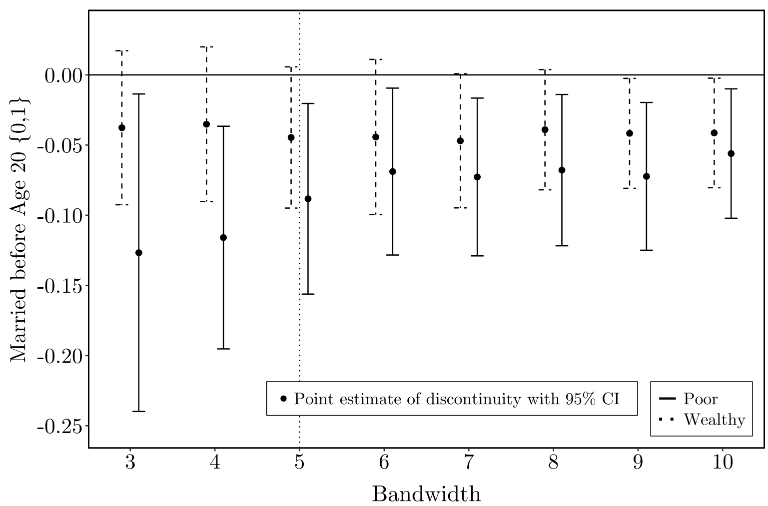


Figure B.19: Second-Stage Estimate by Bandwidth, Married before Age 20 {0,1}

Notes: Identically produced as Figure B.15 but replacing the outcome variable.

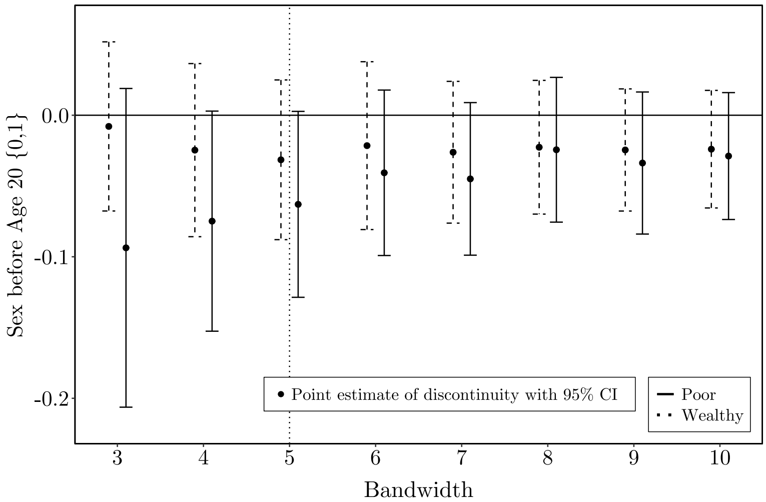


Figure B.20: Second-Stage Estimate by Bandwidth, Sex before Age 20 {0,1}

Notes: Identically produced as Figure B.15 but replacing the outcome variable.

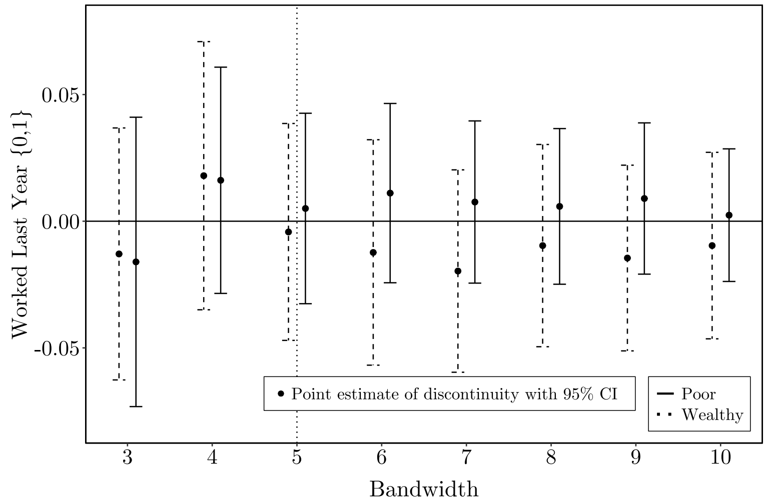


Figure B.21: Second-Stage Estimate by Bandwidth, Worked Last Year {0,1}

Notes: Identically produced as Figure B.15 but replacing the outcome variable.

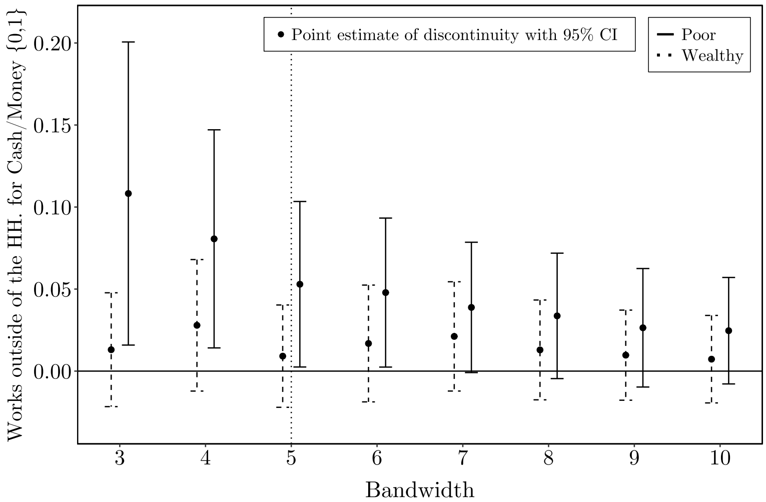


Figure B.22: Second-Stage Estimate by Bandwidth, Works outside of HH. for Cash {0,1}

Notes: Identically produced as Figure B.15 but replacing the outcome variable.

Chapter 4

Regional Market Integration and Household Welfare: Spatial Evidence from the East African Community[1]

Abstract. The distributional consequences of trade liberalization in Africa are under-researched. In this paper, I investigate the differential impact of the East African Community (EAC) on household welfare using three distinct sets of longitudinal, geo-referenced household-level surveys from the three founding members Kenya, Tanzania and Uganda. I thereby treat the re-establishment of the EAC in 2001 – and the expansion to a customs union and common market in 2005 and 2010, respectively – as a regional policy intervention having differential effects on individual households governed by their geo-spatial location within the countries, a prediction I derive formally from a canonical New Economic Geography (NEG) model. The model accommodates a distinct spatial feature of the EAC, namely that all its members host economically dominating agglomerations that are tucked away geographically from the regions bordering the respective community members. To empirically test the hypotheses drawn from the model, I employ a reduced form, difference-in-differences specification with treatment intensity given by households' road distance to internal EAC border crossings, effectively comparing household welfare between "remote" and "border" households (first difference) before and after the intervention (second difference), while allowing for a corresponding differential trend of outcomes over time in the pre-existing economic hubs. Results reveal that households located closer to the internal EAC border did not experience relative welfare effects compared to other regions following the re-establishment. Rather, the findings hint at increased agglomeration in the "core", as measured by an increase in consumption and population density in the pre-existing economic hubs. I observe mixed evidence on extensive as well as intensive labor market outcomes.

JEL Classification: F14, F15, R12, O15, O55

Keywords: FTA, East African Community, New Economic Geography, Difference-in-Differences

[1] *Acknowledgements.* Funding from the DFG (project number 390713894) is gratefully acknowledged. I thank the participants at the American Economic Association Annual Meeting 2024 (San Antonio, USA), at the Centre for the Study of African Economies Conference 2024 (Oxford, UK), at the EPCS Annual Meeting 2024 (Vienna, Austria), at the African Studies the IPORA-EXC52 Workshop 2023 (Bordeaux, France), at the 2023 German Economic Association (VFS) Annual Meeting (Regensburg, Germany), at the German Development Economics Conference 2023 (Dresden, Germany), at the 2023 BEST conference (Brisbane, Australia), at the African Studies Association of Africa Bi-Annual Meeting 2022 (Cape Town, South Africa), at the LaTrobe-Bayreuth Roundtable 2021 (Melbourne, Australia), at the EXC52 Cluster of Excellence Annual Conference 2023 and 2021 (Bayreuth, Germany), and at the Biannual Graduate Seminar of the faculty of economics at the University of Bayreuth 2022 and 2021 for their insightful comments and remarks at various stages of the project. I also thank Lara Bieske and Dimitrios Theodoridis for their extensive assistance in the gathering of data. Last, I thank Axel Dreher for his specific comments on an early draft. All remaining errors are my own.

4.1 Introduction

> "[...] the Community shall ensure: [...] the strengthening and consolidation of co-operation in agreed fields that would lead to equitable economic development within the Partner States and which would in turn, raise the standard of living and improve the quality of life of their populations."
>
> Treaty of the East African Community (EAC 1999; 12)

Regional economic integration is widely regarded as welfare-enhancing and has been a specifically popular policy intervention in developing economies (Schiff and Winters 2003). Particularly in Africa, deepened cooperation and trade has been suggested as ways to alleviate several barriers to development such as landlockedness, fragmented national markets as well as poor transport- and communications infrastructure (United Nations Development Programme 2011; World Bank 2020). Research on the aggregate effects of trade and integration generally supports such sentiments and point to largely positive (long-run) effects of trade liberalization (see e.g. Frankel and Romer 1999; Feyrer 2019). However, donor agencies have been emphasizing the potentially inequality-enhancing impact of trade within countries (e.g. World Bank 2009), and there now exists a well-established literature which studies these distributional concerns and provides evidence for them (for an overview see Pavcnik 2017). Heterogeneities may form along factors such as the composition of labor markets (e.g. import-competing vs. export-oriented), the income and consumption patterns of households, worker and capital mobility, and the nature of the distortions affected, among others (Winters et al. 2004; Goldberg and Pavcnik 2007; Winters and Martuscelli 2014). One aspect which has received particular attention is the spatial consequence of trade liberalization, i.e. the question what happens to countries' internal economic geography in response to external trade liberalization (for an overview see Brülhart 2011; Redding 2022).

Regarding developing economies, the evidence on such distributive effects mainly stem from liberalization experiences in Asia or the Americas, with Mexico and India forming prominent country-cases (for an overview see Engel et al. 2021; Barros and Martínez-Zarzoso 2022). In Africa, similar assessments have not been explored until recently and are split along analyzing household-level evidence without specific regard for (intra-national) space (see Erten et al. 2019; McCaig and McMillan 2020; Giovannetti et al. 2022)[2], or for a spatially motivated analysis, rely on the use of economic proxies such as light emitted by night (e.g. Cadot et al. 2015; Brülhart et al. 2017; Eberhard-Ruiz and Moradi 2019).

In this paper, using a distinct set of geo-referenced household-level surveys, I provide novel evidence on the distributional effects of regional trade liberalization in Africa by

[2] Here, exposure is typically defined at an administrative boundary and differentiated by the relative composition of specific industries within these regions.

combining the spatial considerations of market integration with a household-level analysis. I thereby treat the re-establishment of the East African Community (EAC) by Kenya, Tanzania and Uganda in 2001 – and the expansion to a customs union and common market in 2005 and 2010, respectively – as a regional policy intervention having differential effects on individual households governed by their geo-spatial location within the countries. I derive this prediction from a canonical New Economic Geography (NEG) model which fits the distinct spatial layout of the EAC. As such, and lending myself to previous variants of NEG, the model is put together to accommodate the stark spatial heterogeneity present in all EAC countries, given by the presence of economically dominating primate cities (Nairobi, Dar es Salaam and Kampala), and further incorporates the fact that these pre-existing economic agglomerations are tucked away geographically as seen by their relative positioning to their EAC counterparts. A robust result from the simulations of the model is that as trade costs among member countries decrease, the relative welfare in regions closer to the new markets, i.e. border regions, increase, which corroborates previous theoretical as well as empirical results. Given that all three countries have long hosted the preeminent economic agglomerations away from the borders, regional market integration in the EAC is thereby predicted to act as a dispersion force and to decrease previous spatial inequalities. These predictions are brought to an empirical test using a distinct set of geo-referenced household surveys before and after trade liberalization. Specifically, I employ a reduced-form difference-in-differences specification with treatment intensity given by households' road distance to internal EAC border crossings, effectively comparing household welfare between "remote" and "border" households (first difference) before and after the intervention (second difference), and also test for a similar differential trend across these time periods for households living in "core" agglomerations.[3]

The empirical results show that households living closer to the internal EAC border did not experience the expected relative welfare gains following the establishment of the EAC as measured by an array of consumption measures as well as intensive and extensive labor market outcomes. Rather, I observe that regional market accession within the EAC had a statistically significant and economically relevant effect on households living in the preeminent interior economic hubs. For instance, in comparison to pre-EAC periods, households surveyed in the core increased their consumption of consumer durables by 12% relative to those of all other regions. Further, they show a relevant decrease in the occurrence of lived poverty given by deprivations in basic consumption such as food, water and medicine by 32%. I observe mixed evidence on labor market outcomes from the cross-sectional data employed. However, the panel dataset shows that individuals in core agglomerations had an 11% increase in the likelihood of working in more skill-intensive occupations and also, show evidence of having salaried work. Corresponding to these relative shifts in welfare, I also document a strong increase in population density by up to one standard deviation in the years after these observed welfare effects, which

[3] From now on, the predominant economic hubs are interchangeably referred to as "core" or "interior".

provides evidence on increased agglomeration following the establishment of the EAC. The results are robust to an array of empirical modifications, extensions as well as validity tests.

As such, these findings go against the general prediction of the present theoretical simulations and also against the hypothesis prominently outlined in Krugman and Elizondo (1996), which predicted a dispersion of the concentrated economic activity of developing countries following liberalization. My results are also in contrast to other recent empirical findings, which have regularly documented regions closer to the new market (potential) to profit from the less costly access to them (for an overview see Brülhart 2011). Most notably, the present results are at odds with Eberhard-Ruiz and Moradi (2019) who provide (remote sensing) evidence that the EAC's re-establishment led to a positive (one-off) effect on city growth for cities closer to the internal EAC borders. While in disagreement with the general prediction of the model and these previous contributions, the model allows a potential insight into the reasons why such results may come about. In short, the presence of economically dominating interiors in all partner countries potentially weakens the dispersion force of market integration to the point where a regional pattern outlined by stark inequalities as existent prior to the EAC remains a possible outcome even after regional trade liberalization. Such explanations remain contentious, though, as the reduced-form empirics are ultimately not apt to inform on specific parameter coefficients which would permit isolating a key factor driving these observed effects.

The remainder of the paper is structured as follows. Section 4.2 introduces the relevant empirical and theoretical literature. Section 4.3 provides the institutional background of regional market integration in the EAC. Section 4.4 builds and simulates a New Economic Geography model which incorporates the stylized facts of the EAC. The spatial corollaries of regional integration drawn from the model motivate the empirical strategy introduced in Section 4.5. Section 4.6 introduces the data. Section 4.7 provides the main empirical results, robustness checks and validity tests, as well as several extensions to the results. Section 4.8 offers concluding remarks.

4.2 Related Literature

The paper relates to the body of research investigating the impact of trade on households and welfare on the one hand, and the literature analyzing the spatial consequences of trade liberalization on the other.

4.2.1. Trade and Household Welfare

Increased availability of detailed survey data has aided the growth of the literature assessing the link between trade liberalization and household welfare (for an overview see Goldberg and Pavcnik 2007; Winters and Martuscelli 2014; Pavcnik 2017; Barros and Martínez-Zarzoso 2022). This body of research confirms the notion that trade does not unequivocally increase the welfare of all households within a country, i.e. produces winners and losers. An analytical starting point in thinking about these heterogeneous effects is given by a stylized production-consumption schedule of households and may be encapsulated by $\Delta W = (q_i - c_i)\Delta p_i$, whereby welfare changes are explicitly moderated by (trade-induced) price changes (see Deaton 1997; Winters et al. 2004). Depending on whether the household is a net consumer (c_i) or producer (q_i) of product i, a given price change Δp_i will either lead to net benefits or net losses. In his seminal paper, Porto (2006) extends such partial equilibrium statics to a general equilibrium model of trade, taking into account the simultaneous changes in prices of non-traded goods, and subsequently, second-round effects resultant of altered factor- rewards and intensities in specific industries. These dynamics are relevant in cases where specific sectors are facing increased import-competition from international exporters or where export-oriented produces are drawing increased demand from abroad. As such, one and the same trade policy may render very different results across the population depending on the goods affected, households' production and consumption schedule, and subsequent general equilibrium effects.

Porto's approach has been subsequently extended and employed to study trade effects in various contexts, including Mexico (Nicita 2009), Brazil (Borraz et al. 2013), India (Marchand 2012; 2019), Tunisia (Martínez-Zarzoso et al. 2016), as well as in six African countries (Nicita et al. 2014). These studies typically employ changes in (non-)traded goods prices together with their respective income- and consumption share reported in household surveys. To assess the overall welfare impact, these changes are then compared across the (income or expenditure) distribution to assess the pro-poor or pro-rich character of a trade-policy. Most of these studies provide evidence of a pro-poor effect of trade, some of them showing mixed results, and Nicita (2009) being the only exception in showing a clear "rich-only" impact of trade in Mexico.

A second branch of the literature on trade and household welfare has relied on "Bartik-style" shift-share instruments to identify trade effects.[4] Here, exposure is typically defined at an aggregate level, such as at a particular administrative unit (e.g. districts or regions). The intensity of trade on households living within a specific region is then differentiated by the pre-liberalization concentration of industries and the respective tariff cuts (see Goldberg and Pavcnik 2007; Winters and Martuscelli 2014). For instance, McCaig (2011) shows that the U.S.-Vietnam Bilateral Trade Agreement accelerated poverty decline, as export growth due to tariff removal was largest in the low-skilled labor-intensive apparel and clothing sectors. On the other hand, Topalova (2010) provides evidence that India's trade liberalization of 1991 actually slowed poverty decline in the most affected regions, i.e. the ones intensive in agriculture, given that such sectors faced increased import-competition. Related studies have looked at similar issues in Brazil (Castilho et al. 2012), China (Emran and Hou 2013), India (Edmonds et al. 2010), Indonesia (Kis-Katos and Sparrow 2015), and Vietnam (Fukase 2013; Vo and Nguyen 2020). These studies are mixed in finding both decreases as well as increases in relative poverty

[4] As introduced by Bartik (1991) as well as Blanchard and Katz (1992).

as measured by indicators such as consumption, employment, or wages. In Africa, the evidence on liberalization experiences in this literature is almost universally negative. For instance, drawing on South Africa's trade liberalization of the 1990s, Erten et al. (2019) find decreased formal as well as informal employment for more affected regions and no effects on wages for those remaining employed. Relatedly, McCaig and McMillan (2020) find neither a contraction nor an expansion of industries in neighboring Botswana, which was affected by the same liberalization schedule.[5] Rather, they report higher likelihoods of being employed informally for more intensely affected regions. In the same vein, evidence from Ethiopia suggests increased unemployment levels in regions more exposed to trade liberalization and import competition in light of the Structural Adjustment Programs (SAP) of the early 1990s (Giovannetti et al. 2021). One exemption to these findings is Giovannetti et al. (2022) who provide evidence of a negative effect of protective policies in Egypt shortly after the Spring Revolution. Interestingly, they find neither positive nor negative results of trade liberalization in the preceding decades.

To my knowledge, there exists no study analyzing household-level welfare concerns of trade liberalization in Africa from a spatial point of view.[6] One exception to this is Cali (2014), who assesses Uganda's progressive liberalization policy with Kenya in the 1990s on wage premia, i.e. changing returns to schooling.[7] However, the variation across space is given at a district level (comparable to GADM[8] classification of 2), of which there are a total of 38 and 45 in the study across the two survey rounds, respectively. For comparison, in this paper, households' location is defined by latitude-longitude combinations comparable to GADM3 or finer. As such, I draw from a minimum of 104 and a mean of 324 GPS locations per country per round, or an average of 299, 326, 353 for Uganda, Tanzania and Kenya, respectively. As such, analyzing the spatial response of household welfare to trade liberalization (with higher precision) represents a relevant research gap I aim to fill. Motivating differential trade effects across space requires an overview of the relevant theoretical and empirical findings in this regard, which is provided in the next section.

4.2.2. Spatial Effects of Trade

The second strand of literature to which this paper relates investigates the spatial consequences of trade liberalization. This growing body of research has its roots in New Economic Geography (NEG) and has extended to an active field now better referred to as "quantitative spatial economics" (for an overview see Redding and Rossi-Hansberg 2017; Brakman et al. 2019; Redding 2022).

---

[5] Botswana is a member of the South Africa Customs Union (SACU).
[6] I was not able to identify a study exploiting the geo-referencing of survey locales to study these links on any continent for that matter.
[7] The analysis is motivated by a Hecksher-Ohlin type trade effects, thereby suggesting decreasing wage inequality in a developing country who is labor abundant and human capital scarce.
[8] As given by the Database of a Global Administrative Areas https://gadm.org/.

Krugman's (1991) seminal paper was a crucial expansion on earlier conceptualizations of spatial economic distribution, which mainly concentrated on allocations within cities, such as the von-Thünen model (1826), or the relative size of cities (Henderson 1974, 1982). The advantage of NEG in comparison to these earlier specifications lies in the fact that it can explain the spatial distribution of cities against each other on a featureless plane such that there are not simply "floating islands" (Brakman et al. 2019; 3). Krugman's model is essentially based on new-trade-theory (Krugman 1979, 1980) and combines monopolistic competition (Dixit and Stiglitz 1977) with increasing returns to scale. Most importantly, trade costs factor in between locations, regulating their spatial allocation against each other (Krugman 1991). Hence, the endogenous allocation of activity ultimately boils down to producer- and consumer maximization problems, who optimize over a given set of preferences and production technology, while factoring in trade costs. Agglomeration is then a product of cost (forward) and demand (backward) linkages which produce centripetal forces, while dispersion is a product of increased competition, the costs of urban congestion, or immobile factors of production. For instance, because firms operate under increasing returns to scale and incur transport costs, they benefit from the increased demand in larger locations, i.e. move where demand is highest (demand linkage).[9] And given that consumers have a "love of variety" and will additionally save on higher price tag for shipping, consumers prefer to locate close to (a large number) of producers (cost linkage). However, while large regions offer firms high demand and consumers lower prices, competition as well as costs of congestion (commuting, land rents) are increased which decreases agglomeration tendencies. In the long-run, an equilibrium is given by the balance of these forces, i.e. when the advantages and disadvantages of agglomeration or dispersion, expressed in real wages, are net zero. In this scenario, there exists no incentive for firms or workers to relocate.

This endogenization of the spatial allocation of economic activity has provided a workhorse model and spurred subsequent extensions and applications to questions on how spatial inequalities form and how they may be affected. Importantly, NEG allows the comparative statics examination of what happens to the centrifugal and centripetal tensions in response to changes in *internal transportation costs* or, importantly, *external trade costs* (for a synthesis see Fujita et al. 2001).[10] Concerning the latter, both theoretical and empirical results vary in their prediction of whether liberalization increases or decreases spatial disparities within countries (for an overview see Brülhart 2011). Krugman and Elizondo's initial treatment (1996) famously predicted the dissolving of the "giant Third World metropolis" of developing countries

[9] Note that in large markets, the additional presence of a firm increases demand mechanically, and by being able to pay higher wages, thereby further strengthening the backward linkage.

[10] While the core model of NEG is known for its "bang-bang" property for changes in transport costs, i.e. equilibria between complete spreading or agglomeration, subsequent adaptations have accommodated a wider range of equilibria, using stronger centrifugal (dispersion) forces such as interregional labor immobility (e.g. Krugman and Venables 1995), diminishing returns in the non-traded sector (e.g. Puga 1999), or housing (e.g. Helpman 1998).

in response to external trade liberalization. The model extends the stylized two-region case to a three-region-economy, with two regions situated in the home country, and one region ("the rest of the world") posing as the international market to which trade costs are successively lowered (a 2+1 economy). Krugman and Elizondo (1996) sparked an array of refinements and extensions to this basic setup. Interestingly, however, the prediction from these theoretical advancements is far from uniform. While Behrens et al. (2003, 2007) confirm the original prediction, several adaptations arrive at the contrary result, i.e. that increased trade liberalization sparks intra-national agglomeration. For instance, in the same original 2+1 setup, Paluzie (2001) as well as Brülhart et al. (2004) and Crozet and Koenig (2004) provide evidence of increased agglomeration in response to external trade liberalization. Further studies have extended the setup to 2+2 economies, confirming these predictions (Monfort and Nicolini 2000; Monfort and van Ypersele 2003). The difference among all of these studies is how they chose key elements from the "menu of building blocks" (Redding and Rossi-Hansberg 2017; 25), i.e. how consumer preferences (CES or quasilinear) as well as dispersion forces (immobile workers vs. congestion) are modeled.[11] One particularly interesting adaptation of this literature is to allow for *heterogeneous intra-national space*, i.e. regions (within-countries) to differ from one another ex-ante in terms of their access to foreign markets. For instance, in Mansori (2003), Brülhart et al. (2004), Crozet and Koenig (2004) and Behrens et al. (2006) they additionally test what happens to the prediction if one region has better access to the international market than the other, i.e. poses as a "border" or "gate" region. What this class of models show is that in almost all instances, external trade liberalization leads to increased "draw" to the border, i.e. to the region with the better foreign market access (Crozet and Koenig 2004b). However, depending on the relative size and the export intensity of the home and foreign markets, this draw to the border may be alleviated as the interior acts as a shield to foreign competition (Brülhart et al. 2004). These effects may be further mediated by varying intra-national transport costs which regulate the pass-through of changes in international trade costs towards the interior as well as the symmetry of the foreign country (Behrens et al. 2006). These initial refinements to asymmetric regions were first steps into what is now more richly embodied in "quantitative spatial economics" whereby first-nature characteristics (e.g. local endowments such as productivity, amenities or floor space) are paired with the "classical" second-nature agglomeration and dispersion forces, which are produced by the endogenous relative position of agents against each other (see for a distinction Redding 2022).

The empirical evidence reflects the ambiguity shown across these models. While evidence from cross-country settings lean towards the convergence of economic activity in response to trade liberalization, within-country evidence has shown increasing inequalities for various settings (see for an overview Brülhart 2011). However, a rather robust result across the

[11] The menu additionally consists of choices on building blocks such as production technology, trade costs, externalities, labor mobility, as well as an endowment structure across regions, among others.

empirical literature is that regions with relatively better access to foreign markets, often border regions or regions near the coast, generally stand to benefit comparatively more. This mirrors the theoretical results of the class of models with heterogeneous intra-national space. Naturally, whether this leads to convergence or divergence of economic activity within countries naturally depends on the pre-liberalization diffusion of economic activity. For instance, convergence is found to occur in settings where market access is higher in the historically economically weaker border regions, as was the case in Austria (Brülhart et al. 2012) or Germany (Redding and Sturm 2008).[12] On the other hand, divergence is found somewhat more frequently, as documented by the increasing activity to the already industrialized U.S.-Mexican border following NAFTA (Hanson 1994, 1997), or in China, where trade has benefitted the already more developed coastal areas (Kanbur and Zhang 2005). Next to singular country cases, a growing field of literature employs large-scale evidence from satellite imagery, where lights emitted by night serve as a proxy for economic activity to assess factors contributing to spatial (within-country) inequality.[13] So far, much of the evidence has a tendency for trade to increase within-country inequality, and particularly so in developing regions (Ezcurra and Rodríguez-Pose 2014; Hirte et al. 2020; Ezcurra and Del Villar 2021).[14]

Within-country evidence for Africa is scarce and is mostly conflated with the mentioned large-scale studies of all world regions. Particular country-case investigations in Africa so far have also exclusively relied on nighttime lights as a source of data across space. For instance, Cadot et al. (2015) look at the influence of improved trade on the border shadow in sub-Saharan Africa. Similarly, Brülhart et al. (2017) estimate this border shadow for Uganda and Rwanda in specific. Analyzing trade facilitation measures such as "one-stop-border-posts", both papers show a tendency for trade to decrease previous spatial inequalities, i.e. benefit the regions now better connected. Lastly, and most relevant to the present study, Eberhard-Ruiz and Moradi (2019) analyze the impact of the East African Community on city growth within Kenya, Tanzania and Uganda. Using remote-sensing data from before as well as after the re-establishment of the EAC, the authors provide evidence for a short-whiled, relative increase in observed nightlight growth for cities closer to internal borders.

As noted in the previous section, there exists no study for the African continent which exploits the geo-referencing of household surveys to analyze (spatial) effects of trade. However, evidently, there are distinct benefits in using household-level data to measure such distributive effects. First, it allows to directly analyze consumption of households, which has been argued to provide a better measure of overall welfare if one assumes consumption smoothing (Deaton 1997). In a similar vein, trade policies tend to alter prices in a non-uniform way which

[12] Redding and Sturm (2008) show evidence for population movements west, increasing regional inequality in Germany. The reason for my conclusion is that the population movement was induced by market loss of border regions, rather than market gain, which would vice-versa lead to the opposite result.
[13] See e.g. Gibson et al. (2020) on the various uses of night light data in economics.
[14] One exception is Brülhart et al. (2019) who show that trade expansion reduces "border shadows".

(differentially) affects both income and consumption, the effect of which is better captured by consumption as an outcome of both (Goldberg and Pavcnik 2007). Third, the use of household surveys may also allow us to analyze migration patterns and to test for a potential selection into specific regions (see Subsection 4.7.2). And while household surveys also have some distinct weaknesses such as lack of depth in variables, consistency, and comparability across time, or (non-)response rates, they may also provide a countercheck on the results drawn from using remote sensing data, e.g. lights emitted by night. Recent research on the quality of nightlight data has cautioned of the quick application, particularly for developing contexts. Results have suggested that precisely the areas relevant to development economists, i.e. low density, rural (agricultural) areas, are due to non-negligible measurement issues (e.g. Bickenbach et al. 2016; Gibson et al. 2020, 2021). Most relevant to our case, studies have shown that nightlight-to-GDP elasticities may differ largely between rural and urban areas, which may lead to conflating a systematic measurement error with policy impacts (Bluhm and McCord 2022).

4.3 Institutional Background

The East African Community (EAC) was originally founded by the Republics of Kenya, Tanzania, and Uganda in 1967. Placed around Lake Victoria in East Africa, the three countries share two common borders each and economic and political cooperation between the countries has historical roots. In pre-independence periods, roughly from 1900-1960, they shared large infrastructure outlays such as railways, telecommunication, postal service and a common currency (Hazlewood 1979; EAC 1999). However, not soon after the first formal treaty towards the establishment of an East African Community was signed in 1967, questions on sovereignty, and particularly the "disproportionate sharing of benefits of the Community among the Partner States" arose (EAC 1999; 1). While attempts at redistribution of benefits were made, it was deemed insufficient by the member states and trade restrictions were levied between them even while formally in union (Mugomba 1978; Hazlewood 1979). Next to a "lack of strong political will" (EAC 1999; 1), these are often cited reasons for the ultimate demise of the original EAC in 1977 when it was formally dissolved. However, the mutual interest in working together in a union was kept alive in the decades thereafter, as seen by the gradual move towards the modern EAC for instance by the establishment of the "Permanent Tripartite Commission for East African Cooperation" in 1993 or the "East African Cooperation Development Strategy" in 1997, which focused on closer co-ordination in economic, political, fiscal, immigration, infrastructural as well as social and cultural arenas (EAC 1999).

The institutional establishment of the modern day East African Community was initiated with the treaty of 1999, which was ratified on July 7th of 2000, and the new EAC began to operate as a free trade area on January 15th of 2001 (EAC 1999; Kaahwa 2003). Hence, it was not before 2001 after which the substantial lowering of tariff rates by member states was initiated.[15] The EAC consistently moved towards deeper integration in the years thereafter, with the protocol for a customs union operational from the 1st of January 2005 followed by a transitional period to a common market on the 1st of July in 2010. While member states have since ratified the move towards a monetary union in 2013, a common currency has not been implemented as of yet. Figure 4.1 below depicts these developments quantitatively, by plotting the simple (unweighted) average tariffs among the EAC founding members together with their total merchandise and manufacturing trade in mUSD from 1995 to 2020 (UNCTAD 2022; UNSD 2022).[16]

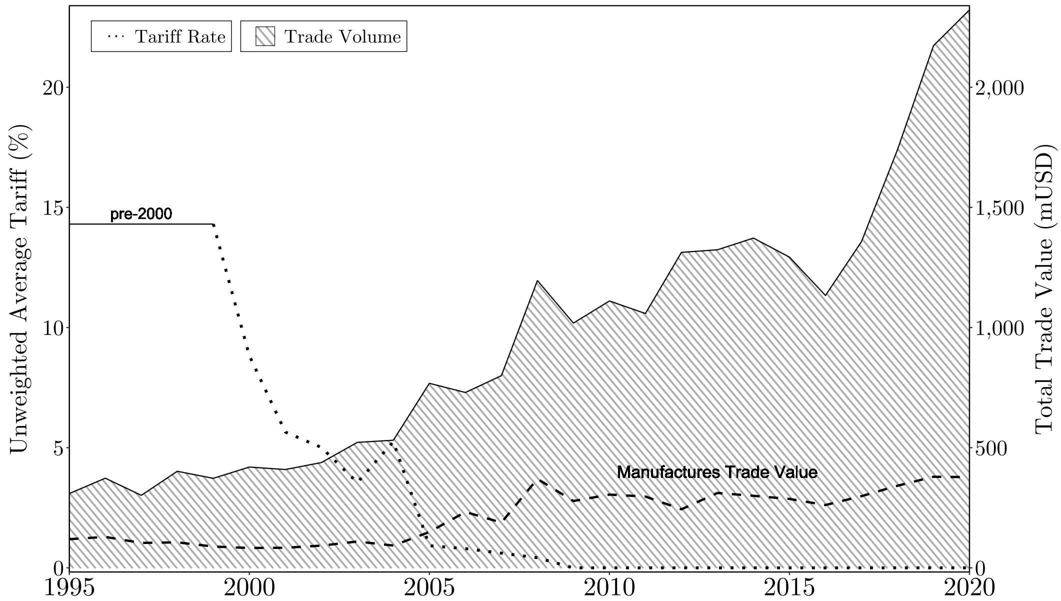


Figure 4.1: Tariffs and Trade in the East African Community (EAC)

Notes: The plot displays trade volumes and tariff rates among the three EAC founding members Kenya, Tanzania and Uganda. The unweighted average tariff on the dotted line represents the "AHS simple average tariff" and comes from the United Nations Conference on Trade and Development (UNCTAD) Trade Analysis Information System (TRAINS). The values between 1995 and 2000 are irregularly reported and stem from a single country only, hence why the average is taken. The trade volume depicted by the shaded area represents the total USD value of all goods imported within the EAC member states (net of re-imports) and stems from the United Nations Commodity Trade Statistics Database. "Manufactures Trade Value" is comprised of trade of merchandise classified under Section 6 of the SITC Revision 3. Both are facilitated via the World Integrated Trade Solution (WITS) platform.

The EAC has also expanded outwards to contiguous countries of the region, with the accession of Burundi and Rwanda in 2007, South Sudan in 2016, the Democratic Republic of Congo in 2022 and also Somalia most recently in 2023. However, the three founding members

[15] For instance, Tanzania postponed many substantial tariff line removals to the budgetary year beginning July 1st 2001 and for sugar even until July 1st of 2002, which was the 7th highest valued import in the years between 1996 and 2001 of all 96 chapters in the H1 nomenclature (UNCTAD 2022). See also Eberhard-Ruiz and Moradi (2019) for a more detailed account on the tariff structure around the implementation.

[16] The numbers reported reflect current dollar values of the respective year. We use import values as there are some gaps in the reporting of exports values.

still account for the overwhelming majority of economic activity with over 70% of the EAC's total GDP in 2022. As such and given that I aim to evaluate effects over the entire timeline of the modern EAC, this paper concentrates primarily on evidence drawn from the founding members Kenya, Tanzania, and Uganda.[17] Since the establishment of the EAC in 2001 the three economies have grown economically by a total of 111%, 86% and 47%, respectively. However, with a per capita GDP (PPP) of 2,624$ and 2,280$ two of the three countries, Tanzania and Uganda, are still categorized as "low-income-countries".[18] Only Kenya has graduated to a "lower-middle-income country" as of 2014, with a current GDP per capita of 4,882$. Importantly, population growth within these countries over the same time period roughly doubled from 90 million in 2000 to 166m in 2022.[19] Concerning the economic structure of the countries, they are heavily reliant on agriculture and services. In Kenya, the service sector makes up a total of 48% of the GDP, followed by agriculture (38%) and manufacturing (9%). Services also dominate in Tanzania (40% of GDP), who hosts a large tourism sector, with agriculture making up 32%. Manufacturing is not as important in Tanzania with a contribution to total GDP of 6%. In Uganda, the respective figures total to 27%, 52% and 9% (WTO 2019).

Concerning trade, merchandise exports display a relevant contribution to the economies' GDP with shares of 25%, 28 %, and 29%, in Kenya, Tanzania and Uganda, respectively (World Bank 2022). However, regarding the direction of trade, the large majority of merchandise is still sourced and exported to markets outside of the continent, with extra-African export and import shares of 40-50% and 80-90%, respectively. China, India, as well as markets in the Middle East and the EU have been the predominant trading partners within the last decade (WTO 2019).[20] As such, the share of intra-community ("intra-EAC") trade has been relatively low, hovering around 10% of total trade since its establishment in 2000 (UNU-CRIS 2019).[21] Some of the reasons for the relatively low volumes of (documented) intra-regional trade in the EAC are outlined by the substitutability of goods produced, several non-tariff barriers of trade, infrastructural shortcomings but also the importance of informal cross-border trade (WTO 2019). However, compared to the other eight officially recognized regional economic communities (RECs) on the continent, the EAC has the second highest intra-regional trade share, trailing only the Southern African Development Community (SADC) whose members' intra-regional trade account for 20% of total trade.[22] Further, there is a significant asymmetry

[17] Also owing to data availability, which is particularly scant for Burundi and Rwanda in the decades preceding and following the re-establishment of the EAC.
[18] GDP figures are expressed in constant 2017 international (PPP) USD.
[19] The population of the three countries grew from 31m to 54m in Kenya, 34m to 65m in Tanzania and from 24m to 47m in Uganda.
[20] The main goods exported are primary products (mainly agriculture produce) which make up 60%, 61%, and 43% of total export value across the three countries respectively, and those declared manufactures 28%, 18% and 18% (WTO 2019).
[21] Between 10 to 20% when including the trade in services (IMF 2023).
[22] The eight RECs have an average intra-regional trade share of 6% (UNU-CRIS 2019).

in the pattern of trade, as only 6% of imports are sourced within the EAC, but 20% of countries' exports are directed to markets of the EAC (WTO 2019). The predominant type of goods traded within the original EAC members are comprised of primary products such as mineral fuels and oils, gemstones as well as cereals but also manufactured goods such as rolled iron, steel and steel products, vehicles and electrical machinery, plastic goods, processed food and beverages as well as pharmaceuticals (UNCTAD 2022).

One particularly pertinent aspect of the three economies is their spatial configuration, which is outlined by exceedingly high levels of urban *economic* primacy.[23] While most of the population is dependent on agriculture and lives in rural environments (70% in 2022, down from 80% in 2000), the large majority of the countries' economic activity is concentrated in the three geographically limited primate cities Nairobi, Dar Es Salaam, and Kampala, respectively (World Bank 2022). For instance, around the time of the EAC's establishment in 2001, Dar es Salaam hosted only 7% of the country's population but 51% of formal employment of the private sector and contributed to over 57% of the total wage bill (Tanzania National Bureau of Statistics 2004, 2006).[24] Considering that the administrative region of Dar es Salaam makes up a mere 0.16% of Tanzania's total land area, this describes a large intra-national (spatial) discrepancy in economic activity. To compare, the next largest contributors to the wage bill in 2001 were Kilimanjaro, Arusha and Dodoma with 6%, 5% and 5%, respectively, and land shares of 1.5% and 4% and 5%. This pattern has continued to persist and is particularly pronounced in the high value-added manufacturing sector. For instance, in 2008, Dar es Salaam hosted 55% of manufacturing establishments (30% of the manufacturing labor force) while contributing over 51% of the country's total value added (Tanzania National Bureau of Statistics 2010). In the latest available survey of 2016, Dar es Salaam still contributed to over 41% to total value added, albeit hosting a smaller share of 27% of all manufacturing establishments, and 32% of the manufacturing workforce, which is however, well over twice the amount the next largest region Morogoro.[25] The structure of the EAC partner countries evinces the same spatial pattern. Concerning Kenya, Nairobi accounted for 46% of (formal) wage employment in 2001 and for 51% of the total wage bill among main towns (Kenya Central Bureau of Statistics 2003; Kenya National Bureau of Statistics 2011).[26] In 2009, almost a decade later, these figures were virtually unaltered. Together with the second largest industrial hub, Mombasa, these figures increase to over 63% and 69% in 2009 for the employment and wage bill, respectively. Again,

[23] Note the important difference from the "primacy" based purely on urban population shares most commonly discussed.

[24] For instance, around the time of the EAC's establishment in 2002, Dar es Salaam hosted only 7% of the country's mainland population but contributed to over 40% of the total wage bill and hosted 57% of total employment in the private sector (Tanzania National Bureau of Statistics 2006, 2007).

[25] Dar es Salaam is also a hub for large firms, hosting over 33% of all firms sized over 100 employees and 13 out of the 44 firms over 500 employees. The second largest region Pwani, which encloses Dar es Salaam geographically, hosts a mere 7% of such (Tanzania National Bureau of Statistics 2018).

[26] Earnings in informal sector and rural small scale agriculture as well as pastoralists activities are excluded (Kenya National Bureau of Statistics 2011; 236).

to compare, Nairobi makes up only 0.12% of the total land area and 7 (8%) of the population as per the census of 1999 (2009) (Kenya Central Bureau of Statistics 2001; Kenya National Bureau of Statistics 2011). Concerning the industrial structure, the main sectors clustered in Nairobi are manufacturing, construction, and financial services, and in 2009 Nairobi hosted 49% of all manufacturing employment and 51% of the total manufacturing wage bill (Kenya National Bureau of Statistics 2011, 2013).[27] And lastly, in Uganda, Kampala hosted 45% of all formal businesses establishments in 2001 and 2006, followed by Mbarara and Wakiso as the second largest industrial cities with a share of 5% each (Uganda Bureau of Statistics 2003, 2007). If one includes the "Central" region of Uganda which encloses Kampala geographically, the figure increases to 63% in 2001 and 65% in 2006.[28] Similar to Nairobi and Dar Es Salaam, Kampala contains the majority of the high value-added manufacturing sector with 42% of all firms operating in Kampala and 61% together with the central region in 2006. As such, the Kamapala region contributed to 47% of value added in 2006 and over 77% when including the central district (Uganda Bureau of Statistics 2006).[29] Similar to Nairobi and Dar es Salaam, Kampala only makes up 0.09% of the total land area and 5% (4%) of the population in 2002 (2014) (Uganda Bureau of Statistics 2016).[30]

Note that high rates of urbanization are a not a particularly surprising characteristic of developing countries and a research topic which has received increased attention as of late (Glaeser 2014; Jedwab and Vollrath 2015; Gollin et al. 2016). However, for the present study, next to its obvious connection to NEG models as introduced in Subsection 4.2.2, the urban primacy as described in the EAC is particularly relevant in at least two further aspects. For one, all three urban centers (Nairobi in Kenya, Dar es Salaam in Tanzania and Kampala in Uganda) are geographically tucked away as viewed from the common borders connecting the respective EAC partner state(s).[31] In the data, the average road distance to internal EAC border crossings for the three cities is 395km, 922km, and 269km, respectively.[32] The connection via road is particularly relevant for intra-EAC trade, as over 95% of the regional trade in the area is transported via the road network, and only 5% via rail (see Figure C.8 and Nathan Associates (2011)). Second, based purely on urban population shares, Kenya (30%) Tanzania (36%) and Uganda (26%) show for relatively low levels of urban *population primacy*, compared to most other (low & middle income) countries (World Bank 2022).[33] Paired with the skewed

[27] The respective figures for construction and financial services are 75% and 64% (Kenya National Bureau of Statistics 2011).
[28] Establishments with 5 employees or more. If one includes informal businesses, Kampala has contained 30% and 29% of all businesses in 2001 and 2011 and 60% and 59% when including the central region, respectively (Uganda Bureau of Statistics 2003, 2012).
[29] Kampala also hosts the majority of large firms with 40% of firms with 100 employees or more in 2006 (Uganda Bureau of Statistics 2007). The central region also had the largest increase in manufacturing businesses, with a 40% increase between 2001 to 2006.
[30] Together with the Central region, this increases to 20%.
[31] See Figure 4.2 of Subsection 4.4.2 for a stylized depiction of this spatial structure.
[32] The minimum distances to the nearest EAC border crossings are 152km, 389km and 185km, respectively.
[33] The share of urban population in high-income countries is 80%, the average of low- and middle-income countries is 52%. Latin America

economic activity towards primate cities of the EAC, this is somewhat of a deviation from the "urbanization without growth" hypothesis (Fay and Opal 2000). Lastly, recent research has suggested natural resource-dependency to cause the emergence of "consumption" rather than "production" cities (Gollin et al. 2016). Interestingly, in the last two decades, with the exception of Uganda, Kenya and Tanzania have ranked below the 10% cutoff of natural resource rents to GDP suggested in the paper (World Bank 2022). Hence, studying the household effects of trade liberalization where the countries opening up to trade are outlined by the described spatial structure provides a promising case which may inform the literature on the causes and consequences of urbanization in developing contexts.

The key takeaway of the section is that the three countries joining in on a common trade union are all outlined by stark regional economic inequalities, given by the presence of economic hubs positioned away from the common borders of the community. To which extent this prior spatial configuration has bearings on the effects of regional market integration is formalized and discussed in the following section, in which I build a quantitative spatial equilibrium that accommodates these stylized facts.

4.4 A Four Region Economy

To lay the theoretical groundwork on which to analyze the exposure of regional market integration in the EAC across space, this section develops a canonical, four-region spatial equilibrium model, which combines aspects from the models discussed in Subsection 4.2.2. The model is built on Krugman's (1991) core fundamentals while adding an external economy as introduced by Krugman and Elizondo (1996). However, rather than the 2+1 cases in which the external economy acts as one region (e.g. Paluzie 2001), I extend the foreign economy to two regions (2+2) as in Monfort and Nicolini (2000) and Zeng and Zhao (2010). Finally, the model is rendered unique as I tweak the structure of intra-national transport costs borrowing from the 2+1 models of Crozet and Koenig (2004) and Brülhart et al. (2012) such that regions within both countries are outlined by differential access to foreign markets, i.e. the model encapsulates heterogeneous intra-national space in both economies. The model thereby allows to additionally analyze the potential implications of *foreign* economic (in)equality on the *domestic* distribution of activity in the context of increasing regional integration.[34] As anticipated in the previous section, this accommodates the stylized facts spatial features in how trade may play out among the EAC member countries. Note that I refrain from

[34] To my knowledge, (Behrens et al. 2006) were the first to introduce such heterogeneous space in a > 3 region model and to analyze the potential effect of the foreign economic distribution on domestic outcomes. One point of departure to the present model is that they do not explicitly model both countries as being "gated", i.e. only one of the countries outlined by heterogeneous intra-national space. And second, they make us of a nonhomothetic preferences structure with a quasilinear utility introduced by (Ottaviano et al. 2002). Given an implied decreasing expenditure share on manufactures, this has been deemed potentially inappropriate in developing contexts (Fujita and Mori 2005) and I elect the standard utility function of the Cobb-Douglas / CES type of the original models.

computationally more involved multi-region approaches, as the 2+2 case with heterogeneous intra-national space encapsulates the stylized facts of the EAC and keeps the model tractable.

As such, we set-out with a four-region world economy consisting of R locations denoted by $r = \{1, 2, 3, 4\}$, where regions 1 and 2 are assigned to be in the "home" (from now interchangeably referred to as "domestic") country and regions 3 and 4 in the "foreign country". As will be specified in Subsection 4.4.2, regions 1 and 4 pose as the "core" (from now on interchangeably referred to as "interior") economic hubs while regions 2 and 3 are placed at the border to formalize the stylized facts introduced above. Note that for most of the analysis conducted across this section, I refer to effects on regions 1 and 2 of the domestic country. However, by symmetry between the two countries, the results readily translate into the view from the "foreign" country also, i.e. for regions 3 and 4.

Moving on with the model, there are two sectors in the economy, manufacturing, and agriculture. The latter sector is characterized by perfect competition and produces the homogenous agricultural good "food", F, under constant returns to scale using the immobile, inelastically supplied input "farmers". The modern manufacturing sector is characterized by monopolistic competition and thereby produces a variety of differentiated goods, "manufactures", using the input factor "workers" or "labor". Farmers and workers within each country are drawn from a total population mass L of which $L^{M(anufactures)} = \delta \cdot L$ are engaged in manufacturing and the rest $L^{F(ood)} = (1-\delta)L$ in agriculture, hence $0 < \delta < 1$. Manufacturing workers are mobile between regions but not across sectors or countries, i.e. only mobile between regions 1 and 2 or 3 and 4, respectively. As such, the total manufacturing workforce within countries is fixed, but workers allocate themselves endogenously across regions over time in response to real wage differentials. The respective shares of manufacturing of each region are given by λ_r, which satisfies $\sum_{i=r}^{R} \lambda_r = 1$. We make the simplification that $(\lambda_1 + \lambda_2) = (\lambda_3 + \lambda_4)$, such that the total manufacturing workforce of the two countries is equal, albeit with the potential to be unequally distributed within. The distribution of the immobile agricultural farmers is exogenously fixed and spread evenly across all regions such that their respective shares across regions are given by $\phi_1 = \phi_2 = \phi_3 = \phi_4 = 0.25$. To ease notification, we set the total population mass of the economy L to $L = 2$ and assume countries to be of equal size, i.e. $L_{F(oreign)} = L_{H(ome)} = 1$. As we will see later, this allows us to express the share of the manufacturing workforce for each region in a country by a λ, which is bound between zero and one. This facilitates the interpretation of λ as a measure of the relative economic disparity within a country and further eases interpretation down the line.

4.4.1. Consumer Preferences and Behavior

As in classical NEG models, a consumer decides how to spend her income Y with a preference assumed to be of Cobb-Douglas type. In fact, all consumers have a preference representation of Cobb-Douglas which combines a utility derived from the consumption of the agricultural good, F, as well as a Dixit-Stiglitz (Dixit and Stiglitz 1977) Constant-Elasticity-of-Substitution (CES) sub-utility for manufactures, M:

$$U = F^{1-\delta} \cdot M^{\delta} \tag{4.1}$$

$$M = \left[\sum_{i=1}^{n} c_i^{\rho} \right]^{\frac{1}{\rho}} \tag{4.2}$$

$$with \quad 0 < \delta < 1 \qquad and \qquad 0 < \rho < 1$$

Whereby δ denotes the share of income spent on consumption of the manufacturing variety such that the share of income not spent on manufactures $(1-\delta)$ is spent on the consumption of food. Note that δ also represents the share of the population engaged in manufacturing. This represents a normalization and is without consequence in the model as the shares are both exogenously set parameters (Brakman et al. 2020). c_i specifies the level of consumption of manufacturing variety i of a total of n varieties, among which the consumer chooses with elasticity ρ. ρ is chosen to be constrained between 0 and 1 such that varieties are substitutable but not perfect substitutes. ρ is set to $\varepsilon = \frac{1}{1-\rho}$ such that epsilon represents the elasticity of substitution. From (4.2) it is immediate that M is increasing more strongly in n than in c which reflects the well-known “love of variety” property, the strength of which is regulated by ε. The consumer problem is then given by maximizing utility U subject to the budget constraint:

$$Y = p^F F + \sum_{i=1}^{n} p_i c_i \tag{4.3}$$

Solving the consumer problem thereby involves first finding an optimal allocation of income Y on F and M, and then, maximizing the sub-utility derived from consumption of the composite index M subject to the budget constraint for such manufacturing varieties which follows from the first optimization problem. Hence, our first optimization problem is given by:

$$max\, U = F^{1-\delta} \cdot M^{\delta}$$

$$s.t. \;\; Y = p^F F + \sum_{i=1}^{n} p_i c_i$$

Some algebra leads to the well-known result that consumers spend share δ of income Y on manufactures, and $(1-\delta)Y$ on food:

$$p^F F = (1-\delta)Y \tag{4.4}$$

$$\sum_{i=1}^{n} p_i c_i = \delta Y \tag{4.5}$$

The next step involves finding the optimal spending *among* manufacturing varieties n, which is encapsulated by the following optimization problem:

$$\max M = \left[\sum_{i=1}^{n} c_i^{\rho}\right]^{\frac{1}{\rho}}$$

$$s.t. \quad \sum_{i=1}^{n} p_i c_i = \delta Y$$

Taking the ratio of first order conditions for a pair of varieties, the maximization problem yields the equality of marginal rates of substitution to price ratios:

$$\frac{c_i^{p-1}}{c_j^{p-1}} = \frac{p_i}{p_j}$$

$$\text{or} \quad c_i = p_i^{-\varepsilon} \cdot p_j^{\varepsilon} c_j \tag{4.6}$$

Once we substitute this result into the budget constraint for manufactures (4.5) we get:

$$\sum_{i=1}^{n} p_i c_i = \sum_{i=1}^{n} p_i \cdot (p_i^{-\varepsilon} \cdot p_j^{\varepsilon} c_j) = p_j^{\varepsilon} c_j \cdot \sum_{i=1}^{n} p_i^{1-\varepsilon} = c_j = p_j^{-\varepsilon} \cdot I^{\varepsilon-1} \cdot \delta Y \tag{4.7}$$

$$using\ I \equiv \left[\sum_{i=1}^{n} p_i^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}} \tag{4.8}$$

Hence, the expenditure needed to attain M is:

$$M = \left[\sum_{j=1}^{n} c_j^{\rho}\right]^{\frac{1}{\rho}} = \left[\sum_{j=1}^{n} (p_j^{-\varepsilon} \cdot I^{\varepsilon-1} \cdot \delta Y)^{p}\right]^{\frac{1}{\rho}} = I^{\varepsilon-1} \cdot \delta Y \left[\sum_{j=1}^{n} \left(p_j^{-\varepsilon p}\right)\right]^{\frac{1}{\rho}},$$

$$M = I^{\varepsilon-1} \cdot \delta Y \cdot I^{-\varepsilon} \tag{4.9}$$

Where we made use of that $-\varepsilon p = 1 - \varepsilon$, and $\frac{1}{p} = \frac{-\varepsilon}{1-\varepsilon}$, given that $\varepsilon = \frac{1}{1-p}$. Given that I multiplied by the quantity composite manufacturing consumption M is equal to expenditure δY, I is also known as the price index, which measures the minimum cost of purchasing manufacturing goods bundle M. Consumer demand functions are thereby:

$$F = \frac{(1-\delta)Y}{p^F} \tag{4.10}$$

$$M = \frac{\delta Y}{I} \tag{4.11}$$

Plugging these utility-maximizing consumption levels of F and M into (4.1) leads to the indirect utility function:

$$U = \delta^{\delta}(1-\delta)^{1-\delta} \cdot Y \cdot I^{-\delta}(p^F)^{-(1-\delta)} \tag{4.12}$$

Hence, the maximum attainable welfare is a function of the income Y weighted by the cost of living as given by price indices I and p^F together with their relative consumption shares δ and 1 − δ (Fujita et al. 2001).

4.4.2. Transport Costs and Heterogeneous intra-national Space

All manufacturing varieties can be consumed in each home or foreign location. However, evidently, a variety locally consumed but not produced needs to be imported, which entails transport costs. As is standard in NEG models, these transport costs are encapsulated by the *Samuelson-Von Thünen* iceberg-type, which envisions only a fraction of the goods to arrive at a destination, i.e. goods “melting” in transit (von Thünen 1826; Samuelson 1952). Thereby, a producer located in region 1 has to dispatch an additional amount together with the demanded amount, summing to T, for $1/T$ to arrive at the destination. For instance, if 20% of the dispatched goods regularly melt away en-route between regions i and j, iceberg transport costs are given by $T_{ij} = 1.25$. In other words, for one-unit of a good produced in region i to arrive at region j, suppliers located in region 1 have to dispatch 1.25 units of the good. Note at this point that we assume food to be transported costlessly across all national and international regions, which represents a standard simplification of NEG models.

As anticipated above, the present model is outlined by heterogeneous intra-national space, which will be operationalized by a specific transport cost structure. The reason for this adjustment is added realism on the one hand, but more importantly, because the spatial layout of the EAC as anticipated in Section 4.3 lends itself naturally to this modification. Note again that all three countries are outlined by economically dominating urban centers (Nairobi in Kenya, Dar es Salaam in Tanzania and Kampala in Uganda) which are geographically tucked away from the common borders connecting the respective EAC partner state(s) (see Figure 4.2 below for a stylized depiction).[35] To operationalize this specific spatial layout of the EAC in the model, I assume that among the two regions within each country, one of the regions has better access to the foreign market, i.e. is a "border" or "gated" region (Crozet and Koenig 2004b; Behrens et al. 2006). As such, shipping goods from a non-border region to a foreign location means transiting through this region, i.e. higher trade costs.[36] This effectively places the four regions on a line with regions 1 and 4 at the end of the spectrum and regions 2 and 3 connecting the two home and two foreign countries.

As expected, regions 1 and 4 represent the economic hubs of the countries, i.e. Nairobi, Dar es Salaam and Kampala, and are denoted as "interior" or "core" regions. As in Brülhart et al. (2012), I formalize this transport structure by accumulating all transport costs which accrue throughout the transit, i.e. simply multiply all types iceberg transport costs T which lie between the origin and the destination region. For instance, for region 1, which is an interior region, sending (importing) goods to (from) regions 2, 3, and 4 entails total iceberg transport costs of $T_{12} = T_{12}$, $T_{13} = T_{12} \cdot T_{23}$, and $T_{14} = T_{12} \cdot T_{23} \cdot T_{34}$. I additionally assume that intra-national transport costs in the home and foreign country are identical and that transport costs are symmetric, such that $T_{12} = T_{34} = T_{D(omestic)}$, $T_{23} = T_{32} = T_{F(oreign)}$ and $T_{ij} = T_{ji}$. Note at this point, the distinction made between domestic transport costs, T_D and foreign transport costs T_F. Domestic transport costs represent the general cost of transporting goods across distance (e.g. logistics and freight), while foreign transport costs additionally entail tariffs, and further frictions such as processing costs (at borders), as well as non-tariff barriers or even cultural differences (Behrens et al. 2007). Finally, note that transport costs are zero when consuming a variety produced within the same region, i.e. $T_{ij} = 1$, for all $i = j$. As such, the transport costs of trading goods between the four sending regions of our model $i = \{1, 2, 3, 4\}$ and arrival regions $j = \{1, 2, 3, 4\}$ can be summarized by the following five types of total trade costs across regions.

[35] In the data, the average road distance to EAC border crossings for the three cities is 395km, 922km, and 269km, respectively. These road distances are particularly relevant for intra-EAC trade, as over 95% of the regional trade in the area is in fact transported via the road network, and only 5% via rail (see Figure C.8 and Nathan Associates 2011).

[36] Note that this is assumed to hold both for the export and import of goods.

$$1 = T_{ij} \begin{cases} for\ i = i\ and\ j = i \\ for\ i = j\ and\ j = j \end{cases}$$

$$T_D = T_{ij} \begin{cases} for\ i = 1\ and\ j = 2 \\ for\ i = 3\ and\ j = 4 \end{cases}$$

$$T_F = T_{ij} \begin{cases} for\ i = 2\ and\ j = 3 \\ for\ i = 3\ and\ j = 2 \end{cases}$$

$$T_{DF} \equiv T_D \cdot T_F = T_{ij} \begin{cases} for\ i = 1{,}2\ and\ j = 3{,}4 \\ for\ i = 3{,}2\ and\ j = 1{,}4 \end{cases}$$

$$T_{DFD} \equiv T_D \cdot T_F \cdot T_D = T_{ij} \begin{cases} for\ i = 1\ and\ j = 4 \\ for\ i = 4\ and\ j = 1 \end{cases}$$

Figure 4.2 depicts this spatial cost structure for the 2+2 model illustratively using Uganda and Kenya as a stylized example. The dashed line in the countries depicts the main trade route between the countries called the "northern corridor" (Nathan Associates 2011).[37] The vertically dotted line illustrates the border.

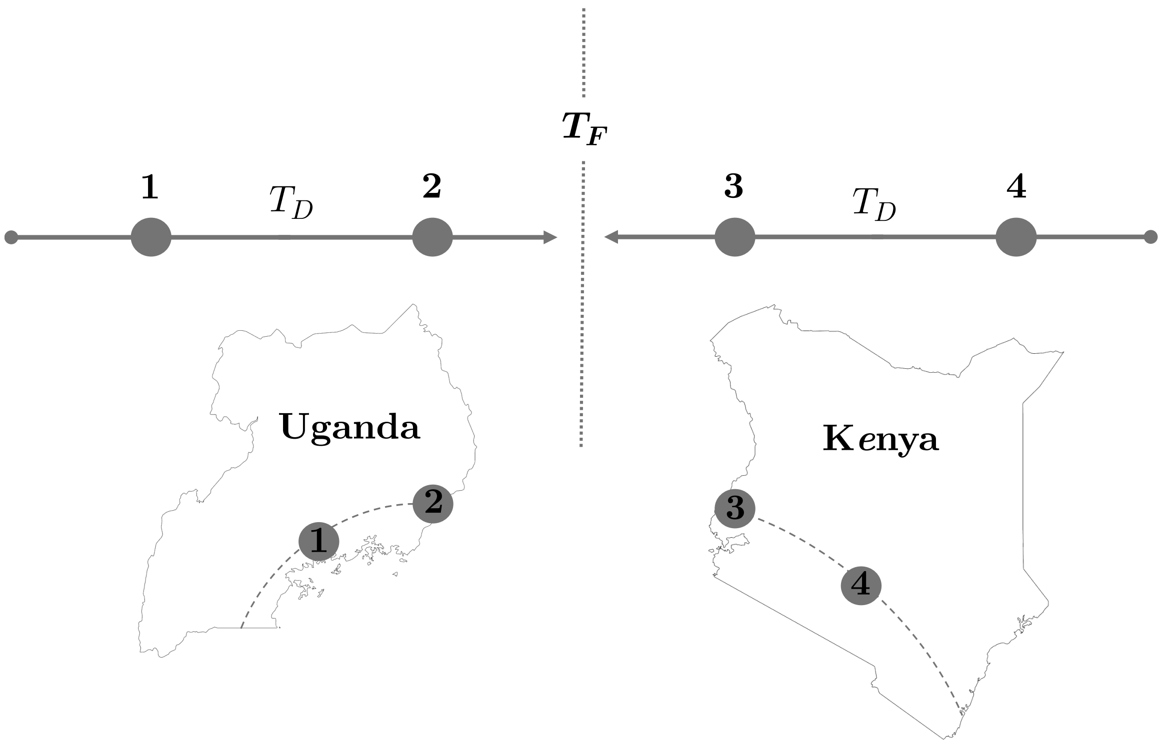


Figure 4.2: Transport Cost Structure in the four-region Economy

[37] Note that the figure is not drawn up to scale and serves as a stylized model of the spatial trade structure, only (see Figure C.7 of the Appendix for a more accurate depiction of the geography as well as a depiction of the "central corridor" which connects the countries via Tanzania). Country outline maps are from vemaps.com.

As is depicted in Figure 4.2, regions 1 and 4 represent the economic hubs Kampala and Nairobi, respectively, with 2 and 3 posing as the "border" regions.[38] Note that Tanzania borders the depicted countries to the south, respectively, and given the position of Dar es Salaam, creates a similar spatial pattern.[39] As such, this transport cost structure is assumed to be symmetric and thereby extends to the two other trade pairs, Tanzania-Uganda and Tanzania-Kenya, analogously. Granted that this a simplification of the spatial realities on the ground, including varying absolute and relative distances, differing processing times etc., this transport cost structure is nonetheless useful because it easily lets us operationalize the comparative statics of a change in regional market integration and the subsequent effect on (pre-existing) regional disparities in Subsection 4.4.5 by solely altering the costs of moving goods between regions 2 and 3, i.e. by altering T_F.

Carrying on with the model, these transport costs imply that the delivered price is T_{ij} higher than the f.o.b. price.[40] A standard assumption I follow is that all transport costs are incurred by consumers such that the total cost of consuming one-unit of variety produced in i in region j increases to $p_i = p_i T_{ij}$. Note that given (4.7), the demand for a variety produced in region i, consumed in location j is now given by:

$$c_j = (p_i T_{ij})^{-\varepsilon} \cdot I_j^{\varepsilon - 1} \cdot \delta Y_j \tag{4.13}$$

Note that this necessitates the simplifying assumption that one manufacturing variety is produced at one location only, which follows from internal economies of scale, and also, that all varieties n produced in this respective location are produced using the same technology and, therefore, price. The total price index I of region j is then given by:

$$I_j = \left[\sum_{i=1}^{R} n_i \cdot (p_i T_{ij})^{(1-\varepsilon)} \right]^{\frac{1}{1-\varepsilon}} \tag{4.14}$$

To arrive at the total sales of a given variety i, we sum demand for this variety over all regions R using (4.13) and note that the supply incurs shipping T_{ij} units of i. Hence, we arrive at:

$$q_i = \delta \sum_{j=1}^{R} Y_j \cdot (p_i T_{ij})^{-\varepsilon} \cdot I_j^{\varepsilon - 1} \cdot T_{ij} \tag{4.15}$$

[38] Malaba is the main border-crossing connecting these countries (Nathan Associates 2011).
[39] See detailed maps of these routes, i.e. the "northern corridor" as well as the "central corridor" connecting the larger region in Figure C.7 in the Appendix from Nathan Associates (2011).
[40] The "mill" or "f.o.b.", free on board, price, is the price charged at the "mill", the production location, not incurring shipping costs.

This encapsulates that total demand of a variety q_i is decreasing in the price of the good p_i and the transport cost incurred T_{ij} for the respective importing region. Demand is increasing in income Y_j and price index I_j of regions as well as in the share spent on manufactures δ.

4.4.3. Producer Behavior

As defined previously, food is produced under constant returns to scale as well as under the assumption of perfect competition. Given that we have just assumed food to be traded costlessly across all regions, the price of food is equal everywhere, and so is the wage given that farmers are paid their marginal product. We then set the technology coefficient of food production to 1 such that $w^F = p^F = 1$ and the agricultural good acts as the *numeraire* throughout the analysis. In the manufacturing sector, production technology is of increasing returns to scale. It thereby involves a fixed cost of production F and marginal costs per unit c. Given that labor is our only input factor, the production of a quantity q of a variety i produced in location i is given by labor input requirement:

$$l = F + cq \tag{4.16}$$

and this is assumed to be the same technology for all varieties. Given increasing returns to scale, consumer preference for variety, firms will choose to produce a variety, not produced by any other firm such that a variety is produced only in one location by one firm.[41] This has the result that the number of available varieties is equal to the number of firms. The profit of a specific firm producing at location i with a given wage rate w_i, and an f.o.b. price p_i is:

$$\pi_i = p_i q_i - w_i(F + cq_i) \tag{4.17}$$

Making the simplification $q = B{p_i}^{-\varepsilon}$ (see Brakman et al. 2020) and differentiating w.r.t. price and setting equal to zero leads to the $f.o.c.$:

$$(1-\varepsilon)B{p_i}^{1-\varepsilon} + \varepsilon w_i \cdot cB{p_i}^{-\varepsilon-1} = 0 \tag{4.18}$$

Rearranging leads us to the well-known result that prices are a combination of f.o.b. price, which are given by marginal costs $w_i c$, and a mark-up, determined by the elasticity of substitution ε:

$$p_i(1 - \tfrac{1}{\varepsilon}) = w_i c, \text{ or}$$

[41] Where an additional assumption is that the number of varieties goes to infinity (Fujita et al. 2001).

$$p_i = \frac{cw_i}{\rho} \tag{4.19}$$

Given that we assume free entry and exit, profits are driven to zero. Using the new pricing rule (4.19) in the profit function (4.17) and setting to zero leads:

$$\pi_i = \frac{cw_i}{\varepsilon - 1}(q_i - \frac{F(\varepsilon - 1)}{c}) \tag{4.20}$$

Hence, equilibrium output by any active firm i is the constant:

$$q^* = \frac{F}{c}(\varepsilon - 1) \tag{4.21}$$

and the required labor input producing this amount is then given by plugging (4.21) into the production technology used (4.16):

$$l^* = F + c\left(\frac{F}{c}(\varepsilon - 1)\right), or$$

$$l^* = F\varepsilon \tag{4.22}$$

Which carries the result that the number of varieties n produced in a location i, and thereby the number of manufacturing firms, is directly proportional to the manufacturing population at this location, $\lambda_i \delta L$:

$$n_i = \frac{\lambda_i \delta L}{F\varepsilon} \tag{4.23}$$

4.4.4. Short-run Equilibrium

In equilibrium, the output of firms must match demand by consumers. Using (4.15) we have:

$$q_i^* = \delta \sum_{j=1}^{R} Y_j \cdot p_i^{-\varepsilon} \cdot T_{ij}{}^{1-\varepsilon} \cdot I_j{}^{\varepsilon-1} \tag{4.24}$$

In other words, firms break even if the price they charge equals:

$$p_i^{\varepsilon} = \frac{\delta}{q_i^*} \sum_{j=1}^{R} Y_j \cdot T_{ij}^{1-\varepsilon} \cdot I_j^{\varepsilon-1} \tag{4.25}$$

Plugging in the pricing rule (4.19), leads to the well-known wage equation:

$$w_i = \left(\frac{\varepsilon - 1}{\varepsilon c}\right) \left(\frac{\delta}{q_i^*} \sum_{j=1}^{R} Y_j \cdot T_{ij}^{1-\varepsilon} \cdot I_j^{\varepsilon-1}\right)^{\frac{1}{\varepsilon}} \tag{4.26}$$

To arrive at real wages, ω, we simply have to divide nominal wages (4.26) by the cost of living, which is a combination of the manufacturing price index of the region (4.14) and food prices:

$$\omega_i = w_i \cdot I^{-\varepsilon} \cdot (p^F)^{-(1-\delta)} \tag{4.27}$$

It is convenient to use some normalizations to simplify analysis (Fujita et al. 2001). Hence, we redefined the marginal labor requirement is:

$$c = \frac{\varepsilon - 1}{\varepsilon} = \rho \tag{4.28}$$

Then, (4.19) turns to:

$$p_i = w_i \tag{4.29}$$

Also, we set a unit of measurement for the number of firms n, such that the fixed input requirement F is given by:

$$F = \frac{\delta}{\varepsilon} \tag{4.30}$$

Remember that the number of firms in each location is directly proportional to the manufacturing labor force in this location $\lambda_i \delta L$, such that (4.23) reduces to:

$$n_i = \frac{\lambda_i \delta L}{F\varepsilon} = \lambda_i L \tag{4.31}$$

From this, the price index (4.14) as well as the wage equation can be simply expressed as:

$$I_j = \left[\sum_{i=1}^{R} \lambda_i L (w_i \cdot T_{ij})^{(1-\varepsilon)}\right]^{\frac{1}{1-\varepsilon}} \tag{4.32}$$

$$w_i = \left(\sum_{j=1}^{R} Y_j \cdot \ T_{ij}{}^{1-\varepsilon} \cdot I_j{}^{\varepsilon-1}\right)^{\frac{1}{\varepsilon}} \tag{4.33}$$

These constitute the first two of three equations that characterize the short-run equilibrium. What is missing is the income-determining equation, which is defined by the sum of wage income from manufacturing workers in the region $\lambda_i \delta L$ as well as from farmers $\phi_i(1-\delta)L$.[42] Hence, the income of a region i is given $Y_i = \ \lambda_i \cdot w_i \cdot \delta L + \phi_i(1-\delta)L$. Taking into account our initial simplifications, namely that the manufacturing workforce is immobile across countries and exogenously set to $\phi = 0.25$, that the distributions of the manufacturing workforce is given by $\sum_{i=1}^{4} \lambda_r = 1$, that the total mass of population is set to $L = 2$ and that the two countries are of equal size lets us write the income equation in our four region case as:

$$Y_i = \ \lambda_i \cdot w_i \cdot \delta + \frac{(1-\delta)}{2}, \qquad 0 \leq \lambda_i \leq 1 \tag{4.34}$$

Where we additional use of our assumption $\lambda_1 + \lambda_2 = \lambda_3 + \lambda_4$. And similarly, the price index simplifies to:

$$I_j = \left[\sum_{i=1}^{R} \lambda_i (w_i \cdot T_{ij})^{(1-\varepsilon)}\right]^{\frac{1}{1-\varepsilon}}, \qquad 0 \leq \lambda_i \leq 1 \tag{4.35}$$

Given that manufactures can be traded across all regions, and our economy is made up of four regions in total, the short-run equilibrium relationship is expressed by 12 equations (3 for each region) given in (4.36) through (4.47):

$$Y_1 = \ \lambda_1 \cdot w_1 \cdot \delta + \frac{(1-\delta)}{2} \tag{4.36}$$

$$Y_2 = \ \lambda_2 \cdot w_2 \cdot \delta + \frac{(1-\delta)}{2} \tag{4.37}$$

[42] Note that given constant returns to scale and perfect competition, the wages for agricultural labor are equal everywhere and set as the numeraire as seen before.

$$Y_3 = \lambda_3 \cdot w_3 \cdot \delta + \frac{(1-\delta)}{2} \quad (4.38)$$

$$Y_4 = \lambda_4 \cdot w_4 \cdot \delta + \frac{(1-\delta)}{2} \quad (4.39)$$

$$I_1 = [\lambda_1 \cdot {w_1}^{1-\varepsilon} + \lambda_2(w_2 \cdot T_D)^{(1-\varepsilon)} + \lambda_3(w_3 \cdot T_{DF})^{(1-\varepsilon)} + \lambda_4(w_4 \cdot T_{DFD})^{(1-\varepsilon)}]^{\frac{1}{1-\varepsilon}} \quad (4.40)$$

$$I_2 = [\lambda_1(w_1 \cdot T_D)^{(1-\varepsilon)} + \lambda_2 \cdot {w_2}^{1-\varepsilon} + \lambda_3(w_3 \cdot T_F)^{(1-\varepsilon)}\lambda_4 + (w_4 \cdot T_{DF})^{(1-\varepsilon)}]^{\frac{1}{1-\varepsilon}} \quad (4.41)$$

$$I_3 = [\lambda_1(w_1 \cdot T_{DF})^{(1-\varepsilon)} + \lambda_2(w_2 \cdot T_F)^{(1-\varepsilon)} + \lambda_3 \cdot {w_3}^{1-\varepsilon} + \lambda_4(w_4 \cdot T_D)^{(1-\varepsilon)}]^{\frac{1}{1-\varepsilon}} \quad (4.42)$$

$$I_4 = [\lambda_1(w_1 \cdot T_{DFD})^{(1-\varepsilon)} + \lambda_2(w_2 \cdot T_{DF})^{(1-\varepsilon)} + \lambda_3(w_3 \cdot T_D)^{(1-\varepsilon)} + \lambda_4 \cdot {w_4}^{1-\varepsilon}]^{\frac{1}{1-\varepsilon}} \quad (4.43)$$

$$w_1 = (Y_1 \cdot {I_1}^{\varepsilon-1} + Y_2 \cdot {T_D}^{1-\varepsilon} \cdot I_2^{\varepsilon-1} + Y_3 \cdot {T_{DF}}^{1-\varepsilon} \cdot I_3^{\varepsilon-1} + Y_4 \cdot {T_{DFD}}^{1-\varepsilon} \cdot I_4^{\varepsilon-1})^{\frac{1}{\varepsilon}} \quad (4.44)$$

$$w_2 = (Y_1 \cdot {T_D}^{1-\varepsilon} \cdot I_1^{\varepsilon-1} + Y_2 \cdot I_2^{\varepsilon-1} + Y_3 \cdot {T_F}^{1-\varepsilon} \cdot I_3^{\varepsilon-1} + Y_4 \cdot {T_{DF}}^{1-\varepsilon} \cdot I_4^{\varepsilon-1})^{\frac{1}{\varepsilon}} \quad (4.45)$$

$$w_3 = (Y_1 \cdot {T_{DF}}^{1-\varepsilon} \cdot I_1^{\varepsilon-1} + Y_2 \cdot {T_F}^{1-\varepsilon} \cdot I_2^{\varepsilon-1} + Y_3 \cdot I_3^{\varepsilon-1} + Y_4 \cdot {T_D}^{1-\varepsilon} \cdot I_4^{\varepsilon-1})^{\frac{1}{\varepsilon}} \quad (4.46)$$

$$w_4 = (Y_1 \cdot {T_{DFD}}^{1-\varepsilon} \cdot I_1^{\varepsilon-1} + Y_2 \cdot {T_{DF}}^{1-\varepsilon} \cdot I_2^{\varepsilon-1} + Y_3 \cdot {T_D}^{1-\varepsilon} \cdot I_3^{\varepsilon-1} + Y_4 \cdot I_4^{\varepsilon-1})^{\frac{1}{\varepsilon}} \quad (4.47)$$

These 12 equilibrium conditions formalize the notion of centripetal (demand and cost linkages) as well as centrifugal forces (competition) anticipated in Subsection 4.2.2. Take first, the price index given in equations (4.40) through (4.43). Consumer prices at one particular location can be seen as a weighted average of all source location sizes (λ) and prizes (which, given (4.19) are directly proportional to the wage rate w) with weights given by the distance to these exporting locations (T), respectively.[43] As such, the price index is lower in those regions, where a higher share of demand is sourced from large (high λ), low wage (low w) and importantly, nearby locations (low T); and of course, most cheaply sourced locally, i.e. when $T = 1$. In other words, locations with large shares of own or close by manufacturing employment have lower price indices given that a smaller share of the total consumption needs to be imported; this is the "price index effect" analytically derived in Fujita et al. (2001). These dynamics describe the cost (forward) linkage described in Subsection 4.2.2 whereby a larger home market provides lower consumer prices. As such, moving to a region, i.e. making it larger, displays a self-reinforcing centripetal force.

[43] Consumer prices at one particular location can be seen as a weighted average of all source locations and their prizes (which, given (4.19) are directly proportional to the wage rate w) as well as distance T with weights given by the relative size of these locations λ.

The wage equations given in (4.44) through (4.47) can be interpreted similarly. In essence, wages are higher in regions where income Y, and thereby expenditure, is high or in regions where these larger markets are more proximate (low T). Put simply, firms are able to pay higher wages if they have better access to large markets. This describes the demand (backward) linkage anticipated before and indicates that a larger number of workers, and thereby, consumers, increase the local demand which increases the wage firms are able to pay. Similarly, as for the cost linkage, this attracts more workers to this region, and also firms, thereby acting as a self-reinforcing centripetal force. This is described by the "home market effect" (for the full derivation, see Fujita et al. 2001). Importantly, the wage equation also encapsulates a centrifugal force which is given by its positive dependence on the price index I. As just established, the price index is lower in larger regions, i.e. those with a higher number of manufacturing workers, and thereby, varieties. And given that the number of manufactures is regulated not by output per firm, but by the number of firms, a lower price index automatically indicates a larger number of competing firms, which exerts a downward pressure on the wages a firm is able to pay.[44] As a results, firms may seek to relocate in order to shelter from competition allowing them to pay higher wages, which may also draw workers.

In the end, the relative strength of these centrifugal and centripetal forces can be handily manifested in real wage differentials across regions, which combine the effects on nominal wages and prices. Formally, the real wages ω of regions are given by dividing the total wage income w by the consumer price index of both manufactures I and food F together with their relative consumption shares δ, hence $\omega_i = w_i \cdot I^{-\delta}(p^F)^{-(1-\delta)}$. Note that we are able to dismiss the component of the agricultural good, as it is set as the numeraire. Real wages of all four regions are then expressed by:

$$\omega_1 = w_1 \cdot I_1^{-\delta} \tag{4.48}$$

$$\omega_2 = w_2 \cdot I_2^{-\delta} \tag{4.49}$$

$$\omega_3 = w_3 \cdot I_3^{-\delta} \tag{4.50}$$

$$\omega_4 = w_4 \cdot I_4^{-\delta} \tag{4.51}$$

Where the values of the right-hand side are given by the simultaneous solution to the 12 short-run equilibrium conditions (4.36) through (4.47). In the long run, we assume that workers respond to the real wage differential across regions by migrating such that the share of manufacturing workers within the two home and foreign economies, λ_1 and λ_2, as well as λ_3 and λ_4, are endogenously determined. I assume workers to move between regions with the following dynamics:

[44] This also be validated in (4.7) or (4.15), where demand of an individual firm is inversely related to the price index.

$$\frac{d\lambda_i}{dt} = \gamma \begin{cases} \dfrac{\omega_i}{\omega_j} - 1 & if\ 0 < \lambda_i < 1 \\ min\left\{0, \dfrac{\omega_i}{\omega_j} - 1\right\} & if\ \lambda_i = 1 \\ max\left\{0, \dfrac{\omega_i}{\omega_j} - 1\right\} & if\ \lambda_i = 0 \end{cases} \tag{4.52}$$

Hence, for a given real wage differential and spatial configuration λ_i, workers move between regions across regions with a particular speed γ. We now have all the ingredients we need to define a *long-run* equilibrium as well. By (4.52), the first type of long-run equilibrium can be described by a spatial configuration for which real wages across regions are equalized, i.e. a situation in which workers have no incentive to move. Formally, this is given by a $\lambda \epsilon$ [0,1] for which $\omega_i/\omega_j = 1$. One specific case of such is the equal spreading of workers, i.e. for our four-region model $\lambda_{1,2} = \lambda_{3,4} = 0.5$ and $\omega_{1,3}/\omega_{2,4} = 1$. This is also called the "symmetric" or spreading" equilibrium. The model also admits a second type of a long-run equilibrium, one in which real wages are not equalized. In these cases, all of the manufacturing workforce is agglomerated in one of the regions, which represents a corner solution. Formally, such an equilibrium is given by $\lambda_{1,4} = 1$ and $\lambda_{2,3} = 0$, and often referred to as an "agglomerated" or "core-periphery" equilibrium. To complete the discussion on long-run equilibria, one important distinction to make is whether such an equilibrium is also a *stable* one. In general, the stability of an equilibria depends on whether a small perturbation in the manufacturing workforce at this spatial configuration triggers dynamics which reinstates the just-left allocation of workers or not. For the first type of equilibria, the stability is thereby defined by a second condition which is that the derivative of the real wage differential w.r.t. an infinitesimal change in the manufacturing workforce is smaller or equal to zero, i.e. formally whether $d(\omega_i/\omega_j)/d\lambda_i \leq 0$. Put simply, if migrating from region j to region i increases the real wage differential ω_i/ω_j, then the previous equilibrium was not a stable one. In the second type of equilibria, the stability condition entails that the real wage differential is skewed in favor of the agglomerated region, such that for $\lambda_1 = 1$, $\frac{\omega_1}{\omega_2} \geq 1$ and for $\lambda_1 = 0$ and $\frac{\omega_1}{\omega_2} \leq 1$. I analyze the stability of these two types of equilibria more thoroughly in Appendix C.1.

4.4.5. Spatial Equilibria and Regional Trade Liberalization

The four-region model and the short- as well as long-run equilibrium conditions just developed lends itself to the comparative static examination of what happens to the forces inducing agglomeration or dispersion once trade across countries in the EAC is liberalized. In particular, we can use the solutions to the simultaneous equilibrium conditions (4.36) through (4.47) as inputs to compute the real-wage differential arising for a given spatial structure and transport costs regiment which dictates the dynamic process towards a stable long-run equilibrium described in (4.52).

The analysis in this section thereby entails tracking what happens to the (short-run) real wage differential across regions inside the countries once the costs connecting the two economies T_F are lowered from a former prohibitive level (i.e. autarky) down to levels which mirror those incurred within the respective countries, i.e. $T_F = T_D$.[45] This will effectively allow us to analyze how the process of trade liberalization affects the (stability of) specific (long-run) equilibrium allocations of workers across regions. Note, however, given that the real wage differential ω_i/ω_j depends on twelve simultaneous non-linear equations, the real wage differential is not a simple function of λ_i. As such, and as is common in the NEG literature, I will analyze the dynamics of the spatial equilibria mainly via numerical simulations. This is most efficiently done by plotting the real wage differential ω_1/ω_2 across the full range of potential manufacturing distributions $\lambda \in [0,1]$ which may be realized at any point in time. To nonetheless provide some analytical insights into the numerical results, Appendix C.1 provides a "sustain" and "break" analysis in the vein of Fujita et al. (2001), which revolves around assessing the stability of the two specific types of equilibria described above, i.e. "agglomeration" and "spreading". Some of the key results of this analysis are discussed in this section as well.

As a final remark on the approach of this section's analysis, it turns out to be instructive to compare the results of the model to a more general version of it. To be specific, I will conduct the simulations additionally for a four-region model with *homogenous* (or *symmetric*) intra-national space. This model mirrors the one described by equations (4.36) through (4.47), but with a tweak regarding the transport cost structure. This is done by simply setting the three different types of external iceberg trade costs equal, i.e. $T_F = T_{DF} = T_{DFD}$, such that the two home regions have identical (international) trade costs to both foreign regions. In the vein of Figure 4.2, one can think of regions 1 and 2 as well as 3 and 4 in this adjusted model as placed on a line parallel, rather than perpendicular to the border with roads diagonally connecting the home and foreign regions, respectively. Note that the model thereby reduces to the one studied in Monfort and Nicolini (2000), and their conclusions apply analogously. However, comparing the results from such a model to the present one provides an intuitive reference and helps in evaluating the additional role *heterogeneous* intra-national space (from now on interchangeably referred to as *asymmetric* space), and thereby unequal access to the newly integrated foreign markets, plays, particularly for an integration process of two highly unequal countries, which describes the (stylized) setting of the East African Community's re-establishment.

Figure 4.3 below initiates our analysis and plots the real wage differential between regions 1 and 2 across the full range of possible manufacturing distributions $\lambda \in [0,1]$ as well as for

[45] In this scenario, the cost associated with trading goods across borders mirrors those incurred when shipping goods intra-nationally, i.e. $T_F = T_D$.

three levels of international trade costs T_F, respectively. Note that solving for this set necessitates a choice on the exogenous parameter values given by δ and ε, and the intra-national trade costs T_D. I use values commonly employed in the literature which are given below the figures. Table C.1 in the Appendix provides additional sensitivity tests using a wider range of values and includes a discussion thereof. The main interpretations remain. Notice that although our model entails four endogenous parameters, λ_1, λ_2, λ_3, and λ_4, the plots in Figure 4.3 only depict two at a time, i.e. is two-dimensional. This is done by setting $\lambda_3 = \lambda_4 = 0.5$, i.e. by initially assuming an equal distribution of manufacturing in the foreign country. This assumption is relaxed later in Figure 4.5, when we assess the moderating influence of foreign economic inequality. Note lastly, that as anticipated earlier, we focus our view on the home country, i.e. regions 1 and 2, but given symmetry, the results and intuitions apply identically for regions 4 and 3 Figure C.3 and Figure C.4 of the Appendix also provide the full three-dimensional plots, which effectively combine the results depicted in Figure 4.3 and Figure 4.5.

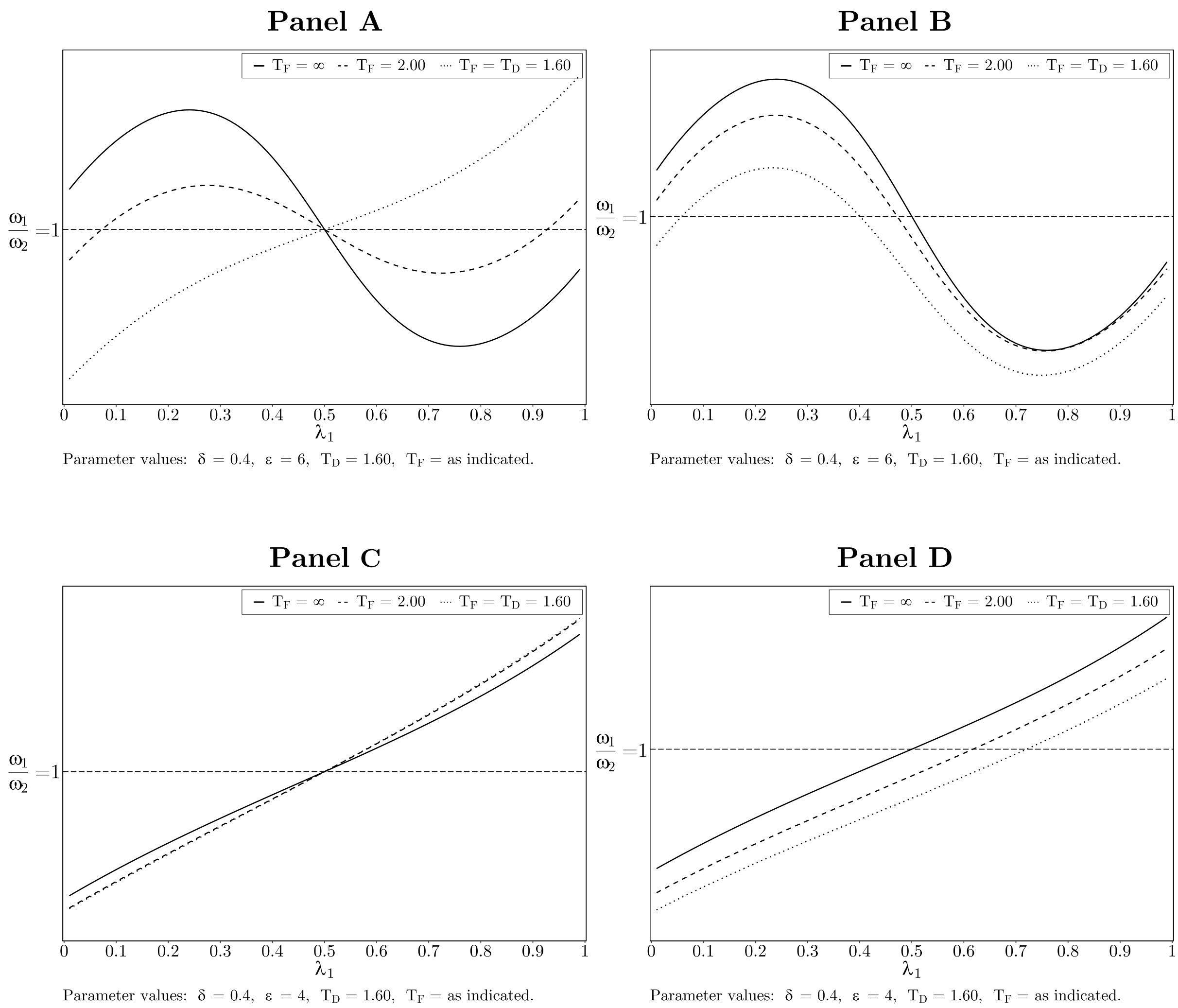


Figure 4.3: Trade Liberalization and Spatial Equilibria

The numerical simulations depicted in Figure 4.3 provide the main insights into the process of trade liberalization across a four-region economy, i.e. lowering the intra-national trade costs T_F. It depicts the results for two spatial setups (the symmetric and asymmetric intra-national space) for two sets of parameter values, respectively.[46] More specifically, Panels A and C represent the symmetric case (homogenous intra-national space) for values of the elasticity of substitution $\varepsilon = 6$ and $\varepsilon = 4$. And Panels A and D present the results analogously for the asymmetric case (heterogeneous intra-national space).

We focus first on the real wage differentials in autarky, i.e. where international trade costs are prohibitively high $T_F = \infty$, as depicted by the solid line. In the case of low product differentiation ($\varepsilon = 6$), Panels A and B, we notice that there exists a long-run stable symmetric equilibrium where the workforce is equally spread across the two home regions for both models, as can be seen by the negative slope passing-through real wage parity. While this equilibrium also exists for the case where product differentiation is high (Panel C and D), this equilibrium is not stable anymore, as can be depicted by the positive slope through the point where $\lambda_1 = 0.5$. What happens to this type of equilibria in the home country when the external trade costs to regions 3 and 4 are lowered? This is depicted by the new short-run equilibrium real wage differentials given by the dashed ($T_F = 2.00$) and dotted lines ($T_F = T_D = 1.60$).[47] As a first pass through the Panels, and as shown in previous results, lowering the costs to trade with an external market increases agglomerating tendencies, i.e. increases intra-national inequality (e.g. Monfort and Nicolini 2000; Paluzie 2001). This can be seen by a general attenuation of the slopes passing through the symmetric equilibria. Most starkly, in Panel A, the slope concludes a full rotation from negative to positive values from autarky to free trade. Hence, when the countries are liberalized, the former stable equilibrium for equal distribution of manufacturing activity turns out to be unstable. As we defined in the previous section, this is so because an infinitesimal small shock (increase) to the manufacturing workers in any direction would also cause a higher real wage skewed towards this region, which would not induce workers to move back to the symmetric equilibrium. As such, once trade is liberalized, the strength of the force holding together the equal spreading, i.e. the costs of serving remote markets, weakens. This may therefore set in motion a cumulative causation for a small increase of consumers in region 1, leading to full agglomeration in region 1, and vice versa, for region 2 if initially moved in the opposite direction. However, this effect on the slope is generally not as pronounced in the model with heterogeneous intra-national space. For instance, in Panel B, while the slope is reduced for higher values of trade liberalization, there still exists a stable equilibrium not leading to a full core-periphery pattern as it would in Panel A. Remarkably, this long-run stable equilibrium is brought about at an unequal distribution of the workforce within the home country. That is, we observe a shift of the curve which cuts the constant parity line parallel to the left. This

[46] The two parameter configurations represent the two most common cases seen across the simulations.
[47] We thereby implicitly assume a change in the ad valorem tariff of crossing international borders down to 25% and 0%.

effectively indicates a stable equilibrium at an unequal distribution across the home regions. Hence, applied to the case of the EAC, there now exists an increased draw to the border regions, given that λ_1 reduces from 0.5 in autarky to 0.4 in the free trade scenario, which indicates that now over half of the manufacturing activity is operating at the border. This is similar to the result provided in Behrens et al. (2006) and (Crozet and Koenig 2004b), albeit in a 2+1 setup.[48]

If we move our view to the results in Panel C and D, this result is further corroborated. In this scenario, the centripetal forces are accentuated as can be seen by positive values of the slope of ω_1/ω_2 throughout. Given that the only amendment is a lower elasticity of substitution ε, higher product differentiation causes the strength of scale economies to increase (Fujita et al. 2001; Paluzie 2001). The dynamics are seen in (4.40) through (4.47) in how the strength of the centripetal and centrifugal forces depends on the parameter ε. For one, in the price indices, a lower elasticity of substitution (ε) increases the strength of the love of variety, such that for any increase in the low-cost access of goods (high λ, as well as a low w and T), the price index is lower than for higher values of ε.[49] Intuitively speaking, the higher the differentiation between varieties, the higher the added utility gain of (increased availability of) a further variety to consumers. Hence, lowering ε causes the forward (cost) linkage to intensify. Note however, as established above, that this also automatically also leads to stiffer product market competition among varieties, as I is reduced.[50] Secondly, in the wage equation, this means that lower values of ε have a negative effect on the wage firms are able to afford, which displays a centrifugal force. However, counteracting this, remember that any increase in market access (high Y and low T) also increases the wages firms are able to afford, and lowering ε causes the backward (demand) linkage to intensify.

As seen by the comparison between top and bottom panels, the forward and backward linkages are strengthened, i.e. centripetal forces dominate the centrifugal forces caused by a decrease in ε.[51] As such, a core-periphery pattern is more likely at any level of intra- or international trade costs. This is seen by positive slopes in both Panel C and D. Again, while decreases in T_F cause only minor changes in the slope for Panel C, it significantly alters the equilibrium configuration in the model with heterogeneous intra-national space. However, different from the case in B, where the cut point with the parity line has shifted to the left, it has now shifted to the right. This is reconciled with the previous results showing a relative shift to the border by interpreting this change as decreasing the basin of attraction, which would

[48] Note that setting $\lambda_3 = \lambda_4 = 0.5$ is not equal to the case of a foreign country with one region when the two foreign regions are accessed with different transport costs.
[49] To validate this, note that the negative exponent of the entire bracket in (4.40) through (4.43) gets larger, while the negative exponents of w and T get smaller. Further, from (4.8) and noting that varieties are produced with the same technology in all locations, which renders the price index as $I = p \cdot n^{1/1-\varepsilon}$. It is easily seen that I is more strongly decreasing in n (varieties) for lower values of ε (Fujita et al. 2001).
[50] This can also be confirmed in (4.7), i.e. $c_j = p_j^{-\varepsilon} \cdot I^{\varepsilon-1} \cdot \delta Y$. As established in footnote 49, increasing n decreases I which thereby lowers the demand for any variety. This is also seen in (4.21) and (4.23), whereby a decrease in ε leads to an increase in varieties n at each location.
[51] Table C.1 shows that this is the case for all tested parameter configurations.

lead to full agglomeration in region 1. In other words, in autarky, it suffices to increase the manufacturing workforce in region 1 an infinitesimal amount above 50%, for dynamics to unfold which lead to full agglomeration in 1. In the free trade scenario close to 73% of the manufacturing workforce would have to be in region 1 for this cumulative causation mechanism to kick in. Again, Crozet and Koenig (2004) as well as Brülhart et al. (2004) show qualitatively similar results, albeit for a 2+1 setting.[52] This is the second noteworthy departure from the model with symmetric intra-national space and highlights that relevant different conclusions arise when the access to foreign regions is unequal.

Hence, the results until now paint two consistent insights for a process of regional market integration in a spatial layout such as the one present in the EAC. For one, liberalizing trade across the two countries increases internal agglomeration tendencies. And secondly, that this agglomeration is more likely to occur in the region bordering the newly accessed markets. For additional insights into (the moderating forces of) these dynamics, see Section C.1 of the Appendix. Figure 4.4 reproduces part of the analytics carried out in C.1. As such, it plots the results of the "sustain" analysis, which essentially evaluates the stability of an agglomerated equilibrium, i.e. depicts the range of intra-national transport costs T_D for which the agglomerated equilibrium in the core economic hubs, e.g. region 1, proves sustainable. Remember that the stability condition of this equilibrium at $\lambda_1 = 1$ requires $\omega_1/\omega_2 \geq 1$. Given that we have seen an increased draw to the border (region 2), we are interested in which range of values a sustainable agglomeration in the interior (region 1) can be upheld. Again, Panels A and C provide the case for the symmetric case, while Panels B and D conduct the analysis for our main model again for the three levels of international transport costs T_F. The point where the line crosses the real-wage differential from below is called the "sustain" point, $T(S)$ and describes the maximum level of intra-national transport costs for which agglomeration is still sustainable, i.e. for which $\omega_2/\omega_1 \leq 1$ (note the reversal). This equates to a real wage differential at λ_1 in Figure 4.3 which stays above the parity line. Beyond this point, agglomeration is not sustainable anymore, i.e. a case where the line is below parity in Figure 4.3. What happens when trade is liberalized? In Panel A, the sustain point shifts to the right, which indicates that agglomeration is able to be upheld for a wider range of domestic transport costs T_D. By design, this result mirrors the one in Monfort and Nicolini (2000) and also what we have seen in Figure 4.3 Panel A, i.e. that decreased cross-border trade costs increase the agglomeration forces. Note that this is mainly due to a decrease in the slope of the ascending part of the plotted lines. At these levels, an infinitesimally small increase in T_D increases ω_2/ω_1 and liberalization thereby seems to influence the centrifugal forces (competition) to a larger degree than the centripetal forces (cost and demand linkages), as seen by the positive slope (Crozet and Koenig 2004a).

[52] Brülhart et al. (2004) additionally departs from CES and uses a quasilinear consumer utility in the vein of (Ottaviano et al. 2002).

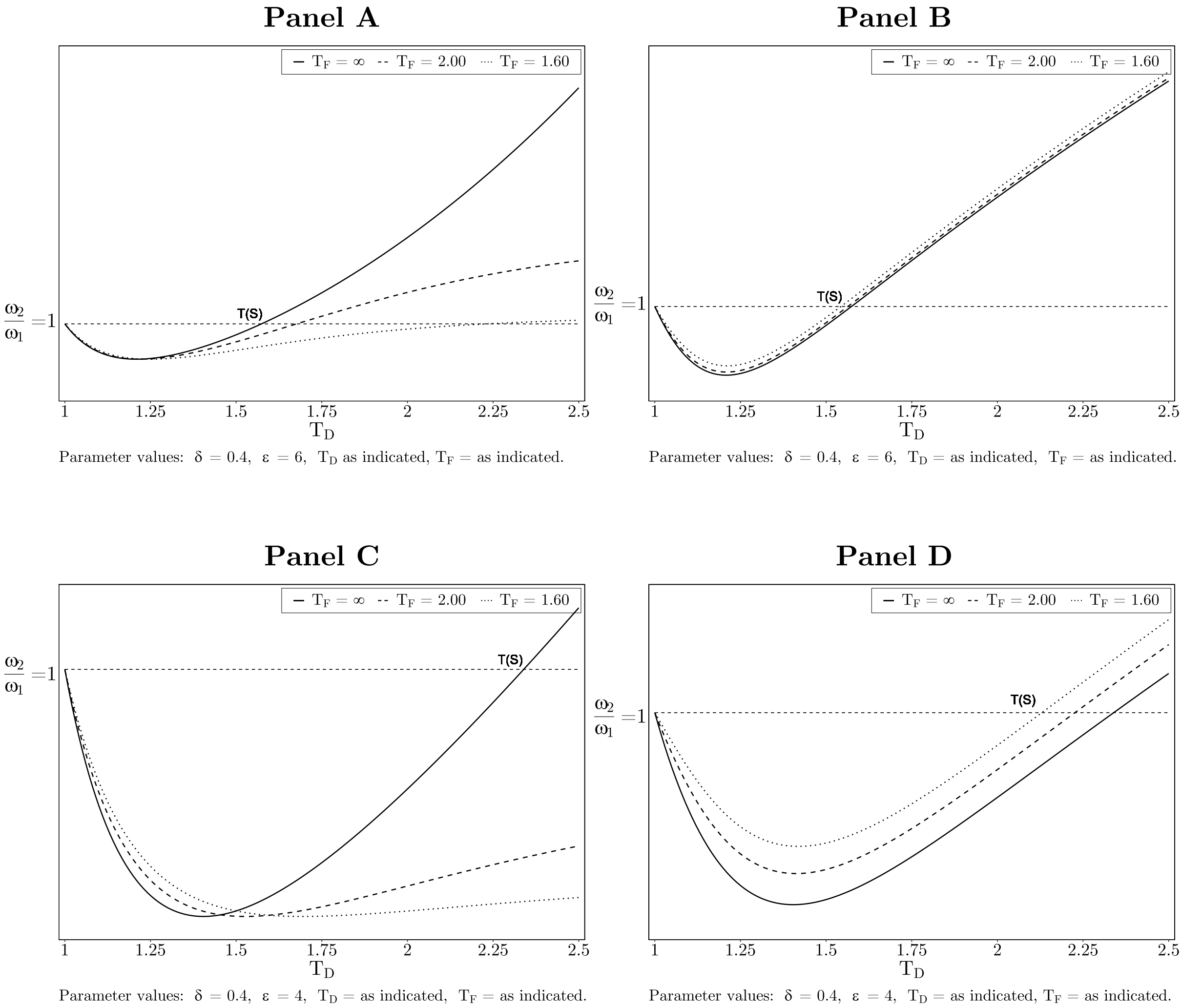


Figure 4.4: Internal Transport Costs and Sustainable Agglomeration

As such, a decrease in international trade costs mainly modulates the strength of the dispersion forces. In Panel C, these centripetal strengths dominate by reasons given above, such that the core-periphery pattern is upheld for a larger range of intra-national transport costs. So much so that in full trade scenario, there exists no sustain point and agglomeration in region 1 is never broken. Again, the results for our main model with heterogeneous intra-national space provide different conclusions. While the range of transport costs for which a core-periphery pattern is upheld also decreases in ε (compare Panels B and D), trade liberalization works towards the opposite, i.e. puts negative pressure on the full agglomeration in region 1, as seen by the negative shift of the sustain point $T(S)$ to the left. As such, regional market integration decreases the range of values for which agglomeration away from the border region can be upheld. This time, the change in the slope occurs mainly for the descending part, indicating that trade liberalization decreases centripetal forces for region 1.[53] These results essentially confirm analytically what is depicted numerically Figure 4.3, i.e. that there is an increased draw to the border with increased trade liberalization. However, one interesting aspect is that, for any level of regional trade integration T_F, agglomeration in region 1 is more likely to be upheld in the case of high product differentiation. Next to stronger scale effects, one possible way to interpret this result may be to refer to the heightened competitive pressures of an increased number of firms from abroad, keeping domestic firms in the interior, where they are sheltered (Crozet and Koenig 2004b).[54] See the discussion of Section C.1.3 of the Appendix.

So far, we have confirmed the results of previous symmetric $2+2$ settings (Monfort and Nicolini 2000) and extended those from asymmetric $2+1$ layouts to an economy with four regions. What is left to investigate in our unique $2+2$ setting is the influence of *foreign* economic inequality on these initial results, given that we have so far set $\lambda_3 = \lambda_4 = 0.5$. We now relax this assumption and discuss additional results for the full range of spatial configurations $\lambda \in [0,1]$ in the foreign country. Given this added dimension, Figure 4.3 turns three-dimensional which makes it a bit cumbersome to evaluate at first sight (see Figure C.3 and Figure C.4 of the Appendix). To make it more accessible, for the moment, we restrain ourselves to assessing the influence of a varying foreign manufacturing distribution on our two types of long-run equilibria depicted in Figure 4.3. As such, Figure 4.5 plots the combinations of λ_1 and λ_4 where the real wage differential ω_1/ω_2 is equalized, i.e. give the contour lines of the plane spanned by the two endogenous variables as given in Figure C.3.[55] The arrows in the graphs indicate the stability of the equilibrium, not depictable in contour lines. Arrows pointing towards the line indicate a stable equilibrium, given that economic forces (the real wage differential) lead consumers back to the original allocation, and vice versa for arrows pointing away from the line. As is now common, Panel A and C shows the result for the homogenous $2+2$ model whereas Panel B and D depicts our asymmetric case. Given that the foreign spatial configuration can only exert influence when trade costs are not prohibitive, we restrict ourselves to analyze the case for which $T_F = 1.60$.[56] To no surprise, Panel A and C shows a vertical line at $\lambda_1 = 0.5$. This is because when the home country is equally spread, and both regions have equal access to the foreign market via T_F, there is no difference in the relative real wages of the two regions. Hence, shifting shares of the workforce in the foreign regions doesn't affect the existence of such an equilibrium.[57] Panels B and D paint a wholly different picture, and show

---

[53] Note importantly, that given the unequal intra-national space, a reversal of the analysis, i.e. evaluating the stability of an agglomerated equilibrium at the border with $\lambda_1 = 0$ and $\omega_1/\omega_2 \leq 1$ would render the shift of the lines as in Panels A and C, i.e. would move the sustain point to the right.

[54] This is seen in (4.23) as an increase in product differentiation increases the number of varieties at a location, solely by an incerase in the number of producers (at a location).

[55] Remember that $\lambda_2 = (1-\lambda_1)$ and $\lambda_3 = (1-\lambda_4)$.

[56] The results for all three levels of international transport costs are provided in Figure C.5 and Figure C.6.

[57] However, what may in fact happen, as shown in Monfort and Nicolini (2000), for a specific parameter configuration, is that the *stability* of this spreading equilibrium depends on the foreign distribution (interdependent equilibrium).

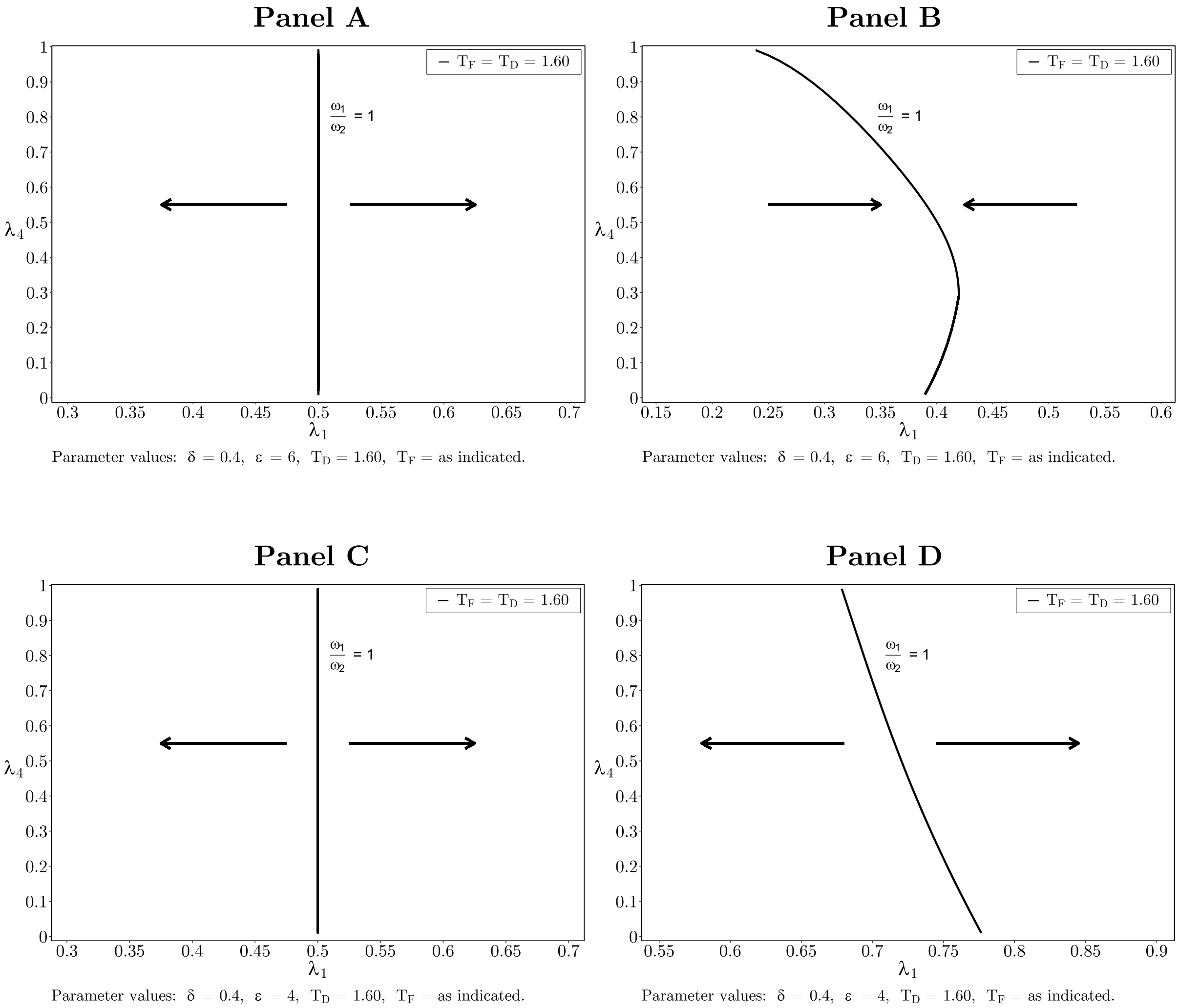


Figure 4.5: Foreign economic Inequality and Spatial Equilibria

that in the presence of heterogeneous intra-national space, foreign spatial inequality has a moderating impact on the domestic (agglomeration vs. dispersion) forces induced by trade liberalization. This result was first shown by (Behrens et al. 2006), albeit for a setup in which a "gate" exists in only one of the countries and for a different setup of consumer preference structures (see footnote 34).

For the case where a long-run stable equilibrium exists ($\varepsilon = 6$), we see that a changing share of workers in the interior of the foreign country modulates the domestic allocation for which this equilibrium is reached. Panel D corroborates this view for the case where trade liberalization has led to a core-periphery pattern as the only stable equilibrium. In general, the higher the foreign spatial distribution is skewed towards the interior (a higher λ_4), the lower the share of workers in region 1 needed for both types of equilibria depicted in Panel B and D in Figure 4.3. What is the implication of this result? In the case of less intensive scale effects ($\varepsilon = 6$), the draw to the border is further increased (Panel B). On the other hand, when scale effects are large and full agglomeration is the only stable equilibrium ($\varepsilon = 4$), this result is reversed (Panel D): while there is still an increased draw to the domestic border in free trade, as can be seen by values for λ_1 above 0.5 on the x-axis, this draw is decreasing in λ_4. Hence, given a stark regional inequality in the foreign country which is outlined by an economically strong interior (high λ_4), the basin of attraction needed for domestic agglomeration at the border is decreased when product differentiation is high. This is intuitively plausible, as it may signal a decreased relevance of sheltering from foreign competition further away from the border, the effect of which is particularly pronounced for environments with a larger number of competing firms (low ε). In the contrary scenario (high ε), market access considerations may prevail over this competition effect, leading to increased agglomeration at the border. Note that a similar point is made (Behrens et al. 2006), who show that for the case in which one of the countries is outlined by heterogeneous intra-national space, the general agglomeration tendency in the domestic country is increased when the foreign country hosts a lower number of firms at the border. See Section C.1 for a more thorough discussion on these possible mechanisms.

In sum, the NEG model developed in this section and the counterfactual exercise of an increased market integration performed through it holds three main insights. First, given heterogeneous intra-national space, progressive trade liberalization draws economic activity to the border, i.e. the real wages at border regions are relatively higher than in the interior when compared to autarky. Second, general agglomeration tendencies increase the freer trade is, although the agglomeration is more likely to occur at the border. And third, foreign spatial inequality has non-negligible impacts on these domestic effects, such that a core-periphery pattern in the foreign economy may attenuate or reinforce the first and second results. This result holds relevant insights for the case of the East African Community, where the countries integrating are in fact all outlined by large interior hubs positioned away from internal borders.

With these deliberations in mind, the next section tests for reduced form evidence on these potential spatial effects of regional market integration in the EAC.

4.5 Empirical Strategy

The theoretical exercise of the previous section motivates the empirical strategy. As is seen from the simulations, lowering trade costs among EAC members is predicted to increase the draw to the border, i.e. to increase relative welfare in the regions with better access to the new markets. As discussed in Section 4.2.2, this result is corroborated both by previous theoretical models as well as by empirical evidence from both developed and developing settings.[58] However, what we also saw in the simulations is that this draw may be attenuated or reinforced depending on the economic (in)equality present in the foreign economy. Given the particular spatial layout of the EAC member countries anticipated in Section 4.3 and Subsection 4.4.2,

[58] Note also that a pure reference to new trade theory is not strictly necessary to render an increased impact of trade closer to borders. It has been shown in other developing settings that price pass through is highest directly at the border and decays perpendicular to it (e.g. Nicita 2009; Cali 2014; Atkin and Donaldson 2015).

trade liberalization among them presents a fitting empirical case on which to study these dynamics, i.e. testing whether the re-establishment of the EAC did increase the relative welfare of border regions in comparison to the pre-existing economic hubs, and also provides a setting with which to identify household welfare effects of trade integration.

Note that so far, we have measured the attractiveness of border regions with real wages, and particularly, real wage differentials. However, we can readily translate these real wages into household welfare, as first established using indirect utility in (4.12), by simply noting that food was set as the numeraire.[59] As such, the real wages discussed broadly in Section 4.4 encapsulate what we envision as household welfare in the simplest form, which is, the income consumers earn and the prices they face (see e.g. Deaton 1997; Fujita et al. 2001; Winters 2002; Brülhart et al. 2012). The comparative statics tested theoretically thereby translate to the empirics in a reduced-form manner and revolve around assessing what happens to households' welfare (indirect utility) across space following a change in the trade costs among EAC members from a former prohibitive level down to levels of trade costs that mirror those of the type within the domestic country, i.e. only given by the geographic distance between locations.[60] To operationalize this, I employ a difference-in-differences (DiD) specification comparing the changes in welfare of households living *relatively* closer to borders, ω_2, with those of households living *relatively* closer to the economic hubs ω_1, before and after the re-establishment of the EAC in 2001.[61] To flexibly allow for treatment across space a the potential draw to the border, I model this relationship nonparametrically, i.e. employ a continuous treatment intensity which is captured by households' (road) distances to nearest EAC border crossings. The estimating equation therefore reads:

$$Y_{ict} = \alpha + \beta_1 Border_i + \beta_2 Core_i + \sum_{t=\{EAC\}}^{\{CM\}} \beta_{3,t}(\gamma_t \times Border_i) + \sum_{t=\{EAC\}}^{\{CM\}} \beta_{4,t}(\gamma_t \times Core_i) + X'_i + \delta_{ct/i/h} + e_{ict} \qquad (4.53)$$

$Y_{i,c,t}$ represents the respective welfare indicator of individual i living in country c, surveyed at survey-sampling period t. $Border_i$ is the inverse, relative within-country distance to the nearest EAC border crossing [0,1], such that a value of 1 indicates individuals in the sample living closest to the border in the sample, and value of 0 those furthest away. $Core_i$ is a dummy {0,1} indicating individuals living within 50 kilometers of the three preeminent interior agglomerations, namely Nairobi in Kenya, Dar es Salaam in Tanzania and Kampala in

[59] For a given set of exogenous parameter values, the indirect utility function (4.12) then reduces to a function of income, which is varying across space only in the nominal manufacturing wages w (see (4.36) through (4.39), and consumer prices for manufactures I.
[60] Note that transport costs between regions *within* countries are assumed constant throughout the same time period.
[61] The empirical specifiction is similar to Eberhard-Ruiz and Moradi (2019), however with distinct regard for urban vs. rural differences.

Uganda.[62] γ_t is an indicator for the respective integration period i.e. switching to 1 for the free trade period (EAC) between 2001 and 2004, the customs union period (CU) between 2005 and 2009, as well as the common market period (CM) after 2010, respectively.[63] Therefore, under the assumptions that no other concomitant policy or shock has induced a differential trend in outcomes between these regions during the same time period, β_3 and β_4 represent estimates of the effect of the EAC on border regions as well as on interior agglomerations. Specifically, they give estimates of the differential effect of households living at the border compared to those living furthest away, and those living in interior agglomerations to those living in the auxiliary (rest of the country), both compared before and after the EAC was established. β_3 and β_4 can thus be seen as a test on the theoretical predictions, i.e. whether the EAC led to larger relative increases in welfare in border regions, i.e. $\Delta\omega_1/\Delta\omega_2 < 1$, given by a $\beta_3 \neq 0$ which also satisfies $\beta_3 > \beta_4$, rather than the opposite, i.e. in preexisting interior agglomerations, $\Delta\omega_1/\Delta\omega_2 > 1$, given by $\beta_4 \neq 0$ for which $\beta_3 < \beta_4$.[64] These estimates therefore also indicate if we should expect *dispersion* of the previously concentrated economic activity rather than *concentration* as proposed by the endogenous adjustment process in (4.52). $\boldsymbol{X}$ represents a matrix of individual-level control variables which allows us to account for all influences potentially conflating the relationship between access to (new) markets and household welfare. δ captures country-time fixed effects such that identification in the cross-sectional datasets comes from variation within individual member countries in specific survey-periods in time. For the results produced from the household panel, identification comes from changes within household/individuals over time, such that δ represents household-, respectively, individual fixed effects. Standard errors are constructed by allowing for spatial correlation of errors, i.e. Conley standard errors are used (Conley 1999, 2010). I additionally check for the clustering of errors at the level of the survey enumeration area, i.e. at the survey cluster level.[65] Binary dependent variables are estimated with a simple Linear Probability Model (LPM) specification.[66]

4.6 Data

I employ a distinct set of longitudinal, geo-referenced household-level surveys that were sampled in all three founding members before and after the establishment of the EAC. First, I make

[62] I also test the lower distance thresholds 25km and 10km in the robustness tests of Subsection 4.7.2.

[63] In the sample, the first post-EAC responses we measure come from the year 2002, which is indicated in the regression results.

[64] Of course, this statement is true only for the maximum effect (change) of border distance, i.e. going from the largest distance to the smallest distance in the sample. In-sample prediction, such as an interquartile range bound between 0 and 1 is arguably more appropriate as a comparison, where β_3 must be more than infinitesimally larger than β_4.

[65] The cut-off for Conley standard errors is chosen by the function of the “fixest” package in R, which ensures a large enough sample size within a certain distance cutoff and is additionally robust to sub-sampling. The results are robust to sensitivity checks of the Conley standard errors by increasing and reducing the distance cutoff. Results can be obtained from the author.

[66] Results for binary dependent variables estimated via *Probit* yields qualitatively identical and quantitatively similar marginal effects.

use of the complete set of available Demographic and Health Surveys (DHS). DHS are cross-sectional, household-based surveys which are representative at both the national- as well as regional level, and collect a broad array of information on topics such as demographics, education, employment and occupation, as well as fertility and family planning (Croft et al. 2018).[67] The main respondents are women of reproductive age (15-49), but DHS also provides information on men and children living in the sampled households, as well as on household-specifics such as consumer durables and wealth assets in possession. To increase the variable size and sample space, I pool these additional data from the Men- and Household recodes also, leading to a main sample of 227,860 individuals living in 140,943 households across 7,962 survey locales interviewed between 1999 and 2020. Later extensions and robustness checks expand this sample to include non-GPS survey rounds sampled from 1988 onwards, leading to an extended sample size of 332,725 individuals living in 203,150 households.[68] Note that the main sample also includes special survey rounds such as the AIDS Indicator Survey (AIS), the Malaria Indicator Survey (MIS), as well as the Knowledge, Attitudes and Practices Survey (KAP), to gather a higher frequency of survey years. While all variables used across the paper are consistently available in these survey years, too, I provide an additional robustness that drops these survey waves in Subsection 4.7.2.[69]

Second, I make use the geo-referenced Afrobarometer (AFB) survey rounds, which span a timeframe of 18 years (from 1999 to 2017) across seven survey waves, i.e. rounds 1 through 7 (Afrobarometer 2019).[70] Afrobarometer surveys are representative at the national level, and the main respondents are adults of the sampled households. They carry individual- and household-level information on basic characteristics, socio-demographics as well as own (economic) living conditions, household assets, and additionally, provide information on individuals' sentiments as well as opinions towards the economy, democracy, governance and society. Afrobarometer fits geo-coordinates (latitude and longitude) to respondents at the level of their respective enumeration area, and the sampling procedure aims for eight individuals/households per EA (BenYishay et al. 2017). The Afrobarometer adds information on 38,644 individuals (households) living in 3,414 geo-referenced localities across Kenya, Uganda and Tanzania to the sample, and additionally, provides the opportunity to test specific sentiments and attitudes towards free trade, which I turn to in the section on robustness and extensions (4.7.2).

Lastly, I supplant the analysis with information from the Kagera Health and Development

[67] More precisely, the first level administrative subdivision, most often referred to as regions, districts, provinces or states.
[68] The surveys were sampled in Kenya in 1989, 1993, 1998, 2003, 2008-09, 2014, 2015 (MIS), and 2020 (MIS), in Tanzania in 1991-92, 1994 (KAP) 1995, 1996, 1999, 2003-04 (AIS), 2004-05, 2007-08 (AIS), 2010, 2011-12 (AIS), 2015-16, and 2017, and in Uganda in 1988-89, 1995-96, 1995, 2000-01, 2006, 2009 (MIS), 2011, 2011 (AIS), 2014-15 (MIS) 2016 and 2018-19 (MIS).
[69] With the exception of *Employed Work* {0,1}.
[70] Surveys were sampled in 2000-2001 (only Tanzania and Uganda), 2002-03, 2005, 2008, 2011-12, 2014-15 and 2017.

Survey (KHDS) (World Bank and University of Dar es Salaam 1994, 2004, 2010). The KHDS is a representative panel originally sampled from Kagera, a GADM-1 administrative region of Tanzania bordering Uganda in the northwest. The panel collected detailed information on households' and individuals' wealth and poverty dynamics, such as employment, salary, (non-) durable assets as well as food- and non-food consumption, all for which values in constant (deflated) Tanzanian Shilling are provided (Beegle et al. 2006; De Weerdt et al. 2010). The KHDS also includes information on migration decisions of individuals as well as community-level variables such as the price of commodities in local markets. The KHDS set out by interviewing 6,356 individuals living in 915 households spread across 51 sampling clusters in the first four yearly survey waves between 1991 and 1994 (round 1).[71] All of the initially sampled households (rather, the individuals living within those households) were sought to be re-contacted in the succeeding two survey rounds in 2004 and 2010 (rounds 2 and 3), respectively. The tracking of individuals was highly successful, the sample evincing re-contact rates of 80% for individuals and over 90% for singular households.[72] Importantly, the number of administered households (and individuals) grew significantly over the sample timespan, as all members residing in (new) households of original respondents were fully included in the sample, and also tracked in later survey years.[73] As such, the KHDS is able to add information on 21,696 distinct individuals – interviewed a minimum of one-, and a maximum of six times – whose households are spread across 2,019 survey locales in Tanzania and Uganda. Importantly, out of the 6,356 original survey respondents sampled between 1991-1994, 4,430 individuals were successfully (re)-interviewed in 2004, and 3,848 were surveyed in all three rounds, including 2010.

Figure 4.6 visually depicts this distinct set of geo-referenced data by plotting the sample enumeration areas of households from each of these three sources across East Africa. Notice that the map also depicts enumeration areas of contiguous EAC-accession as well as non-accession countries. These data will be employed in the extensions and robustness tests of Subsection 4.7.2.

4.6.1. Dependent Variables

As anticipated above, the real wages analyzed in theory are proxied by household welfare i.e. the income consumers earn and the prices they face (Deaton 1997; Winters 2002). Given the usual data restrictions of household surveys, i.e. a lack of precise wage and price data, I proxy welfare by a set of intensive and extensive labor market outcomes (work, employment, and income) as well as consumption measures (food and non-food consumption, durable as well as non-durable assets).

[71] Not all of the households were interviewed in all of the first four waves.
[72] And over 90% for those cases where at least one of the original household members was aimed to be re-interviewed.
[73] As such, the number of singular households contained in the survey expanded from 915 in the first round (1991 to 1994) to 2,719 in 2004 and 3,314 in 2010 (De Weerdt et al. 2010).

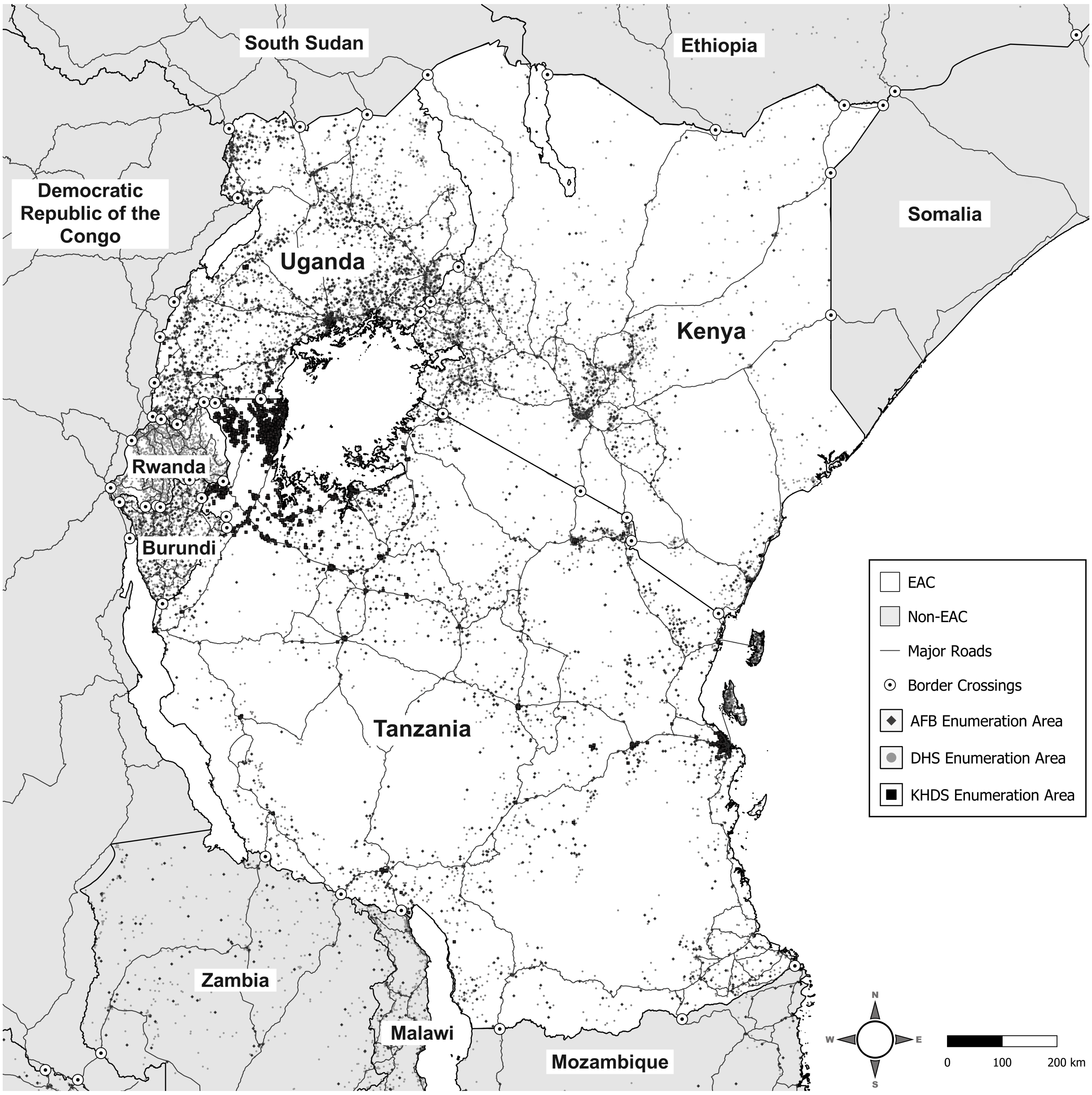


Figure 4.6: Sample Coverage

To measure the levels of consumption in the Demographic and Health Surveys, I make use of the *Wealth Index* {1,5}, which is a DHS-constructed index placing households on a relative scale of wealth within their respective sample. The construction of the index is based on an array of consumer durables, such as the construction of dwelling, sanitation facilities as well as possessions such as a TV, motor vehicles etc. (Rutstein and Johnson 2004). I countercheck these results with the *Comparative Wealth Index* established by the DHS, which facilitates the comparison of the wealth scores underlying the wealth indices across countries and samples (Rutstein and Staveteig 2014). I additionally construct the *International Wealth Index (IWI)* as established in Smits and Steendijk (2015) as a further attempt to make household's wealth ranking more comparable across surveys.[74] Concerning labor market outcomes as the second dimension, I test the variable *Employed Work* {0,1}, which indicates whether the respondent worked for someone outside of the household (conditionally on having work). In later extensions for these labor market results, I also test *Worked in last Year* {0,1}, which is the baseline measurement indicating whether survey respondents were pursuing some activity aside from housework within the last calendar year, on which 2) is conditioned on, and test whether this activity was also remunerated, i.e. *Paid in Cash* {0,1}. Lastly, I test the *Occupational Type* {1,3} of work, which places all activities categorized within the survey schedule from "agrarian" (1), "worker" (2) and "professional" (3) activities.[75] This may be regarded as a test on the skill-intensity of the respondent's occupation.[76]

Concerning the Afrobarometer, the level of (basic) household consumption is measured by *Lived Poverty* [0,4], which is constructed by averaging individuals' responses to the three separate questions: "Over the past year, how often, if ever, have you or anyone in your family gone without: Enough clean water for home use" / "[…]: Enough food to eat" / "[…]: Medicines or medical treatment?".[77] The response values range from "never" (0), "just once or twice" (1), "several times" (2), "many times" (3) and "always" (4). Similar to the DHS, the primary test on individuals' labor market outcome is measured via *Employed Work* {0,1}. Again, I test whether individuals *Worked Last Year* {0,1} as well as their *Occupation Type* {1,3} to explore an extended set of labor market outcomes. Relatedly, I employ *Cash Income Deprivation* {0,4} which is asked in the same way as the components of the lived poverty index and indicates how often individuals within a household have gone without a cash income within the past year. One of the key characteristics of the Afrobarometer surveys is the component containing opinions, attitudes, and sentiments on individual, political, as well as domestic and international economic topics. As such, I am able to test the variable *Support for: Regional Integration* {1,5} which evaluates the strength of the support for free movement and trade.[78] Further, I test whether individuals living closer or further from the border assess the *Ease of Crossing Borders* {1,4} as more or less difficult, how much they evaluate the *EAC* as well as the *African Union (AU)*, as the supranational trade facilitator of the continent, in helping their

[74] The IWI exploits information from the entire universe of developing countries household surveys to construct factor loadings of specific household wealth items.
[75] For instance, "workers" are occupations such as traders, artisans, or unskilled manual labor. "Professional" is comprised of lawyers, accountants and teachers.
[76] A simple regression of *Occupation Level* {1,3} on the individual characteristics age, age squared, years of education as well as a female dummy and country-time fixed effects shows that each year of education increases the index by 0.06 units and a standard error of 0.0045.
[77] Contrary to the Afrobarometer's use, I don't include the question on *Cash Income* droughts, as I test it as a separate outcome for the labor market outcomes. Further, I am unable to include the response to having enough fuel to cook food, as responses to this variable are not available in pre-EAC survey rounds.
[78] The question asked in Round 6 of the Afrobarometer probes the support by agreeing with either of the two following mutually exclusive statements. Statement 1: "People living in the sub-region should be able to move freely across international borders to trade or work". Statement 2: "because foreign workers take away jobs, and foreign traders sell their goods at very low prices, governments should protect their own citizens and limit the cross-border movement of people and goods" (Afrobarometer 2019).

country {0,3}, and evaluate whether they would like having an *Immigrant as a Neighbor* {1,5}. I supplant these variables with further subjective assessments of the *Present vs. Past: Life Satisfaction* {1,4}, as well as their *Present vs. Past: Living Standards of People* {1,4}.

The Kagera Health and Development Survey (KHDS) allows us more detailed access into the consumption and income dimensions of households and the individuals therein. First, I test changes in the *Annual per capita Household Consumption* which expresses aggregate food and non-food consumption in constant, i.e. deflated, 2010 Tanzanian Shilling ('000 TZS).[79] Food items are constituted of both purchased as well as home-produced food, non-food items are comprised of expenses on items such as clothing, schooling, services like haircuts, or utilities.[80] I also test for differences in *Food*- and *Non-Food Consumption* separately, to identify potential systematic differences across the two and to countercheck the results on lived poverty measured in the Afrobarometer more precisely. Secondly, I test for changes in household wealth similar to the DHS using the *Value of Durable Assets* as well as the *Value of the occupied Dwelling* of a survey household.[81] Both measures are given in deflated, 2004 Tanzanian Shilling ('000 TZS) as this component was last administered in the survey wave of 2004. Concerning income, I proceed in similar to the previous two surveys and ask whether the respondent has *Employed Work* {0,1}, whether they have *Salaried Work* {0,1} and the *Occupation Type* {1,3}. In extensions, I also test the overall likelihood of having any kind of work via *Worked last Year* {0,1}, as well as employed individuals' *Monthly Salary* in deflated 2004 Tanzanian Shilling ('000 TZS). Similarly to the Afrobarometer, the KHDS contains an array of subjective assessments and evaluations. As such, I test for the main motives of migrating to the current place of residence, and whether movers found *Paid (formal) Employment* {0,1} right after migrating to the current location. Again, these measures are supplanted by assessments of their S*ubjective HH. Wealth* {1,5} both at the survey time (2004) as well as 10 years before (1994) and the general *Life Satisfaction on a Ladder* {1,9}.

As a last empirical investigation into the spatial corollaries anticipated in Section 4.4, I test for changes in the extent of agglomeration at respondents' geographic locations (enumeration areas) across all three surveys. To do this, I merge granular population data provided by the Gridded Population of the World (CIESIN 2017) to evaluate the EAC's effect on *Population Density* (sdz.), a measure of the total number of persons per square kilometer at the specific geography, to all unique enumeration areas of the respective samples. [82] Assessing changes in population density across space before and after the EAC is a direct test of the long-run dynamics induced by the agglomeration vs. dispersion forces discussed in Section 4.4. In another vein, population density may be useful as a more general indicator suggested to

[79] The total annual household consumption is distributed equally across all household members.

[80] For the full information on the construction of the aggregate consumption, see https://microdata.worldbank.org/index.php/catalog/2251/download/ 34035.

[81] Durable goods include e.g. radios, refrigerators, telephones. The figure for the value of the occupied dwelling represents an estimate of the head of the household.

[82] To facilitate interpretation, I standardize the value at a mean of 0 and a standard deviation of one.

capture “underlying differences in productivity and quality of life” (Breinlich et al. 2014; 733), which displays a relevant outcome for policy interventions such as market integration.[83]

4.6.2. Independent Variables

The main explanatory variable of interest *Border* [0,1] is measured by calculating the shortest road distance from each household’s enumeration area to the nearest (within country) internal EAC border crossing (depicted in Figure 4.6).[84] To circumvent endogeneity in the construction of roads, I only use major roads, i.e. motorways, trunk- and primary roads as provided by OpenStreetMap (OSM 2022), which can be tracked back to the pre-EAC era. Border crossings are defined as points where these major roads connect to both sides of the border. To assess the sensitivity of the results to the specific distance calculation, I also construct beeline (as the crow flies) distances from all enumeration areas to both the border crossings as well as to the nearest possible point on the borderlines spanned by two EAC country pairs. Shapefile data for country administrative areas, i.e. the boundaries of which I use come from the Center for Spatial Sciences at the University of California (GADM 2020). To test the effects on the pre-existing economic hubs, I employ *Core* {0,1}, which is a dummy indicator switching to one for households located within 50km of the country’s respective economic hub (i.e. Dar es Salaam, Nairobi, and Kampala).[85] To also provide a test on the general tendency of agglomeration derived in Section 4.4, irrespective of border distance, I employ *Agglomeration* {0,1}, which switches to 1 for households living within 50km of an “urban center” demarcated as such in the year 2000. The list of urban centers I use is provided by the European Commission (2019). Lastly, I also make use of the *Urban* {0,1} dummy, which is a survey-specific classification of the level of urbanization at the specific location.[86]

I also include an array of controls to account for influences which potentially conflate the relationship between household welfare and spatial aspects over time. As such, I include the individual-level covariates *Age*, *Age squared*, a dichotomous indicator of gender, *Female* {0,1}, as well as individuals’ *Educational Attainment*. I additionally account for potentially correlated geographic influences of development across distance and closely follow Henderson et al. (2018) with a set of important physical geographic features. I therefore include the location’s *Elevation* (Farr et al. 2007), *Ruggedness* (Nunn and Puga 2012) as well as agricultural characteristics such as the number of *Growing Days* (Ramankutty et al. 2002) as well as average long-term

[83] See for instance Rappaport and Sachs (2003).

[84] I measure distances using the projection of coordinates along the earth’s ellipsoid (using WGS 84, EPSG 7030).

[85] Later robustness tests relax this distance cutoff and also test areas within 25km and 10km, respectively. Notice that for the KHDS survey, there is no data for households living in these hubs in the pre-EAC era, as the survey was initially sampled in Kagera only. For these earlier cases, the dummy switches to 1 for individuals living in “Bukoba” the urban capital of Kagera.

[86] Urban stems either from country census information (DHS) or on the assessment of sample enumerators (Afrobarometer). See e.g. https://www.idhsdata.org/idhs-action/variables/URBAN#comparability_section and https://www.afrobarometer.org/wp-content/uploads/2022/07/AB_R9.-Survey-Manual_eng_FINAL_20jul22.pdf.

Monthly Temperature and *Monthly Rainfall* (Fick and Hijmans 2017). In later robustness checks, I additionally control for the location's *Malaria Ecology* (Sachs et al. 2004), *Absolute Latitude*, as well as household's distance to *Navigable Rivers*, as well as *Major Lakes* and *Major Harbors*.[87] Lastly, I add country-year fixed effects to control for time-specific influences as well as country-specific influences at specific points in time, such as the Kenyan Post-Election Crisis of 2007-2008, and additionally include household-, respectively, individual fixed effects for estimations using the Kagera Health and Development Survey.

Table C.2 of the Appendix provides summary statistics for the dependent- and independent variables categorized by sample source. The table also provides first insights into the distribution of outcome variables across space by grouping values into (border) distance quartiles and reporting a separate mean of outcomes within the three core agglomerations (see panel b). As expected from the facts outlined in Section 4.3 the countries are highly polarized. For instance, compared to the overall mean in the sample, individuals living in the core have between 0.8 and 2 additional years of schooling, which may provide a reason for the higher occupation type which is also shifted 0.5 units in favor of in the core. Also, individuals living in the economic hubs have a 20 percentage points higher likelihood of employed work, and a 24% decrease in lived poverty.

4.7 Results

4.7.1. Main Results

Table 4.1 presents the first set of results estimated via regression equation (4.53). Importantly, these results set out the analysis by reporting an aggregate "pre vs. post" difference-in-differences effect across all post-integration periods estimated by setting $\gamma_t = 1$ for all responses collected in or after 2002.[88] The effects differentiated across the specific post-integration periods, i.e. the test on the temporal evolution of the EAC towards a Customs Union (CU) as well as a Common Market (CM), are established from Table 4.2 onwards. Note lastly, that the results reported in all tables of this section are restricted to the reporting of the two DiD estimates β_3 and β_4, only. This is because we are primarily concerned with the relative development of border vs. interior agglomerations, and because the identification strategy is ultimately bound to estimate the differential effects of regional market integration only.[89]

We first focus on results produced from the two nationally representative household-

[87] The criteria for rivers' "navigability" as well as the importance of lakes ("major") is defined as in Henderson et al. (2018), i.e. I select all natural rivers within size categories 1-5 (scale 1-7) as defined in Natural Earth (2019) and lakes with a surface area of over 5,000 sq. kilometers (Lehner and Döll 2004). Concerning harbors, I define all large and medium sized ports listed in the World Port Index (WPI) as "major harbors" (NGA 2019) as in (Wild and Stadelmann 2022), who provide recent individual-level evidence that access to such harbors is a robust predictor of living standards and household welfare.

[88] 2002 is the earliest available post-EAC survey year for the samples employed.

[89] Given that the control group is merely treated less intensely, however effectively located in the same intervention jurisdiction, the estimation differences out any effect that is common to both regional entities.

surveys, the Afrobarometer and the DHS for which the table depicts estimates on three specific outcomes, each representing a distinct dimension of and contributing to spatial inequality. To facilitate the comparison of effects across samples, the dependent variables are consistently grouped into consumption, income and agglomeration categories. Concerning the consumption

Table 4.1: Aggregate Difference-in-Differences Estimates

	Dependent Variable					
	Afrobarometer			DHS		
	Consumpt.	Income	Agglom.	Consumpt.	Income	Agglom.
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[3.14]	[0.19]	[0.00]
Border [0,1] $\times$ EAC 1[t $\geq$ 2002]	0.114	0.076*	0.111	-0.479**	-0.060	0.069
	(0.278)	(0.041)	(0.094)	(0.202)	(0.041)	(0.094)
Core {0,1} $\times$ EAC 1[t $\geq$ 2002]	-0.362***	-0.045	0.758**	0.363***	0.023	0.799***
	(0.069)	(0.062)	(0.336)	(0.121)	(0.040)	(0.118)
Individual Controls	YES	YES	NO	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	183,250	71,738	7,692
R-Squared	0.13	0.14	0.25	0.34	0.16	0.29
R-Squared -Within	0.10	0.10	0.23	0.33	0.11	0.28

Notes: The results in each column are produced by a separate regression. In columns (1) through (3), data come from the Kenya, Uganda and Tanzania Afrobarometer surveys rounds 1 through 7 sampled between 2000 and 2020. In columns (4) through (6), data come from the Kenya, Uganda and Tanzania Demographic and Health surveys (DHS) sampled between 1999 and 2020. The sample mean of the respective dependent variable is given in brackets above the estimates. For the DHS variable measured at the household level, *Wealth Index* in column (4), the answer from the main survey respondents (women) is used. For the coefficients on *Population Density* in columns (3) and (6), the level of observation is given by the unique sample enumeration areas of the respective survey. EAC 1[t $\geq$ 2002] switches to one for individuals sampled from 2002 and onwards. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

dimension reported in columns (1) and (4), the results paint a stark picture. Both the AFB and the DHS provide no evidence of relative increases in household welfare in regions closer to internal EAC borders following the establishment of the EAC. In fact, the results show that consumption was differentially affected in the negative direction, as given by a statistically significant, relative shift in household wealth by -0.115 (=-0.479$\times$0.24), when comparing households located at the median distance in the sample (236 km, which corresponds to a value of 0.76 for Border [0,1]) with households living closest to the border. This represents a reduction in wealth quintiles by 3.7% relative to the overall sample mean. Concerning the complementary DiD estimate for the core agglomerations, we observe the opposite effect. Compared to the rest of the country, households living in Nairobi, Dar es Salaam or Kampala increased their position on the wealth quintile by 0.363 units (12%) and reduced the occurrence of lived poverty by 0.362 (32%). Concerning income, the results do not provide systematic evidence for a change in the likelihood of having employed work, only showing a marginally significant increase (at the 10% level) by 1.9 percentage points (8%) on the interquartile range for border households in the Afrobarometer. Turning to the evidence on agglomeration patterns (columns 3 and 6), we see that the population density of preexisting core agglomerations is increased starkly following the re-establishment of the EAC by three-quarters of a standard deviation, compared to household locations in the rest of the country.

In sum, this first set of results does not provide evidence in favor of the main prediction of Section 4.4, i.e. that relative household welfare (real wages) is increased in border regions. Households and individuals living closer to borders did not experience (greater) relative welfare gains after the EAC was established, compared to individuals living further away. This result is consistent throughout the outcome variables tested in both the Afrobarometer as well as the DHS samples. Inasmuch as trade liberalization weakens agglomeration tendencies in the primate cities, which was the second prediction of Section 4.4, the findings strongly negate this notion, as we can observe large differential increases in household consumption and population density in the core agglomerations predating the EAC.

I now turn to a more nuanced assessment of these aggregate effects and estimate the full set of period-specific difference-in-differences estimates anticipated in equation (4.53). Table 4.2 reports these results, which effectively expand the simple DiD effect of Table 4.1 to three separate estimates for border and interior regions, which compare outcomes across space in the initial free trade regiment (EAC), the customs union (CU), and the common market era (CM) all relative to the same pre-EAC period. As such, the estimates are directly comparable and allow an insight into the temporal evolution of the mean increases in spatial inequality across the EAC shown in Table 4.1. While these temporally differentiated effects confirm the average effect shown in Table 4.1, the results provide three interesting additional insights. First, border regions did not seem to have benefitted differentially more compared to more distant regions in any of the EAC's time periods. Importantly, this result is true even in the early years following the re-establishment, which goes against Eberhard-Ruiz and Moradi (2019) who show that growth of nightlights in the EAC was differentially higher for cities closer to the borders in the initial periods. In fact, the present results show that it is more likely that regions closer to the border experienced negative relative welfare reductions following re-establishment. Three out of the four significant estimates show reductions in the DHS wealth index as well as employed work opportunities (columns 4 and 5). The only positive effect depictable is a 9.8

Table 4.2: Difference-in-Differences across three Integration Periods

	Dependent Variable					
	Afrobarometer			DHS		
	Consumpt.	Income	Agglom.	Consumpt.	Income	Agglom.
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[3.14]	[0.19]	[0.00]
Border [0,1] × EAC 1[2002-2004]	-0.083	0.050	0.098	-0.284	-0.129	0.104
	(0.296)	(0.058)	(0.071)	(0.312)	(0.089)	(0.136)
Border [0,1] × CU 1[2005-2009]	0.172	0.060	0.064	-0.372*	-0.138**	0.084
	(0.292)	(0.065)	(0.081)	(0.224)	(0.054)	(0.089)
Border [0,1] × CM 1[t ≥ 2010]	0.142	0.098**	0.119	-0.547***	-0.034	0.060
	(0.282)	(0.039)	(0.101)	(0.202)	(0.040)	(0.095)
Core {0,1} × EAC 1[2002-2004]	-0.261***	-0.013	-0.191***	0.433***	0.073*	0.146
	(0.077)	(0.050)	(0.054)	(0.141)	(0.040)	(0.278)
Core {0,1} × CU 1[2005-2009]	-0.470***	-0.006	0.728**	0.314***	0.037	0.770***
	(0.097)	(0.045)	(0.332)	(0.105)	(0.039)	(0.177)
Core {0,1} × CM 1[t ≥ 2010]	-0.338***	-0.076	1.083***	0.361**	0.006	0.956***
	(0.065)	(0.069)	(0.375)	(0.142)	(0.041)	(0.111)
Individual Controls	YES	YES	NO	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	183,250	71,738	7,692
R-Squared	0.13	0.14	0.27	0.34	0.16	0.30
R-Squared -Within	0.11	0.10	0.25	0.33	0.11	0.29

Notes: The results in each column are produced by a separate regression. In columns (1) through (3), data come from the Kenya, Uganda and Tanzania Afrobarometer surveys rounds 1 through 7 sampled between 2000 and 2020. In columns (4) through (6), data come from the Kenya, Uganda and Tanzania Demographic and Health surveys (DHS) sampled between 1999 and 2020. The sample mean of the respective dependent variable is given in brackets above the estimates. For the DHS variable measured at the household level, *Wealth Index* in column (4), the answer from the main survey respondents (women) is used. For the coefficients on *Population Density* in columns (3) and (6), the level of observation is given by the unique sample enumeration areas of the respective survey. EAC 1[2002-2004] switches to one for individuals sampled from 2002 to and including 2004, CU 1[2005-2009] for individuals sampled from 2005 and including 2009, and CM 1[t ≥ 2010] for individuals sampled from 2010 onwards. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

percentage point increase in employed work in the CM period, which may or not be suggestive evidence of a common market protocol, i.e. free movement of labor potentially benefiting individuals closer to new markets. Second, while the positive welfare effect for households living in the core agglomerations is present for all years, the effect is either reduced (column 4) or constant (column 1) over the progressive deepening of the EAC, and even vanishes for one of the estimates (column 5). The estimates for changes in population density show the opposite dynamic, which constitutes the third noticeable insight of Table 4.2. While in the early EAC period, the results either show non-identifiable or even negative differential effects (column 6 and 3, respectively) in population density for survey locales close to the interior hubs, the estimates on later periods provide evidence of growing agglomeration tendencies. Expressed quantitatively, and in comparison to the rest of the country, the agglomeration in the economic hubs grew more strongly by a magnitude of about one standard deviation when measuring differences in the common market period to years preceding the EAC. Reconciling this observation with the positive welfare effects of agglomerations (columns 1 and 4), which predate these positive responses of population density in time, the finding provides suggestive evidence for individuals responding with the dynamic proposed in (4.52) of Section 4.4, i.e. respond to positive welfare differentials with migration inflows.[90] Given that estimated welfare differences stagnate or even decrease in the periods after these large population inflows (column 4), these findings may also suggest that by moving, individuals endogenously regulate welfare differences (downward). The loss of significance on the positive employed work outcome (column 5) from the CU period onwards may provide some (weak) evidence for this in the form of an increase in the elasticity of labor supply due to migration. One interesting aside on these first set of results is their connection to the discussed issue of "urbanization without growth" (Fay and Opal 2000). The effects show for increased consumption and agglomeration tendencies in primate cities, without concomitant expansion of labor market opportunities, they provide evidence on "urbanization without growth", i.e. favor the notion of "consumption-" rather than "production cities" associated with the urbanization of developing economies. Given their discussed low shares of resource dependency, this goes against the results of (Gollin et al. 2016), who illustrate resource dependency as a driving difference between these two distinct processes.

Extended Labor Market Results. The results on the remaining set of labor market outcomes introduced in the Section 4.6 are presented in Table C.9 of the Appendix. The table shows evidence on the differential development of having any type of activity outside of household work within the last year (columns 1 and 4), whether conditionally on such an activity, one is paid in cash compared to in-kind or no compensation, and, at which level of the occupational skill dimension, i.e. *Occupation Type* {1,3} the work is situated.

Overall, the results across variables and samples are mixed and provide no systematic evidence in favor of or against spatial inequalities following the establishment of the EAC. The most robust result is a positive change in the likelihood of having work outside of household activities for individuals closer to border regions. However, the effect is decreasing in absolute terms and in statistical significance over time (column 4). Also, evidence on the type or

[90] The effects on consumption are strongest in the EAC and CU years (columns 1 and 4), the effects on population density in the periods directly succeeding these eras (columns 3 and 6).

remuneration of this work shows that the likelihood of it being low-skilled and unpaid work is larger (compare columns 5 and 6). Concerning core agglomerations, there is some weak evidence on having work (column 4) and being remunerated for it, household's occurrence of experiencing monetary droughts increased in the CU period (column 3). However, given the inconsistency of these results, further interpretation of these results remains contentious. The next section looks at these issues arguably more fittingly, using the panel dataset of the sample. As anticipated above, one related possibility which cannot be isolated by the empirics is that a quick inflow of population attenuates a potentially positive labor market effect given the oversupply of it. This issue is better assessed with the panel dataset of the sample with which one may partly entangle these effects and to which we turn to next.

Kagera Health and Development Survey. Table 4.3 provides the final set of main results, reporting estimates produced from the Kagera Health and Development Survey (KHDS). A few notes towards the interpretation of results in comparison to the Afrobarometer and the DHS. First, given the timeframe of the survey, I am unable to provide longer-term evidence as the third and final sampling round was conducted in 2010 and Table 4.3 thereby reports estimates on the effect of the EAC and CU periods only. Second, some survey items were not administered in 2010, such that values cannot be estimated for these time periods either. Third, concerning the interpretation of the *Core* {0,1} dummy. Given the spatial limitations of the KHDS survey in the first waves between 1991-1994, I am unable to measure outcomes for households in Dar es Salaam prior to 2004. Hence, the DiD estimate on *Core* {0,1} is given by comparing differences of individuals over time living in Kagera's urban capital "Bukoba" in addition to those individuals who have later moved to (or were initially sampled in) Dar es Salaam or Kampala in the second and third rounds. The robustness checks in Subsection 4.7.2 removes the latter group and thereby provide a test on differential spatial sorting and on the general tendency of agglomeration predicted in Section 4.4.[91] Fourth and last, as a panel, the KHDS allows the inclusion of household-, respectively, individual fixed effects, and identification thereby stems from changes within households or individuals across time.[92]

Columns (1) and (2) depict the first set of results which test for differential changes in per capita consumption of food and non-food items as well as the value of durable items in the household. Both figures are expressed in deflated Tanzanian Shilling (TZS) priced in constant 2010 and 2004 levels, respectively. Households living in core agglomerations increased their consumption following the establishment of the EAC compared to households in the auxiliary, while households living closer to the regions bordering the new markets do not evince statistically significant differences in either integration period. Quantitatively, each individual living in households in core agglomerations consumes an extra of over 170,320 TZS worth of

[91] I pick up concerns about spatial sorting in Subsection 4.7.2 on Robustness Checks, Validity Tests, and Extensions.
[92] Depending on whether questions are administered at the household- or individual-level.

food and non-food items more than in the pre-EAC period, compared to the development of the rest of the country over the same time span. In relative terms, this indicates an increase of

Table 4.3: Difference-in-Differences using the KHDS Survey

	Dependent Variable					
	Kagera Health and Development Survey (KHDS)					
	Consumption		Income			Agglom.
	Annual p.c. Consumpt. (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
Border [0,1] × EAC 1[2004]	-853.376	187.152	0.581*	0.071	0.107	-0.848**
	(1072.010)	(754.481)	(0.331)	(0.049)	(0.261)	(0.384)
Border [0,1] × CU 1[2010]	-1317.845		0.432		-0.010	-1.631***
	(1154.386)		(0.305)		(0.404)	(0.586)
Core {0,1} × EAC 1[2004]	170.320***	706.607***	-0.013	0.026***	0.134***	0.891***
	(42.328)	(192.323)	(0.023)	(0.008)	(0.021)	(0.070)
Core {0,1} × CU 1[2010]	275.036***		-0.039		0.069	0.891***
	(57.368)		(0.025)		(0.067)	(0.099)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,492	2,695	24,972	14,254	15,685	2,292
Observations - Fixed Effects	3,816	2,363	12,747	6,988	6,253	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.43
R-Squared -Within	0.12	0.09	0.04	0.01	0.14	0.41

Notes: The results in each column are produced by a separate regression. Data come from the Kagera Health and Development Surveys (KHDS) collected in four waves across 1991-1994, as well as one wave in 2004 and 2010, respectively. In columns (1) and (2) outcome variables represent aggregate household information provided by the head of the household provided by the head of the household, in columns (3) through (5) they are administered on an individual level. For the coefficient on *Population Density* in column (6), the level of observation is given by the unique sample enumeration areas available in the sample. Certain indicators were not sampled in the survey wave of 2010, which is why there is no estimate given for these columns. The sample mean of the respective dependent variable is given in brackets above the estimates. Border [0,1] is the inverse, relative within country distance to the nearest Border crossing. Core {0,1} is a dummy indicating individuals living in the core agglomeration of their respective country (i.e. Dar es Salaam and Kampala). For the initial KHDS survey waves "Bukoba" - the capital of Kagera represents the core agglomeration. EAC 1[2004] switches on for individuals (re-)sampled in 2004. CU 1[2010], switches on for individuals (re-)sampled in 2010, the second re-interview period of the KHDS. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The regressions also include an indicator whether the household is living in proximity to (former) refugee camps. The regressions testing household-level outcomes, columns (1) and (2), include household fixed effects, the regressions testing individual-level outcomes, columns (3) through (5), include individual fixed effects. All regressions include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

roughly 31% compared to the average of the sample mean. This figure grows to almost 50% in the post-EAC years. Households also increase the consumption of consumer durables, owning a stock of goods in value of over 706,607 TZS higher than the comparison group. While I cannot identify positive effects on a higher likelihood of having employed work in agglomerations, there is a weakly significant effect at borders. A change from the median distance to borders in the sample to the minimum distance leads to an increase in the likelihood of employed work by 5.0 ($=0.581\times0.086$) percentage points. However, the remaining labor market outcomes paint a more consistent picture compared to the cross-sectional results. That is, individuals in core agglomeration profit differentially more from the establishment of the EAC both in terms of extensive and intensive labor market outcomes. Living in core agglomerations compared to the periphery increases the likelihood of having salaried work by 2.6 percentage points and raises the occupation type in which the individual works by 0.134 units. In other words, the skill gap in work done by individuals in capital cities versus peripheries increased by another 11% relative to the sample mean.

As a further exploration into these dynamics Table C.10 of the Appendix provides the findings on the remaining labor market outcomes introduced in the data Section 4.6. The findings show no differential effects across space for the likelihood of having worked in the past year or showing an increase in monthly salary.[93] Table C.10 also tests an expanded set of consumption components, i.e. tests for differences in food vs. non-food consumption and tests the value of the occupied dwelling of the household. With the exception of row 3 of column (2), all results confirm the evidence shown in Table 4.3, namely that households proximate to borders do not gain differentially more than interior regions but those living in the core agglomerations, which show strong, positive effects on consumption and wealth. Concerning the last dimension of and contributing to spatial inequality, population density, we see quantitatively highly similar and qualitatively identical results compared to the cross-sectional samples. The disparity in population density is growing, by up to half a standard deviation as measured in 2010. Evincing strengthened agglomeration patterns.[94] In sum, the results from the panel setting support the evidence produced from the cross-sectional samples.

4.7.2. Robustness Checks, Validity Tests and Extensions

Robustness Checks. As a first insight into the stability of the presented results, I conduct an array of robustness tests summarized in Table C.3 through Table C.8 of the Appendix and briefly discussed here. As before, all of the results are produced using regression specification (4.53), unless otherwise indicated. The tables also report the baseline coefficients from Table

[93] Interestingly, regressions relaxing the strict individual fixed effects show a strong differential increase of wages in agglomerations and a decrease of them at borders (significant at the 10 and 5 percent level).
[94] An explanation of the larger and statistically more significant effect in the initial EAC period is that this estimate measure changes in the tail end of the EAC period (i.e. in 2004), in contrast to the earlier samples, which average changes between 2001 and 2004.

4.2 in the top rows with which to compare the results of the adapted estimations. Given the lack of effects for border regions throughout the samples and variables tested, the table confines the reporting of the three difference-in-difference coefficients on *Core* {0,1}. The full results including the DiD estimates for *Border* [0,1] can be accessed in the source tables which are referred to in these summary tables at each instance of the respective test.

To begin, (a) allows for the clustering of standard errors at the enumeration area level instead of implementing Conley standard errors. (b) Removes all individual and geographic controls from the regression.[95] (c) Adds the extended set of geographic controls anticipated in Subsection 4.6.1, namely locations' *Absolute Latitude*, *Malaria Ecology* (Sachs et al. 2004), *Navigable Rivers*, *Major Lakes* and *Major Harbors* to control for other trade-related influences and adds the dummy *non-EAC* $\leq$ *100km* {0,1} as well as the interaction with all EAC period dummies to net out effects potentially stemming from a change (loss) in market access at non-EAC borders. (d) Employs the sample survey weights provided by the Afrobarometer and DHS, accounting for the pooling across countries and years by standardizing the weights for each country-survey round pair.[96] (e) Excludes low-precision localities. For the Afrobarometer, this is implemented by dropping all observations for which the AidData precision code is above 2 (AidData 2017).[97] In the DHS survey, I drop all observations for which coordinates are not generated from a GPS receiver used by the fieldworker.[98] This test cannot be conducted for the KHDS as there is no distinction in the precision of GPS locales. (f) Replaces the *Core* {0,1} dummy in specification (4.53) with *Agglomeration* {0,1} which switches to 1 for households living within 50km of an "urban center" demarcated as such in the year 2000. The list of urban centers I use are provided by the European Commission (2019).[99] This can be seen as a general test on the agglomeration tendency of the results in Subsection 4.4.5 irrespective of border distance or proximity to the capital cities. As a further test on this, I also try the *Urban* {0,1} dummy attached to the surveys.[100] (g) Reduces the spatial cut-off criteria for living in core agglomerations to 25km and (h) to 10km, respectively. (i) Splits the CM period into two dummies, namely *CM 1[2010-2014]* and add a post-CM time period *post-CM 1[t* $\geq$ *2015]* to provide a test on the hypothesized transitory shift to a welfare-equalizing equilibrium as discussed in the previous section.[101] (j) Excludes the individuals in the sample which did not live in the survey location at least three years before the establishment of the EAC (before

---

[95] For the KHDS survey, I cluster observations at a specific geographic delineation, which is based on 2 decimal places of latitude-longitude combination, i.e. raster of slightly larger than 1 square km. at the equator.
[96] I.e. transform the weights such that they sum to 1 for each pair.
[97] The scale ranges from 1-8. Using precision code 1 leads to a loss of data in the range of 40-50%.
[98] See https://dhsprogram.com/Methodology/upload/MEASURE-DHS-GPS-Data-Format.pdf.
[99] The definition reads: "The spatially-generalized high-density clusters of contiguous grid cells of 1 km2 with a density of at least 1,500 inhabitants per km2 of land surface or at least 50% built -up surface share per km2 of land surface, and a minimum population of 50,000." (European Commission 2019; 13). For my purposes, I use a minimum population threshold of 100,000.
[100] Results can be requested from the author.
[101] Notice that "post-CM" has no further meaning other than nomenclature. The common market of the EAC has continued to persist.

1999), i.e. excludes “post-EAC Migrants”.[102] (k) Excludes DHS’ special survey rounds including the AIS, KAP and MIS surveys, leading to a sample of 141,879 individuals living in 88,196 households located across 5,110 survey locales and interviewed between 1999 and 2016. (m) “logs” relevant dependent variables, i.e. transforms the variables in the way log(1+y), for outcomes expressed in constant Tanzanian Shilling (only Table C.5).

While the general upshot of all these sensitivity checks is that all previous conclusions and interpretations hold, there are some interesting takeaways for two specific tests which also have bearings on the theoretical results. For one, by splitting up the CM years into two periods (j), we notice that the convergence anticipated in Table 4.2 is at least partly corroborated. While the Afrobarometer shows seemingly unaltered results compared to the years 2010-2014, estimates produced with the DHS survey, which include three more survey years from 2017 to and including 2020, drop both in size as well as significance. Most notably, the coefficient on population density is halved compared to the “early” CM period. And secondly, when exchanging the *Core* {0,1} dummy with a wider selection of urbanities (f), the effects are either non-significant, significantly weaker, or point in the opposite direction. Particularly interesting are the effects on population density which are insignificant for most of the estimates, and significantly lower in magnitude (by an order of 4) when compared to the developments in the core agglomeration. Only for the KHDS are results similar in magnitude and significance. This may be reconciled by the fact that the dummy is a test on “urbanities” rather than economic hubs existent in 2000, and that “urbanities” are endogenous to the outcome. However, given that endogenous formation of agglomeration is precisely what theory dictates, I take this observation seriously and evaluate this effect in the context of the main theoretical model, i.e. with heterogeneous intra-national space. I do so by estimating a triple-difference specification which tests for a differential effect for (endogenous) agglomerations at border regions, which is effectively done by interacting the previous treatment interaction *Border* [0,1] $\times$ *EAC [t ≥ 2002]* with the *Urban* {0,1} dummy provided in the Afrobarometer and DHS. The results are shown in Table C.11 of the Appendix. The combined effect in row three provides no convincing evidence that agglomeration newly establishes at borders, and that relative welfare increases at those agglomerations are differentially higher. Only the DHS wealth index is significant at common levels (column 4), providing evidence of reduced relative wealth of households located in urbanities at the border. Panel b) tests the extended set of labor market outcomes. Again, most results are insignificant or strongly negative (columns 5 and 6). This confirms the previous results which did not show a draw to the border.

Next to this battery of tests on the stability of the DiD coefficients on *Core* {0,1} I also provide relevant robustness test on *Border* [0,1], which are summarized in Table C.6 through Table C.8. (a) Replaces distances calculated over road with a simple beeline (“as the crow flies”) distance to EAC border crossings. (b) Approaches in similar but estimates effects for the

[102] Because of missing migration information, this test is only possible for the DHS and KHDS sample.

nearest (beeline) distance on the entire borderline formed by two EAC member countries. (c) "logs" the original road distance measure, in the fashion log(1+x). Relatedly, Table C.82 and Table C.83 provide results for a more flexible distance specification by entering the continuous distance to internal EAC border crossings and core agglomerations as well as their squared value as explanatory variables. (d), (e) and (f) replace the continuous treatment indicator *Border* [0,1] with dummy indicators of living within 100km, 50km, and 25km to the Border Crossings. In sum, the previous findings of non-identifiable, or negative, differences across border distance before and after the EAC's re-establishment remains robust.[103] The only slight deviations from the results presented before are shown in the border dummy specifications (d) through (f) which show evidence of positive effects on consumption for the narrowest distance bandwidth ($\leq$ 25km). We explore some caveats with these effects in the extensions below.

One last test on the robustness of the results is provided in Table C.20 and Table C.21 where I test for the composite items of the two indices employed, i.e. *Lived Poverty* [0,4] and *Occupation Type* {1,3}. As expected, the results are robust to the decomposition. Interestingly, in line with positive effects at borders in the CM period we see a shift of occupations away from agrarian jobs to semi-formal engagements, at least for the Afrobarometer and KHDS surveys, respectively. However, this effect is also apparent for individuals in core agglomerations, and larger in size. The last noteworthy comment regards the alternative consumption and wealth indices, the *International Wealth Index* and the C*omparative Wealth Index*. What is more strongly confirmed in these results is that the effect of increased consumption in agglomerations is a one-off effect which is decreased strongly across time. Again, this provides suggestive evidence on population inflows to equalize welfare differences due to negative agglomeration externalities (e.g. congestion, housing market).

Validity Tests. The fundamental assumption behind the employed difference-in-differences design requires that absent policy change, the spatial disparity in households' welfare within the EAC countries would have evolved "in parallel" i.e. continued their relative pre-intervention trajectories. In other words, for our estimates to represent a causal relationship, nothing other than the policy of regional market integration should have induced a differential welfare change across space in the timespan (shortly) the before and after the re-establishment. While in practice ultimately never verifiable, I provide three distinct pieces of evidence that may strengthen our confidence in this assumption.

Before more formal tests are discussed, there needs to be an initial check on other (concomitant) policy measures with the potential to influence the economic geography of the respective countries. In the timespan between 1995 and 2010, the most relevant policies I was able to identify mainly concentrated on trade facilitation. For instance, the Northern Corridor Transport Improvement Plan (NCIP) of 2004 aimed to improve transport infrastructure to

[103] One exception is the now positive effect on *Occupation Type* {1,3} for border regions in the CM period.

facilitate trade integration. I do not deem this investment as undermining the results, as the completion of project goals aimed within the NCIP were temporally lagging the main results presented here by a large margin (see World Bank 2016). A related concern is that the facilitation of "one-stop-border-posts" (OSBPs) may lead to differential success of integration across border. However, as in the case of the NCIP most OSBPs were erected many years after the large increases first set in in our results (Cadot et al. 2015; EAC 2015).[104] Another initial concern may be displayed by "Export Processing Zones" (EPZs) or "Special economic Zones" (SEZs) in or near the core agglomerations defined in the paper. While all of the three member countries actively promote SEZs, the timing as well as the spatial pattern relief concerns of an effect entirely attributable to such developments. For instance, in Kenya, the majority of EPZs are outside of Nairobi, many of them in the port of Mombasa not included in the *Core* Definition.[105] In Tanzania, the operation of EPZs is possible since the ratification of the EPZ act of 2002 and the SEZ act of 2006, respectively. However, while data on firms operating under such licenses is untransparent with several contradicting reports on the absolute number, they agree that the general impact to industrialization was small (Andreoni et al. 2022). Concerning the spatial dimension, most recent data suggests that the majority of firms operating under an SEZ license are outside of Dar es Salaam (Kinyondo et al. 2016; Andreoni et al. 2022).[106] Lastly, the distribution of EPZs in Uganda does in fact evince a stark regional disparity skewed towards the Central region enclosing Kampala (UFZA 2022). However, SEZ only began operating after 2014, which is the year the Ugandan Free Zones Act was ratified, rendering our results robust to this development.

Placebo Tests. The first formal test against the difference-in-differences assumption is presented in Table 4.4. The results shown are produced in the identical way as they were in Table 4.1, but use data on contiguous "placebo countries", i.e. estimate a differential change across pre and post-EAC time periods for non-EAC countries bordering Kenya, Tanzania and Uganda. The countries available in the data are Malawi, Mozambique, and Zambia in the Afrobarometer. The DHS expands the countries to include Ethiopia and Rwanda.[107] In these countries *Border* [0,1] represents the inverse, relative within-country road distance to the nearest major road crossings with an EAC country and *Core* {0,1} identifies individuals living within the respective core agglomeration (i.e. Addis Abeba in Ethiopia, Kigali in Rwanda, Lilongwe in Malawi, Lusaka in Zambia and Maputo in Mozambique). The results provide

[104] I also estimate a potential heterogeneity across border regions in Table C.17 of the Appendix, which is discussed at the end of this subsection. The findings provide no evidence of such an influence.
[105] While many of them are in the Machakos county, next to Nairobi, road and beeline distances are above the commonly used threshold of 50km. Also note that the results are robust to the narrower spatial delineation (i.e. 25 and 10km), if one assumes laborers to commute to these EPZs.
[106] Even though the largest of the SEZ zones are located in the Dar es Salaam-Bagamoyo corridor. Bagamoyo is in similar distance to Dar es Salaam as Machakos to Nairobi in Kenya.
[107] We include data from Rwanda only until 2005, given that the country joined the EAC in 2007.

Table 4.4: Placebo Tests – DiD in contiguous non-Intervention Countries

	Dependent Variable					
	Afrobarometer			DHS		
	Consumpt.	Income	Agglom.	Consumpt.	Income	Agglom.
Panel a)	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.25]	[0.24]	[0.00]	[3.20]	[0.17]	[0.00]
Border [0,1] × EAC 1[t ≥ 2002]	0.179	0.062	-0.031	0.682***	0.015	-0.176
	(0.143)	(0.075)	(0.159)	(0.137)	(0.039)	(0.128)
Core {0,1} × EAC 1[t ≥ 2002]	-0.142	-0.040	0.773**	0.130	-0.067	-0.111
	(0.103)	(0.041)	(0.303)	(0.167)	(0.087)	(0.400)
Observations	28,541	18,994	2,329	234,346	126,587	9,026
R-Squared	0.09	0.19	0.23	0.30	0.14	0.27
R-Squared -Within	0.05	0.14	0.22	0.29	0.13	0.25
	Afrobarometer			DHS		
	Income					
Panel b)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[0.49]	[1.62]	[2.12]	[0.67]	[1.48]	[0.44]
Border [0,1] × EAC 1[t ≥ 2002]	0.022	0.027	-0.579***	-0.033	-0.055	-0.106
	(0.074)	(0.138)	(0.141)	(0.081)	(0.104)	(0.120)
Core {0,1} × EAC 1[t ≥ 2002]	-0.038	-0.016	-0.361***	0.083*	0.048	0.040
	(0.055)	(0.068)	(0.073)	(0.048)	(0.068)	(0.091)
Observations	19,191	12,715	28,274	310,644	187,498	196,153
R-Squared	0.18	0.31	0.11	0.17	0.24	0.18
R-Squared -Within	0.10	0.27	0.06	0.15	0.22	0.13
Individual Controls	YES	YES	NO/YES	YES	YES	NO/YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table conducts a "placebo" analysis by testing for a spatially differentiated effect across contiguous, non-EAC countries within the time frame of the EAC's establishment and expansion. As such, in columns (1) through (3), data come from the Malawi, Mozambique and Zambia Afrobarometer surveys rounds 1 through 7 sampled between 1999 and 2018. In columns (4) through (6), data come from the Ethiopia, Malawi, Mozambique, Rwanda, and Zambia Demographic and Health surveys (DHS) sampled between 2000 and 2019. The sample mean of the respective dependent variable is given in brackets above the estimates. EAC 1[t ≥ 2002] switches to one for individuals sampled from 2002 onwards. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The results in each column and panel are produced by a separate regression. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

suggestive evidence that the differential welfare change in economic hubs was not a larger, regional trend within the timeframe under evaluation. More generally, the results do not show any consistent evidence of spatial inequalities increasing or decreasing over time. While there are positive effects on population density in the Afrobarometer, the arguably more precise spatial estimates produced from the DHS does not confirm this finding. Further, while there some improvements in wealth at borders (column 4 of panel a), column 3 in panel b) shows that the occurrence of cash income droughts was similarly strongly decreased in core agglomerations.[108] The rest of the results show no hint towards a differential development following the EAC, which strengthens the confidence in attributing the previous estimates to the EAC. For completeness, Table C.12 of the Appendix shows the companion results which splits the *EAC 1[t ≥ 2001]* dummy into the three integration periods. The results from this more nuanced regression do not provide interpretations different than the ones just discussed. One interesting aspect, however, is that, in the DHS sample, many of the significant effects for core agglomerations are only apparent in the EAC period. Given that they almost universally imply a negative impact and show no accompanied increase at borders, I deem these results as non-indicative of other region-wide processes influencing the results from the EAC countries.[109]

Pre-Trends. One weakness of the datasets employed is the narrow timeframe of pre-integration periods. While a difference-in-difference approach implemented on data shortly before and shortly after the intervention may alleviate concerns of other concomitant policies driving the results, it excludes the possibility to net out potentially longer-term, unit-specific time trends from the identified treatment effects. Also, it does not allow a test on the pre-intervention evolvement of relative welfare within the countries. Given that this is a crucial test concerning the validity of the results, I try to circumvent this data restriction by drawing on region-based estimates from the country, i.e. include non-GPS coded survey rounds of the DHS which extend back to 1988.[110] While there is no finely gridded information on the location of respondents available, DHS provides regional-based information, which identifies individuals' residence on the GADM-1 level.[111] I use this information and construct a new *Core Region* {0,1} dummy indicating whether individuals live in the capital city region. Note that while many of the surveys in pre-GPS years do not provide granularity of the measure employed previously, given their political and economic importance, Nairobi and Dar es Salaam were nonetheless demarcated as their own region at finer levels.[112] As such, the use of this *Core Region* {0,1} dummy likely captures much of what is also measured by the GPS-based *Core*

[108] Going from the median distance of the sample to the border (912km) renders an effect size of -0.257 (=0.597×0.43).
[109] Four out of five estimates imply negative welfare changes at core agglomerations.
[110] See section on Data sources 4.6.
[111] See https://spatialdata.dhsprogram.com/boundaries/.
[112] For Kampala in Uganda, this was only done in the 1988-1989 survey for non-GPS surveys. After this year and until 2002, Kampala was located in the "central" region from which point on it was its own district again.

{0,1} dummy used in the main estimations.[113] Concerning the definition of border regions, the matter is not as straightforward. Many of the regions defined in the DHS, particularly in the pre-GPS years, could be considered both border as well as interior regions given their vast extent to the inland of countries (see footnote 111). Hence a dummy categorization as used for capital cities will not likely suffice to capture true border households. I therefore try to improve upon a simple dummy with the following steps. First, I assign households in all survey rounds a regional correspondence for which boundaries are consistently available from as early on as possible. This yields 7, 20 and 4 consistently demarcated regions for the years 1988 until 2020 for Kenya, Tanzania and Uganda, respectively.[114] Second, using all available GPS samples, I retrieve the mean, modal as well as median values of road distances within these boundaries as central tendencies of the distribution within them. Third, I assign these values to all households nested within a specific region, especially now also for the non-GPS households. Finally, I encode those households living in regions ranking in the 10th percentile of these boundary-based distances within their country as a 1 in the dummy *Border Region* {0,1}.[115] Under the assumption that the population distribution within regions has not dramatically changed between 1988 and the latter two decades, this arguably allows a more precise ordering of border to non-border regions within the sample.

Table 4.5 presents the result using these two region-based indicators, using the mean as the central tendency for the *Border Region* {0,1} dummy. Panel a) relates to Table 4.1 and shows the aggregate difference-in-differences estimates before and after the establishment of the EAC, with the omitted time period, i.e. the reference group now defined from 1988-2000.[116] Panel b) explicitly tests for pre-trends by introducing a *pre-EAC 1[1996-2000]* dummy which tests for differential changes in border-, respectively, capital city regions over the time period before the establishment of the EAC. The reference group for all estimates shown in this panel has therefore changed to pre-EAC 0[1988-1995]. Panel a) confirms the result shown in Table 1 from the GPS-based measurements. There are no indications of a positive relative household welfare change across all tested outcomes.[117] And further, we see the same strongly positive effects of welfare for households living in the core agglomerations.[118] Turning to the test of parallel trends in panel b), we see no robust indication of an unequal trend of core vs. peripheral regions. Only one of the estimates is significant at conventional 5%-levels (column 6). However,

---

[113] Indeed, for survey rounds where GPS information is available, the Pearson correlation coefficient between *Core* {0,1} ≤ 50km (the main dummy) and *Core Region* {0,1} is 0.694, for Core {0,1} ≤ 25km it increases to 0.749 and for Core {0,1} ≤ 10km it correlates to 0.681.

[114] For one round, the 1991-92 Tanzania DHS, I am confined to 6 regions.

[115] For survey rounds where GPS information is available, 61.6% (54.5%) of individuals scoring a 1 for this dummy live within 100km of the (granularly) calculated road distance. To compare, a simple border categorization leads to a value of 37.5%. The results are also robust to using the 20th percentile.

[116] Standard errors are clustered at the "region" level and are further robust to clustering at the "cluster" level also.

[117] Indeed, column 6 shows evidence of a negative effect of cash employment.

[118] Note that *Population Density*, as location-based outcome variable, is constructed in the same fashion as the region-based distances.

Table 4.5: Region-Based Estimates and Pre-Tests

	Dependent Variable					
	DHS (Region-based)					
	Consumpt.	Income	Agglom.	Income		
Panel a)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked Last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.12]	[0.19]	[0.00]	[0.76]	[1.52]	[0.53]
Border Region {0,1} × EAC 1[t ≥ 2002]	0.021	-0.033	-0.104	-0.082	-0.042	-0.133**
EAC 0[1988-2000]	(0.170)	(0.022)	(0.137)	(0.063)	(0.051)	(0.055)
Core Region {0,1} × EAC 1[t ≥ 2002]	0.386***	0.090***	2.203*	-0.037	0.089*	-0.002
EAC 0[1988-2000]	(0.122)	(0.022)	(1.223)	(0.054)	(0.050)	(0.030)
Observations	258,820	104,440	11,841	236,646	140,613	136,163
R-Squared	0.28	0.16	0.54	0.21	0.25	0.18
R-Squared -Within	0.27	0.10	0.53	0.18	0.16	0.08
Panel b)						
Border Region {0,1} × pre-EAC 1[1996-2000]	0.098	0.024	0.084	-0.296**	0.014	-0.063
pre-EAC 0[1988-1995]	(0.125)	(0.029)	(0.082)	(0.121)	(0.060)	(0.049)
Border Region {0,1} × EAC 1[t ≥ 2002]	0.088	-0.017	-0.068	-0.304**	-0.031	-0.171**
EAC 0[1988-1995]	(0.200)	(0.032)	(0.120)	(0.134)	(0.078)	(0.065)
Core Region {0,1} × pre-EAC 1[1996-2000]	0.440*	0.022	0.131	-0.245*	0.099	-0.185***
pre-EAC 0[1988-1995]	(0.223)	(0.026)	(0.261)	(0.121)	(0.102)	(0.052)
Core Region {0,1} × EAC 1[t ≥ 2002]	0.652***	0.102***	2.255*	-0.211	0.147	-0.094**
EAC 0[1988-1995]	(0.229)	(0.023)	(1.294)	(0.137)	(0.087)	(0.045)
Observations	258,820	104,440	11,841	236,646	140,613	136,163
R-Squared	0.28	0.16	0.54	0.21	0.25	0.18
R-Squared -Within	0.27	0.10	0.53	0.18	0.16	0.08
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table makes use of the non-GPS survey rounds of the Demographic and Health Surveys (DHS) sampled before 1999 and additionally conducts “pre-tests” towards the difference-in-differences approach. The data thereby come from the full sample of Kenya, Tanzania, and Uganda DHS surveys sampled between 1988 and 2004, making use of AIS, KAP and MIS rounds as well. The sample mean of the respective dependent variable is given in brackets above the estimates. Border Region {0,1} switches to one for individuals living in a region with a median road distance to Border crossings below the 10th percentile of all (within-country) GPS-border distances in the sample. Core Region {0,1} is a dummy indicating individuals living in the region which hosts the core agglomeration of their respective country (i.e. Nairobi, Dar es Salaam and Kampala). EAC 1[t ≥ 2002] switches to one for individuals sampled from 2002 onwards. Pre-EAC 1[1996-2000] switches to one for individuals sampled in survey years between 1996 and including 2000. As such, in panel a), the reference group of the estimates are comprised of individuals sampled in the full pre-EAC period, i.e. from 1991 to 2000, while in panel b), the reference group is formed by individuals sampled between 1988 and including 1995. Hence, the DiD estimate on “pre-EAC” in panel b) represents the pre-test. The results in each column and panel are produced by a separate regression. All regressions include individual-level controls for respondents’ age, gender, as well as education. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for clustering at the “region” level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

the coefficient signals a sign opposite to what is expected, and then attenuates for the EAC estimate which would be consistent with a positive welfare effect in core agglomerations. I conduct the same estimations with the placebo-countries analyzed in Table 4.4, and the results are reported in Table C.13. As expected from the earlier findings of the contiguous countries, the results do not provide consistent evidence of a growing spatial inequality, which further negates the notion of an ongoing trend towards more pronounced economic inequality in the larger East African region. Importantly, the results are fully robust to using the median or mode road distances to the border as well. They are additionally robust to using the simpler border categorization. Appendix Table C.14 and Table C.15 also provide the results on the temporally disaggregated DiD effects. All of the previous conclusions remain. The added temporal dimension of the region-based samples allows for one further (placebo) test which is presented in Table C.16, Figure C.9 and Figure C.10.[119] The estimates shown are produced from the identical regression specification as for Table 4.1, but now the single temporal treatment dummy switches to one for varying (placebo) intervention years which are plotted on the x-axis, respectively. The estimates therefore test for differential changes across space when one would not expect such changes if it was indeed the re-establishment of the EAC that has influenced the spatial pattern during the *true* intervention years. Importantly, the sample used for the estimation is re-centered at the respective (placebo) intervention year such that (actual) post-EAC data does not influence the estimates too strongly. The estimation always uses the earliest possible survey waves for which entries of the respective outcome are variable but truncates the data to the right. For example, in the case of the placebo treatment year 1994 the sample includes household-level data from 1988 to 1997. For the placebo treatment year 1995, the sample is expanded to include 1998, and so on. The results shown are in accordance with the true intervention kicking in the early 2000s. There are almost no significant differential trends for border or core agglomerations throughout the 1990s. For the few that are, coefficient stability may be a concern, and is likely due to the rather infrequent sampling in the (very) early sampling years.[120] Effects become significant from 2000 onwards with a peak of point coefficient size in 2002.

Columns (4) through (6) of Table C.16, as well as Figure C.9 and Figure C.10 provide the last piece of evidence in this regard, by plotting these same difference-in-difference estimates together with their 95% confidence interval against those estimated from the sample of placebo-countries defined above.[121] Note that for legibility, the variables in the plots are transformed into z-scores and can therefore not be directly compared to Table C.16 concerning coefficient

[119] To improve legibility, the figures are split in that Figure C.9 analyzes the effects on core agglomerations, and Figure C.10 on border regions.
[120] For instance, for the *Wealth Index* the first available data is provided in the Kenya DHS 1993, with the next available samples available in Uganda in 1995, Tanzania in 1996 and again, Kenya in 1998.
[121] Note that the plot starts with the 1997 estimate given that data for the placebo countries as only few samples provide the main dependent variables before this date. Again, *Wealth Index* is available only in Malawi in 1992 Rwanda in 1992, and Zambia in 1996. 1997 is the first year for which repeat observations in a country are available.

size, albeit against each other. As can be depicted in Figure C.9, placebo estimates do not evince the same "hump" as those from the EAC countries in the first years following the establishment of the EAC. Figure C.10 shows that a discontinuous border effect is apparent neither in EAC- nor in placebo countries. What is interesting in the results on placebo countries is that they show evidence of a positive effect on consumption in borders in the early EAC timeframe paired with negative effects at their core agglomerations following these years. A thorough investigation of this is out of the scope of this paper, but needs to be mentioned, nonetheless. In sum, these results strengthen our confidence that our estimates reflect the effect of regional market integration and do not conflate them with longer-term trends over time or region-wide effects at a specific point in time.

Spatial Sorting. The last validity test offered addresses concerns about the potential spatial sorting of individuals within the intervention years, which would systematically influence the measured treatment. Specifically, there exists the possibility that skilled individuals positively select, i.e. move, into border-, respectively, capital city regions within treatment years in attempt to profit from the policy or other complementarities such as more productive firms, or joint productivity (Kremer 1993). This would lead to an upward bias of the observed results. Given that most of the results are drawn from cross-sectional households, entirely excluding this possibility is not possible. However, I offer four-part evidence that this may be unlikely. First, using regression specification (4.53), I test for the duration of having lived at the current locale using the KHDS survey to test for differential migration into border or capital city regions following the establishment of the EAC as well as the common market (CM) protocol. The results are shown in column 6, panel b) of Table C.10 in the Appendix and provide no evidence on differential migration movements of sample respondents across regions induced by the EAC. Importantly, individuals are not more likely to have recently migrated to the primate city before and after the EAC compared to other locales. Second, as seen in the robustness checks, the results are (highly) robust to excluding "post-EAC migrants", i.e. individuals that moved to the respective region of residence less than 3 years before the EAC was operational. Importantly, this is also the case for the results from the KHDS panel, which excludes the possibility that "movers" drive the aggregate effects observable in this sample. Third, using the KHDS panel survey once more, I specifically test for spatial sorting of high-skilled individuals into the capital cities of Dar es Salaam and Kampala by regressing an indicator of having moved to these cities onto personal characteristics such as education, gender and age as well as country-time fixed effects.[122] I also regress a dummy of living in these capital cities on this set of covariates. In both cases, I find no systematic evidence of selection into economic hubs, with only the coefficient on education marginally significant (p values of [0.097] and [0.078], respectively) and small effect sizes in the area of 0.5 and 0.9 percentage points (likelihood

[122] There are no individuals in the KHDS which have moved to Nairobi over the survey timeframe.

increases) for each year of schooling.[123] Fourth, note that I consistently include relevant individual-level controls such as age, gender and education which may conflate an EAC effect with differential spatial sorting. Note also that the inclusion of which do not alter the results dramatically (compare the baseline coefficients to panel (b) in Table C.3 through Table C.8, respectively). In addition, as is discussed in the extensions provided in the next paragraph, I test for the main motivation of migrating to the current place of residence for KHDS sample respondents. The effect does not show evidence for more or less economically motivated migration before and/or after the EAC's re-establishment.

In sum, these results do not provide evidence in favor of a negative selection out of, and a positive selection into, border- as well as capital city regions, respectively, which could drive the results identified at the aggregate.[124] And while none of the validity tests can completely eradicate concerns about a potential ongoing trend, they do not undermine the findings to a degree which would cast significant doubts about the nature of the main results, nor the validity of the identification strategy employed.

Extensions. I provide three extensions to the existing set of results, which provide further insights into the nature of the findings. First, given the partly significant coefficients using border dummies (see panels (e) to (f) in Table C.6 through Table C.8), I explore a heterogeneous effect across the different EAC border segments. I do so by replacing the continuous measure of border distance *Border* [0,1] with dummies switching to 1 for households living within 50km of border distance to the three country-border pairs from both country directions. As such, I again estimate regression equation (4.53), albeit with separate difference-in-differences effects for *TZA-UGA* {0,1}, *TZA-KEN* {0,1}, and *KEN-UGA* {0,1}, together with *Core* {0,1}. Note that this also constitutes a test on a parametric specification and identification of effects. The results are presented in Table C.17. The findings provide mixed evidence. Concerning the Tanzania-Uganda border section, Afrobarometer as well as DHS do not agree on the direction of the effect. Note that this particular border represents a special case, as it hosted a large inflow of refugees from Rwanda in the decade before the EAC to the south, i.e. in Kagera (Tanzania), which first caused a decrease in average living standards in this region and an increase in the years thereafter. The suggested reasons for positive turnaround in welfare over time include directly targeted programs (e.g. infrastructure) and agglomeration externalities as given by the longer-term settlement of the refugee population (Maystadt and Duranton 2019). In an additional robustness test (not shown here), I control for the distance to these (former) refugee camp sites and observe that the effects for these border dummies diminish significantly.[125] Paired with the fact that the panel estimates of the

[123] Results can be obtained from the author.
[124] Of course, many of these tests rely on the recall of individuals, which constitutes a general weakness of household surveys.
[125] Results can be obtained from the author.

main results, drawn mainly from the Kagera region, provide no evidence for a differential effect across border distance, these partly significant border dummies at this segment likely do not represent EAC-induced effect across distance.[126] Mostly negative effects on consumption together with some income measures are observed at the Tanzania-Kenya border (columns 1, 7, 8 and 11) and mixed effects are depictable for the income components at the Kenya-Uganda border section (columns 4, 8 and 9) as well as positive effect on population density which is, however, less than 7% of the differential effect when compared to the core.

The second extension makes use of the opinion polling of the Afrobarometer and KHDS surveys to evaluate (potentially altered) sentiments towards free trade and individuals' subjective wellbeing. Table C.18 presents these results showing the component of the Afrobarometer in panel a) and the results drawn from the KHDS survey in panel b). Concerning subjective wellbeing, there is no uniform direction of differential effects across space. While individuals at borders evaluate their life satisfaction as worse than 5 years prior, they assess the general standard of living as higher.[127] The rest of the variables tested are only administered in one of the survey rounds, which is why there is no DiD estimate possible. In these cases, we test simple differences between border- and core agglomeration regions to the rest of the country, having arguably benefited more from trade integration. While individuals living in border regions deem it as easier to cross international borders in order to work and trade in foreign countries, the rest of the tested opinions show for non-significant differences in border regions.[128] Most importantly, individuals situated at borders, or in capital cities at that, do not support the free movement of labor across countries (column 3), nor do they assess the EAC or the African Union (AU) as trade-facilitating entities as more or less helpful to their country (panel a), columns 5 and 6).[129] And lastly, individuals in capital cities show for a higher

[126] Another concern may be the sample size available for these border dummies in the Afrobarometer, for which most of the effects appear. For instance, in the pre-EAC periods, the TZA-UGA dummy ($\leq$ 50km) contains only 67 individuals, 73 for the KEN-UGA segment, and 92 for TZA-KEN. In the DHS, this increases to 200, 275 and 343, respectively.

[127] Specifically, it sets a prior to compare the current living standards to the former military rule. The survey question reads: "We are going to compare our present system of government with the former system of military rule. Please tell me if the following things are better or worse now than they used to be. People have an adequate standard of living." The response values range from 'Much worse' (1), 'Somewhat worse' (2), 'No change' (3), 'Somewhat better' (4) and 'Much better' (5). I remove the observations valued 'Don't know'. Regarding the question in column (1), the survey question reads: "When you look at your life today, how satisfied do you feel compared with five years ago?". The response values range from 'Much less satisfied' (1), 'Slightly less satisfied' (2), 'About the same' (3), 'Slightly more satisfied' (4) and 'Much more satisfied' (5). I remove the observations valued 'Don't know'. In Tanzania, this question is asking the respondents to compare their life to one year ago.

[128] The survey question reads: "In your opinion, how easy or difficult is it for people in [West/South/East/North/Central] Africa to cross international borders in order to work or trade in other countries, or haven't you heard enough to say?". The response values range from 'Very difficult' (1), 'Difficult' (2), 'Easy' (3), 'Very Easy' (4) and 'Never try' (7). I remove the observations valued 'Never Try'.

[129] The survey question of column (3) reads: "Which of the following statements is closest to your view? Choose Statement 1 or Statement 2. Statement 1: People living in [West/South/East/North/Central] Africa should be able to move freely across international borders in order to trade or work in other countries. Statement 2: Because foreign migrants take away jobs, and foreign traders sell their goods at very cheap prices, governments should protect their own citizens and limit the cross-border movement of people and goods." The response

tendency to dislike immigrants or foreign workers as neighbors.[130] This may be evidence for increased experience with competition on the labor market, but given the general insignificance of the other results, this result is not further interpreted. Moving to the results on simple differences in the Kagera Health and Development Survey, we see some indication of what is found in the main results.[131] That is, individuals in border regions have a lower likelihood of having had paid (formal) employment directly after migrating there. Going from the median border distance in the sample (103 km) to respondents directly at the border decreases the likelihood of paid (formal) employment after the move by 5.6 (4.1) percentage points. The rest of the results indicating higher or lower subjective household life satisfaction or wealth are generally insignificant with the exception of a positive effect on the current assessment of the households as rich rather than poor in core agglomerations.

The third and last extension tries to shed some light on the non-differential effects of the trade policy across border distance. Without reference to NEG theory, studies have shown that following trade, the pass-through of price changes following trade liberalization decays strongly in border distance (see e.g. Nicita 2009 for the case of Mexico in NAFTA). However, there is also evidence that borders in Africa are particularly “thick”, and that trade agreements may not result in improvements from such deviations in the law of one price, depending on local characteristics such as differences in ethnic makeup, language, or informal credit practices across borders (see e.g. Versailles 2012; Aker et al. 2014; Brenton et al. 2014). To test for a distance penalty in price pass-through, I make use of the price questionnaire in the Kagera Health and Development Survey of rounds 1 and 2 which provide market prices of various food items. The prices are expressed per unit (e.g. per kilogram) and averaged across rainy and dry seasons to control for periodic fluctuations. To test for the border pass-through, I estimate regression equation (4.53), and assess the logged price of four homogenous, heavily consumed and, of course, traded goods across the three countries and thereby check for nominal price differences before and after the establishment of the EAC with regards to border distance, and as usual, core agglomerations. The results are presented in Appendix Table C.19. As is seen, there are almost no differential effects of price changes across border distance. Only the price for Millet (a cereal grain), decreased slightly more at survey locales closer to borders. Finally, we see the now common effect for core agglomerations, which show statistically significant price

---

values range from ‘Agree very strongly with Statement 1’ (1), ‘Agree with Statement 1’ (2), ‘Agree with Statement 2’ (3), ‘Agree very strongly with Statement 2’ (4), ‘Agree with Neither’ (5) and ‘Don’t know’ (7). I recode 5 to represent the median value. I remove the observations valued ‘Don’t know’.
The survey question on the variable in column (4) and (5) reads: In your opinion, how much do each of the following do to help your country, or haven’t you heard enough to say? [EAC/African Union]. The response values range from ‘Don’t help’ (0), ‘Help a little’ (1), ‘Help somewhat’ (2), ‘Help a lot’ (3) and ‘Don’t know’ (7). I remove the observations valued ‘Don’t know’.

[130] The survey question reads: “For each of the following types of people, please tell me whether you would like having people from this group as neighbors, dislike it, or not care: Immigrants or foreign workers.” ‘Strongly dislike’ (1), ‘Somewhat dislike’ (2), ‘Would not care’ (3) and ‘Somewhat like’ (4), ‘Strongly like’ (5) and ‘Don’t know’ (9). I remove the observations valued ‘Don’t know’.

[131] There are weak findings for an increased likelihood of any type of activity besides housework.

decreases for all items tested. The relative size of the effects is in the range of 2-8%. Why there is no positive price change at borders is up for debate. What can be said from previous literature (see above) is that integration of markets across (East) Africa, is that borders remain a hinderance.

Reconciliation with Theory. As a final note on the empirical results observed, I attempt to reconcile them with the (contrary) predictions of the theoretical simulations in Section 4.4. Importantly, while I provide some deliberations on the possible (parameter) configurations that may bring about such results, they remain speculative. The reduced-form empirics employed in this chapter are not apt to inform on specific parameter coefficients which would permit isolating the key factor driving these observed results. Nonetheless, keeping in mind the theoretical predictions, the results shown in this section hint towards three broad possibilities.[132] The first trivial one is that the cross-border activity triggered by regional market integration of the EAC did not induce a large enough (market access) shock to render proximity to new markets as relevant enough to break a general agglomeration tendency towards interior hubs. Second, the interior economic hubs display too strong of an attraction that even in the event of a fully integrated market, agglomeration in Nairobi, Dar Es Salaam and Kampala is sustained. This latter scenario can be intuitively displayed by referring to Panel D in Figure 4.3, where a large enough share of either previous manufacturing distribution or population inflows renders a fully agglomerated equilibrium of interiors possible even for the scenario of full trade liberalization, i.e. a $T_F = T_D = 1.60$. Depending on the degree of integration, the necessary share of manufacturing positioned in hubs prior to liberalization setting in motion such an equilibrium has a minimum of 50% and a maximum of 73% in the full free trade scenario. Importantly, this is the case only for the foreign market which is equally balanced. As is established in Panel D of Figure 4.5, this necessary labor share is decreasing in foreign spatial inequality, where 78% depicts the scenario when all foreign activity is situated at the border and less than 68% when the foreign economy is fully agglomerated in the interior.[133] As such, given the significant polarization of all member countries before the establishment of the EAC (see Section 4.3) one can imagine a scenario in which liberalization does not suffice to break agglomeration, because countries open up to similarly spatially unequal foreign markets.

And third, the dynamics indicated by the dwindling differences of AFB and DHS consumption and agglomeration measures over time displayed throughout the results, particularly in Table 4.2, Table C.4, Table C.16 and Table C.21 may also point towards a transitory shift to a new welfare equalizing equilibrium. The mechanics can be displayed in Panel B of Figure 4.3 and are as follows. With a starting point of equal (manufacturing) labor distribution, trade liberalization sets in, which lowers the real wage differential of core

[132] Of course, only if the assumptions of the difference-in-differences employed hold. These assumptions have been tested in Subsection 4.7.2.
[133] For results in all three trade scenarios, see Figure C.4.

agglomerations. Labor immediately responds with outflow as seen in column (3) of Table 4.2. However, these outflows are too large such that they quickly offshoot the long-run stable equilibria between 40% and 50% and render real wages in interior hubs as exceeding those at borders. The subsequent response is a migration inflow back into core agglomerations until welfare equalization at a new (lower) equilibrium share of labor is reached. Three factors go against this possibility, however: First, an equal or less than equal spatial distribution of labor and welfare in favor of region 2 was not a likely initial position of all three countries before the EAC. Second, population inflows into border regions are not observed in Table 4.2. And third, most importantly, while differential changes in welfare and population decrease across time, they do remain positive in all periods post-liberalization, which is only possible in a move towards a fully agglomerated equilibrium. One last possibility, of course, is that there is no practical spatial heterogeneity, i.e. border regions do not have better access to the new markets.[134] In such a scenario, depicted in Panel A and C in Figure 4.3, the result is fully robust with trade setting in motion full agglomeration, as depicted by the rotating lines in all parameter configurations. Agglomeration is more likely in the regions which hosted the majority of workers in pre-EAC years. Note that there is yet another scenario which may generate the observed results. That is, if market integration among the three partner countries increased trade diversion, i.e. trade with the rest of the world, instead of intra-regional trade, one may observe clustering in the interior hubs, which also host the significant links to these international markets (Marchand 2012). The increasing agglomeration after the customs union protocol may be a hint at this. However, this is only vague evidence, and needs to be thoroughly investigated against the specific external tariff structure of the EAC. Exploring this is beyond the scope of the present paper.

4.8 Conclusion

This paper investigates the impact of the re-establishment of the East African Community (EAC) on household welfare using three distinct sets of longitudinal, geo-referenced household-level surveys from the three founding members Kenya, Tanzania and Uganda. I formally derive the potential impact of the EAC on households from a canonical New Economic Geography (NEG) model with heterogeneous intra-national space and test the predictions through a reduced-form difference-in-differences specification with treatment intensity given by households' road distance to internal EAC border crossings. I therefore treat the re-establishment of the EAC in 2001 – and the expansion to a customs union and common market in 2005 and 2010, respectively – as a regional policy intervention having differential effects on individual households governed by their geo-spatial location within the countries.

The results show that households and individuals living closer to the internal EAC border did not experience a relative increase in welfare following the re-establishment, as measured by

[134] One may think of shifting transport types i.e. using airports or harbors rather than roads.

an array of consumption indices as well as intensive and extensive employment outcomes. Rather, the results hint at the strengthened concentration of activity in the pre-existing, interior economic hubs as evinced by the strong differential and economically relevant increases in household welfare across the measured dimensions as well as subsequent inflows of population. As such, these findings go against the general prediction of the theoretical simulations and also against the hypothesis prominently outlined in Krugman and Elizondo (1996), which predicted a dispersion of the concentrated economic activity of developing countries following liberalization. My results are also in contrast to other recent empirical findings, which have regularly documented regions closer to the new market (potential) to profit from the less costly access to them (for an overview see Brülhart 2011). Most notably, the present results are at odds with Eberhard-Ruiz and Moradi (2019) who provide (remote sensing) evidence that the EAC's re-establishment led to a positive (one-off) effect on city growth for cities closer to the internal EAC borders.

While in disagreement with the general prediction of the model and these previous contributions, the model allows a potential insight into the reasons why such results may come about. The theoretical results show that the spatial configuration of the foreign economy has a non-negligible moderating force on the strength of dispersion induced by trade liberalization in the domestic country. As such, the presence of economically dominating interiors in all partner countries potentially weakens the dispersion force of market integration to the point where a regional pattern outlined by stark inequalities as existent prior to the EAC remains a possible outcome even after regional trade liberalization. Such explanations remain contentious, however, as the reduced-form empirics are ultimately not apt to inform on specific parameter coefficients, which would permit isolating key factors driving these observed effects.

One last, potentially policy-relevant contribution of the present paper is the connection to the widely discussed issue of "urbanization without growth" (Fay and Opal 2000). Given that the results show for increased consumption and agglomeration tendency in primate cities, without a concomitant expansion of labor market opportunities, they provide evidence in favor of developing countries' metropoles growing to "consumption-" rather than "production cities" (Gollin et al. 2016), and that this process may be influenced by regional market integration.

4.9 References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. ‘The Colonial Origins of Comparative Development: An Empirical Investigation’. *American Economic Review* 91 (5): 1369–1401.

Afrobarometer, [Kenya, Tanzania, Uganda] [Round 1, 2, 3, 4, 5, 6, 7]. 2019. ‘Afrobarometer Data’.

AidData. 2017. ‘Geocoding Methodology, Version 2.0.1.’ Williamsburg, VA: AidData at William & Mary.: AidData Research and Evaluation Unit.

Aker, Jenny C., Michael W. Klein, Stephen A. O’Connell, and Muzhe Yang. 2014. ‘Borders, Ethnicity and Trade’. *Journal of Development Economics* 107 (March): 1–16.

Andreoni, Antonio, Julian Boys, and Ole Therkildsen. 2022. ‘The Political Economy of “Specialism” in Tanzania: How to Make Export Processing Zones Work via Conditional Special Licensing’. In . Anti-Corruption Evidence Working Paper 046. London, UK.: SOAS University of London. Anti-Corruption Evidence (ACE) Research Consortium.

Atkin, David, and Dave Donaldson. 2015. ‘Who’s Getting Globalized? The Size and Implications of Intra-National Trade Costs’. w21439. Cambridge, MA: National Bureau of Economic Research.

Barros, Laura, and Inmaculada Martínez-Zarzoso. 2022. ‘Systematic Literature Review on Trade Liberalization and Sustainable Development’. *Sustainable Production and Consumption* 33 (September): 921–31.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute.

Beegle, Kathleen, Joachim De Weerdt, and Stefan Dercon. 2006. ‘Kagera Health and Development Survey 2004 Basic Information Document’. The World Bank, Washington, DC.

Behrens, Kristian, Carl Gaigné, Gianmarco I P Ottaviano, and Jacques-François Thisse. 2003. ‘Inter-Regional and International Trade: Seventy Years After Ohlin’. In *CEPR Discussion Paper No. 4065*. Paris & London: CEPR Press.

Behrens, Kristian, Carl Gaigné, Gianmarco I.P. Ottaviano, and Jacques-François Thisse. 2006. ‘Is Remoteness a Locational Disadvantage?’ *Journal of Economic Geography* 6 (3): 347–68.

———. 2007. ‘Countries, Regions and Trade: On the Welfare Impacts of Economic Integration’. *European Economic Review* 51 (5): 1277–1301.

Bickenbach, Frank, Eckhardt Bode, Peter Nunnenkamp, and Mareike Söder. 2016. ‘Night Lights and Regional GDP’. *Review of World Economics* 152 (2): 425–47.

Blanchard, O., and Lawrence F. Katz. 1992. ‘Regional Evolutions’. *Brooking Papers on Economic Activity* 23 (1): 1–76.

Bluhm, Richard, and Gordon C. McCord. 2022. ‘What Can We Learn from Nighttime Lights for Small Geographies? Measurement Errors and Heterogeneous Elasticities’. *Remote Sensing* 14 (5): 1190.

Borraz, Fernando, Daniel Ferrés, and Máximo Rossi. 2013. ‘Assessment of the Distributive Impact of National Trade Reforms in Brazil’. *The Journal of Economic Inequality* 11 (2): 215–35.

Brakman, Steven, Harry Garretsen, and Charles van Marrewijk. 2019. ‘New Economic Geography After Thirty Years’. In *Handbook of Regional Science*, edited by Manfred M. Fischer and Peter Nijkamp, 1–23. Berlin, Heidelberg: Springer Berlin Heidelberg.

Brakman, Steven, Harry Garretsen, and Charles van Marrewijk. 2020. *An Introduction to Geographical and Urban Economics: A Spiky World.* Third edition. Cambridge, United Kingdom: Cambridge University Press.

Breinlich, Holger, Gianmarco I.P. Ottaviano, and Jonathan R.W. Temple. 2014. 'Regional Growth and Regional Decline'. In *Handbook of Economic Growth*, 2:683–779. Amsterdam.

Brenton, Paul, Alberto Portugal-Perez, and Julie Régolo. 2014. 'Food Prices, Road Infrastructure, and Market Integration in Central and Eastern Africa'. *World Bank Policy Research Working Paper 7003*, Policy Research Working Papers, , August.

Brülhart, Marius. 2011. 'The Spatial Effects of Trade Openness: A Survey'. *Review of World Economics* 147 (1): 59–83.

Brülhart, Marius, Olivier Cadot, and Alexander Himbert. 2017. 'Trade Integration and Spatially Balanced Development: Implications for Uganda and Rwanda'. International Growth Centre Working Paper, no. F-43406-CCP-1.

———. 2019. 'Let There Be Light: Trade and the Development of Border Regions'. FERDI Working Paper, No. P251.

Brülhart, Marius, Céline Carrère, and Federico Trionfetti. 2012. 'How Wages and Employment Adjust to Trade Liberalization: Quasi-Experimental Evidence from Austria'. *Journal of International Economics* 86 (1): 68–81.

Brülhart, Marius, Matthieu Crozet, and Pamina Koenig. 2004. 'Enlargement and the EU Periphery: The Impact of Changing Market Potential'. *The World Economy* 27 (6): 853–75.

Cadot, Olivier, Alexander Himbert, and Marie-Agnès Jouanjean. 2015. 'Trade Facilitation and Concentration: Evidence from Sub-Saharan Africa'. *Shaping Policy for Development - ODI Research Reports*.

Cali, Massimiliano 2014. 'Trade Boom and Wage Inequality: Evidence from Ugandan Districts'. *Journal of Economic Geography* 14 (6): 1141–74.

Castilho, Marta, Marta Menéndez, and Aude Sztulman. 2012. 'Trade Liberalization, Inequality, and Poverty in Brazilian States'. *World Development* 40 (4): 821–35.

CIESIN. 2017. 'Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11'. Center For International Earth Science Information Network-CIESIN-Columbia University. Socioeconomic Data and Applications Center (SEDAC).

Conley, T.G. 1999. 'GMM Estimation with Cross Sectional Dependence'. *Journal of Econometrics* 92 (1): 1–45.

Conley, Timothy G. 2010. 'Spatial Econometrics'. In *Microeconometrics*, edited by Steven N. Durlauf and Lawrence E. Blume, 303–13. London: Palgrave Macmillan UK.

Crozet, Matthieu, and Pamina Koenig. 2004a. 'Trade Liberalization and the Internal Geography of Countries'. In *Multinational Firms' Location and the New Economic Geography*, by Jean-Louis Mucchielli and Thierry Mayer, 3311. Edward Elgar Publishing.

———. 2004b. 'EU Enlargement and the Internal Geography of Countries'. *Journal of Comparative Economics* 32 (2): 265–79.

De Weerdt, Joachim, Kathleen Beegle, Helene B. Lilleor, Stefan Dercon, Kalle Hirvonen, Martina Kirchberger, and Sofya Krutilov. 2010. 'Kagera Health and Development Survey 2010: Basic Information Document. Study Paper 46'. Copenhagen: Rockwool Foundation Working.

Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Baltimore, MD: Published for the World Bank [by] Johns Hopkins University Press.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. 'Monopolistic Competition and Optimum Product Diversity'. *The American Economic Review* 67 (3): 297–308.

EAC. 1999. 'Treaty for the Establishment of the East African Community'. *African Yearbook of International Law Online / Annuaire Africain de Droit International Online* 7 (1): 419–509.

———. 2015. 'Report of the Committee on Communication, Trade and Investments for the on-Spot Assessment on the EAC One Stop Border Posts (OSBPs) in EAC Partner States'. Arusha, Tanzania: East African Legislative Assembly.

Eberhard-Ruiz, Andreas, and Alexander Moradi. 2019. 'Regional Market Integration in East Africa: Local but No Regional Effects?' *Journal of Development Economics* 140 (September): 255–68.

Edmonds, Eric V., Nina Pavcnik, and Petia Topalova. 2010. 'Trade Adjustment and Human Capital Investments: Evidence from Indian Tariff Reform'. *American Economic Journal: Applied Economics* 2 (4): 42–75.

Emran, M. Shahe, and Zhaoyang Hou. 2013. 'Access to Markets and Rural Poverty: Evidence from Household Consumption in China'. *Review of Economics and Statistics* 95 (2): 682–97.

Engel, Jakob, Deeksha Kokas, Gladys López Acevedo, and Maryla Maliszewska. 2021. The Distributional Impacts of Trade: Empirical Innovations, Analytical Tools, and Policy Responses. Trade and Development. Washington, DC: World Bank.

Erten, Bilge, Jessica Leight, and Fiona Tregenna. 2019. 'Trade Liberalization and Local Labor Market Adjustment in South Africa'. *Journal of International Economics* 118 (May): 448–67.

European Commission. 2019. *Description of the GHS Urban Centre Database 2015: Public Release 2019 : Version 1.0.* LU: European Commission Joint Research Centre. Publications Office.

Ezcurra, Roberto, and Alba Del Villar. 2021. 'Globalization and Spatial Inequality: Does Economic Integration Affect Regional Disparities?' *The Annals of Regional Science* 67 (2): 335–58.

Ezcurra, Roberto, and Andrés Rodríguez-Pose. 2014. 'Trade Openness and Spatial Inequality in Emerging Countries'. *Spatial Economic Analysis* 9 (2): 162–82.

Farr, Tom G., Paul A. Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, et al. 2007. 'The Shuttle Radar Topography Mission'. *Reviews of Geophysics* 45 (2): RG2004.

Fay, Marianne, and Charlotte Opal. 2000. *Urbanization without Growth: A Not so Uncommon Phenomenon*. Vol. 2412. World Bank Publications.

Feyrer, James. 2019. 'Trade and Income—Exploiting Time Series in Geography'. *American Economic Journal: Applied Economics* 11 (4): 1–35.

Fick, Stephen E., and Robert J. Hijmans. 2017. 'WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas'. *International Journal of Climatology* 37 (12): 4302–15.

Frankel, Jeffrey A., and David Romer. 1999. 'Does Trade Cause Growth?' *American Economic Review* 89 (3): 379–99.

Fujita, Masahisa, Paul R. Krugman, and Anthony J. Venables. 2001. *The Spatial Economy: Cities, Regions, and International Trade*. 1. paperback ed. Cambridge, Mass.: MIT Press.

Fujita, Masahisa, and Tomoya Mori. 2005. 'Frontiers of the New Economic Geography*'. *Papers in Regional Science* 84 (3): 377–405.

Fukase, Emiko. 2013. 'Export Liberalization, Job Creation, and the Skill Premium: Evidence from the US–Vietnam Bilateral Trade Agreement (BTA)'. *World Development* 41 (January): 317–37.

GADM. 2020. 'Global Administrative Areas [Digital Geospatial Data]'. Center for Spatial Sciences at the University of California, Davis.

Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger. 1999. 'Geography and Economic Development'. *International Regional Science Review* 22 (2): 179–232.

Gibson, John, Susan Olivia, and Geua Boe-Gibson. 2020. 'NIGHT LIGHTS IN ECONOMICS: SOURCES AND USES [1]'. *Journal of Economic Surveys* 34 (5): 955–80.

Gibson, John, Susan Olivia, Geua Boe-Gibson, and Chao Li. 2021. 'Which Night Lights Data Should We Use in Economics, and Where?' *Journal of Development Economics* 149 (March): 102602.

Giovannetti, Giorgia, Enrico Marvasi, and Arianna Vivoli. 2021. 'The Asymmetric Effects of 20 Years of Tariff Reforms on Egyptian Workers'. *Economia Politica* 38 (1): 89–130.

Giovannetti, Giorgia, Marco Sanfilippo, and Arianna Vivoli. 2022. 'Import Tariff Liberalization, Employment, and Gender in Ethiopia'. *The Journal of Development Studies* 58 (11): 2366–92.

Glaeser, Edward L. 2014. 'A World of Cities: The Causes and Consequences of Urbanization in Poorer Countries'. *Journal of the European Economic Association* 12 (5): 1154–99.

Goldberg, Pinelopi Koujianou, and Nina Pavcnik. 2007. 'Distributional Effects of Globalization in Developing Countries'. *Journal of Economic Literature*, 75.

Gollin, Douglas, Remi Jedwab, and Dietrich Vollrath. 2016. 'Urbanization with and without Industrialization'. *Journal of Economic Growth* 21 (1): 35–70.

Hanson, Gordon. 1994. 'Localization Economies, Vertical Organization and Trade'. *American Economic Review* 86 (5).

Hanson, Gordon. 1997. 'Increasing Returns, Trade and the Regional Structure of Wages'. *The Economic Journal* 107 (440): 113–33.

Hazlewood, Arthur. 1979. 'THE END OF THE EAST AFRICAN COMMUNITY: WHAT ARE THE LESSONS FOR REGIONAL INTEGRATION SCHEMES ?' *JCMS: Journal of Common Market Studies* 18 (1): 40–58.

Helpman, Elhanan. 1998. 'The Size of Regions'. *Topics in Public Economics: Theoretical and Applied Analysis*, 33–54.

Henderson, J. Vernon. 1974. 'The Sizes and Types of Cities'. The American Economic Review 64 (4): 640–56.

———. 1982. 'Systems of Cities in Closed and Open Economies'. Regional Science and Urban Economics 12 (3): 325–50.

Henderson, Vernon J., Tim Squires, Adam Storeygard, and David Weil. 2018. 'The Global Distribution of Economic Activity: Nature, History, and the Role of Trade'. *The Quarterly Journal of Economics* 133 (1): 357–406.

Hirte, Georg, Christian Lessmann, and André Seidel. 2020. 'International Trade, Geographic Heterogeneity and Interregional Inequality'. *European Economic Review* 127 (August): 103427.

IMF. 2023. 'Direction of Trade Statistics (DOTS)'. 2023.

Jedwab, Remi, and Dietrich Vollrath. 2015. 'Urbanization without Growth in Historical Perspective'. *Explorations in Economic History* 58 (October): 1–21.

Kaahwa, Wilbert T. K. 2003. *East African Community: EAC Treaty and Challenges to the Community.* Occasional Papers (East African Community Secretariat), no. 3. Arusha, Tanzania: East African Community Secretariat in conjunction with the German Agency for Technical Co-operation.

Kanbur, Ravi, and Xiaobo Zhang. 2005. 'Fifty Years of Regional Inequality in China: A Journey Through Central Planning, Reform, and Openness: FIFTY YEARS OF REGIONAL INEQUALITY IN CHINA'. *Review of Development Economics* 9 (1): 87–106.

Kenya Central Bureau of Statistics. 2001. ‘Statistical Abstract: 2001’. Nairobi, Kenya.: Kenya Central Bureau of Statistics, Ministry of Finance and Planning.

———. 2003. ‘Statistical Abstract: 2003’. Nairobi, Kenya.: Kenya Central Bureau of Statistics, Ministry of Finance and Planning.

Kenya National Bureau of Statistics. 2011. ‘Statistical Abstract: 2011’. Nairobi, Kenya.: Kenya National Bureau of Statistics.

———. 2013. ‘Basic Report on The 2010 Census of Industrial Production (CIP)’. Nairobi, Kenya.: Kenya National Bureau of Statistics, Ministry of Industrialization.

Kinyondo, Abel, Carol Newman, Finn Tarp, and UNU-WIDER. 2016. *The Role and Effectiveness of Special Economic Zones in Tanzania*. 122nd ed. Vol. 2016. WIDER Working Paper. UNU-WIDER.

Kis-Katos, Krisztina, and Robert Sparrow. 2015. ‘Poverty, Labor Markets and Trade Liberalization in Indonesia’. *Journal of Development Economics* 117 (November): 94–106.

Kremer, Michael. 1993. ‘The O-Ring Theory of Economic Development’. *The Quarterly Journal of Economics* 108 (3): 551–75.

Krugman, Paul, and Anthony. J. Venables. 1995. ‘Globalization and the Inequality of Nations’. *The Quarterly Journal of Economics* 110 (4): 857–80.

Krugman, Paul. 1980. ‘Scale Economies, Product Differentiation, and the Pattern of Trade’. *The American Economic Review* 70 (5): 950–59.

———. 1991. ‘Increasing Returns and Economic Geography’. *Journal of Political Economy* 99 (3): 483–99.

Krugman, Paul, and Raul Livas Elizondo. 1996. ‘Trade Policy and the Third World Metropolis’. *Journal of Development Economics* 49: 137–50.

Krugman, Paul R. 1979. ‘Increasing Returns, Monopolistic Competition, and International Trade’. *Journal of International Economics* 9 (4): 469–79.

Lehner, Bernhard, and Petra Döll. 2004. ‘Development and Validation of a Global Database of Lakes, Reservoirs and Wetlands’. *Journal of Hydrology* 296 (1–4): 1–22.

Mansori, Kashif S. 2003. ‘The Geographic Effects of Trade Liberalization with Increasing Returns in Transportation’. *Journal of Regional Science* 43 (2): 249–68.

Marchand, Beyza Ural. 2012. ‘Tariff Pass-through and the Distributional Effects of Trade Liberalization’. Journal of Development Economics 99 (2): 265–81.

———. 2019. ‘Inequality and Trade Policy: The Pro-Poor Bias of Contemporary Trade Restrictions’. Review of Income and Wealth 65 (S1). Martínez-Zarzoso, Inma, Leila Baghdadi, and Hendrik Wiard Kruse. 2016. ‘Is Tunisian Trade Policy Pro-Poor?’ In *1033*.

Maystadt, Jean-François, and Gilles Duranton. 2019. ‘The Development Push of Refugees: Evidence from Tanzania’. *Journal of Economic Geography* 19 (2): 299–334.

McCaig, Brian. 2011. ‘Exporting out of Poverty: Provincial Poverty in Vietnam and U.S. Market Access’. *Journal of International Economics* 85 (1): 102–13.

McCaig, Brian, and Margaret S. McMillan. 2020. ‘Trade Liberalisation and Labour Market Adjustment in Botswana’. *Journal of African Economies* 29 (3): 236–70.

Monfort, Philippe, and Rosella Nicolini. 2000. ‘Regional Convergence and International Integration’. *Journal of Urban Economics* 48 (2): 286–306.

Monfort, Philippe, and Tanguy van Ypersele. 2003. ‘Integration, Regional Agglomeration and International Trade’. In *CEPR Discussion Paper No. 3752*, 32. Paris & London: CEPR Press.

Mugomba, Agrippah T. 1978. ‘Regional Organisations and African Underdevelopment: The Collapse of the East African Community’. *The Journal of Modern African Studies* 16 (2): 261–72.

Nathan Associates. 2011. ‘Corridor Diagnostic Study of the Northern and Central Corridors of East Africa’. In *Volume 1: Main Report.* Alexandria, VA, USA.

Natural Earth. 2019. ‘Natural Earth. Free Vector and Raster Map Data.’

NGA, Natioal Geospace-Intelligence Agency. 2019. ‘World Port Index 27th Edition’. United States Government.

Nicita, Alessandro. 2009. ‘The Price Effect of Tariff Liberalization: Measuring the Impact on Household Welfare’. *Journal of Development Economics* 89 (1): 19–27.

Nicita, Alessandro, Marcelo Olarreaga, and Guido Porto. 2014. ‘Pro-Poor Trade Policy in Sub-Saharan Africa’. *Journal of International Economics* 92 (2): 252–65.

Nunn, Nathan, and Diego Puga. 2012. ‘Ruggedness: The Blessing of Bad Geography in Africa’. *Review of Economics and Statistics* 94 (1): 20–36.

OSM. 2022. *OpenStreetMap Database*. OpenStreetMap contributors. Cambridge, United Kingdom: OpenStreetMap Foundation.

Ottaviano, Gianmarco, Takatoshi Tabuchi, and Jacques-Francois Thisse. 2002. ‘Agglomeration and Trade Revisited’. *International Economic Review* 43 (2): 409–35.

Paluzie, Elisenda. 2001. ‘Trade Policy and Regional Inequalities*’. *Papers in Regional Science* 80 (1): 67–85.

Pavcnik, Nina. 2017. ‘The Impact of Trade on Inequality in Developing Countries’. w23878. Cambridge, MA: National Bureau of Economic Research.

Porto, Guido G. 2006. ‘Using Survey Data to Assess the Distributional Effects of Trade Policy’. *Journal of International Economics* 70 (1): 140–60.

Puga, Diego. 1999. ‘The Rise and Fall of Regional Inequalities’. *European Economic Review*, 32.

Ramankutty, Navin, Jonathan A. Foley, John Norman, and Kevin McSweeney. 2002. ‘The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change: *Cropland Sensitivity to Climate Change*’. *Global Ecology and Biogeography* 11 (5): 377–92.

Rappaport, Jordan, and Jeffrey D. Sachs. 2003. ‘The United States as a Coastal Nation’. *Journal of Economic Growth* 8: 5–46.

Redding, Stephen J. 2022. ‘Trade and Geography’. In *Handbook of International Economics*, 5:147–217. Elsevier.

Redding, Stephen J., and Esteban Rossi-Hansberg. 2017. ‘Quantitative Spatial Economics’. *Annual Review of Economics* 9 (1): 21–58.

Redding, Stephen J, and Daniel M Sturm. 2008. ‘The Costs of Remoteness: Evidence from German Division and Reunification’. *American Economic Review* 98 (5): 1766–97.

Rodrik, Dani. 2002. ‘Institutions, Integration, and Geography: In Search of the Deep Determinants of Economic Growth’. In *In Search of Prosperity: Analytical Country Studies on Growth, Princeton, NJ.* Princeton University Press.

Rutstein, Shea O., and Kiersten Johnson. 2004. ‘The DHS Wealth Index’. In *DHS Comparative Reports No. 6.* Calverton, Maryland, USA: Macro International.

Rutstein, Shea O., and Sarah Staveteig. 2014. ‘Making the Demographic and Health Surveys Wealth Index Comparable’. ICF International Rockville, Maryland, USA.

Sachs, Jeffrey D. 2003. ‘Institutions Don’t Rule: Direct Effects of Geography on Per Capita Income’. w9490. Cambridge, MA: National Bureau of Economic Research.

Sachs, Sonia Ehrlich, Andrew Mellinger, Pia Malaney, Jeffrey Sachs, Anthony Kiszewski, and Andrew Spielman. 2004. ‘A Global Index Representing the Stability of Malaria Transmission’. The American Journal of Tropical Medicine and Hygiene 70 (5): 486–98.

Samuelson, Paul A. 1952. ‘The Transfer Problem and Transport Costs: The Terms of Trade When Impediments Are Absent’. *The Economic Journal* 62 (246): 278.

Schiff, Maurice W., and L. Alan Winters. 2003. *Regional Integration and Development*. Washington, DC : [s.l.]: World Bank ; Oxford University Press.

Smits, Jeroen, and Roel Steendijk. 2015. 'The International Wealth Index (IWI)'. *Social Indicators Research* 122 (1): 65–85.

Tanzania National Bureau of Statistics. 2004. 'Employment and Earnings Survey: 2001 Analytical Report'. Dar es Salaam, Tanzania: National Bureau of Statistics, Ministry of Planning, Economy and Empowerment. The United Republic of Tanzania.

———. 2006. 'Population Census: 2002 Analytical Report'. Dar es Salaam, Tanzania: National Bureau of Statistics, Ministry of Planning, Economy and Empowerment. The United Republic of Tanzania.

———. 2007. 'Employment and Earnings Survey: 2002 Analytical Report'. Dar es Salaam, Tanzania: National Bureau of Statistics, Ministry of Planning, Economy and Empowerment. The United Republic of Tanzania.

———. 2010. 'Annual Survey of Industrial Production and Performance: 2008 Statistical Report'. Dar es Salaam, Tanzania: National Bureau of Statistics, Ministry of Industry, Trade and Marketing. The United Republic of Tanzania.

———. 2018. 'Annual Survey of Industrial Production: 2016 Statistical Report'. Dar es Salaam, Tanzania: National Bureau of Statistics, Ministry of Industry, Trade and Investment. The United Republic of Tanzania.

Thünen, Johann H. von. 1826. *Der Isolierte Staat in Beziehung Auf Landwirtschaft Und Nationalökonomie*. Hamburg: Perthes.

Topalova, Petia. 2010. 'Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India'. *American Economic Journal: Applied Economics* 2 (4): 1–41.

UFZA. 2022. 'Uganda Free Zone Authority Annual Report 2021/2022'. Kampala, Uganda.

Uganda Bureau of Statistics. 2003. 'Uganda Business Register: 2001/2002, Report'. Kampala, Uganda: Uganda Bureau of Statistics, Directorate of Business Industry, Agriculture and Energy Statistics.

———. 2006. 'Uganda Business Inquiry: 2000/2001'. Entebbe, Uganda: Uganda Bureau of Statistics.

———. 2007. 'Uganda Business Register: 2006/2007, Report'. Kampala, Uganda: Uganda Bureau of Statistics, Directorate of Business Industry, Agriculture and Energy Statistics.

———. 2012. 'Statistical Abstract: 2012'. Kampala, Uganda: Uganda Bureau of Statistics.

———. 2016. 'The National Population and Housing Census 2014 – Main Report'. Kampala, Uganda: Uganda Bureau of Statistics.

UNCTAD. 2022. 'Trade Analysis and Information System (TRAINS)'. United Nations Conference on Trade and Development.

United Nations Development Programme. 2011. 'Regional Integration and Human Development: A Pathway for Africa'. New York, NY: United Nations Development Programme.

UNSD. 2022. 'International Merchandise Trade Statistics (COMTRADE)'. United Nations Statistics Division.

UNU-CRIS. 2019. 'Regional Integration Knowledge System (RIKS)'.

Versailles, Bruno. 2012. 'Market Integration and Border Effects in Eastern Africa', 48.

Vo, Thang T., and Dinh X. Nguyen. 2020. 'Impact of Trade Liberalization on Household Welfare: An Analysis Using Household Exposure-to-Trade Indices'. *Social Indicators Research*, September.

Wild, Frederik, and David Stadelmann. 2022. ‘Coastal Proximity and Individual Living Standards: Econometric Evidence from Georeferenced Household Surveys in sub-Saharan Africa’. *Review of Development Economics* 26 (4): 1883–1901.

Winters, L. Alan. 2002. ‘Trade Liberalisation and Poverty: What Are the Links?’, 29.

Winters, L. Alan, and Antonio Martuscelli. 2014. ‘Trade Liberalization and Poverty: What Have We Learned in a Decade?’ *Annual Review of Resource Economics* 6 (1): 493–512.

Winters, L. Alan, Neil Mcculloch, and Andrew Mckay. 2004. ‘Trade Liberalization and Poverty: The Evidence So Far’. *Journal of Economic Literature*, 64.

World Bank. 2009. *World Development Report 2009: Reshaping Economic Geography.* 2009. Washington, D.C.: World Bank.

———. 2016. ‘Implementation Completion and Results Report (IDA-39300 IDA-45710) for the Northern Corridor Transport Improvement Project’. Report ICR00002771. Washington, D.C: The World Bank.

———. 2020. *The African Continental Free Trade Area: Economic and Distributional Effects.* Washington, DC: World Bank.

———. 2022. ‘Data.Worldbank.Org’. 2022.

World Bank, and University of Dar es Salaam. 1994. *Kagera Health and Development Survey 1991-1994 (Wave 1 to 4 Panel). Ref. TZA_1991_KHDS_v01_M. Dataset Downloaded from the World Bank Microdata Library (Www.Microdata.Worldbank.Org).* World Bank and University of Dar es Salaam.

———. 2004. *Kagera Health and Development Survey 2004. Ref. TZA_2004_KHDS_v01_M.* World Bank and University of Dar es Salaam.

———. 2010. *Kagera Health and Development Survey 2010.* World Bank and University of Dar es Salaam.

WTO. 2019. ‘East African Community (EAC) Trade Policy Review: Report by the Secreteriat’. WT/TPR/S/384. Geneva, Switzerland: World Trade Organization.

Zeng, Dao-Zhi, and Laixun Zhao. 2010. ‘Globalization, Interregional and International Inequalities’. *Journal of Urban Economics* 67 (3): 352–61.

Appendix C

Regional Market Integration and Household Welfare: Spatial Evidence from the East African Community

C.1 Analytical Insights to the Model

As anticipated in Subsection 4.4.5, this section provides analytical insights into the simulation results of the main text. As in the NEG tradition, the analysis revolves around checking the stability of two specific equilibria of the model, namely the "spreading" equilibrium, i.e. where the real wage differential $\omega_1/\omega_2 = 1$ and $\lambda_1 = \lambda_2 = 0.5$, and the "agglomerated" equilibrium, where all manufacturing is concentrated in one of the regions such that $\lambda_1 = 1$ and $\omega_1/\omega_2 \geq 1$ or $\lambda_1 = 0$ and $\omega_1/\omega_2 \leq 1$. The analytical evaluation is thereby concerned with assessing the stability of both equilibria to varying internal transport costs, i.e. at which level of T_D spreading is broken and agglomeration sustainable, hence the name "sustain" and "break" analysis. As such, the treatment closely follows the exposition in Chapter 5 of Fujita et al. (2001), albeit for a modified spatial layout and focusing on the influence of external trade costs rather than internal ones only. Note that the analysis is constrained to these two cases because at these points, the system of non-linear equations reduces to a more tractable set which simplifies the analysis. However, given that λ varies all the way from 0 to 1, and choices have to be made on the other parameter values, (the stability of) further equilibria may depend on many such combinations, the main analysis of this paper relies on the numerical simulations of Subsection 4.4.5 in order to give a full picture of the long-run dynamics. As first introduced in Subsection 4.4.5, we compare the analytical results for the setting with *heterogeneous* (asymmetric) intra-national space to the ones drawn from a $2+2$ setting with *homogenous* (symmetric) intra-national space. Note, that as mentioned in Subsection 4.4.5, the analysis of the latter set-up mirrors the one fist shown in (Monfort and Nicolini 2000).

C.1.2 Symmetry Breaking

We start by analyzing the robustness of a symmetric equilibrium, that is, the configuration in which $\lambda_1 = \lambda_2 = 0.5$ and $\omega_1/\omega_2 = 1$. From the discussion in Subsection 4.4.4, and visually depictable in Figure 4.3, we know that this equilibrium is stable if migrating in either direction leads to a lower real wage in the destination region than in the origin. Stated more generally, for the symmetric equilibrium to be a stable one, the slope of the total differential with respect to $d\lambda_{\mathrm{i}}$ has to satisfy:

$$\frac{d\frac{\omega_i}{\omega_j}}{d\lambda_i} \leq 0. \tag{C.1}$$

Before we start deriving an expression for (C.1), notice that Figure 4.3, specifically, the differences across Panels A and B as well as C and D, already hold the insight insofar as a symmetric equilibrium can be upheld during trade liberalization within our heterogeneous $2+2$ setting. The simulations show that any move away from autarky ($\tau \neq \infty$) also entails a move away from the symmetric distribution of manufacturing as an equilibrium. This is observable by the shift of the cut point to the left (Panel B) and to the right (Panel D). Hence, contrary to Panels A and C, the relative share of the manufacturing workforce across regions in the first type of equilibrium is dependent on external transport costs. As such, for the "symmetry breaking" analysis, we are limited to the $2+2$ setting with homogenous intra-national space. In this setup, the equal distribution is always a possible equilibrium, independent of the (external) transport costs. This is explicated in the following steps. First note that when $\lambda^{H(ome),F(oreign)} = 0.5$, income $Y^{H,F} = 0.5$ and from this, the wage reduces to $w^{H,F} = 1$, and this is true for all (home and foreign) regions, hence the drop of the indices.[135] We can confirm this by plugging these values into (4.36) through (4.44), and solving.

$$Y = \frac{\delta}{2} + \frac{(1-\delta)}{2} = 0.5 \tag{C.2}$$

$$I = [0.5 + 0.5T_D^{1-\varepsilon} + T_F^{1-\varepsilon}]^{\frac{1}{1-\varepsilon}} \tag{C.3}$$

Plugging these two results into the wage equation leads to

$$w = (0.5I^{\varepsilon-1} + 0.5I^{\varepsilon-1}T_D{}^{1-\varepsilon} + I^{\varepsilon-1}T_F{}^{1-\varepsilon})^{\frac{1}{\varepsilon}},\ or$$

$$w = \left(\frac{0.5 + 0.5T_D{}^{1-\varepsilon} + T_F{}^{1-\varepsilon}}{I^{1-\varepsilon}}\right)^{\frac{1}{\varepsilon}} = 1 \tag{C.4}$$

$$given\ that\ I^{1-\varepsilon} = 0.5 + 0.5T_D^{1-\varepsilon} + T_F^{1-\varepsilon}$$

Note that the subscripts can be dropped as income, price indices and therefore wages are equal in all regions at the symmetric equilibrium. Note that the existence of this equilibrium configuration does not depend on foreign trade costs T_F.[136] From this set of manipulations, specifically (C.3) and (C.4), it is easily seen that real wages across regions are equal. With these results in mind, we are able to proceed with a crucial simplification in the derivation of the total differential. Namely, at this symmetric equilibrium, a change in one of the endogenous variables for one region requires the identical change for the other region in the opposite direction (Fujita et al. 2001). This can be confirmed, for instance, by computing the total derivatives of the two income equations for regions 1 and 2, plugging in the equilibrium values, and checking whether $dY_1 + dY_2 = 0$.

---

[135] Note that the superscripts H and F indicate identical values for all home (1 and 2) and foreign regions (3 and 4), respectively.

[136] Accordingly, the existence of a symmetric equilibrium in the homogenous $2+2$ region case is independent of the foreign labor distribution.

$$dY_1 = d\lambda_1 w_1 \delta + dw_1 \lambda_1 \delta \tag{C.5}$$

$$dY_2 = -d\lambda_1 w_2 \delta + dw_2 (1-\lambda_1)\delta \tag{C.6}$$

Where we made use that $\lambda_2 = (1-\lambda_1)$. If we now plug in the equilibrium values derived in (C.2) through (C.4) and assuming analogously that $dw_1 = -dw_2$, gives

$$dY_1 = d\lambda_1 \delta + dw_1 0.5\delta \; and \; dY_2 = -d\lambda_1 \delta - dw_1 0.5\delta \tag{C.7}$$

$$\text{which satisfies } \; dY_1 + dY_2 = 0. \tag{C.8}$$

This confirms that $dY^H \equiv dY_1 = dY_2$. Hence, the total derivate of the income at the symmetric equilibrium can be written as

$$dY^H = \; d\lambda\delta + \frac{\delta}{2} \; dw \tag{C.9}$$

And equally for the foreign country such that $dY \equiv dY^H = dY^F$. This operation can be confirmed for all other equilibrium equations, i.e. for price indices dI and wages dw.[137] Importantly, the same intuition applies to the total differential of the real wage equations also, such that it suffices to assess only the change in the real wage of one of the two foreign or home regions, and (C.1) effectively boils down to $d\omega_i^{H,F}/d\lambda_i^{H,F}$, e.g. given by

$$\frac{d\omega_i/d\omega_j}{d\lambda_i} \; \equiv \; \frac{d\omega}{d\lambda} \; = \; \frac{dw \cdot I^{-\delta} - dI \cdot w \cdot I^{-(1+\delta)} \cdot \delta}{d\lambda} \tag{C.10}$$

After some manipulations, which involves plugging in (C.9) into the equations for $dI^{H,F}$ and $dw^{H,F}$, and solving these four equations as a system, we arrive at expressions to plug into (C.10) which are solely dependent on the exogenous parameter values δ and ε as well the iceberg trade costs T_D and T_F.[138] As this expression hinges on the totals differentials from all four regions, the expressions is unwieldly compared to a 2-region NEG model. Hence, to facilitate interpretation, the results from this “break” analysis are provided graphically in Figure C.1 for a given set of parameter values and the three levels of external transport costs

[137] I do not show them here because they grow relatively large as they are additionally dependent on changes in the endogenous variables of the foreign country.

[138] Also using $I = \; \left[0.5 + 0.5T_D^{(1-\varepsilon)} + T_F^{(1-\varepsilon)}\right]^{\frac{1}{1-\varepsilon}}$.

T_F. Note that we are now also able to assess the stability of this equilibrium to a range of internal transport costs which is not simply set to $T_D = 1.60$ as previously. As in the NEG tradition, Table C.1 additionally provides the "break" values $T_D(B)$ together with the ones derived in b), i.e. the "sustain" values $T_D(S)$, for a range of parameter values δ and ε and the three levels of trade liberalization T_F. Figure C.1 provides the results to the break analysis for the four-region model with homogenous intra-national space, plotting $(d\omega_2/d\omega_1)/d\lambda_2$ across an increasing value of intra-national trade costs T_D for three different values of international trade costs T_F, separately. Note that by symmetry, we are free to choose the direction of effects in any of the two home or foreign regions. We focus on the real wage differential of region 2, however, as it facilitates the comparison with the sustain analysis in b) and given that we have seen the increased draw towards border regions in the simulations discussed in Section 4.4.

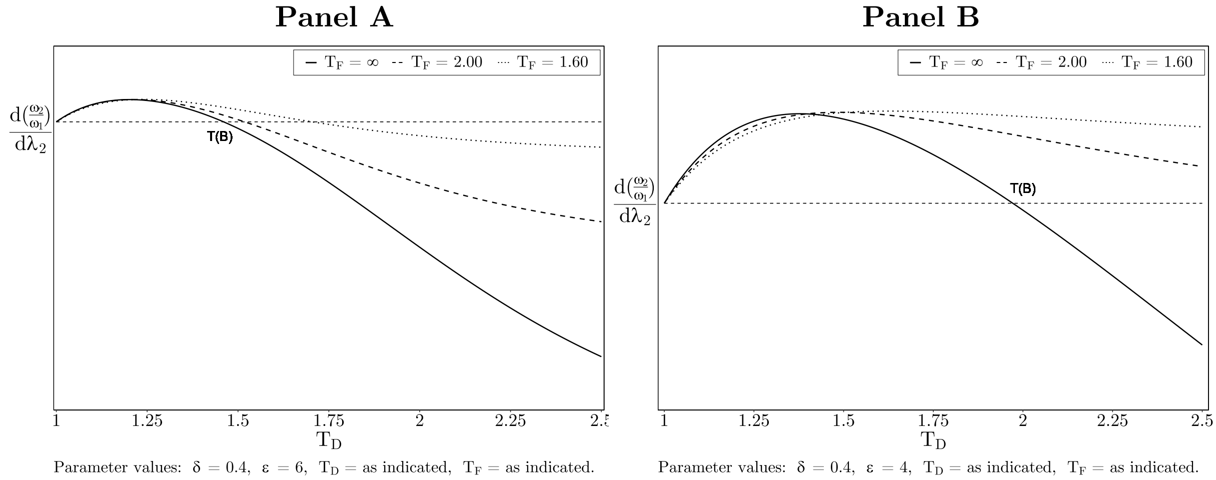


Figure C.1: Internal transport costs and symmetric equilibrium

Note first, that with zero transport costs, i.e. $T_D = 1$, there is no difference in the (real) wage across regions such that the total differential is zero at the origin.[139] Increasing the intra-national transport costs from this point on in Panel A shows that up until a level of transport costs of $T_D = 1.47$, the symmetric equilibrium is unstable, given that a move away from region 1 increases real wages at the destination, i.e. in region 2, i.e. the total differential is positive.[140] In other words, in this scenario, the cost- and demand linkages of agglomerating are strong enough to render the cost of serving the demand of region 1 at a distance as profitable. For any increase in transport costs beyond this point, this is not true anymore and manufacturing activity spreads out. The dashed and dotted curves show the effect of trade liberalization, which is to increase the range of transport costs within which a symmetric equilibrium is

[139] The basic NEG setup hinges on positive transport costs, i.e. a $T_D \neq 1$ which essentially means that economies are not rendered identical in all respects.
[140] You can retrieve the precise value from Table C.1.

unstable. Why is this the case? We need to analyze a couple of (countervailing) effects step by step in order to interpret the likely effects (e.g. Crozet and Koenig 2004a; Brülhart et al. 2004; Brülhart 2011). Note at first that liberalizing trade, i.e. decreasing T_F down from prohibitive levels, causes the components in both the price index (4.40)-(4.43) as well as the wage equation (4.44)-(4.47) that are dependent on external markets 3 and 4 to make up a larger component of the overall I and w at the respective location. This results in several dynamics. First, it lowers producers' need to locate close to consumers in the home country as a larger share of their sales come from abroad, i.e. the demand linkage is lowered. Secondly, it analogously decreases consumer's need to locate near producers in the home country as a larger share of their demand now stems from abroad, i.e. the cost linkage is also lowered (Crozet and Koenig 2004b). As such, lowering T_F essentially reduces agglomerating tendencies inside the domestic country which is the well-known result put forward in Krugman and Elizondo (1996). Note that this also means that the moderating force of these cost and demand linkages as given by the internal transport costs T_D is weakened also, as can be depicted by an attenuation of the slopes in Figure C.1. But why is agglomeration in this present model more likely then? There are two crucial differences to the model in Krugman and Elizondo (1996) which turn this result around. For one, as Crozet and Koenig (2004a) point out, Krugman and Elizondo (1996) do not model an immobile agricultural sector, the demands of which act as a spreading force, and secondly, they explicitly model congestion costs of agglomerations (such as rent or commuting). These congestion costs are independent of trade costs, hence decreasing trade costs does not lower the centrifugal tendency of them. On the contrary, the dispersion force of the type of model I as well as (Monfort and Nicolini 2000) employ, immobile farmers, is crucially dependent on the trade costs to serve them. Together with the key result of the original core NEG model (Krugman 1991), i.e. that the strength of the centrifugal force given by these farmers falls faster in (international) transport costs than the strength of the centripetal force, this may display one reason why the result is turned towards agglomeration in our case (see also the discussion in Brülhart 2011). However, there is one further potential reason why agglomeration tendencies may be increased by opening up to external markets, which is increased competition of firms from abroad (Crozet and Koenig 2004b). Remember from Section 4.4 that the dispersion force stems from increased competition given by the positive relationship between the price index I and the break-even wage rate firms are able to afford to pay, as seen in (4.44) through (4.47). Thereby, in a similar line of argument as above, from the point of producers, decreasing T_F lowers the relative importance of domestic competition, such that sheltering away from local firms is less important given the new competition foreign firms pose (Crozet and Koenig 2004a; Brülhart et al. 2004). By looking at the results in Figure 4.3 above as well as Figure C.1, it seems that this decreased competition effect dominates the decreased agglomeration forces. Additionally, when looking at Panel B in Figure C.1, we see this effect amplified up to a point where the does not even exist a break point anymore, i.e. $T_D = \emptyset$, and agglomeration is the only long-term stable equilibrium in the case of free trade. Note that the difference between Panel A and B makes intuitive sense, given that the only change between the two is the reduced elasticity of substitution, from $\varepsilon = 6$ to $\varepsilon = 4$. This change increases the forward and backward linkages, by reasons given in Subsection 4.4.5, leading to an increased agglomerating force (Fujita et al. 2001).

Notice how these results compare to the simulations in Figure 4.3, in which we have set $T_D = 1.60$. At this point on the x-Axis (Figure C.1), spreading is never sustainable for an $\varepsilon = 4$, as is confirmed by the positive slope in Figure 4.3 Panel B. And for an $\varepsilon = 6$ only stable when trade liberalization has not fully concluded yet, e.g. a value $2.00 \geq T_F > 1.60$ (Figure 4.3 Panel A). Table C.1 encapsulates these results at one glance. We see that a decrease in the elasticity of substitution consistently shifts the break point to the right, i.e. increases the range of values for which spreading is unsustainable. Notice, also that an increase in δ, i.e. an increase in the share of income devoted to manufactures, has the identical effect.

How do these deliberations compare to the model with heterogeneous intra-national space? Panel B in Figure 4.3 shows that for the case with a lower product differentiation ($\varepsilon = 6$), an equilibrium where economic activity is spread out is more likely at higher degrees of trade liberalization than in the homogenous case; albeit with higher shares of economic activity placed at the border. In this case, it seems that the competition effect from abroad does not yet seem to fully dominate the local one and its spreading tendency.[141] Intuitively, firms and consumers now also profit from increased agglomeration, but there is a bias towards agglomerating in the vicinity of the newly accessed markets, i.e. in region 2. This notion is further confirmed seen in Panel D of Figure 4.3, where, as in the case with homogenous intra-national space, decreased product differentiation ($\varepsilon = 4$) causes the curve to reverse, rendering full agglomeration as the only stable, long-run equilibrium, but now this is more likely to happen at the border region 2, compared to the interior region 1.

C.1.3 Sustainable Agglomeration

We now turn to the "sustain" analysis. In 4.4.4, we have already established that the stability of this equilibrium trivially depends on the condition $\omega_i/\omega_j \geq 1$ if $\lambda_i = 1$. As such, we need to derive an expression for the real wage differential at this point which, as in C.1.2, depends only on the parameter values δ, ε as well as, importantly, the different types of iceberg trade costs T_D, T_F, T_{DF}, T_{DFD}. For this analysis, we are able to derive analytical solutions for both spatial layouts, i.e. the homogenous as well as the heterogeneous layout of trade costs. As in C.1.2, the first step entails plugging in the equilibrium values for λ_i, i.e. $\lambda_1 = \lambda_4 = 1$ and correspondingly, $\lambda_2 = \lambda_3 = 0$, and noting that the wage equations of regions 1 and 4 reduce to

141 Although, notice the slope does in fact decline in Figure 4.4 Panel B slightly with progressing trade liberalization.

$w_1 = w_4 = 1$.[142] To see this, note that the income equations (4.36) through (4.39) in this spatial configuration are given by

$$Y_{1,4} = \frac{(1+\delta)}{2} \tag{C.11}$$

$$Y_{2,3} = \frac{(1-\delta)}{2} \tag{C.12}$$

Note, from this set of four equations, income in region 1 is always higher, which represents the demand (backward linkage) introduced in Subsection 4.2.2. Correspondingly, the price indices reduce to

$$I_{1,4} = [1 + T_F^{1-\varepsilon}]^{\frac{1}{(1-\varepsilon)}} \tag{C.13.a}$$

$$I_{2,3} = [T_D^{1-\varepsilon} + T_F^{1-\varepsilon}]^{\frac{1}{(1-\varepsilon)}} \tag{C.14.a}$$

And for the model with heterogeneous intra-national space:

$$I_{1,4} = [1 + T_{DFD}^{1-\varepsilon}]^{\frac{1}{(1-\varepsilon)}} \tag{C.13.b}$$

$$I_{2,3} = [T_D^{1-\varepsilon} + T_{DF}^{1-\varepsilon}]^{\frac{1}{(1-\varepsilon)}} \tag{C.14.b}$$

The first summand of the price index equations is equal for both spatial layouts (C.13.a) and (C.13.b). The difference lies in the second summand in the price indices. In the model with homogenous intra-nation space (C.13.a) and(C.14.a), the cost of living in the peripheral region is at best equal to that of the agglomerated region, but for transport which satisfy $T_D > 1$ are always higher than in the agglomerated region, given that $\varepsilon > 1$ which we assume by default given our CES utility structure. This is the cost (forward) linkage as described in Subsection 4.2.2. However, in (C.13.b) and (C.14.b), given that $T_{DF} < T_{DFD}$, we cannot readily determine whether cost of living is higher or lower in region 1. Moving on, we plug (C.11) through (C.14.a) into the wage equations:

[142] This can be confirmed by guessing $w_1 = 1$, working out (4.36) through (4.43) using this value for W_1 and seeing that (4.44) is indeed 1. Notice that in our heterogeneous 2+2 case, this also entails assuming an equal distribution in the foreign country, i.e. full agglomeration as for instance given by $\lambda_4 = 1$ together with $W_4 = 1$.

$$w_{1,4}=\left[\frac{Y_{1,4}(1+T_F^{1-\varepsilon})}{1+T_F^{1-\varepsilon}}+\frac{Y_{2,3}(T_D^{1-\varepsilon}+T_F^{1-\varepsilon})}{T_D^{1-\varepsilon}+T_F^{1-\varepsilon}}\right]^{\frac{1}{\varepsilon}}, or$$

$$w_{1,4}=\left[\frac{(1+\delta)}{2}+\frac{(1-\delta)}{2}\right]^{\frac{1}{\varepsilon}}=1 \tag{C.15.a}$$

And similarly, for heterogeneous intra-national space

$$w_{1,4}=\left[\frac{Y_{1,4}(1+T_{DFD}^{1-\varepsilon})}{1+T_{DFD}^{1-\varepsilon}}+\frac{Y_{2,3}(T_D^{1-\varepsilon}+T_{DF}^{1-\varepsilon})}{T_D^{1-\varepsilon}+T_{DF}^{1-\varepsilon}}\right]^{\frac{1}{\varepsilon}}, or$$

$$w_{1,4}=\left[\frac{(1+\delta)}{2}+\frac{(1-\delta)}{2}\right]^{\frac{1}{\varepsilon}}=1 \tag{C.15.b}$$

Where we made use of a similar manipulation as in C.1.2, i.e. that $I^{-1-\varepsilon}\equiv[I]^{-1}$. As such, wages in the interior are always 1. Now, real wage equations reduce to

$$\omega_{1,4}=\frac{1}{\left[(1+T_F^{1-\varepsilon})^{\frac{1}{1-e}}\right]^{\delta}} \tag{C.16.a}$$

$$\omega_{1,4}=\frac{1}{\left[(1+T_{DFD}^{1-\varepsilon})^{\frac{1}{1-e}}\right]^{\delta}} \tag{C.16.b}$$

Note that, technically, the (real) wage equations for regions 2 and 3 are only implied functions, as there is no actual manufacturing wage in this spatial configuration given by the absence of manufacturing workers, i.e. $\lambda_2=\lambda_3=0$. One can think of these implied wages as the maximum wage that firms moving to this location would be able to pay (Fujita et al. 2001). The derivation of $\omega_{2,3}$ follow the same type of manipulations just made (C.15.a) through (C.16.b) and lead to expressions only dependent on the parameter values δ, ε and transport costs T_D, T_F, T_{DF}, T_{DFD}.[143] We now have all the ingredients for an expression of the real wage differential within the home or foreign economy, i.e. $\omega_i^{H,F}/\omega_j^{H,F}$. As in C.1.2, to assess the analytical results, we plot the real wage differential in the home economy against the intra-national transport costs T_D and for our three levels of external transport costs T_F. This is done in Figure C.2. Importantly, note the change in the y-axis; we now express the real wage differential from the point of view of the peripheral region, i.e. plot ω_2/ω_1 to facilitate a comparison to the break

[143] Results are not reported here.

analysis in C.1.2. More precisely, the range of transport costs T_D for which this type of equilibrium is sustainable also lies below the constant. Again, we analyze the dynamics from the point of view of region 2, i.e. assess when full agglomeration in region 1 is unsustainable. In contrast to C.1.2 we are now also able to discuss peculiarities of the model with heterogeneous intra-national space analytically.

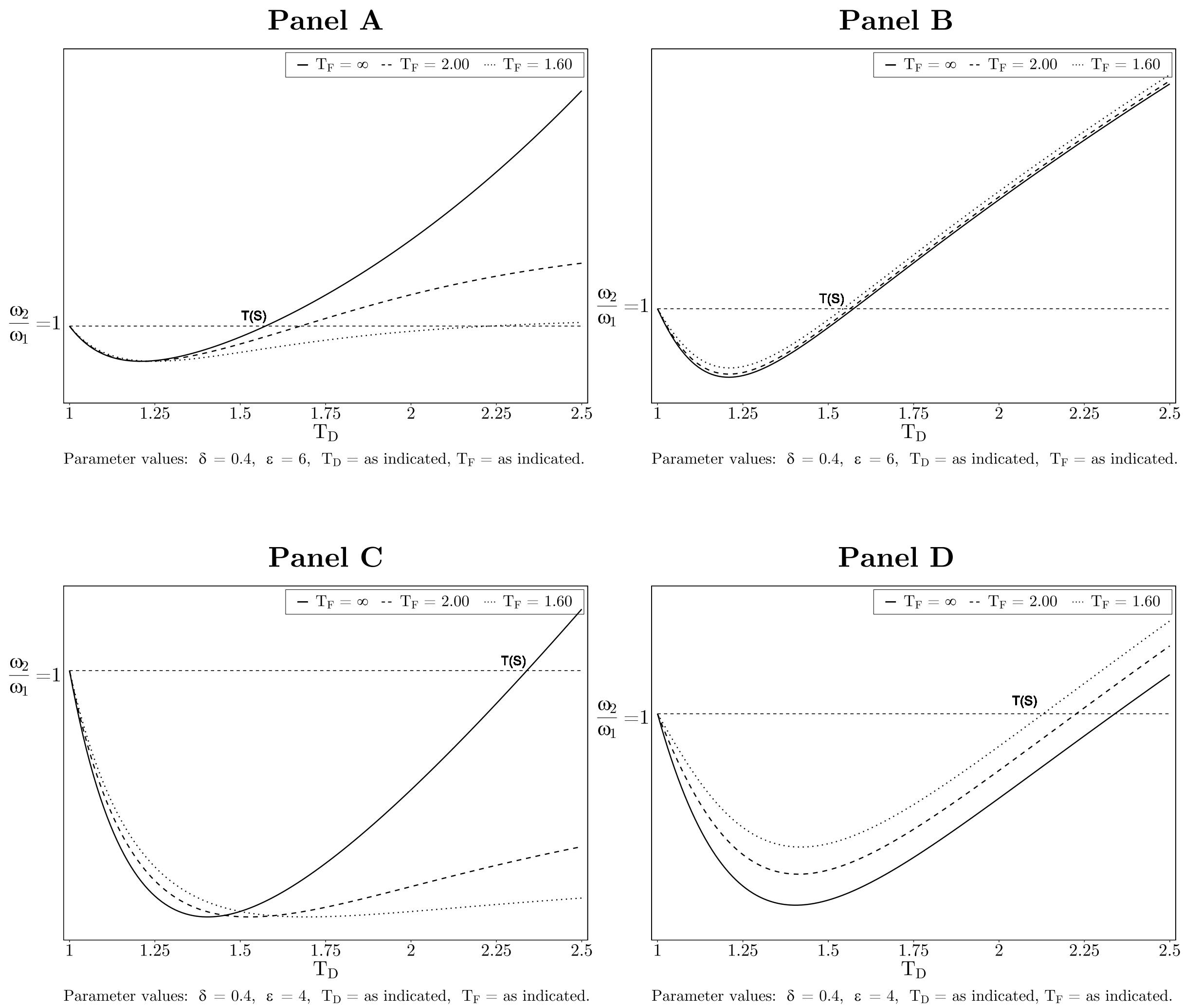


Figure C.2: Internal Transport costs and Sustainable Agglomeration

The results shown in Figure C.2 show similar tendencies as in C.1.2, that is, increased trade liberalization increases the range of values for which a full agglomeration equilibrium is more likely (note the shift in the sustain point $T(S)$ to the right in Panel A). By construction, this mirrors the result in Monfort and Nicolini (2000). Also, by reasons given in Subsection 4.4.5, a lower level of ε cause agglomeration forces to be strengthened, up to the point where a sustain point does not exist for low transport costs across countries (Panel C). Hence, for a level of $T_F \leq 2.00$ and $\varepsilon = 4$, there does not exist a level of internal transport costs for which agglomeration becomes unsustainable. In other words, the existence of external markets renders the costs of serving domestic markets from a distance negligible and it increasingly pays to agglomerate given reduced international trade costs. Notice, however, the stark difference to Panels B and D, i.e. in the case of heterogeneous intra-national space. Here, external trade liberalization causes a decrease in the range of intra-national transport costs T_D for which agglomeration in region 1 is sustainable, and for all parameters tested, there exists a level where agglomeration in region 1 is unsustainable (see Table C.1). Whence the difference? We need to latch on to the discussion in C.1.2 where we discussed the relative influence of centrifugal and centripetal forces of a changing T_F. For heterogeneous intra-national space, there is now an additional component which mediates the relative strength of these two forces, namely, the differential exposure to the external markets, initially shown by Crozet and Koenig 2004b). To the former (centrifugal forces), while agglomeration tendencies are lowered, one may expect an increased draw of firms and consumers to the border so as to benefit from better access to new demand and supply, respectively. To the latter (centripetal forces), there is the possibility that the dispersion force is further amplified, which pushes economic activity towards region 1 given that its larger distance to the border provides an increased level of protection from new foreign competition.

What are the implications for our present results? For all of the parameter configurations of ε and δ tested (see Table C.1), we see a falling range of intra-national transport costs T_D for which agglomeration in region 1 is sustainable.[144] Hence, sheltering in the interior regions does not seem to happen to a larger degree than the draw to the border when international trade costs are decreased. This goes against Crozet and Koenig (2004b) where a push to the interior happens at intermediate international trade costs. The difference in these results most likely stems from the setup of the foreign economy and the moderating force of this. In their model, the foreign economy is larger than the domestic one, which has arguably larger bearings on the competition effect just described.[145] However, contrary to Behrens et al. (2006), what Crozet and Koenig (2004b) have not analyzed in their $2+1$ setup, is the influence of the *relative* size of foreign *regions* on these dynamics. Figure 4.5 in Subsection 4.4.5 provides the main results of this analysis, where we have seen that the draw to the border may be lower or higher when foreign economic activity agglomerates in the interior but nonetheless exists for any distribution of home or foreign activity. Hence, for two equally sized economies, foreign economic inequality cannot turn around our main results, which is that the draw to the border dominates any benefit by sheltering from foreign competition. What we can say, however, is that this effect

[144] What is also confirmed is that for higher shares of δ (consumption share of manufactures) and lower levels of ε (higher product market competition) agglomeration is supported for a wider range of transport costs, i.e. agglomeration forces are strengthened.

[145] Of course, this also depends on the structure of the economies, i.e. whether the two economies are complementary in their trade or whether one of the countries dominates in either imports or exports. These effects are analyzed in Brülhart et al. (2004), albeit for a different model set-up concerning the utility function and thereby not directly comparable to the one in this paper.

may be moderated by foreign economic inequality. As such, from the results in Figure 4.5, it seems as if sheltering in the interior is generally more important when much of the foreign activity is agglomerated at the border, i.e. in region 3 (a low λ_4). Intuitively plausible, this effect seems to be more relevant for the case of increased product market competition (compare Panels B and D).

Table C.1 Sustain and Break points across δ, ε and T_F

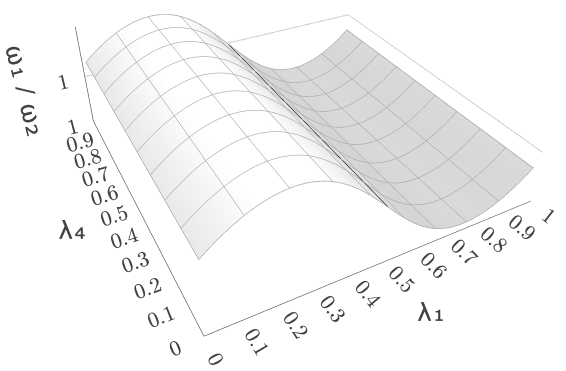
			Sustain and Break Values					
			$\delta = 0.4$		$\delta = 0.5$		$\delta = 0.6$	
			$T_D(B)$	$T_D(S)$	$T_D(B)$	$T_D(S)$	$T_D(B)$	$T_D(S)$
$T_F = \infty$	$\varepsilon = 4$	Symmetric	1.97	2.34	2.47	4.00	3.30	14.62
		Asymmetric	-	2.34	-	4.00	-	14.62
	$\varepsilon = 5$	Symmetric	1.63	1.81	1.90	2.52	2.30	5.00
		Asymmetric	-	1.81	-	2.52	-	5.00
	$\varepsilon = 6$	Symmetric	1.46	1.57	1.64	2.00	1.90	3.16
		Asymmetric	-	1.57	-	2.00	-	3.16
$T_F = 2$	$\varepsilon = 4$	Symmetric	3.58	∅	∅	∅	∅	∅
		Asymmetric	-	2.23	-	3.69	-	12.50
	$\varepsilon = 5$	Symmetric	1.83	2.28	2.69	∅	∅	∅
		Asymmetric	-	1.78	-	2.45	-	4.78
	$\varepsilon = 6$	Symmetric	1.52	1.68	1.79	3.83	2.48	∅
		Asymmetric	-	1.56	-	3.83	-	3.11
$T_F = 1.60$	$\varepsilon = 4$	Symmetric	∅	∅	∅	∅	∅	∅
		Asymmetric	-	2.13	-	3.45	-	10.93
	$\varepsilon = 5$	Symmetric	3.02	∅	∅	∅	∅	∅
		Asymmetric	-	1.73	-	2.37	-	4.49
	$\varepsilon = 6$	Symmetric	1.70	2.23	∅	∅	∅	∅
		Asymmetric	-	1.54	-	1.94	-	3.01

Notes: The values in this table represent the intra-national iceberg transport costs at which agglomeration turns "sustainable" [T(S)] and where the symmetric (spreading) equilibrium is "broken" [T(B)], i.e. at which real wages in the agglomeration exceed those in the periphery and a migration towards one of the regions leads to real wage gains, respectively. For more details on the derivation, see Section C.1 of the Appendix.

C.1.4 Full Simulations

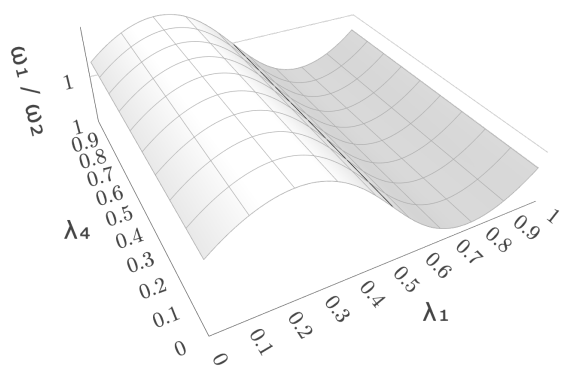
As anticipated in Section 4.4, this subsection provides the full set of simulations, of which selected results are presented and discussed in the main text. I thereby provide the three-dimensional depictions of the simulations, which were discussed as simpler, two-dimensional illustrations before. Figure C.3 and Figure C.4 plot the plane of real wage differentials ω_1 / ω_2 spanned by all possible home and foreign spatial configurations, given by relative shares of the home and foreign workforces λ_1 and λ_4, respectively for a given set of parameter values. Additionally, I provide the full set of corresponding contour lines in Figure C.5 and Figure C.6 which depict the changing influence of foreign economic inequality for stable and unstable equilibria. As established in Subsection 4.4.5, I plot the results for all three levels of external transport costs and additionally, compare results from the main 2+2 setting against the more general 2+2 setting with homogeneous intra-national space.

Panel A



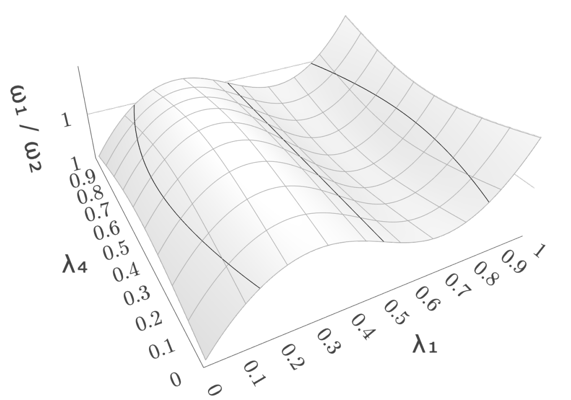
Parameter values: $\delta = 0.4$, $\varepsilon = 6$, $T^d = 1.60$, $T^f = \infty$.

Panel B



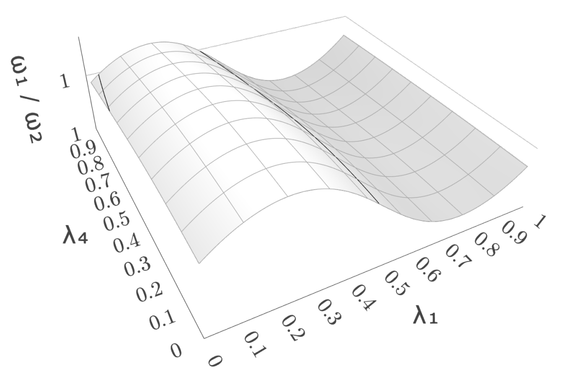
Parameter values: $\delta = 0.4$, $\varepsilon = 6$, $T^d = 1.60$, $T^f = \infty$.

Panel C



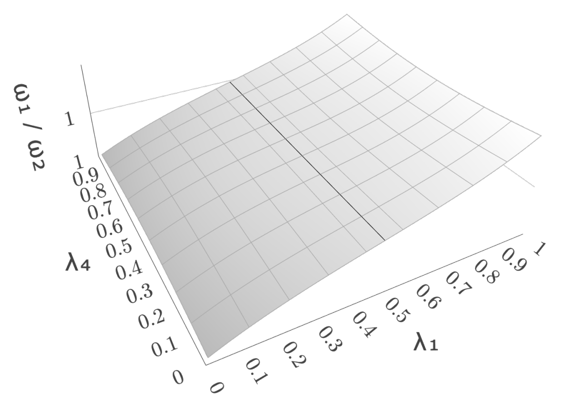
Parameter values: $\delta = 0.4$, $\varepsilon = 6$, $T^d = 1.60$, $T^f = 2.00$.

Panel D



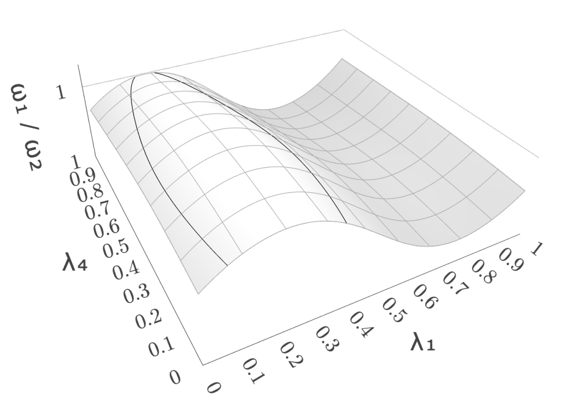
Parameter values: $\delta = 0.4$, $\varepsilon = 6$, $T^d = 1.60$, $T^f = 2.00$.

Panel E



Parameter values: $\delta = 0.4$, $\varepsilon = 6$, $T^d = 1.60$, $T^f = 1.60$.

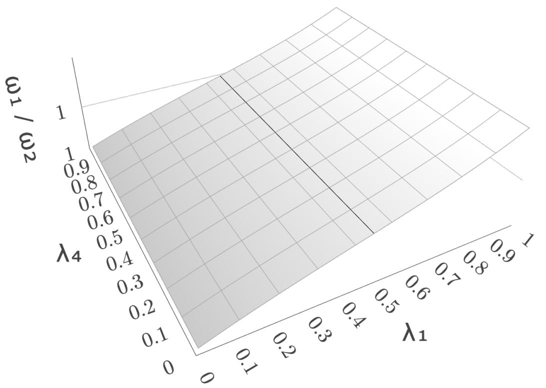
Panel F



Parameter values: $\delta = 0.4$, $\varepsilon = 6$, $T^d = 1.60$, $T^f = 1.60$.

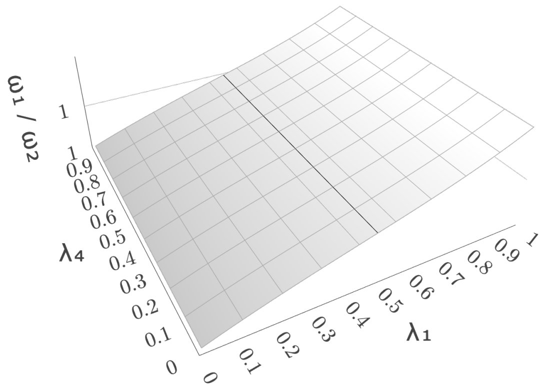
Figure C.3: Three-dimensional Depiction of Spatial Equilibria ($\boldsymbol{\varepsilon = 6}$)

Panel A



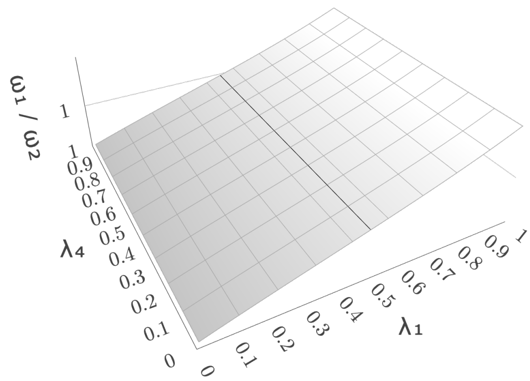
Parameter values: $\delta = 0.4$, $\varepsilon = 4$, $T^d = 1.60$, $T^f = \infty$.

Panel B



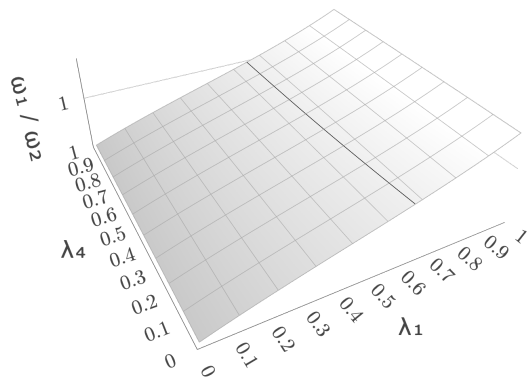
Parameter values: $\delta = 0.4$, $\varepsilon = 4$, $T^d = 1.60$, $T^f = \infty$.

Panel C



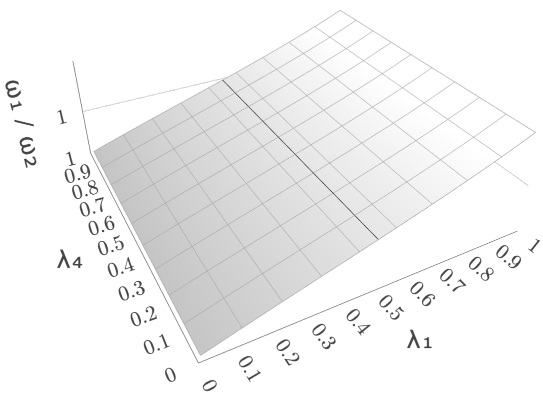
Parameter values: $\delta = 0.4$, $\varepsilon = 4$, $T^d = 1.60$, $T^f = 2.00$.

Panel D



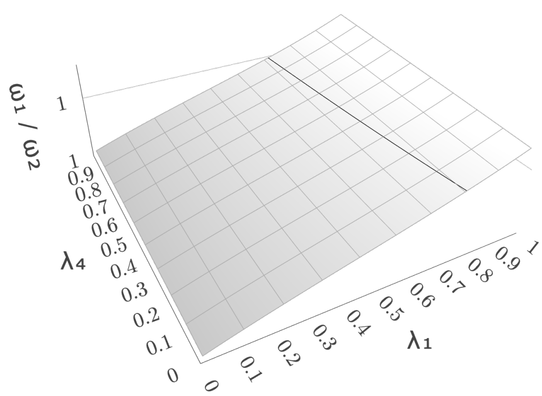
Parameter values: $\delta = 0.4$, $\varepsilon = 4$, $T^d = 1.60$, $T^f = 2.00$.

Panel E



Parameter values: $\delta = 0.4$, $\varepsilon = 4$, $T^d = 1.60$, $T^f = 1.60$.

Panel F



Parameter values: $\delta = 0.4$, $\varepsilon = 4$, $T^d = 1.60$, $T^f = 1.60$.

Figure C.4: Three-dimensional Depiction of Spatial Equilibria ($\boldsymbol{\varepsilon = 4}$)

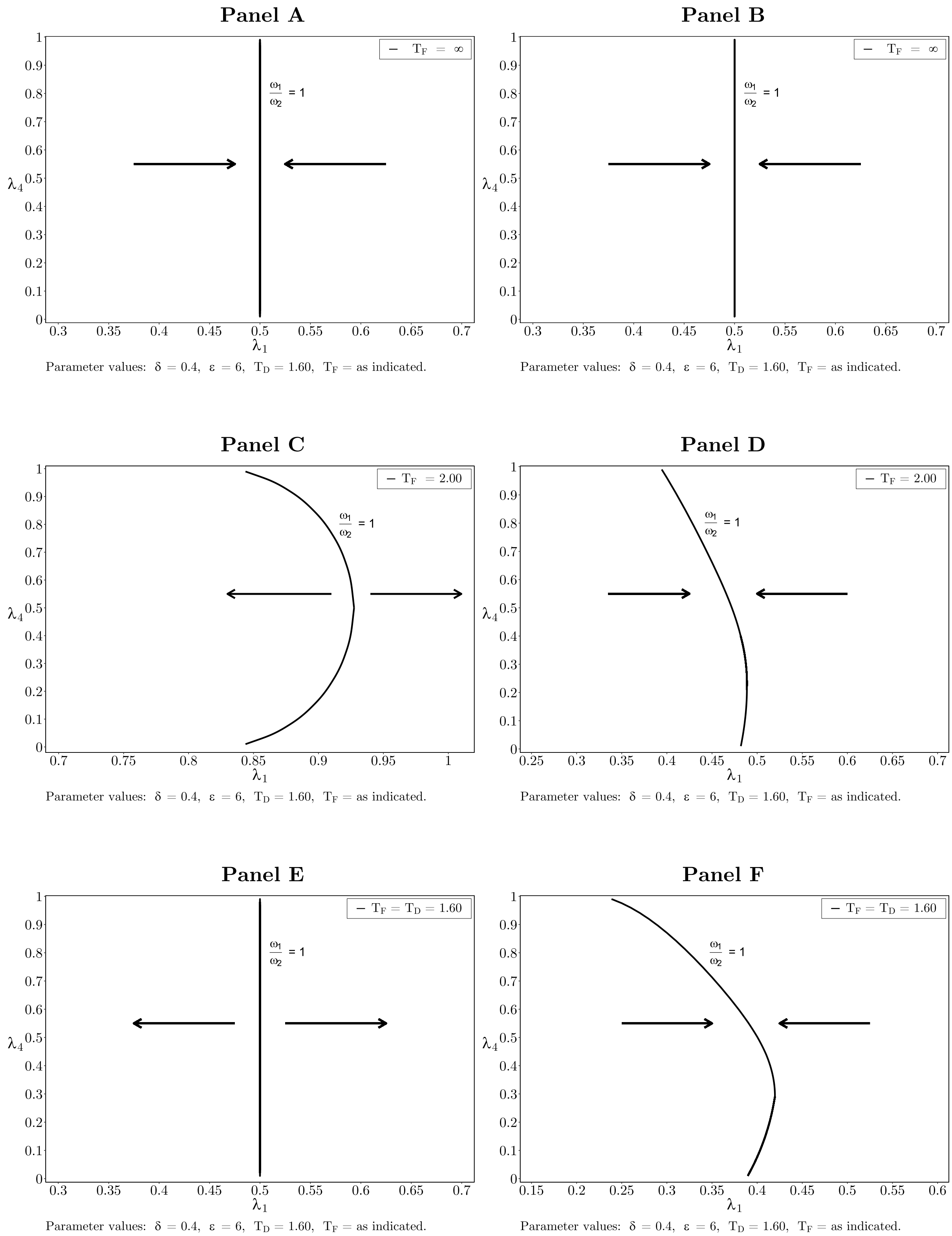


Figure C.5: Contour Lines of Spatial Equilibria ($\boldsymbol{\varepsilon = 6}$)

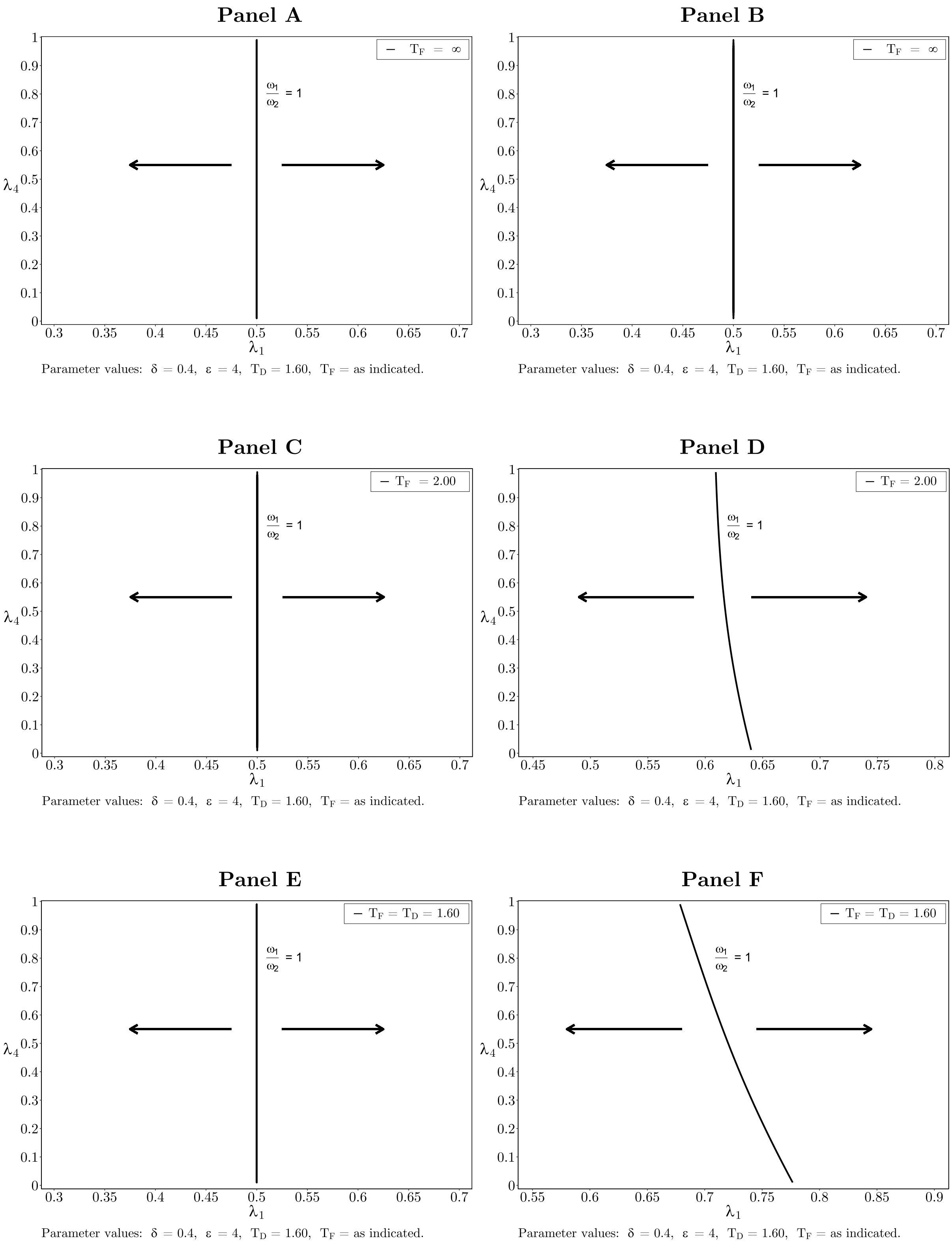


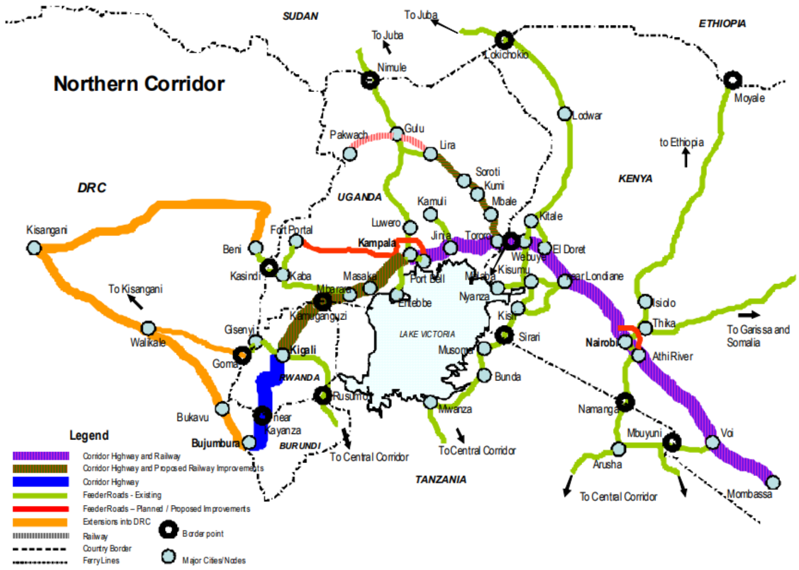
Figure C.6: Contour Lines of Spatial Equilibria ($\boldsymbol{\varepsilon} = \mathbf{4}$)

C.2 Further Empirical Results and Documentation

Figure C.7: East African (Northern and Central) Trade Corridor

6

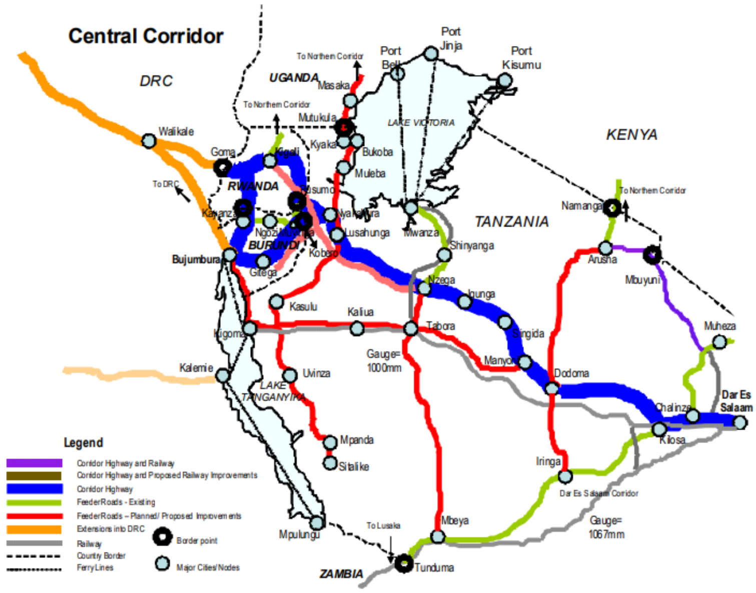
Figure 2-1
Northern Corridor Network



Source: Nathan Associates Inc.

31

Figure 3-1
Central Corridor Network



Source: Nathan Associates Inc.

Source: Figures taken from Nathan Associates (2011) pages 6 and 31, respectively.

Figure C.8: East African Trade, Means of Transport

Table 4-4

Northern and Central Corridor Traffic by Type and Mode, 2009 (000 tons)

Type of Traffic	Road	Rail	Total	Rail Share (%)
NORTHERN				
Transit	5,509	417	5,926	7%
Regional	2,974	151	3,125	5%
Domestic	11,817	622	12,439	5%
Total	20,300	1,190	21,490	6%
CENTRAL				
Transit	357	111	468	24%
Regional	658	32	690	5%
Domestic	5,617	296	5,913	5%
Total	6,632	439	7,071	6%
Total	26,932	1,629	28,561	6%

Source: Nathan Associates Inc.

Source: Table taken from Nathan Associates (2011) page 63.

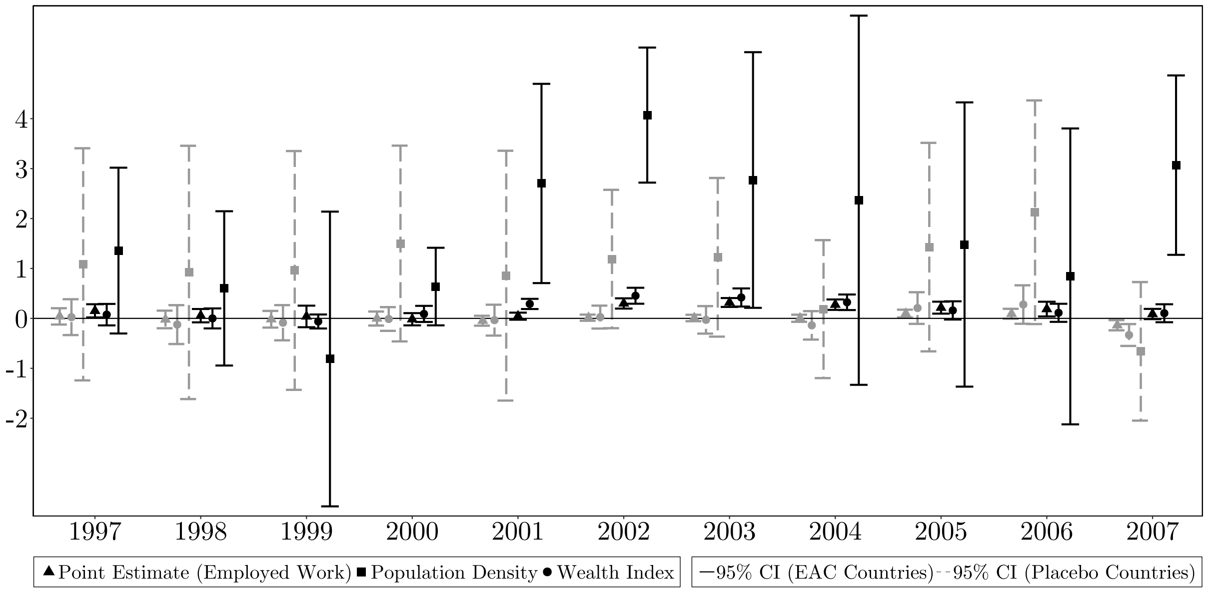


Figure C.9: DiD in Core Regions with varying Intervention Year

Notes: The "Point Estimates" depicted represent difference-in-difference estimates for the "Core Region {0,1}", including a 95% confidence interval, for each of the three dependent variables. The estimates are constructed for each (placebo) intervention year, separately, by expanding the data past the respective cut-off. E.g. for the estimate on the placebo intervention year 1997, the regressions use data from first available data for all dependent variables and includes survey years up until 2000. For the intervention year 1998, the data frame is expanded to include the year 2001, for 1999 it includes 2002, and so on. The plot is restricted to estimates from 1997 onwards, given that data for outcome variables in Placebo Countries before this date is only sparsely available. See the Table C.16 for a longer range of estimates. for the main outcome variables in placebo countries only few samples provide the main dependent variables before this date. Again, Wealth Index is available only in Malawi in 1992 Rwanda in1992, and Zambia in 1996. 1997 is the first year for which repeat observations in a country are available. The data used for the estimations stem from the Women, Men and Household Recodes of the Demographic and Health Surveys (DHS), including KAP, MIS and AIS rounds. For the "EAC Countries" the total sample includes the Kenya survey rounds of 1988-89, 1993, 1998, 2003, 2008, the Tanzania survey rounds of 1991-92, 1994, 1996, 1999, 2003-04, 2007-08, 2010, and the Uganda survey rounds of 1988-89, 1995, 1995-96, 2000-01, 2006, 2009, 2011. For the "Placebo Countries" the total sample includes the Ethiopia survey rounds of 2000,2005, 2011, the Mozambique survey rounds of 1997, 2003, 2009, 2011, the Malawi survey rounds of 1992, 2000, 2004, 2010, the Rwanda DHS survey rounds of 1992,2000, 2005, and the Zambia survey rounds of 1996, 2001-02 and 2007. Standard errors are clustered at the region level. All dependent variables represent z-scores, i.e. are standardized with a mean of zero and a standard deviation of one.

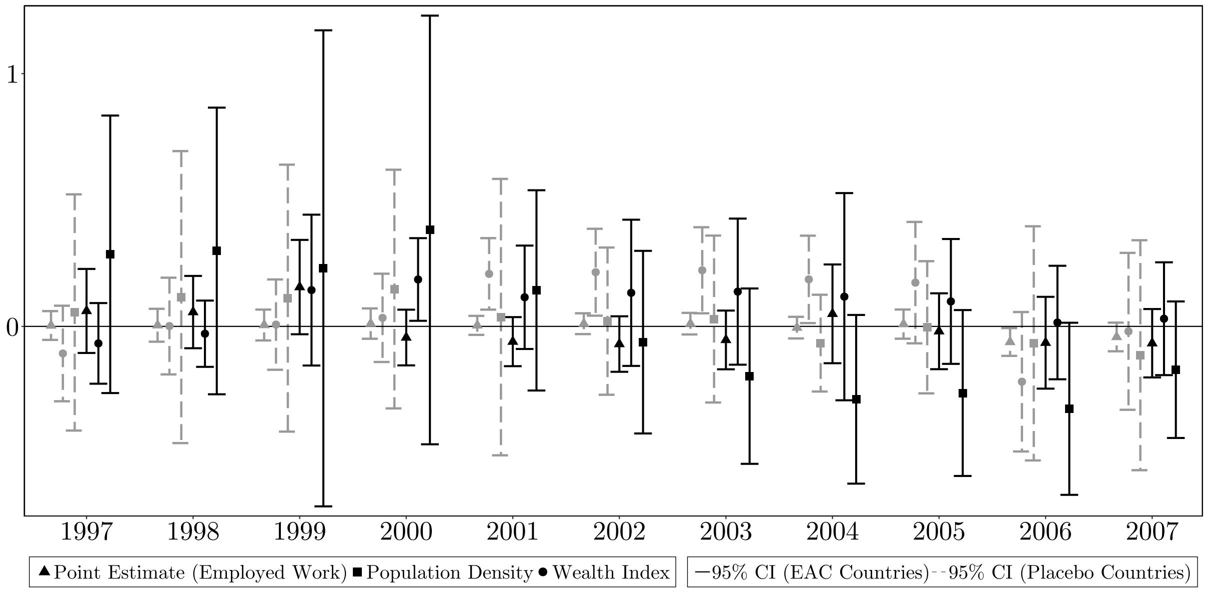


Figure C.10: DiD in Border Regions with varying Intervention Year

Notes: The results depicted in this table are estimated in the same way as those shown in Figure C.9 such that most of the respective notes apply here. However, not the following important amendments: The "Point Estimates" depicted represent difference-in-difference estimates for the Border Region {0,1}, including a 95% confidence interval, for each of the three dependent variables. See full notes below Figure C.9.

Table C.2: Summary Statistics

Panel a)	Mean	St.Dev.	Min.	1st Quart.	Median	3rd Quart.	Core	Max.	N
	Distribution across Sample								
Basic Characteristics									
Afrobarometer (AFB)									
Age	36	14	17	25	33	43	34	101	38,322
Education (Level)	3.3	1.8	0.0	2.0	3.0	4.5	4.2	9	38,556
Female {0,1}	0.5	0.5	0.0	0.0	1.0	1.0	0.5	1.0	38,322
Demographic and Health Survey (DHS)									
Age	29	10	15	20	27	36	28	60	332,725
Education (Years)	5.83	4.00	0.00	2.00	7.00	8.00	8.71	26	330,979
Female {0,1}	0.8	0.4	0.0	1.0	1.0	1.0	0.8	1.0	332,725
Kagera Health and Development Survey (KHDS)									
Age	22	19	0	7	16	31	23	105	48,075
Education (Years)	5.32	3.30	0.00	3.00	6.00	7.00	6.12	22	30,154
Female {0,1}	0.5	0.5	0.0	0.0	1.0	1.0	0.5	1.0	48,119
Main Covariates									
AFB									
Core {0,1}	0.13	0.34	0.00	0.00	0.00	0.00	1.00	1.00	38,644
Urban {0,1}	0.26	0.44	0.00	0.00	0.00	1.00	0.73	1.00	38,644
Agglomeration {0,1}	0.28	0.45	0.00	0.00	0.00	1.00	0.64	1.00	38,644
Road Distance to nearest Border Crossing (km)	315	264	0.8	127	236	407	240	1362	38,644
Inv. rel. Distance to nearest Border Crossing [0,1]	0.71	0.22	0.00	0.61	0.76	0.87	0.77	1.00	38,644
DHS									
Core {0,1}	0.09	0.29	0.00	0.00	0.00	0.00	1.00	1.00	210,747
Urban {0,1}	0.22	0.42	0.00	0.00	0.00	0.00	0.73	1.00	245,118
Agglomeration {0,1}	0.25	0.43	0.00	0.00	0.00	1.00	0.42	1.00	210,747
Road Distance to nearest Border Crossing (in km)	342	283	0.1	132	262	437	208	1373	205,050
Inv. rel. Distance to nearest Border Crossing [0,1]	0.69	0.23	0.00	0.58	0.74	0.87	0.77	1.00	205,050
KHDS									
Core {0,1}	0.21	0.41	0.00	0.00	0.00	0.00	1.00	1.00	9,288
Urban {0,1}	0.22	0.41	0.00	0.00	0.00	0.00	0.98	1.00	9,288
Road Distance to nearest Border Crossing (in km)	148	107	0	80	103	170	119	1134	9,288
Inv. rel. Distance to nearest Border Crossing [0,1]	0.87	0.10	0.00	0.85	0.90	0.93	0.90	1.00	9,288
Geographical Covariates									
AFB									
Distance to Harbor (in km)	702	322	1	441	738	956	537	1259	38,644
Distance to Navigable River (in km)	305	271	1	73	198	492	398	963	38,644
Distance to Major Lake (in km)	180	180	0	46	117	251	238	700	38,644
Elevation (in m)	1213	507	4	1072	1205	1519	1	3914	38,644
Terrain Ruggedness (standardized)	0.00	1.00	-0.77	-0.56	-0.30	0.13	-0.49	8.63	38,644
Average Monthly Temperature (in Celsius)	24	3	8	22	24	25	24	32	38,644
Average Monthly Rainfall (in mm)	101	27	17	86	102	116	102	214	38,644
Growing Days {0,365}	292	71	14	247	304	358	310	365	38,644
Malaria Ecology Index	7.0	6.4	0.0	1.5	5.6	10.5	8.2	31.1	38,604
DHS									
Distance to Harbor (in km)	696	319	0	446	731	927	744	1267	210,747
Distance to Navigable River (in km)	313	269	1	84	207	502	204	961	210,747
Distance to Major Lake (in km)	183	185	0	43	115	254	102	706	210,747
Elevation (in m)	1189	514	2	1059	1193	1472	1	3248	210,747
Terrain Ruggedness (standardized)	0.00	1.00	-0.75	-0.56	-0.30	0.12	-0.46	12.7	210,747
Average Monthly Temperature (in Celsius)	22	3	11	20	22	23	21	30	210,747
Average Monthly Rainfall (in mm)	94	26	15	77	95	109	98	219	210,747
Growing Days {0,365}	286	72	10	229	295	358	329	365	210,747
Malaria Ecology Index	4.9	8.0	0.0	0.0	1.1	5.5	0.1	36.0	171,881
KHDS									
Distance to Harbor (in km)	899	159	0	913	919	938	806	1189	9,288
Distance to Navigable River (in km)	70	157	1	19	31	51	139	963	9,288
Distance to Major Lake (in km)	51	104	0	5	17	66	77	665	9,288
Elevation (in m)	1274	250	3	1188	1259	1392	1	4249	9,287
Terrain Ruggedness (standardized)	0.00	1.00	-0.80	-0.73	-0.48	0.35	-0.36	5.85	9,288
Average Monthly Temperature (in Celsius)	21	1	5	20	21	21	21	27	9,287
Average Monthly Rainfall (in mm)	113	31	48	83	112	142	152	168	9,287
Growing Days {0,365}	316	37	154	292	324	353	341	365	9,288
Malaria Ecology Index	4.2	3.2	0.0	2.6	3.9	4.3	6.1	23.0	9,260

Notes: The table is continued on the next page.

Panel b)	Mean	St.Dev.	Min.	Distribution across Distance (Quartiles) 1st	2nd	3rd	4th	Core	N
Consumption									
AFB									
How often: Gone without (Lived Poverty) [0,4]	1.12	0.95	0	1.05	1.02	1.16	1.25	0.85	36,371
How often: Gone without Food {0,4}	0.98	1.12	0	1.01	0.91	1.00	1.00	0.75	36,335
How often: Gone without Water {0,4}	1.13	1.31	0	0.92	1.00	1.18	1.38	0.86	36,353
How often: Gone without Medical Care {0,4}	1.26	1.23	0	1.23	1.17	1.28	1.37	0.94	36,300
DHS									
Wealth Quintile {1,5}	3.14	1.47	1	3.17	3.42	3.03	2.59	4.61	251,684
International Wealth Index (IWI)	23.44	2.71	18	23.37	23.67	23.46	22.89	25.02	332,725
Comparative Wealth Index (CWI)	-0.67	0.85	-2	-0.66	-0.38	-0.60	-0.91	0.46	104,693
KHDS									
Annual p.c. Consumpt. in 2010 TZS ('000)	554	522	36	558	551	436	650	747	6,855
Annual p.c. non-Food Consumpt. in 2010 TZS ('000)	204	299	6	199	199	149	261	293	6,921
Value of Occupied Dwelling in 2004 TZS ('000)	650	3527	0	624	1062	477	356	1425	6,076
Value of Durable Assets in 2004 TZS ('000)	112	1060	0	93	164	45	140	287	6,079
Income & Work									
AFB									
Worked last Year {0,1}	0.55	0.50	0	0.52	0.56	0.56	0.55	0.66	26,880
Employed Work {0,1}	0.23	0.42	0	0.24	0.29	0.23	0.17	0.36	25,775
Occupation Level: Agr.-Worker-Prof. {1,3}	1.66	0.81	1	1.66	1.83	1.65	1.53	2.18	21,482
How often: Gone without Cash Income {0,4}	2.09	1.24	0	2.08	2.02	2.12	2.15	1.75	36,304
DHS									
Worked last Year {0,1}	0.76	0.42	0	0.76	0.75	0.79	0.82	0.72	230,419
Employed Work {0,1}	0.19	0.40	0	0.21	0.30	0.18	0.08	0.47	105,528
Occupation Level: Agr.-Worker-Prof. {1,3}	1.51	0.64	1	1.62	1.70	1.51	1.37	2.00	143,027
Paid in Cash {0,1}	0.52	0.50	0	0.52	0.59	0.50	0.42	0.77	136,866
KHDS									
Worked last Year {0,1}	0.26	0.44	0	0.26	0.24	0.19	0.33	0.33	20,259
Employed Work {0,1}	0.12	0.33	0	0.10	0.12	0.13	0.14	0.13	32,017
Occupation Level: Agr.-Worker-Prof. {1,3}	1.20	0.49	1	1.22	1.22	1.11	1.25	1.41	20,446
Salaried Work {0,1}	0.01	0.10	0	0.01	0.01	0.01	0.01	0.02	19,465
Monthly Salary in 2004 TZS ('000)	28.8	55.6	0.0	29.0	33.6	20.2	27.8	33.7	2,118
Agglomeration & Migration									
AFB									
Population Count	1.26	4.34	0	0.81	2.64	1.02	0.55	6.30	38,644
Population Density	1.47	5.05	0	0.96	3.07	1.21	0.65	7.34	38,644
DHS									
Population Count	1.18	3.45	0	0.97	1.93	1.49	0.35	5.33	188,990
Population Density	1.39	4.03	0	1.16	2.26	1.75	0.41	6.23	188,990
At region of residence before EAC {0,1}	0.44	0.50	0	0.61	0.68	0.67	0.74	0.74	69,349
KHDS									
Population Count	1.01	1.87	0	0.79	0.80	1.01	1.44	747	9,288
Population Density	0.64	2.16	0	0.29	0.61	0.22	1.45	1.89	9,288
At region of residence before EAC {0,1}	0.92	0.28	0	0.90	0.90	0.90	0.86	0.86	38,592
Main reason for Migration: Economic {0,1}	0.09	0.29	0	0.09	0.11	0.05	0.08	0.14	6,946
After moving to curr. Residence: Paid Empl. {0,1}	0.09	0.28	0	0.09	0.07	0.03	0.12	0.14	2,388
Sentiments & Attitudes									
AFB									
Helps your Country: REC / EAC {1,3}	1.66	0.92	0	1.67	1.65	1.67	1.66	1.63	11,750
Support for Regional Integration {1,5}	3.71	1.51	1	3.87	3.85	3.70	3.38	3.75	6,394
Would like as Neighbor: Immigrant {1,3}	3.38	1.30	1	3.43	3.31	3.29	3.46	3.18	11,708
Ease of Crossing Borders to live and Work {1,4}	2.21	0.93	1	2.35	2.20	2.15	2.13	2.23	4,784
Present vs. Past: Living Standards of People {1,4}	2.43	1.03	1	2.71	2.71	2.66	2.75	2.74	23,602
KHDS									
Subjective HH. Wealth: Today (2004) {1,5}	2.6	0.6	1.0	2.7	2.7	2.5	2.6	2.7	3,313
Subjective HH. Wealth: Ten Years ago (1994) {1,5}	2.6	0.7	1.0	2.7	2.7	2.6	2.6	2.7	3,313
Subjective HH Life Satisfaction: Ladder {1,9}	3.8	1.5	1.0	3.9	3.9	3.6	3.8	4.1	3,313
Subjective HH Life Satisfaction: Ladder {1,9}	3.8	1.5	1.0	3.9	3.9	3.6	3.8	4.1	3,313

Notes: The table depicts summary statistics corresponding to the main sample used in the estimations across the paper. See the manuscript for more information on the specific survey samples and survey years. The last set of data stem from the Kagera Health and Development Survey (KHDS). The KHDS summary statistics displayed includes repeat observations from tracked individuals (a maximum of 6 times). Remaining variations in the number of observations sizes stem from differences in response rates of variables as well as changes in questions asked across surveys. The distribution "across distance" in b) calculates mean values of the respective variable within quartiles of road distance to nearest Border crossings.

Table C.3: Robustness Checks Summarized (Afrobarometer)

	Afrobarometer					
	Dependent Variable					
Panel a)	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients						
(see Tables 4.2 and C.9)						
Core {0,1} × EAC	-0.261***	-0.013	-0.191***	0.026	0.021	0.041
	(0.077)	(0.050)	(0.054)	(0.067)	(0.079)	(0.083)
Core {0,1} × CU	-0.470***	-0.006	0.728**	0.002	0.057	-0.234**
	(0.097)	(0.045)	(0.332)	(0.055)	(0.090)	(0.105)
Core {0,1} × CM	-0.338***	-0.076	1.083***	-0.258**	-0.009	0.043
	(0.065)	(0.069)	(0.375)	(0.110)	(0.087)	(0.079)
(a) Standard Errors clustered at EA-Level						
(see Table C.22)						
Core {0,1} × EAC	-0.261**	-0.013	-0.158	0.026	0.021	0.041
	(0.126)	(0.041)	(0.135)	(0.044)	(0.084)	(0.159)
Core {0,1} × CU	-0.470***	-0.006	0.769***	0.002	0.057	-0.234
	(0.125)	(0.044)	(0.214)	(0.046)	(0.080)	(0.155)
Core {0,1} × CM	-0.338***	-0.076**	1.113***	-0.258***	-0.009	0.043
	(0.119)	(0.036)	(0.179)	(0.041)	(0.078)	(0.150)
(b) No Controls						
(see Table C.26)						
Core {0,1} × EAC	-0.392***	-0.011	-0.188***	0.018	0.029	-0.047
	(0.088)	(0.044)	(0.040)	(0.057)	(0.068)	(0.089)
Core {0,1} × CU	-0.607***	0.005	0.738**	0.002	0.103	-0.318***
	(0.116)	(0.048)	(0.362)	(0.053)	(0.103)	(0.114)
Core {0,1} × CM	-0.466***	-0.054	1.094***	-0.253***	0.060	-0.061
	(0.081)	(0.065)	(0.396)	(0.102)	(0.110)	(0.077)
(c) Extended Geographic Controls						
(see Table C.30)						
Core {0,1} × EAC	-0.211***	-0.012	-0.216**	0.019	0.021	0.054
	(0.080)	(0.051)	(0.086)	(0.069)	(0.076)	(0.081)
Core {0,1} × CU	-0.424***	-0.001	0.694**	0.000	0.069	-0.205**
	(0.094)	(0.047)	(0.301)	(0.060)	(0.084)	(0.088)
Core {0,1} × CM	-0.304***	-0.076	1.031***	-0.262**	-0.018	0.046
	(0.073)	(0.070)	(0.353)	(0.109)	(0.082)	(0.083)
(d) Including Survey Weights						
(see Table C.34)						
Core {0,1} × EAC	-0.245***	-0.016	-0.168**	0.016	-0.010	0.039
	(0.079)	(0.048)	(0.069)	(0.069)	(0.092)	(0.082)
Core {0,1} × CU	-0.443***	0.001	0.910***	0.014	0.012	-0.187*
	(0.084)	(0.054)	(0.351)	(0.063)	(0.097)	(0.098)
Core {0,1} × CM	-0.340***	-0.089	0.916**	-0.268**	-0.002	0.032
	(0.067)	(0.078)	(0.388)	(0.114)	(0.096)	(0.082)
(e) Excluding Low-Precision Localities						
(see Table C.36)						
Core {0,1} × EAC	-0.008	0.036	-0.139***	0.074	-0.011	0.336***
	(0.092)	(0.036)	(0.030)	(0.058)	(0.092)	(0.123)
Core {0,1} × CU	-0.238**	0.054	0.757**	0.029	-0.067	0.142
	(0.108)	(0.043)	(0.366)	(0.070)	(0.147)	(0.134)
Core {0,1} × CM	-0.081	0.019	1.041***	-0.217**	-0.090	0.383***
	(0.078)	(0.050)	(0.213)	(0.089)	(0.107)	(0.110)

Notes: Table continued on next page.

	Afrobarometer					
	Dependent Variable					
Panel b)	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients						
(see Tables 4.2 and C.9)						
Core {0,1} × EAC	-0.261***	-0.013	-0.191***	0.026	0.021	0.041
	(0.077)	(0.050)	(0.054)	(0.067)	(0.079)	(0.083)
Core {0,1} × CU	-0.470***	-0.006	0.728**	0.002	0.057	-0.234**
	(0.097)	(0.045)	(0.332)	(0.055)	(0.090)	(0.105)
Core {0,1} × CM	-0.338***	-0.076	1.083***	-0.258**	-0.009	0.043
	(0.065)	(0.069)	(0.375)	(0.110)	(0.087)	(0.079)
(f) Agglomerations vs. Core						
(see Table C.38)						
Agglomeration {0,1} × EAC	0.108	-0.056**	0.022	-0.060*	-0.046	-0.199*
	(0.078)	(0.028)	(0.082)	(0.034)	(0.079)	(0.115)
Agglomeration {0,1} × CU	-0.036	-0.006	0.129	-0.011	0.001	-0.350***
	(0.092)	(0.031)	(0.102)	(0.035)	(0.075)	(0.115)
Agglomeration {0,1} × CM	0.024***	-0.011***	0.242***	-0.063***	0.042***	-0.334***
	(0.080)	(0.027)	(0.150)	(0.045)	(0.071)	(0.117)
(g) Core ≤ 25km						
(see Table C.42)						
Core ≤ 25km {0,1} × EAC	-0.301***	-0.037	-0.208**	0.035	0.009	-0.021
	(0.076)	(0.061)	(0.087)	(0.076)	(0.079)	(0.075)
Core ≤ 25km {0,1} × CU	-0.520***	-0.051	0.728*	-0.040	-0.049	-0.244**
	(0.098)	(0.061)	(0.423)	(0.071)	(0.058)	(0.105)
Core ≤ 25km {0,1} × CM	-0.375***	-0.120***	1.195***	-0.342***	-0.101***	0.070***
	(0.062)	(0.090)	(0.440)	(0.126)	(0.093)	(0.080)
(h) Core ≤ 10km						
(see Table C.46)						
Core ≤ 10km {0,1} × EAC	-0.309***	-0.090	0.098	0.007	-0.192***	-0.105
	(0.095)	(0.060)	(0.132)	(0.057)	(0.066)	(0.076)
Core ≤ 10km {0,1} × CU	-0.496***	-0.055	0.914*	-0.055	-0.209***	-0.341***
	(0.102)	(0.044)	(0.529)	(0.041)	(0.066)	(0.112)
Core ≤ 10km {0,1} × CM	-0.368***	-0.126	1.554***	-0.351***	-0.313***	0.091
	(0.068)	(0.079)	(0.597)	(0.097)	(0.114)	(0.082)
(i) Post-CM Development						
(see Table C.50)						
Core {0,1} × EAC	-0.261***	-0.013	-0.189***	0.026	0.021	0.042
	(0.077)	(0.050)	(0.052)	(0.067)	(0.078)	(0.082)
Core {0,1} × CU	-0.470***	-0.006	0.731**	0.003	0.056	-0.233**
	(0.097)	(0.045)	(0.335)	(0.055)	(0.091)	(0.105)
Core {0,1} × CM 1[2010-201	-0.344***	-0.044	1.263***	-0.256**	-0.081	0.053
	(0.066)	(0.080)	(0.409)	(0.123)	(0.088)	(0.094)
Core {0,1} × 1[t ≥ 2015]	-0.332***	-0.097	0.897**	-0.259**	0.037	0.027
	(0.066)	(0.070)	(0.393)	(0.104)	(0.071)	(0.082)

Notes: This table offers an array of robustness tests on the main results of the paper for the difference-in-difference effect for individuals living in core agglomerations. As such, the results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. The full results (including border estimates) are given in the table referred to below the description of each test. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used, if not indicated otherwise. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table C.4: Robustness Checks Summarized (DHS)

	Demographic and Health Survey (DHS)					
	Dependent Variable					
Panel a)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked Last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients **(see Tables 4.2 and C.9)**						
Core {0,1} × EAC	0.433***	0.073*	0.146	0.067*	-0.046	-0.002
	(0.141)	(0.040)	(0.278)	(0.038)	(0.062)	(0.026)
Core {0,1} × CU	0.314***	0.037	0.770***	0.014	0.058	0.053
	(0.105)	(0.039)	(0.177)	(0.043)	(0.111)	(0.033)
Core {0,1} × CM	0.361**	0.006	0.956***	0.061*	0.007	0.029
	(0.142)	(0.041)	(0.111)	(0.031)	(0.088)	(0.034)
(a) Standard Errors clustered at EA-Level **(see Table C.23)**						
Core {0,1} × EAC	0.435***	0.073*	0.146	0.068***	-0.046	-0.002
	(0.110)	(0.040)	(0.201)	(0.021)	(0.058)	(0.038)
Core {0,1} × CU	0.315***	0.037	0.770***	0.014	0.058	0.053
	(0.097)	(0.032)	(0.241)	(0.020)	(0.052)	(0.035)
Core {0,1} × CM	0.345***	0.006	0.956***	0.049***	0.007	0.029
	(0.091)	(0.029)	(0.185)	(0.017)	(0.047)	(0.032)
(b) No Controls **(see Table C.27)**						
Core {0,1} × EAC	0.334*	0.077	0.071	0.092	-0.098	-0.065
	(0.173)	(0.050)	(0.320)	(0.060)	(0.094)	(0.046)
Core {0,1} × CU	0.386***	0.052	0.708***	0.031	0.060	0.029
	(0.149)	(0.042)	(0.170)	(0.062)	(0.136)	(0.049)
Core {0,1} × CM	0.375*	0.018	0.914***	0.077*	-0.024	0.009
	(0.224)	(0.047)	(0.104)	(0.044)	(0.116)	(0.043)
(c) Extended Geographic Controls **(see Table C.31)**						
Core {0,1} × EAC	0.483***	0.066*	0.014	0.042	-0.040	0.016
	(0.182)	(0.036)	(0.255)	(0.029)	(0.069)	(0.027)
Core {0,1} × CU	0.383***	0.042	0.721***	-0.003	0.071	0.079**
	(0.126)	(0.034)	(0.145)	(0.037)	(0.107)	(0.032)
Core {0,1} × CM	0.403**	0.007	0.886***	0.046*	0.008	0.045
	(0.157)	(0.041)	(0.094)	(0.026)	(0.091)	(0.030)
(d) Including Survey Weights **(see Table C.35)**						
Core {0,1} × EAC	0.433***	0.073*	0.146	0.067*	-0.046	-0.002
	(0.141)	(0.040)	(0.278)	(0.038)	(0.062)	(0.026)
Core {0,1} × CU	0.314***	0.037	0.770***	0.014	0.058	0.053
	(0.105)	(0.039)	(0.177)	(0.043)	(0.111)	(0.033)
Core {0,1} × CM	0.361**	0.006	0.956***	0.061*	0.007	0.029
	(0.142)	(0.041)	(0.111)	(0.031)	(0.088)	(0.034)
(e) Excluding Low-Precision Localities **(see Table C.37)**						
Core {0,1} × EAC	0.439***	0.072*	-0.130	0.065*	-0.039	-0.003
	(0.147)	(0.040)	(0.303)	(0.039)	(0.069)	(0.030)
Core {0,1} × CU	0.326***	0.038	0.734**	0.015	0.063	0.052
	(0.107)	(0.038)	(0.293)	(0.043)	(0.113)	(0.035)
Core {0,1} × CM	0.342**	0.006	0.950***	0.035*	0.058	0.042
	(0.157)	(0.043)	(0.103)	(0.019)	(0.075)	(0.041)
(f) Agglomerations vs. Core **(see Table C.39)**						
Agglomeration {0,1} × EAC	0.102	-0.044	0.256	0.015	-0.147**	-0.067
	(0.153)	(0.043)	(0.226)	(0.025)	(0.060)	(0.046)
Agglomeration {0,1} × CU	-0.033	-0.052*	0.208*	0.034*	-0.117**	-0.029
	(0.108)	(0.027)	(0.111)	(0.019)	(0.050)	(0.041)
Agglomeration {0,1} × CM	-0.239***	-0.041**	0.258	0.002	-0.070**	-0.016
	(0.090)	(0.017)	(0.169)	(0.022)	(0.033)	(0.031)

Notes: Table continued on next page.

	Demographic and Health Survey (DHS)					
	Dependent Variable					
Panel b)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked Last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients						
(see Tables 4.2 and C.9)						
Core {0,1} × EAC	0.433***	0.073*	0.146	0.067*	-0.046	-0.002
	(0.141)	(0.040)	(0.278)	(0.038)	(0.062)	(0.026)
Core {0,1} × CU	0.314***	0.037	0.770***	0.014	0.058	0.053
	(0.105)	(0.039)	(0.177)	(0.043)	(0.111)	(0.033)
Core {0,1} × CM	0.361**	0.006	0.956***	0.061*	0.007	0.029
	(0.142)	(0.041)	(0.111)	(0.031)	(0.088)	(0.034)
(g) Core ≤ 25km						
(see Table C.43)						
Core ≤ 25km {0,1} × EAC	0.411**	0.102**	-0.028	0.061	-0.016	0.020
	(0.172)	(0.047)	(0.293)	(0.041)	(0.076)	(0.035)
Core ≤ 25km {0,1} × CU	0.220**	0.060	0.759***	0.011	0.040	0.070**
	(0.108)	(0.049)	(0.160)	(0.044)	(0.132)	(0.036)
Core ≤ 25km {0,1} × CU	0.296*	-0.004	1.056***	0.061**	-0.023	0.012
	(0.162)	(0.049)	(0.180)	(0.030)	(0.084)	(0.040)
(h) Core ≤ 10km						
(see Table C.47)						
Core ≤ 10km {0,1} × EAC	0.420**	0.031	-0.110	0.062	-0.083	-0.019
	(0.175)	(0.043)	(0.330)	(0.041)	(0.058)	(0.062)
Core ≤ 10km {0,1} × CU	0.156	0.030	0.894***	-0.003	-0.008	0.046
	(0.108)	(0.029)	(0.249)	(0.035)	(0.125)	(0.062)
Core ≤ 10km {0,1} × CM	0.270*	-0.024	1.637***	0.052**	-0.087	-0.023
	(0.149)	(0.027)	(0.600)	(0.024)	(0.054)	(0.075)
(i) "Post-CM" Estimate						
(see Table C.51)						
Core {0,1} × EAC	0.422***	0.066*	0.121	0.064*	-0.065	-0.004
	(0.131)	(0.036)	(0.257)	(0.038)	(0.049)	(0.023)
Core {0,1} × CU	0.310***	0.035	0.755***	0.012	0.047	0.052
	(0.104)	(0.038)	(0.184)	(0.044)	(0.109)	(0.032)
Core {0,1} × CM 1[2010-2014]	0.365***	0.009	1.275***	0.080	-0.061	0.062
	(0.114)	(0.037)	(0.217)	(0.050)	(0.072)	(0.054)
Core {0,1} × post-CM 1[t ≥ 2015]	0.349**	0.001	0.636***	0.041	0.034	0.006
	(0.168)	(0.042)	(0.094)	(0.025)	(0.081)	(0.036)
(j) Excluding post-EAC Migrants						
(see Table C.52)						
Core {0,1} × EAC	0.605***	0.066	-0.104	0.110**	-0.037	0.027
	(0.145)	(0.045)	(0.266)	(0.044)	(0.064)	(0.046)
Core {0,1} × CU	0.375***	0.046	0.760***	0.037	0.034	0.137***
	(0.122)	(0.038)	(0.279)	(0.091)	(0.099)	(0.047)
Core {0,1} × CM	0.256**	-0.021	0.734***	0.062	0.023	0.116**
	(0.114)	(0.060)	(0.126)	(0.041)	(0.070)	(0.053)
(k) DHS Sample (excl. AIS, KAP, and MIS)						
(see Table C.55)						
Core {0,1} × EAC	0.433***	0.073*	-0.133	0.096***	-0.046	0.000
	(0.141)	(0.040)	(0.270)	(0.027)	(0.062)	(0.027)
Core {0,1} × CU	0.314***	0.037	0.665**	0.019	0.058	0.073*
	(0.105)	(0.039)	(0.267)	(0.059)	(0.111)	(0.042)
Core {0,1} × CM	0.361**	0.006	1.045***	0.076**	0.007	0.029
	(0.142)	(0.041)	(0.112)	(0.038)	(0.088)	(0.035)

Notes: This table offers an array of robustness tests on the main results of the paper for the difference-in-difference effect for individuals living in core agglomerations. As such, the results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. The full results (including border estimates) are given in the table referred to below the description of each test. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used, if not indicated otherwise. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table C.5: Robustness Checks Summarized (KHDS)

	Kagera Health and Development Survey (KHDS)					
	Dependent Variable					
Panel a)	Annual p.c. Consumpt. (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients						
(see Table 4.3)						
Core {0,1} × EAC 1|2004|	170.320***	706.607***	-0.013	0.026***	0.134***	0.891***
	(42.328)	(192.323)	(0.023)	(0.008)	(0.021)	(0.070)
Core {0,1} × CU 1|2010|	275.036***		-0.039		0.069	0.891***
	(57.368)		(0.025)		(0.067)	(0.099)
(a) Standard Errors clustered at EA-Level						
(see Table C.24)						
Core {0,1} × EAC 1|2004|	170.320*	706.607	-0.013	0.026	0.134***	0.891***
	(100.162)	(1002.964)	(0.033)	(0.018)	(0.021)	(0.126)
Core {0,1} × CU 1|2010|	275.036**		-0.039		0.069	0.891***
	(111.403)		(0.040)		(0.067)	(0.132)
(b) No Controls						
(see Table C.28)						
Core {0,1} × EAC 1|2004|	122.774***	559.391***	-0.012	0.024***	0.109***	1.112***
	(44.950)	(40.735)	(0.027)	(0.008)	(0.021)	(0.296)
Core {0,1} × CU 1|2010|	212.121***		-0.021***		0.035***	1.346***
	(71.187)		(0.031)		(0.079)	(0.461)
(c) Extended Geographic Controls						
(see Table C.32)						
Core {0,1} × EAC 1|2004|	172.061***	713.341***	-0.012	0.026***	0.143***	0.836***
	(39.808)	(69.683)	(0.024)	(0.009)	(0.025)	(0.030)
Core {0,1} × CU 1|2010|	269.142***		-0.036		0.080	0.846***
	(54.303)		(0.028)		(0.072)	(0.072)
(f) Urbanities vs. Core						
(see Table C.40)						
Urban {0,1} × EAC 1|2004|	180.332***	672.214***	-0.014	0.020**	0.128***	0.619***
	(50.703)	(126.314)	(0.026)	(0.009)	(0.024)	(0.140)
Urban {0,1} × CU 1|2010|	267.907***		-0.055*		0.080	0.578***
	(64.455)		(0.029)		(0.070)	(0.171)
(g) Core ≤ 25km						
(see Table C.44)						
Core ≤ 25km × EAC 1|2004|	170.821***	706.607***	-0.013	0.026***	0.132***	0.930***
	(42.732)	(192.323)	(0.023)	(0.008)	(0.020)	(0.102)
Core ≤ 25km {0,1} × CU 1|2010|	276.569***		-0.037		0.074	0.987***
	(58.235)		(0.026)		(0.068)	(0.148)
(h) Core ≤ 10km						
(see Table C.48)						
Core ≤ 10km {0,1} × EAC 1|2004|	166.469***	694.026***	-0.010	0.027***	0.110***	1.004***
	(42.819)	(149.664)	(0.023)	(0.009)	(0.015)	(0.135)
Core ≤ 10km {0,1} × CU 1|2010|	228.369***		-0.042*		0.009	1.279***
	(58.904)		(0.025)		(0.080)	(0.279)
(j) Excluding post-EAC Migrants						
(see Table C.53)						
Core {0,1} × EAC 1|2004|	119.373***	772.930***	0.009	0.017*	0.143***	0.930***
	(42.378)	(91.216)	(0.022)	(0.010)	(0.026)	(0.092)
Core {0,1} × CU 1|2010|	225.295***		-0.049*		0.141*	0.962***
	(46.662)		(0.026)		(0.074)	(0.102)
(m) Logged Dependent Variables						
(see Table C.56)						
Core {0,1} × EAC 1|2004|	0.216*	0.318	-	-	-	1.185***
	(0.114)	(0.349)				(0.025)
Core {0,1} × CU 1|2010|	0.306***		-		-	1.052***
	(0.113)					(0.032)

Notes: Table continued on next page.

Kagera Health and Development Survey (KHDS)						
	Dependent Variable					
Panel b)	Ann. Food Consumpt. (TZS '000)	Ann. non-Food Consumpt. (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients						
(see Table C.10)						
Core {0,1} × EAC 1|2004|	122.188***	35.341	4013.722***	-0.058	8.109	-1.079
	(38.913)	(36.287)	(1314.084)	(0.047)	(25.258)	(1.424)
Core {0,1} × CU 1|2010|	131.046***	133.815***		-0.008		-0.404
	(39.032)	(39.345)		(0.045)		(1.459)
(a) Standard Erros clustered at EA-Level						
(see Table C.25)						
Core {0,1} × EAC 1|2004|	122.188**	35.341	4013.722	-0.058	8.109	-1.079
	(49.625)	(57.589)	(5982.927)	(0.075)	(27.118)	(1.648)
Core {0,1} × CU 1|2010|	131.046***	133.815*		-0.008		-0.404
	(47.557)	(74.509)		(0.051)		(1.636)
(b) No Controls						
(see Table C.29)						
Core {0,1} × EAC 1|2004|	95.130***	20.114	3358.445***	-0.004	28.256	-0.435
	(33.049)	(28.938)	(574.014)	(0.026)	(21.074)	(1.073)
Core {0,1} × CU 1|2010|	109.186***	96.839***		0.065***		-0.479***
	(36.483)	(42.831)		(0.039)		(0.952)
(c) Extended Geographic Controls						
(see Table C.33)						
Core {0,1} × EAC 1|2004|	120.084***	37.350	3971.477***	-0.064	9.471	-1.209
	(35.210)	(36.610)	(577.551)	(0.041)	(25.328)	(1.489)
Core {0,1} × CU 1|2010|	127.440***	130.428***		-0.011		-0.689
	(37.672)	(35.080)		(0.047)		(1.486)
(d) Urbanities vs. Core						
(see Table C.41)						
Urban {0,1} × EAC 1|2004|	120.529***	46.673	3842.175***	-0.078*	8.754	-0.760
	(35.593)	(38.592)	(1205.886)	(0.040)	(25.455)	(1.370)
Urban {0,1} × CU 1|2010|	142.787***	114.511***		-0.012		-0.299
	(42.518)	(38.933)		(0.045)		(1.513)
(e) Core ≤ 25km						
(see Table C.45)						
Core ≤ 25km × EAC 1|2004|	122.246***	35.774	4013.722***	-0.058	8.109	-0.995
	(38.892)	(36.204)	(1314.084)	(0.047)	(25.258)	(1.439)
Core ≤ 25km {0,1} × CU 1|2010|	131.142***	135.227***		-0.007		-0.351
	(39.002)	(39.843)		(0.044)		(1.459)
(f) Core ≤ 10km						
(see Table C.49)						
Core ≤ 10km {0,1} × EAC 1|2004|	117.794***	36.303	3592.924	-0.058	7.206	-0.380
	(34.197)	(45.837)	(2577.006)	(0.042)	(19.772)	(1.427)
Core ≤ 10km {0,1} × CU 1|2010|	116.420***	101.922**		-0.001		-0.116
	(36.746)	(42.044)		(0.037)		(1.467)
(j) Excluding post-EAC Migrants						
(see Table C.54)						
Core {0,1} × EAC 1|2004|	106.040***	0.311	4599.535***	-0.077	9.488	-0.456
	(36.417)	(16.952)	(1202.895)	(0.071)	(25.859)	(2.988)
Core {0,1} × CU 1|2010|	140.531***	74.285**		0.009		-0.171
	(35.606)	(36.188)		(0.047)		(2.939)
(m) Logged Values						
(see Table C.57)						
Core {0,1} × EAC 1|2004|	0.271*	0.025	0.547	-0.058	0.269	-1.079
	(0.140)	(0.139)	(0.566)	(0.047)	(0.495)	(1.424)
Core {0,1} × CU 1|2010|	0.300**	0.154		-0.008		-0.404
	(0.126)	(0.180)		(0.045)		(1.459)

Notes: This table offers an array of robustness tests on the main results of the paper for the difference-in-difference effect for individuals living in core agglomerations. As such, the results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. The full results (including border estimates) are given in the table referred to below the description of each test. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used, if not indicated otherwise. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table C.6: Robustness Checks Summarized (Afrobarometer)

	Afrobarometer					
	Dependent Variable					
Panel a)	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked Last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients						
(see Tables 4.2 and C.9)						
Border [0,1] × EAC	-0.083	0.050	0.098	0.072	0.017	-0.350
	(0.296)	(0.058)	(0.071)	(0.077)	(0.134)	(0.294)
Border [0,1] × CU	0.172	0.060	0.064	0.091	0.040	-0.234
	(0.292)	(0.065)	(0.081)	(0.073)	(0.119)	(0.312)
Border [0,1] × CM	0.142	0.098**	0.119	-0.001	0.289***	-0.456
	(0.282)	(0.039)	(0.101)	(0.076)	(0.087)	(0.291)
(a) Beeline Distance to Borderpost						
(see Table C.58)						
(Beeline to) Border [0,1] × EAC	0.096	0.022	0.080	0.012	-0.052	-0.421
	(0.295)	(0.054)	(0.062)	(0.068)	(0.131)	(0.325)
(Beeline to) Border [0,1] × CU	0.201	0.097	0.177	0.113	0.127	-0.211
	(0.327)	(0.063)	(0.124)	(0.082)	(0.107)	(0.346)
(Beeline to) Border [0,1] × CM	0.221	0.063	0.264	-0.051	0.256***	-0.433
	(0.259)	(0.041)	(0.168)	(0.075)	(0.088)	(0.288)
(b) Beeline Distance to Borderline						
(see Table C.62)						
(Beeline to) Borderline [0,1] × EAC	-0.045	0.062	0.098*	0.078	0.054	-0.402
	(0.282)	(0.058)	(0.060)	(0.077)	(0.129)	(0.297)
(Beeline to) Borderline [0,1] × CU	0.144	0.051	0.075	0.072	0.025	-0.235
	(0.287)	(0.059)	(0.079)	(0.071)	(0.108)	(0.309)
(Beeline to) Borderline [0,1] × CM	0.213	0.074**	0.131	-0.054	0.262***	-0.432
	(0.258)	(0.037)	(0.106)	(0.071)	(0.080)	(0.284)
(c) Logged Distance (+1)						
(see Table C.66)						
Log(1+BorderDistance) × EAC	0.029	-0.006	-0.018	-0.006	0.015	0.091
	(0.071)	(0.014)	(0.022)	(0.019)	(0.031)	(0.069)
Log(1+BorderDistance) × CU	-0.033	-0.023	-0.014	-0.027	-0.015	0.101
	(0.073)	(0.018)	(0.032)	(0.023)	(0.031)	(0.074)
Log(1+BorderDistance) × CM	-0.009	-0.027***	-0.032	0.007	-0.068***	0.143**
	(0.067)	(0.010)	(0.039)	(0.021)	(0.020)	(0.066)
(d) Border Dummy ≤ 100km						
(see Table C.70)						
Border ≤ 100km {0,1} × EAC	-0.137	0.002	-0.022	0.011	0.007	-0.125
	(0.135)	(0.023)	(0.041)	(0.039)	(0.066)	(0.165)
Border ≤ 100km {0,1} × CU	-0.019	0.034	-0.070	0.075	0.116*	-0.179
	(0.156)	(0.035)	(0.050)	(0.051)	(0.065)	(0.169)
Border ≤ 100km {0,1} × CM	-0.032	0.025	-0.035	-0.017	0.153***	-0.201
	(0.131)	(0.022)	(0.056)	(0.048)	(0.058)	(0.153)
(e) Border Dummy ≤ 50km						
(see Table C.74)						
Border ≤ 50km {0,1} × EAC	-0.190	0.029	0.047*	0.009	-0.093	0.078
	(0.149)	(0.024)	(0.026)	(0.032)	(0.061)	(0.116)
Border ≤ 50km {0,1} × CU	-0.035	0.095***	0.005	0.050	-0.019	0.039
	(0.122)	(0.032)	(0.039)	(0.058)	(0.096)	(0.146)
Border ≤ 50km {0,1} × CM	-0.177	0.077***	0.050*	0.012	0.091	-0.129
	(0.145)	(0.025)	(0.029)	(0.058)	(0.056)	(0.127)
(f) Border Dummy ≤ 25km						
(see Table C.78)						
Border ≤ 25km {0,1} × EAC	-0.478**	-0.020	0.039	-0.034	-0.176	-0.381
	(0.225)	(0.048)	(0.039)	(0.054)	(0.151)	(0.288)
Border ≤ 25km {0,1} × CU	-0.029	0.150***	-0.002	0.093	0.007	-0.324
	(0.297)	(0.050)	(0.049)	(0.066)	(0.156)	(0.269)
Border ≤ 25km {0,1} × CM	-0.215	0.034	-0.019	-0.097	-0.083	-0.380***
	(0.255)	(0.043)	(0.055)	(0.079)	(0.107)	(0.146)

Notes: This table offers robustness tests on the main results of the paper for the difference-in-difference effect with treatment intensity increasing towards the border. The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. The full results (including border estimates) are given in the table referred to below the description of each test.

Table C.7: Robustness Checks Summarized (DHS)

	Demographic and Health Survey (DHS)					
	Dependent Variable					
Panel a)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked Last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients						
(see Tables 4.2 and C.9)						
Border [0,1] × EAC	-0.284	-0.129	0.104	0.114*	-0.184	-0.162
	(0.312)	(0.089)	(0.136)	(0.059)	(0.188)	(0.136)
Border [0,1] × CU	-0.372*	-0.138**	0.084	0.118***	-0.291***	-0.271***
	(0.224)	(0.054)	(0.089)	(0.043)	(0.106)	(0.083)
Border [0,1] × CM	-0.547***	-0.034	0.060	0.081*	-0.149**	-0.141
	(0.202)	(0.040)	(0.095)	(0.046)	(0.075)	(0.092)
(a) Beeline Distance to Borderpost						
(see Table C.59)						
(Beeline to) Border [0,1] × EAC	-0.160	-0.141	-0.042	0.076	-0.179	-0.131
	(0.312)	(0.091)	(0.139)	(0.054)	(0.194)	(0.135)
(Beeline to) Border [0,1] × CU	-0.277	-0.135***	0.067	0.121***	-0.207*	-0.141
	(0.265)	(0.045)	(0.143)	(0.043)	(0.118)	(0.101)
(Beeline to) Border [0,1] × CM	-0.550**	-0.066**	-0.027	0.083*	-0.178**	-0.182**
	(0.232)	(0.033)	(0.138)	(0.045)	(0.071)	(0.092)
(b) Beeline Distance to Borderline						
(see Table C.63)						
(Beeline to) Borderline [0,1] × EAC	-0.248	-0.152**	0.026	0.079	-0.168	-0.152
	(0.296)	(0.077)	(0.132)	(0.055)	(0.172)	(0.123)
(Beeline to) Borderline [0,1] × CU	-0.368	-0.106**	0.084	0.112**	-0.211*	-0.220***
	(0.235)	(0.049)	(0.107)	(0.044)	(0.108)	(0.084)
(Beeline to) Borderline [0,1] × CM	-0.559***	-0.024	-0.005	0.063	-0.102	-0.131
	(0.199)	(0.036)	(0.109)	(0.045)	(0.064)	(0.083)
(c) Logged Distance (+1)						
(see Table C.67)						
Log(1+BorderDistance) × EAC	0.029	0.034**	-0.077*	-0.028*	0.018	0.046*
	(0.076)	(0.016)	(0.041)	(0.016)	(0.035)	(0.026)
Log(1+BorderDistance) × CU	0.083	0.040***	-0.053*	-0.024**	0.048	0.075***
	(0.057)	(0.013)	(0.032)	(0.011)	(0.030)	(0.023)
Log(1+BorderDistance) × CM	0.120**	0.016	-0.066*	-0.015	0.035**	0.048*
	(0.050)	(0.010)	(0.039)	(0.012)	(0.018)	(0.025)
(d) Border Dummy ≤ 100km						
(see Table C.71)						
Border ≤ 100km {0,1} × EAC	-0.421	-0.134***	-0.146*	0.023	-0.181***	-0.204***
	(0.261)	(0.035)	(0.084)	(0.038)	(0.063)	(0.025)
Border ≤ 100km {0,1} × CU	-0.345**	-0.099***	-0.051	0.038	-0.144***	-0.177***
	(0.167)	(0.023)	(0.058)	(0.023)	(0.048)	(0.034)
Border ≤ 100km {0,1} × CM	-0.379**	-0.065***	-0.056	0.024	-0.133***	-0.159***
	(0.150)	(0.015)	(0.056)	(0.025)	(0.034)	(0.049)
(e) Border Dummy ≤ 50km						
(see Table C.75)						
Border ≤ 50km {0,1} × EAC	0.040	-0.119**	-0.055	0.037	-0.086	-0.093***
	(0.214)	(0.050)	(0.078)	(0.041)	(0.061)	(0.033)
Border ≤ 50km {0,1} × CU	-0.048	-0.135***	0.038	0.010	-0.104**	-0.145***
	(0.154)	(0.034)	(0.064)	(0.028)	(0.047)	(0.043)
Border ≤ 50km {0,1} × CM	-0.092	-0.093***	0.013	0.020	-0.117***	-0.121*
	(0.172)	(0.024)	(0.045)	(0.026)	(0.023)	(0.063)
(f) Border Dummy ≤ 25km						
(see Table C.79)						
Border ≤ 25km {0,1} × EAC	0.548***	-0.051***	-0.083	0.090*	-0.010	-0.064
	(0.121)	(0.018)	(0.088)	(0.049)	(0.072)	(0.044)
Border ≤ 25km {0,1} × CU	0.261***	-0.068***	-0.004	0.039	-0.025	-0.158***
	(0.074)	(0.025)	(0.110)	(0.075)	(0.090)	(0.044)
Border ≤ 25km {0,1} × CM	0.244	-0.031	-0.036	0.066	-0.063	-0.088**
	(0.167)	(0.026)	(0.092)	(0.043)	(0.062)	(0.038)

Notes: This table offers robustness tests on the main results of the paper for the difference-in-difference effect with treatment intensity increasing towards the border. The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. The full results (including border estimates) are given in the table referred to below the description of each test.

Table C.8: Robustness Checks Summarized (KHDS)

	Kagera Health and Development Survey (KHDS)					
	Dependent Variable					
Panel a)	Annual p.c. Consumpt. (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients **(see Table 4.3)**						
Border [0,1] × EAC	-853.376	187.152	0.581*	0.071	0.107	-0.848**
	(1072.010)	(754.481)	(0.331)	(0.049)	(0.261)	(0.384)
Border [0,1] × CU	-1317.845		0.432		-0.010	-1.631***
	(1154.386)		(0.305)		(0.404)	(0.586)
(a) Beeline Distance to Borderpost **(see Table C.60)**						
(Beeline to) Border [0,1] × EAC	-816.913	69.400	0.563**	0.051	0.078	-0.710**
	(1112.207)	(715.743)	(0.279)	(0.044)	(0.238)	(0.312)
(Beeline to) Border [0,1] × CU	-1283.327		0.435		-0.033	-1.567***
	(1208.457)		(0.278)		(0.388)	(0.545)
(b) Beeline Distance to Borderline **(see Table C.64)**						
(Beeline to) Borderline [0,1] × EAC	-757.039	122.074	0.490*	0.041	0.062	-0.570*
	(998.849)	(577.684)	(0.261)	(0.044)	(0.234)	(0.318)
(Beeline to) Borderline [0,1] × CU	-1199.598		0.389		-0.043	-1.360***
	(1090.390)		(0.255)		(0.374)	(0.500)
(c) Logged Distance (+1) **(see Table C.68)**						
Log(1+Borderdistance) × EAC	-14.571	-305.802	-0.051	-0.020*	-0.071	-0.404*
	(96.504)	(499.702)	(0.043)	(0.011)	(0.052)	(0.225)
Log(1+Borderdistance) × CU	24.227		-0.033		-0.040	-0.330
	(115.177)		(0.038)		(0.071)	(0.221)
(d) Border Dummy ≤ 100km **(see Table C.72)**						
Border ≤ 100km {0,1} × EAC	-43.826	21.219	0.033	0.009	0.017	0.038
	(85.443)	(83.145)	(0.042)	(0.010)	(0.045)	(0.063)
Border ≤ 100km {0,1} × CU	-105.265		0.022		0.127	-0.085
	(108.035)		(0.041)		(0.089)	(0.102)
(e) Border Dummy ≤ 50km **(see Table C.76)**						
Border ≤ 50km {0,1} × EAC	49.969	-2.821	-0.004	0.035***	0.047	-0.143*
	(292.688)	(77.036)	(0.035)	(0.009)	(0.051)	(0.078)
Border ≤ 50km {0,1} × CU	-13.598		-0.033		-0.022	-0.164*
	(306.353)		(0.033)		(0.149)	(0.096)
(f) Border Dummy ≤ 25km **(see Table C.80)**						
Border ≤ 25km {0,1} × EAC	1046.331***	-242.913	-0.162	0.010	0.188	0.034
	(310.431)	(436.378)	(0.122)	(0.030)	(0.210)	(0.066)
Border ≤ 25km {0,1}× CU	1126.525***		-0.037		0.583*	-0.080
	(354.396)		(0.059)		(0.312)	(0.128)

Notes: Table continued on next page.

	Kagera Health and Development Survey (KHDS)					
	Dependent Variable					
Panel b)	Ann. Food Consumpt. (TZS '000)	Ann. non-Food Consumpt. (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients						
(see Table C.10)						
Border [0,1] × EAC	-177.749	-670.444	7592.679	-0.201	-96.031	5.949
	(237.164)	(1054.004)	(14152.494)	(0.377)	(137.206)	(9.297)
Border [0,1] × CU	-388.178	-911.118		-0.233		6.555
	(236.410)	(1116.362)		(0.344)		(9.162)
(a) Beeline Distance to Borderpost						
(see Table C.61)						
(Beeline to) Border [0,1] × EAC	-111.273	-697.103	4638.227	-0.180	-106.417	3.003
	(210.184)	(1104.755)	(12272.383)	(0.339)	(145.479)	(8.924)
(Beeline to) Border [0,1] × CU	-335.069	-924.540		-0.234		3.458
	(212.127)	(1172.534)		(0.319)		(8.728)
(b) Beeline Distance to Borderline						
(see Table C.65)						
(Beeline to) Borderline [0,1] × EAC	-107.710	-638.292	4075.452	-0.135	-92.091	4.749
	(207.271)	(981.794)	(11014.623)	(0.327)	(133.728)	(8.546)
(Beeline to) Borderline [0,1] × CU	-323.756	-853.098		-0.203		5.328
	(206.881)	(1047.819)		(0.302)		(8.284)
(c) Logged Distance (+1)						
(see Table C.69)						
Log(1+Borderdistance) × EAC	-22.602	13.234	-3084.228	0.014	9.795	-1.344
	(56.181)	(62.018)	(3506.668)	(0.060)	(19.092)	(1.748)
Log(1+Borderdistance) × CU	17.695	8.645		0.025		-1.897
	(62.345)	(68.685)		(0.041)		(1.590)
(d) Border Dummy ≤ 100km						
(see Table C.73)						
Border ≤ 100km {0,1} × EAC	-5.045	-38.379	1207.398	-0.019	-8.716	2.250
	(46.765)	(67.330)	(1921.629)	(0.068)	(26.978)	(1.438)
Border ≤ 100km {0,1} × CU	-36.354	-68.217		-0.040		2.278*
	(50.035)	(86.151)		(0.064)		(1.312)
(e) Border Dummy ≤ 50km						
(see Table C.77)						
Border ≤ 50km {0,1} × EAC	0.159	46.895	1521.906	0.052	13.350	-2.628
	(186.972)	(107.532)	(1408.312)	(0.089)	(24.030)	(2.832)
Border ≤ 50km {0,1} × CU	-33.556	20.242		0.016		-2.064
	(215.623)	(97.817)		(0.058)		(2.874)
(f) Border Dummy ≤ 25km						
(see Table C.81)						
Border ≤ 25km {0,1} × EAC	536.903***	493.731***	3041.987	-0.033	-487.518	-0.755
	(166.385)	(181.678)	(3654.443)	(0.139)	(5036999.421)	(2.051)
Border ≤ 25km {0,1}× CU	663.445***	452.421***		-0.104		0.380
	(233.953)	(150.728)		(0.121)		(2.099)

Notes: This table offers robustness tests on the main results of the paper for the difference-in-difference effect with treatment intensity increasing towards the border. The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. The full results (including border estimates) are given in the table referred to below the description of each test.

Table C.9: Further Labor Market Outcomes (AFB and DHS)

	Dependent Variable					
	Afrobarometer			DHS		
	Income					
	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[0.55]	[1.66]	[2.09]	[0.76]	[1.51]	[0.52]
Border [0,1] × EAC 1[2002-2004]	0.072	0.017	-0.350	0.114*	-0.184	-0.162
	(0.077)	(0.134)	(0.294)	(0.059)	(0.188)	(0.136)
Border [0,1] × CU 1[2005-2009]	0.091	0.040	-0.234	0.118***	-0.291***	-0.271***
	(0.073)	(0.119)	(0.312)	(0.043)	(0.106)	(0.083)
Border [0,1] × CM 1[t ≥ 2010]	-0.001	0.289***	-0.456	0.081*	-0.149**	-0.141
	(0.076)	(0.087)	(0.291)	(0.046)	(0.075)	(0.092)
Core {0,1} × EAC 1[2002-2004]	0.026	0.021	0.041	0.067*	-0.046	-0.002
	(0.067)	(0.079)	(0.083)	(0.038)	(0.062)	(0.026)
Core {0,1} × CU 1[2005-2009]	0.002	0.057	-0.234**	0.014	0.058	0.053
	(0.055)	(0.090)	(0.105)	(0.043)	(0.111)	(0.033)
Core {0,1} × CM 1[t ≥ 2010]	-0.258**	-0.009	0.043	0.061*	0.007	0.029
	(0.110)	(0.087)	(0.079)	(0.031)	(0.088)	(0.034)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	26,563	21,232	35,975	169,875	95,717	98,963
R-Squared	0.23	0.25	0.09	0.23	0.26	0.17
R-Squared -Within	0.23	0.25	0.09	0.23	0.26	0.17

Notes: The results in each column are produced by a separate regression. The sample mean of the respective dependent variable is given in brackets above the estimates. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 4.2.

Table C.10: Further Results (KHDS)

	Dependent Variable					
	Kagera Health and Development Survey (KHDS)					
	Consumpt.		Income			Agglom.
	Ann. Food Consumpt. (TZS '000)	Ann. non-Food Consumpt. (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
Border [0,1] × EAC 1[2004]	-177.749	-670.444	7592.679	-0.201	-96.031	5.949
	(237.164)	(1054.004)	(14152.494)	(0.377)	(137.206)	(9.297)
Border [0,1] × CU 1[2010]	-388.178	-911.118		-0.233		6.555
	(236.410)	(1116.362)		(0.344)		(9.162)
Core {0,1} × EAC 1[2004]	122.188***	35.341	4013.722***	-0.058	8.109	-1.079
	(38.913)	(36.287)	(1314.084)	(0.047)	(25.258)	(1.424)
Core {0,1} × CU 1[2010]	131.046***	133.815***		-0.008		-0.404
	(39.032)	(39.345)		(0.045)		(1.459)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,494	5,524	2,695	16,330	1,782	5,411
Observations - Fixed Effects	3,817	3,830	2,363	11,599	1,190	3,763
R-Squared	0.88	0.83	0.69	0.86	0.94	0.89
R-Squared -Within	0.10	0.09	0.07	0.16	0.18	0.09

Notes: The results in each column are produced by a separate regression. The sample mean of the respective dependent variable is given in brackets above the estimates. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 4.3.

Table C.11: Urbanities in Border Regions – Triple Difference Estimates

	Dependent Variable					
	Afrobarometer			DHS		
	Consumpt.	Income	Agglom.	Consumpt.	Income	Agglom.
Panel a)	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[3.14]	[0.19]	[0.00]
Border [0,1] × EAC	0.032	0.093**	-0.045	-0.159	-0.030	-0.048
	(0.300)	(0.043)	(0.088)	(0.236)	(0.032)	(0.070)
Triple Interaction						
Border [0,1] × EAC × Urban {0,1}	0.498	-0.147	0.509	-0.356	-0.004	0.649*
	(0.368)	(0.092)	(0.448)	(0.313)	(0.071)	(0.343)
Isolated Effect of the EAC on Urbanities in Border Regions						
Combined Effect:						
Border [0,1] × EAC + Triple Interaction	0.530	-0.054	0.464	-0.515**	-0.035	0.601*
	[0.15]	[0.55]	[0.25]	[0.02]	[0.65]	[0.07]
Observations	36,042	25,465	4,156	183,250	71,738	7,692
R-Squared	0.13	0.15	0.29	0.46	0.18	0.34
R-Squared -Within	0.11	0.11	0.27	0.45	0.13	0.33

	Afrobarometer			DHS		
	Income					
Panel b)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[0.55]	[1.66]	[1.12]	[0.76]	[1.51]	[0.52]
Border [0,1] × EAC	0.054	0.207**	-0.378	0.074	-0.045	-0.078
	(0.064)	(0.099)	(0.321)	(0.047)	(0.047)	(0.086)
Triple Interaction						
Border [0,1] × EAC × Urban {0,1}	-0.146	-0.383	0.110	0.058	-0.273*	-0.221**
	(0.147)	(0.284)	(0.349)	(0.096)	(0.141)	(0.103)
Isolated Effect of the EAC on Urbanities in Border Regions						
Combined Effect:						
Border [0,1] × EAC + Triple Interaction	-0.092	-0.177	-0.268	0.131	-0.318**	-0.299***
	[0.53]	[0.51]	[0.39]	[0.13]	[0.02]	[0.00]
Observations	26,563	21,232	35,975	169,875	95,717	98,963
R-Squared	0.23	0.29	0.10	0.23	0.32	0.21
R-Squared -Within	0.08	0.26	0.07	0.20	0.24	0.12
Individual Controls	YES	YES	NO/YES	YES	YES	NO/YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table analyzes the differential effect of Border [0,1] for individuals living in urban regions. Row one shows the uninteracted effect of Border [0,1], i.e. effect of the EAC for individuals at rural border regions, row two shows the differential effect for being in an urban area. Row three depicts the combined effect of the two constituent terms, i.e. the effect of Border [0,1] for individuals in urban regions. The results in each column and panel are produced by a separate regression. The sample mean of the respective dependent variable is given in brackets above the estimates. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 4.2.

Table C.12: Placebo Test – Disaggregated DiD in contiguous Countries

	Dependent Variable					
	Afrobarometer			DHS		
	Consumpt.	Income	Agglom.	Consumpt.	Income	Agglom.
Panel a)	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.25]	[0.24]	[0.00]	[3.17]	[0.17]	[0.00]
Border [0,1] × EAC 1[2002-2004	0.274*	-0.030	-0.199	0.750***	-0.049	-0.467**
	(0.161)	(0.084)	(0.222)	(0.161)	(0.073)	(0.205)
Border [0,1] × CU 1[2005-2009]	0.036	0.165**	-0.120	1.349***	0.080	0.181
	(0.133)	(0.082)	(0.150)	(0.178)	(0.058)	(0.205)
Border [0,1] × CM 1[t ≥ 2010]	0.179	0.056	0.063	0.740***	0.000	-0.165
	(0.143)	(0.076)	(0.164)	(0.140)	(0.037)	(0.120)
Core {0,1} × EAC 1[2002-2004]	-0.114	-0.033	0.580*	-0.757**	-0.245*	-1.551*
	(0.145)	(0.042)	(0.310)	(0.331)	(0.135)	(0.844)
Core {0,1} × CU 1[2005-2009]	-0.025	-0.008	0.482*	-0.004	-0.076	-0.411
	(0.069)	(0.036)	(0.261)	(0.180)	(0.116)	(0.547)
Core {0,1} × CM 1[t ≥ 2010]	-0.159	-0.056	0.942***	0.214	-0.053	0.123
	(0.109)	(0.048)	(0.320)	(0.175)	(0.073)	(0.304)
Observations	28,541	18,994	2,329	200,133	126,587	7,819
R-Squared	0.09	0.19	0.24	0.29	0.14	0.28
R-Squared -Within	0.06	0.15	0.22	0.28	0.14	0.27
	Afrobarometer			DHS		
	Income					
Panel b)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[0.49]	[1.62]	[2.12]	[0.67]	[1.48]	[0.44]
Border [0,1] × EAC 1[2002-2004	-0.117	0.001	-0.575***	-0.058	-0.058	-0.007
	(0.090)	(0.170)	(0.220)	(0.127)	(0.210)	(0.250)
Border [0,1] × CU 1[2005-2009]	0.107	-0.012	-0.839***	-0.063	0.076	-0.032
	(0.081)	(0.156)	(0.149)	(0.073)	(0.108)	(0.126)
Border [0,1] × CM 1[t ≥ 2010]	0.037	0.058	-0.462***	-0.014	-0.081	-0.120
	(0.081)	(0.133)	(0.155)	(0.080)	(0.107)	(0.121)
Core {0,1} × EAC 1[2002-2004]	0.034	0.028	-0.365**	0.130**	-0.268**	-0.246
	(0.058)	(0.059)	(0.169)	(0.057)	(0.129)	(0.160)
Core {0,1} × CU 1[2005-2009]	0.093**	0.001	-0.448***	0.062	0.013	0.016
	(0.041)	(0.076)	(0.044)	(0.060)	(0.073)	(0.096)
Core {0,1} × CM 1[t ≥ 2010]	-0.106	-0.040	-0.320***	0.094**	0.075	0.064
	(0.070)	(0.085)	(0.082)	(0.047)	(0.070)	(0.094)
Observations	19,191	12,715	28,274	286,651	187,498	188,349
R-Squared	0.18	0.31	0.11	0.17	0.24	0.17
R-Squared -Within	0.18	0.31	0.11	0.17	0.24	0.17
Individual Controls	YES	YES	NO/YES	YES	YES	NO/YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table conducts a "placebo" analysis by testing for a spatially differentiated effect across contiguous, non-EAC countries within the time frame of the EAC's establishment and expansion. Border [0,1] × EAC 1[2002-2004] switches to one for individuals sampled from 2002 to and including 2004, CU 1[2005-2009] for individuals sampled from 2005 and including 2009, and CM 1[t ≥ 2010] for individuals sampled from 2010 onwards. See full notes in Table 4.4.

Table C.13: Region-Based Estimates (Placebo Countries)

	Dependent Variable					
	DHS (Region-based)					
	Consumpt.	Income	Agglom.	Income		
Panel a)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked Last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.20]	[0.17]	[0.00]	[0.67]	[1.48]	[0.44]
Border Region {0,1} × EAC 1[t ≥ 2002]	0.230*	-0.004	0.001	-0.022	-0.008	-0.087*
EAC 0[1992-2000]	(0.111)	(0.013)	(0.141)	(0.016)	(0.047)	(0.048)
Core Region {0,1} × EAC 1[t ≥ 2002]	-0.159	-0.056	1.066	0.099**	-0.086	-0.072
EAC 0[1992-2000]	(0.171)	(0.034)	(0.685)	(0.039)	(0.066)	(0.059)
Observations	282,480	167,590	11,391	348,016	225,336	223,908
R-Squared	0.27	0.15	0.37	0.15	0.27	0.20
R-Squared -Within	0.25	0.14	0.35	0.13	0.23	0.13
Panel b)						
Border Region {0,1} × pre-EAC 1[1996-2000]	-0.225	-0.004	0.041	-	0.177	0.014
pre-EAC 0[1992-1995]	(0.152)	(0.020)	(0.128)		(0.130)	(0.062)
Border Region {0,1} × EAC 1[t ≥ 2002]	0.041	-0.009	0.035	-0.022	0.139	-0.077
EAC 0[1992-1995]	(0.217)	(0.017)	(0.202)	(0.016)	(0.116)	(0.045)
Core Region {0,1} × pre-EAC 1[1996-2000]	0.346	0.088	0.780	-0.107**	0.211**	0.188**
pre-EAC 0[1992-1995]	(0.308)	(0.072)	(0.644)	(0.038)	(0.096)	(0.088)
Core Region {0,1} × EAC 1[t ≥ 2002]	0.107	0.011	1.608*	-0.008	0.072	0.078
EAC 0[1992-1995]	(0.260)	(0.050)	(0.868)	(0.025)	(0.049)	(0.085)
Observations	282,480	167,590	11,391	348,016	225,336	223,908
R-Squared	0.27	0.15	0.38	0.15	0.27	0.20
R-Squared -Within	0.26	0.14	0.35	0.13	0.23	0.13
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table conducts a “placebo” analysis by testing for a spatially differentiated effect across contiguous, non-EAC countries within the time frame of the EAC's establishment and expansion. Specifically, the analysis makes use of the non-GPS survey rounds of the Demographic and Health Surveys (DHS) sampled before 1999 and additionally conducts “pre-tests” towards the difference-in-differences approach. The data thereby come from the full sample of Ethiopia, Malawi, Mozambique, Rwanda and Zambia DHS surveys sampled between 1992 and 2019, making use of AIS, KAP and MIS rounds as well. The sample mean of the respective dependent variable is given in brackets above the estimates. Border Region {0,1} switches to one for individuals living in a region with a median road distance to the nearest border crossing of a contiguous EAC country below the 10th percentile of all (within-country) GPS-border distances in the sample. Core Region {0,1} is a dummy indicating individuals living in the region which hosts the core agglomeration of their respective country (i.e. Addis Abeba, Kigali, Lilongwe, Lusaka, Maputo). EAC 1[t ≥ 2002] switches to one for individuals sampled from 2002 onwards. Pre-EAC 1[1996-2000] switches to one for individuals sampled in survey years between 1996 and including 2000. As such, in panel a), the reference group of the estimates are comprised of individuals sampled in the full pre-EAC period, i.e. from 1991 to 2000, while in panel b), the reference group is formed by individuals sampled between 1991 and including 1995. Hence, the DiD estimate on “pre-EAC” in panel b) represents the pre-test. The results in each column and panel are produced by a separate regression. All regressions include individual-level controls for respondents’ age, gender, as well as education. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for clustering at the region level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table C.14: Disaggregated Region-Based Estimates

	Consumpt.	Income	Agglom.	Income		
	Dependent Variable DHS (Region-based)					
Panel a)	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked Last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.12]	[0.19]	[0.00]	[0.76]	[1.52]	[0.53]
Border Region {0,1} × EAC 1[2002-2004]	0.113	-0.037	-0.050	-0.091	-0.051	-0.125**
EAC 0[1988-2000]	(0.199)	(0.024)	(0.145)	(0.092)	(0.054)	(0.058)
Border Region {0,1} × CU 1[2005-2009]	0.207	-0.011	-0.215*	-0.120**	0.109	-0.009
CU 0[1988-2000]	(0.218)	(0.029)	(0.125)	(0.054)	(0.077)	(0.070)
Border Region {0,1} × CM 1[t ≥ 2010]	-0.044	-0.035	-0.086	-0.068	-0.063	-0.155**
CM 0[1988-2000]	(0.178)	(0.032)	(0.144)	(0.058)	(0.065)	(0.064)
Core Region {0,1} × EAC 1[2002-2004]	0.633***	0.114***	3.017***	-0.064	0.101***	-0.005
EAC 0[1988-2000]	(0.116)	(0.022)	(0.517)	(0.065)	(0.032)	(0.075)
Core Region {0,1} × CU 1[2005-2009]	0.399***	0.100***	1.524	-0.094	0.162	0.060**
CU 0[1988-2000]	(0.131)	(0.026)	(1.395)	(0.063)	(0.095)	(0.025)
Core Region {0,1} × CM 1[t ≥ 2010]	0.316**	0.077***	2.195	-0.006	0.062	-0.024
CM 0[1988-2000]	(0.139)	(0.024)	(1.581)	(0.047)	(0.044)	(0.029)
Observations	258,820	104,440	11,841	236,646	140,613	136,163
R-Squared	0.28	0.16	0.55	0.21	0.25	0.18
R-Squared -Within	0.27	0.10	0.53	0.18	0.16	0.08
Panel b)						
Border Region {0,1} × pre-EAC 1[1996-2000]	0.098	0.024	0.084	-0.296**	0.014	-0.063
pre-EAC 0[1988-1995]	(0.125)	(0.029)	(0.082)	(0.121)	(0.060)	(0.049)
Border Region {0,1} × EAC 1[2002-2004]	0.181	-0.022	-0.014	-0.314**	-0.041	-0.163***
EAC 0[1988-1995]	(0.220)	(0.029)	(0.132)	(0.150)	(0.070)	(0.059)
Border Region {0,1} × CU 1[2005-2009]	0.230	-0.001	-0.192	-0.307**	0.009	-0.173**
CU 0[1988-1995]	(0.249)	(0.043)	(0.118)	(0.129)	(0.112)	(0.084)
Border Region {0,1} × CM 1[t ≥ 2010]	-0.003	-0.027	-0.044	-0.290**	-0.058	-0.190**
CM 0[1988-1995]	(0.213)	(0.039)	(0.123)	(0.133)	(0.081)	(0.071)
Core Region {0,1} × pre-EAC 1[1996-2000]	0.440*	0.022	0.131	-0.245*	0.099	-0.185***
pre-EAC 0[1988-1995]	(0.223)	(0.026)	(0.262)	(0.121)	(0.102)	(0.052)
Core Region {0,1} × EAC 1[2002-2004]	0.899***	0.127***	3.069***	-0.238	0.158*	-0.096
EAC 0[1988-1995]	(0.214)	(0.019)	(0.582)	(0.143)	(0.082)	(0.082)
Core Region {0,1} × CU 1[2005-2009]	0.674***	0.110***	2.193	-0.250*	0.232*	-0.016
CU 0[1988-1995]	(0.206)	(0.028)	(1.653)	(0.146)	(0.117)	(0.042)
Core Region {0,1} × CM 1[t ≥ 2010]	0.570**	0.087***	2.020	-0.180	0.099	-0.141***
CM 0[1988-1995]	(0.241)	(0.027)	(1.590)	(0.129)	(0.076)	(0.034)
Observations	255,844	102,005	11,702	232,864	137,542	133,083
R-Squared	0.29	0.16	0.54	0.21	0.25	0.17
R-Squared -Within	0.27	0.10	0.53	0.18	0.16	0.08
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table “disaggregates” the analysis of Table 4.5. See full notes below Table 4.5.

Table C.15: Disaggregated Region-Based Estimates (Placebo Countries)

		Dependent Variable					
		DHS (Region-based)					
		Consumpt.	Income	Agglom.		Income	
		Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked Last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
Panel a)		(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.		[3.20]	[0.17]	[0.00]	[0.67]	[1.48]	[0.44]
Border Region {0,1} ×	EAC 1[2002-2004]	0.306**	0.005	0.031	-0.045	-0.022	-0.057
	EAC 0[1992-2000]	(0.107)	(0.018)	(0.193)	(0.062)	(0.051)	(0.068)
Border Region {0,1} ×	CU 1[2005-2009]	0.252	0.010	0.030	-0.023	0.007	-0.051
	CU 0[1992-2000]	(0.176)	(0.030)	(0.100)	(0.027)	(0.042)	(0.047)
Border Region {0,1} ×	CM 1[t ≥ 2010]	0.218	-0.007	-0.012	-0.019	-0.004	-0.102*
	CM 0[1992-2000]	(0.150)	(0.012)	(0.156)	(0.026)	(0.063)	(0.056)
Core Region {0,1} ×	EAC 1[2002-2004]	-0.051	-0.049	0.615	0.004	-0.021	0.002
	EAC 0[1992-2000]	(0.233)	(0.053)	(0.882)	(0.036)	(0.099)	(0.112)
Core Region {0,1} ×	CU 1[2005-2009]	0.317	0.087*	1.608**	0.058	0.112	0.110
	CU 0[1992-2000]	(0.235)	(0.043)	(0.590)	(0.055)	(0.109)	(0.073)
Core Region {0,1} ×	CM 1[t ≥ 2010]	-0.295	-0.091**	1.048	0.129***	-0.138*	-0.119**
	CM 0[1992-2000]	(0.200)	(0.039)	(0.816)	(0.041)	(0.070)	(0.053)
Observations		282,480	167,590	11,391	348,016	225,336	223,908
R-Squared		0.27	0.16	0.38	0.16	0.27	0.20
R-Squared -Within		0.26	0.15	0.35	0.14	0.23	0.13
Panel b)							
Border Region {0,1} ×	pre-EAC 1[1996-2000]	-0.223	-0.005	0.041	-	0.177	0.014
	pre-EAC 0[1992-1995]	(0.152)	(0.019)	(0.128)		(0.130)	(0.062)
Border Region {0,1} ×	EAC 1[2002-2004]	0.118	-0.002	0.065	-0.045	0.124	-0.048
	EAC 0[1992-1995]	(0.205)	(0.016)	(0.227)	(0.062)	(0.088)	(0.062)
Border Region {0,1} ×	CU 1[2005-2009]	-0.049	-0.043	-0.043	-0.056*	0.076	-0.101***
	CU 0[1992-1995]	(0.319)	(0.026)	(0.150)	(0.031)	(0.075)	(0.027)
Border Region {0,1} ×	CM 1[t ≥ 2010]	-0.005	-0.006	0.040	-0.014	0.149	-0.089
	CM 0[1992-1995]	(0.241)	(0.026)	(0.200)	(0.024)	(0.128)	(0.069)
Core Region {0,1} ×	pre-EAC 1[1996-2000]	0.345	0.088	0.780	-0.098**	0.211**	0.189*
	pre-EAC 0[1992-1995]	(0.306)	(0.071)	(0.644)	(0.036)	(0.096)	(0.088)
Core Region {0,1} ×	EAC 1[2002-2004]	0.215	0.018	1.157	-0.095**	0.137	0.152
	EAC 0[1992-1995]	(0.379)	(0.054)	(0.889)	(0.042)	(0.120)	(0.155)
Core Region {0,1} ×	CU 1[2005-2009]	-0.175	-0.059**	0.417	-0.021	-0.053	-0.014
	CU 0[1992-1995]	(0.281)	(0.027)	(0.459)	(0.028)	(0.040)	(0.067)
Core Region {0,1} ×	CM 1[t ≥ 2010]	0.123	0.034	2.179*	0.033	0.072	0.068
	CM 0[1992-1995]	(0.280)	(0.062)	(1.042)	(0.028)	(0.058)	(0.079)
Observations		254,644	150,220	11,391	309,353	202,333	200,813
R-Squared		0.27	0.16	0.43	0.16	0.26	0.20
R-Squared -Within		0.26	0.15	0.40	0.14	0.23	0.12
Individual Controls		YES	YES	NO	YES	YES	YES
Geographic Controls		NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects		YES	YES	YES	YES	YES	YES

Notes: This table “disaggregates” the analysis of Table 4.5. See full notes below Table 4.5.

Table C.16: Region-Based Estimates with Varying Intervention Year

	Regression Sample					
	DHS (Region-based) - EAC			DHS (Region-based) - Placebo		
	Consumpt.	Income	Agglom.	Consumpt.	Income	Agglom.
	Wealth Index (1-5)	Employed Work {0,1}	Population Density (sdz.)	Wealth Index (1-5)	Employed Work {0,1}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Border Region {0,1} × EAC 1[t ≥ 1994]	0.617*	-0.001	0.186	-	-	-
Border Region {0,1} × EAC 1[t ≥ 1995]	0.464	0.464	0.328	-	-	-
Border Region {0,1} × EAC 1[t ≥ 1996]	-0.094	0.019	-0.036	-0.359	-0.006	-0.081
Border Region {0,1} × EAC 1[t ≥ 1997]	-0.097	0.024	0.285	-0.160	0.004	0.055
Border Region {0,1} × EAC 1[t ≥ 1998]	-0.042	-0.042	0.298	0.002	0.002	0.116
Border Region {0,1} × EAC 1[t ≥ 1999]	0.208	0.061	0.230	0.011	0.006	0.112
Border Region {0,1} × EAC 1[t ≥ 2000]	0.270**	-0.018	0.382	0.050	0.011	0.148
Border Region {0,1} × EAC 1[t ≥ 2001]	0.167	-0.024	0.143	0.306**	0.004	0.037
Border Region {0,1} × EAC 1[t ≥ 2002]	0.192	-0.027	-0.062	0.318**	0.011	0.021
Border Region {0,1} × EAC 1[t ≥ 2003]	0.199	-0.020	-0.197	0.329**	0.011	0.029
Border Region {0,1} × EAC 1[t ≥ 2004]	0.170	0.019	-0.288	0.275*	-0.005	-0.066
Border Region {0,1} × EAC 1[t ≥ 2005]	0.143	-0.008	-0.264	0.256	0.009	-0.003
Border Region {0,1} × EAC 1[t ≥ 2006]	0.022	-0.026	-0.326*	-0.322	-0.062**	-0.067
Border Region {0,1} × EAC 1[t ≥ 2007]	0.044	-0.026	-0.171	-0.029	-0.042	-0.119
Border Region {0,1} × EAC 1[t ≥ 2008]	-0.150	-0.026	-0.218	-0.077	-0.015	-0.086
Border Region {0,1} × EAC 1[t ≥ 2009]	-0.175	-0.003	-0.213	-0.084	-0.015	-0.125
Border Region {0,1} × EAC 1[t ≥ 2010]	-0.184	-0.004	-0.189	0.003	-0.005	-0.130
Core Region {0,1} × EAC 1[t ≥ 1994]	-0.320	-0.137	0.601***	-	-	-
Core Region {0,1} × EAC 1[t ≥ 1995]	-0.291	-0.131	-1.543**	-	-	-
Core Region {0,1} × EAC 1[t ≥ 1996]	0.811***	0.094***	2.017**	0.928***	0.177***	2.798***
Core Region {0,1} × EAC 1[t ≥ 1997]	0.108	0.061	0.601	0.038	0.040	0.921
Core Region {0,1} × EAC 1[t ≥ 1998]	0.000	0.021	0.601	-0.185	-0.021	0.921
Core Region {0,1} × EAC 1[t ≥ 1999]	-0.089	0.015	-0.813	-0.128	-0.017	0.961
Core Region {0,1} × EAC 1[t ≥ 2000]	0.130	-0.007	0.637	-0.017	-0.002	1.500
Core Region {0,1} × EAC 1[t ≥ 2001]	0.418***	0.017	2.702**	-0.051	-0.046	0.856
Core Region {0,1} × EAC 1[t ≥ 2002]	0.655***	0.116***	4.071***	0.039	0.014	1.189
Core Region {0,1} × EAC 1[t ≥ 2003]	0.607***	0.122***	2.771**	-0.043	0.008	1.224
Core Region {0,1} × EAC 1[t ≥ 2004]	0.466***	0.105***	2.366	-0.205	0.003	0.186
Core Region {0,1} × EAC 1[t ≥ 2005]	0.231*	0.084***	1.480	0.307	0.097**	1.428
Core Region {0,1} × EAC 1[t ≥ 2006]	0.162	0.075**	0.841	0.409	0.092*	2.126*
Core Region {0,1} × EAC 1[t ≥ 2007]	0.149	0.034	3.069***	-0.487**	-0.136**	-0.659
Core Region {0,1} × EAC 1[t ≥ 2008]	-0.147	0.033	0.231	-0.283	-0.105***	0.922
Core Region {0,1} × EAC 1[t ≥ 2009]	-0.116	0.032*	0.373	-0.320	-0.105***	0.714
Core Region {0,1} × EAC 1[t ≥ 2010]	-0.120	0.039**	0.572	-0.340**	-0.096***	0.562

Notes: This table makes tests for varying (placebo) intervention years well before and after the EAC's establishment by making use of the non-GPS survey rounds of the Demographic and Health Surveys (DHS) sampled before 1999. For the "EAC Countries" tested in columns (1) through (3), the total sample includes the Kenya survey rounds of 1988-89, 1993, 1998, 2003, 2008, Tanzania 1991-92, 1994 (KAP), 1996, 1999, 2003-04 (AIS), 2004-2005, 2007-08 (AIS), 2010, 2011-12 (AIS), and Uganda 1988-89, 1995, 1995-96, 2000-01, 2006, 2009 (MIS), 2011, 2011 (AIS). For the "Placebo Countries" tested in columns (4) through (6), the total sample includes the Ethiopia survey rounds of 2000, 2005, 2011, Malawi 1992, 1996 (KAP) 2000, 2004, 2010, 2012 (MIS), Mozambique 1997, 2003, 2009 (AIS), 2011, Rwanda DHS 1992, 2000, 2005, and Zambia 1996 (KAP), 2001-02 and 2007. The estimates are constructed for each (placebo) intervention year, separately, by expanding the data past the respective cut-off. E.g. for the estimate on the placebo intervention year 1994, the regressions use data from first available data for all dependent variables (as early as 1988 depending on the variable) and includes survey years up until 1997. For the intervention year 1995, the data frame is expanded to include the year 1998, for 1996 it includes 1999, and so on. Border Region {0,1} switches to one for individuals living in a region with a mean road distance to EAC border crossings below the 10th percentile of all (within-country) GPS-border distances in the sample. Core Region {0,1} is a dummy indicating individuals living in the region which hosts the core agglomeration of their respective country (i.e. Nairobi, Dar es Salaam and Kampala). The standard errors reported allow for clustering at the region level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table C.17: Border Heterogeneity (AFB and DHS)

	Dependent Variable											
	Afrobarometer						DHS					
	Consumpt.	Income	Agglom.	Income			Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked Last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked Last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.76]	[2.09]	[3.14]	[0.19]	[0.00]	[0.76]	[1.51]	[0.52]
TZA-UGA $\leq$ 50km {0,1} $\times$ EAC 1[t $\geq$ 2002]	0.321***	0.113*	0.058	0.227***	0.208***	-0.225***	0.256	-0.115**	-0.078*	0.022	-0.178***	-0.147***
	(0.043)	(0.060)	(0.048)	(0.038)	(0.035)	(0.057)	(0.164)	(0.057)	(0.046)	(0.017)	(0.031)	(0.037)
TZA-KEN $\leq$ 50km {0,1} $\times$ EAC 1[t $\geq$ 2002]	-0.273***	0.035	0.006	-0.002	-0.049	0.042	-0.512***	-0.142***	0.056	0.031	-0.114**	-0.016
	(0.077)	(0.037)	(0.066)	(0.054)	(0.060)	(0.069)	(0.063)	(0.021)	(0.069)	(0.065)	(0.048)	(0.065)
KEN-UGA $\leq$ 50km {0,1} $\times$ EAC 1[t $\geq$ 2002]	-	0.078***	0.043***	-0.051**	-0.052	-	0.020	-0.035*	0.058***	0.004	-0.032	-0.238***
		(0.022)	(0.015)	(0.025)	(0.050)		(0.064)	(0.019)	(0.009)	(0.023)	(0.039)	(0.037)
Core {0,1} $\times$ EAC 1[t $\geq$ 2002]	-0.361***	-0.036	0.767**	-0.136	0.026	-0.059	0.358**	0.016	0.812***	0.052	-0.002	0.015
	(0.074)	(0.062)	(0.336)	(0.103)	(0.065)	(0.082)	(0.151)	(0.044)	(0.121)	(0.036)	(0.099)	(0.030)
Individual Controls	YES	YES	NO	YES	YES	YES	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.13	0.14	0.25	0.23	0.25	0.09	0.34	0.16	0.29	0.23	0.26	0.17
R-Squared -Within	0.10	0.10	0.22	0.07	0.22	0.06	0.32	0.11	0.28	0.20	0.17	0.09

Notes: This table analyzes the heterogeneity across the three Border regions. TZA-UGA {0,1}, TZA-KEN {0,1} and KEN-UGA {0,1} thereby switch to one for individuals living within 50km road distance to the respective border crossing (within either country). The results in each column and panel are produced by a separate regression. The sample mean of the respective dependent variable is given in brackets above the estimates. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 4.1.

Table C.18: Opinion Polling (AFB and KHDS)

	Dependent Variable						
	Afrobarometer						
	DiD		Contemporary Opinion				
Panel a)	Pres. vs. Past: Life Satsfaction {1,4}	Pres. vs. Past: Living Standards {1,4}	Support for: Regional Integration {1,5}	Ease of: Crossing Borders {1,4}	Helps your Country: REC / EAC {1,3}	Helps your Country: African Union {1,3}	Like as Neighbor: Immigr./For. Worker {1,5}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample Mean of Dep. Var.	[2.70]	[2.71]	[3.71]	[2.21]	[1.66]	[1.62]	[3.38]
Border [0,1]	0.041	-1.006***	0.087	0.368***	-0.109	-0.115	-0.245
	(0.130)	(0.245)	(0.249)	(0.083)	(0.095)	(0.092)	(0.160)
Core {0,1}	-0.019	0.172**	0.000	-0.039	-0.033	-0.004	-0.198***
	(0.050)	(0.073)	(0.059)	(0.039)	(0.038)	(0.027)	(0.057)
Border [0,1] × EAC 1[t ≥ 2002]	-0.356**	0.745***	-	-	-	-	-
	(0.146)	(0.255)					
Core {0,1} × EAC 1[t ≥ 2002]	-0.050	-0.209*	-	-	-	-	-
	(0.130)	(0.112)					
Observations	36,213	23,370	6,362	4,766	11,687	13,069	11,647
R-Squared	0.08	0.08	0.10	0.06	0.03	0.03	0.02
R-Squared -Within	0.01	0.02	0.01	0.02	0.00	0.00	0.01

	Kagera Health and Development Survey (KHDS)						
	DiD	Contemporary Opinion					
Panel b)	Main Reason for Migrating: Economic {0,1}	After Migr. to Res.: Paid Empl. {0,1}	After Migr to Res.: Paid form. Empl. {0,1}	Activity in Residence: Working {0,1}	HH. Wealth vs. 10 years ago (1994) {1,5}	HH. Wealth today (2004) {1,5}	HH. Life Satisfaction (2004) {1,9}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample Mean of Dep. Var.	[0.20]	[0.14]	[0.07]	[0.83]	[2.62]	[2.62]	[3.79]
Border [0,1]	-0.631	-0.564**	-0.408**	0.547*	0.274	0.237	0.783
	(5.759)	(0.241)	(0.208)	(0.327)	(0.278)	(0.265)	(0.819)
Core {0,1}	0.674	0.061	-0.069	-0.051	0.160	0.292**	0.444
	(0.927)	(0.073)	(0.072)	(0.103)	(0.188)	(0.143)	(0.415)
Border [0,1] * EAC 1[2004]	0.855	-	-	-	-	-	-
	(3.719)						
Border [0,1] * CU 1[2010]	1.143	-	-	-	-	-	-
	(3.944)						
Core {0,1} * EAC 1[2004]	-0.519	-	-	-	-	-	-
	(0.901)						
Core {0,1} * CU 1[2010]	-0.392	-	-	-	-	-	-
	(0.600)						
Observations	1,347	900	900	774	2,833	2,833	2,833
Observations - Fixed Effects	1,202	2	2	2	2	2	2
R-Squared	0.89	0.16	0.22	0.03	0.02	0.08	0.10
R-Squared -Within	0.096	0.155	0.223	0.024	0.021	0.078	0.10
Individual Controls	YES	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO/YES	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES	YES

Notes: This table analyzes opinions and sentiments of survey respondents. As these data are not available for all tested survey rounds, a difference-in-differences (DiD) estimate can only be conducted for columns (1) and (2) in panel a) as well as column (1) in panel b). The results in each column and panel are produced by a separate regression. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 4.1 and Table 4.3, respectively.

Table C.19: Border Price Pass-Through (KHDS)

	Dependent Variable			
	KHDS			
	Unit Price of Item in Tanzanian Shilling			
	Maize (TZS)	*Millet* (TZS)	*Tea/Coffee* (TZS)	*Sugar* (TZS)
	(1)	(2)	(3)	(4)
Sample Mean of Dep. Var.	[207.25]	[122.83]	[24.60]	[52.50]
Border [0,1] × EAC 1[2004]	-0.019	-1.066*	0.308	-0.048
	(0.090)	(0.563)	(0.928)	(0.094)
Core {0,1} × EAC 1[2004]	-0.036***	-	-0.080*	-0.021***
	(0.006)		(0.047)	(0.004)
Individual Controls	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES
Household Fixed Effects	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES
Observations	3,267	193	948	1,722
Observations - Fixed Effects	1,990	127	554	899
R-Squared	0.99	0.99	0.99	0.99
R-Squared -Within	0.01	0.05	0.05	0.00

Notes: This table tests for a differential price-pass through of traded, homogenous, staple goods, following the establishment of the EAC. The results in each column are produced by a separate regression. Data come from the Kagera Health and Development Surveys (KHDS) consumption components which askes respondents for seasonal prices of goods. These data are collected in four waves across 1991-1994, as well as in 2004. All regressions include country-year fixed effects. The results in each column and panel are produced by a separate regression. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See Table 4.3 for full table notes.

Table C.20: Decomposed Indices (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Lived Poverty			Income		
	Freq. gone without: [Water] {0,4}	Freq. gone without: [Food] {0,4}	Freq. g. without: [Medicine/ Med. Care] {0,4}	Occupation: [Agrarian] {0,1}	Occupation: [Worker] {0,1}	Occupation: [Professional] {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.13]	[0.98]	[1.26]	[0.55]	[0.23]	[0.21]
Border [0,1] × EAC 1[2002-2004]	-0.620*	0.120	0.261	-0.029	0.041	-0.012
	(0.341)	(0.379)	(0.354)	(0.083)	(0.053)	(0.059)
Border [0,1] × CU 1[2005-2009]	-0.268	0.117	0.667*	-0.044	0.048	-0.004
	(0.336)	(0.341)	(0.344)	(0.077)	(0.059)	(0.054)
Border [0,1] × CM 1[t ≥ 2010]	-0.302	0.260	0.466	-0.243***	0.196***	0.047
	(0.333)	(0.301)	(0.332)	(0.056)	(0.048)	(0.042)
Core {0,1} × EAC 1[2002-2004]	-0.167	-0.070	-0.546***	0.006	-0.033	0.027
	(0.157)	(0.106)	(0.086)	(0.066)	(0.059)	(0.022)
Core {0,1} × CU 1[2005-2009]	-0.563***	-0.253**	-0.591***	-0.061	0.065*	-0.004
	(0.100)	(0.116)	(0.121)	(0.062)	(0.035)	(0.030)
Core {0,1} × CM 1[t ≥ 2010]	-0.452***	-0.082	-0.480***	-0.057	0.124**	-0.066**
	(0.082)	(0.080)	(0.086)	(0.066)	(0.055)	(0.030)
Observations	36,024	36,007	35,973	21,232	21,232	21,232
R-Squared	0.08	0.10	0.10	0.25	0.13	0.20
R-Squared -Within	0.06	0.09	0.07	0.20	0.07	0.18

Notes: This table analyzes the components of the two composite indices used throughout the paper. The results in each column and panel are produced by a separate regression. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 4.1.

Table C.21: Decomposed Indices (DHS and KHDS)

	Dependent Variable							
	DHS					KHDS		
	Consumption		Income					
	International Wealth Index	*Comparative Wealth Index*	Occupation: [Agrarian] {0,1}	Occupation: [Worker] {0,1}	Occupation: [Professional] {0,1}	Occupation: [Agrarian] {0,1}	Occupation: [Worker] {0,1}	Occupation: [Professional] {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample Mean of Dep. Var.	[23.40]	-[0.67]	[0.57]	[0.35]	[0.08]	[0.84]	[0.12]	[0.04]
Border [0,1] × EAC 1[2002-2004]	0.842	0.069	0.125	-0.067	-0.059*	-0.030	-0.048	0.077
	(0.762)	(0.171)	(0.185)	(0.186)	(0.030)	(0.224)	(0.210)	(0.077)
Border [0,1] × CU 1[2005-2009]	-0.153	-0.123	0.249***	-0.206**	-0.043	-	-	-
	(0.481)	(0.119)	(0.093)	(0.093)	(0.035)			
Border [0,1] × CM 1[t ≥ 2010]	-0.185	-0.105	0.149**	-0.149**	0.000	-0.113***	0.236***	-0.123
	(0.483)	(0.127)	(0.069)	(0.069)	(0.019)	(0.286)	(0.273)	(0.193)
Core {0,1} × EAC 1[2002-2004]	1.869***	0.724***	0.056	-0.066	0.010	-0.157*	0.180**	-0.023
	(0.375)	(0.159)	(0.058)	(0.055)	(0.011)	(0.021)	(0.023)	(0.015)
Core {0,1} × CU 1[2005-2009]	0.644***	0.320***	-0.025	-0.008	0.033	-	-	-
	(0.228)	(0.093)	(0.119)	(0.131)	(0.021)			
Core {0,1} × CM 1[t ≥ 2010]	0.571***	0.273***	0.025	-0.056	0.032*	-0.099***	0.128***	-0.029***
	(0.111)	(0.094)	(0.073)	(0.060)	(0.017)	(0.057)	(0.056)	(0.023)
Observations	183,253	73,985	95,717	95,717	95,717	15,685	15,685	15,685
Observations - Fixed Effects	-	-	-	-	-	6,253	6,253	6,253
R-Squared	0.17	0.46	0.23	0.13	0.15	0.79	0.72	0.80
R-Squared -Within	0.116	0.437	0.145	0.076	0.100	0.116	0.06	0.08
Individual Controls	YES	YES	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	NO	NO	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES

Notes: This table analyzes the components of the two composite indices used throughout the paper. The results in each column and panel are produced by a separate regression. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 4.2.

Table C.22: Robustness Check: Clustered Standard Errors (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.66]	[2.09]
Border [0,1] × EAC 1[2002-2004]	-0.083	0.050	0.083	0.072	0.017	-0.350*
	(0.170)	(0.045)	(0.058)	(0.057)	(0.115)	(0.194)
Border [0,1] × CU 1[2005-2009]	0.172	0.060	0.045	0.091	0.040	-0.234
	(0.162)	(0.050)	(0.061)	(0.061)	(0.110)	(0.188)
Border [0,1] × CM 1[t ≥ 2010]	0.142	0.098***	0.099	-0.001	0.289***	-0.456***
	(0.147)	(0.037)	(0.063)	(0.049)	(0.092)	(0.176)
Core {0,1} × EAC 1[2002-2004]	-0.261**	-0.013	-0.158	0.026	0.021	0.041
	(0.126)	(0.041)	(0.135)	(0.044)	(0.084)	(0.159)
Core {0,1} × CU 1[2005-2009]	-0.470***	-0.006	0.769***	0.002	0.057	-0.234
	(0.125)	(0.044)	(0.214)	(0.046)	(0.080)	(0.155)
Core {0,1} × CM 1[t ≥ 2010]	-0.338***	-0.076**	1.113***	-0.258***	-0.009	0.043
	(0.119)	(0.036)	(0.179)	(0.041)	(0.078)	(0.150)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,104	26,563	21,232	35,975
R-Squared	0.12	0.13	0.28	0.23	0.25	0.09
R-Squared -Within	0.10	0.10	0.25	0.23	0.25	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: The standard errors are clustered at the enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table C.23: Robustness Check: Clustered Standard Errors (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.19]	[0.00]	[0.76]	[1.51]	[0.52]
Border [0,1] $\times$ EAC 1[2002-2004]	-0.294	-0.129*	0.104	0.110***	-0.184	-0.162*
	(0.296)	(0.070)	(0.094)	(0.041)	(0.134)	(0.097)
Border [0,1] $\times$ CU 1[2005-2009]	-0.377	-0.138***	0.084	0.115***	-0.291***	-0.271***
	(0.274)	(0.036)	(0.085)	(0.037)	(0.088)	(0.059)
Border [0,1] $\times$ CM 1[t $\geq$ 2010]	-0.505**	-0.034	0.060	0.065*	-0.149**	-0.141***
	(0.257)	(0.031)	(0.076)	(0.034)	(0.073)	(0.051)
Core {0,1} $\times$ EAC 1[2002-2004]	0.435***	0.073*	0.146	0.068***	-0.046	-0.002
	(0.110)	(0.040)	(0.201)	(0.021)	(0.058)	(0.038)
Core {0,1} $\times$ CU 1[2005-2009]	0.315***	0.037	0.770***	0.014	0.058	0.053
	(0.097)	(0.032)	(0.241)	(0.020)	(0.052)	(0.035)
Core {0,1} $\times$ CM 1[t $\geq$ 2010]	0.345***	0.006	0.956***	0.049***	0.007	0.029
	(0.091)	(0.029)	(0.185)	(0.017)	(0.047)	(0.032)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	204,991	71,738	7,692	191,616	95,717	98,963
R-Squared	0.34	0.16	0.30	0.22	0.26	0.17
R-Squared -Within	0.33	0.11	0.29	0.22	0.26	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: The standard errors are clustered at the enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table C.24: Robustness Check: Clustered Standard Errors (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
Border [0,1] × EAC 1[2004]	-853.376	187.152	0.581**	0.071	0.107	-0.848***
	(1085.855)	(651.437)	(0.239)	(0.050)	(0.261)	(0.288)
Border [0,1] × CU 1[2010]	-1317.845		0.432*		-0.010	-1.631***
	(1156.327)		(0.246)		(0.404)	(0.289)
Core {0,1} × EAC 1[2004]	170.320*	706.607	-0.013	0.026	0.134***	0.891***
	(100.162)	(1002.964)	(0.033)	(0.018)	(0.021)	(0.126)
Core {0,1} × CU 1[2010]	275.036**		-0.039		0.069	0.891***
	(111.403)		(0.040)		(0.067)	(0.132)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,492	2,695	24,972	14,254	15,685	2,292
Observations - Fixed Effects	3,816	2,363	12,747	6,988	6,253	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.43
R-Squared -Within	0.12	0.09	0.04	0.01	0.14	0.41

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: The standard errors are clustered at the enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table C.25: Robustness Check: Cluster Standard Errors (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
Border [0,1] × EAC 1[2004]	-177.749	-670.444	7592.679	-0.201	-96.031	5.949
	(262.369)	(1034.002)	(12613.087)	(0.393)	(135.964)	(9.453)
Border [0,1] × CU 1[2010]	-388.178	-911.118		-0.233		6.555
	(267.416)	(1093.951)		(0.297)		(8.908)
Core {0,1} × EAC 1[2004]	122.188**	35.341	4013.722	-0.058	8.109	-1.079
	(49.625)	(57.589)	(5982.927)	(0.075)	(27.118)	(1.648)
Core {0,1} × CU 1[2010]	131.046***	133.815*		-0.008		-0.404
	(47.557)	(74.509)		(0.051)		(1.636)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,494	5,524	2,695	16,330	1,782	5,411
Observations - Fixed Effects	3,817	3,830	2,363	11,599	1,190	3,763
R-Squared	0.88	0.83	0.69	0.86	0.94	0.89
R-Squared -Within	0.10	0.09	0.07	0.16	0.18	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: The standard errors are clustered at the enumeration area level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table C.26: Robustness Check: No Controls (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.76]	[2.09]
Border [0,1] $\times$ EAC 1[2002-2004]	-0.420	0.060	0.047	0.103	0.013	-0.524*
	(0.340)	(0.075)	(0.038)	(0.083)	(0.186)	(0.305)
Border [0,1] $\times$ CU 1[2005-2009]	-0.088	0.046	0.023	0.080	0.029	-0.341
	(0.335)	(0.074)	(0.060)	(0.083)	(0.143)	(0.311)
Border [0,1] $\times$ CM 1[t $\geq$ 2010]	-0.151***	0.110	0.065***	0.024***	0.319	-0.599
	(0.307)	(0.042)	(0.072)	(0.087)	(0.114)	(0.297)
Core {0,1} $\times$ EAC 1[2002-2004]	-0.392***	-0.011	-0.188***	0.018	0.029	-0.047
	(0.088)	(0.044)	(0.040)	(0.057)	(0.068)	(0.089)
Core {0,1} $\times$ CU 1[2005-2009]	-0.607***	0.005	0.738**	0.002	0.103	-0.318***
	(0.116)	(0.048)	(0.362)	(0.053)	(0.103)	(0.114)
Core {0,1} $\times$ CM 1[t $\geq$ 2010]	-0.466***	-0.054	1.094***	-0.253***	0.060	-0.061
	(0.081)	(0.065)	(0.396)	(0.102)	(0.110)	(0.077)
Individual Controls	NO	NO	NO	NO	NO	NO
Geographic Controls	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,371	25,775	4,156	26,880	21,482	36,304
R-Squared	0.04	0.06	0.27	0.18	0.09	0.05
R-Squared -Within	0.02	0.02	0.24	0.18	0.09	0.05

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: All individual- and geographic controls are removed from the right side of the regression equation.

Table C.27: Robustness Check: No Controls (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.11]	[0.18]	[0.00]	[0.78]	[1.51]	[0.52]
Border [0,1] $\times$ EAC 1[2002-2004]	-0.275	-0.162	0.096	0.169**	-0.257	-0.180
	(0.390)	(0.134)	(0.150)	(0.074)	(0.232)	(0.187)
Border [0,1] $\times$ CU 1[2005-2009]	-0.104	-0.098*	0.077	0.175***	-0.282**	-0.220**
	(0.307)	(0.059)	(0.086)	(0.056)	(0.128)	(0.103)
Border [0,1] $\times$ CM 1[t $\geq$ 2010]	-0.318	-0.036	0.039	0.133**	-0.216**	-0.152
	(0.246)	(0.044)	(0.095)	(0.052)	(0.085)	(0.094)
Core {0,1} $\times$ EAC 1[2002-2004]	0.334*	0.077	0.071	0.092	-0.098	-0.065
	(0.173)	(0.050)	(0.320)	(0.060)	(0.094)	(0.046)
Core {0,1} $\times$ CU 1[2005-2009]	0.386***	0.052	0.708***	0.031	0.060	0.029
	(0.149)	(0.042)	(0.170)	(0.062)	(0.136)	(0.049)
Core {0,1} $\times$ CM 1[t $\geq$ 2010]	0.375*	0.018	0.914***	0.077*	-0.024	0.009
	(0.224)	(0.047)	(0.104)	(0.044)	(0.116)	(0.043)
Individual Controls	NO	NO	NO	NO	NO	NO
Geographic Controls	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,306	71,757	7,692	169,898	95,740	98,986
R-Squared	0.17	0.11	0.29	0.04	0.17	0.13
R-Squared -Within	0.16	0.06	0.28	0.04	0.17	0.13

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: All individual- and geographic controls are removed from the right side of the regression equation.

Table C.28: Robustness Check: No Controls (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
Border [0,1] × EAC 1[2004]	-244.619	231.719	0.481**	0.065*	0.194	-1.516**
	(739.987)	(230.704)	(0.231)	(0.035)	(0.245)	(0.658)
Border [0,1] × CU 1[2010]	-613.663***		0.449***		0.219***	-1.946***
	(871.268)		(0.246)		(0.414)	(0.731)
Core {0,1} × EAC 1[2004]	122.774***	559.391***	-0.012	0.024***	0.109***	1.112***
	(44.950)	(40.735)	(0.027)	(0.008)	(0.021)	(0.296)
Core {0,1} × CU 1[2010]	212.121***		-0.021***		0.035***	1.346***
	(71.187)		(0.031)		(0.079)	(0.461)
Individual Controls	NO	NO	NO	NO	NO	NO
Geographic Controls	NO	NO	NO	NO	NO	NO
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	6,527	5,856	31,263	19,119	20,120	2,293
Observations - Fixed Effects	4,424	2,956	14,530	8,503	7,570	12
R-Squared	0.86	0.93	0.60	0.69	0.80	0.36
R-Squared -Within	0.02	0.05	0.01	0.00	0.08	0.34

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: All individual- and geographic controls are removed from the right side of the regression equation.

Table C.29: Robustness Check: No Controls (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
Border [0,1] × EAC 1[2004]	54.492	-291.081	3979.185	0.005	-194.034	5.802
	(208.728)	(669.004)	(3760.092)	(0.312)	(319.076)	(7.924)
Border [0,1] × CU 1[2010]	-145.003***	-453.312***		-0.070***		7.066***
	(255.255)	(750.695)		(0.252)		(7.395)
Core {0,1} × EAC 1[2004]	95.130***	20.114	3358.445***	-0.004	28.256	-0.435
	(33.049)	(28.938)	(574.014)	(0.026)	(21.074)	(1.073)
Core {0,1} × CU 1[2010]	109.186***	96.839***		0.065***		-0.479***
	(36.483)	(42.831)		(0.039)		(0.952)
Individual Controls	NO	NO	NO	NO	NO	YES
Geographic Controls	NO	NO	NO	NO	NO	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	6,530	6,583	5,855	19,657	2,036	7,429
Observations - Fixed Effects	4,424	4,451	2,956	13,170	1,374	4,900
R-Squared	0.87	0.82	0.50	0.80	0.83	0.90
R-Squared -Within	0.01	0.01	0.02	0.00	0.03	0.00

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: All individual- and geographic controls are removed from the right side of the regression equation.

Table C.30: Robustness Test: Extended Set of Geographic Controls (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.66]	[2.09]
Border [0,1] × EAC 1[2002-2004]	0.097	0.048	0.039	0.032	0.054	-0.750***
	(0.363)	(0.068)	(0.074)	(0.090)	(0.152)	(0.274)
Border [0,1] × CU 1[2005-2009]	0.281	0.097	0.065	0.105	0.128	-0.548*
	(0.359)	(0.074)	(0.076)	(0.085)	(0.137)	(0.288)
Border [0,1] × CM 1[t ≥ 2010]	0.228	0.124***	0.141	-0.037	0.331***	-0.841**
	(0.349)	(0.044)	(0.099)	(0.090)	(0.107)	(0.242)
Core {0,1} × EAC 1[2002-2004]	-0.211***	-0.012	-0.216**	0.019	0.021	0.054
	(0.080)	(0.051)	(0.086)	(0.069)	(0.076)	(0.081)
Core {0,1} × CU 1[2005-2009]	-0.424***	-0.001	0.694**	0.000	0.069	-0.205**
	(0.094)	(0.047)	(0.301)	(0.060)	(0.084)	(0.088)
Core {0,1} × CM 1[t ≥ 2010]	-0.304***	-0.076	1.031***	-0.262**	-0.018	0.046
	(0.073)	(0.070)	(0.353)	(0.109)	(0.082)	(0.083)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,002	25,441	4,151	26,539	21,209	35,935
R-Squared	0.13	0.14	0.29	0.24	0.26	0.10
R-Squared -Within	0.11	0.10	0.26	0.24	0.26	0.10

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I add the extended set of geographic controls anticipated in the manuscript, namely locations' Absolute Latitude, Malaria Ecology, distance to Navigable Rivers, Major Lakes and Major Harbors and add the dummy non-EAC ≤ 100km {0,1} as well as the interaction of it with all period dummies to net out effects potentially stemming from a change (loss) in market access at non-EAC borders.

Table C.31: Robustness Check: Extended Set of Geographic Controls (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.18]	[0.00]	[0.76]	[1.51]	[0.52]
Border [0,1] $\times$ EAC 1[2002-2004]	-0.098	-0.127	-0.015	0.072	-0.156	-0.071
	(0.333)	(0.080)	(0.138)	(0.060)	(0.177)	(0.163)
Border [0,1] $\times$ CU 1[2005-2009]	-0.079	-0.101**	0.061	0.081**	-0.303***	-0.206**
	(0.289)	(0.052)	(0.108)	(0.040)	(0.096)	(0.088)
Border [0,1] $\times$ CM 1[t $\geq$ 2010]	-0.257	-0.022	0.018	0.021	-0.134*	-0.049
	(0.267)	(0.041)	(0.100)	(0.047)	(0.070)	(0.090)
Core {0,1} $\times$ EAC 1[2002-2004]	0.483***	0.066*	0.014	0.042	-0.040	0.016
	(0.182)	(0.036)	(0.255)	(0.029)	(0.069)	(0.027)
Core {0,1} $\times$ CU 1[2005-2009]	0.383***	0.042	0.721***	-0.003	0.071	0.079**
	(0.126)	(0.034)	(0.145)	(0.037)	(0.107)	(0.032)
Core {0,1} $\times$ CM 1[t $\geq$ 2010]	0.403**	0.007	0.886***	0.046*	0.008	0.045
	(0.157)	(0.041)	(0.094)	(0.026)	(0.091)	(0.030)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.35	0.16	0.30	0.23	0.26	0.18
R-Squared -Within	0.34	0.11	0.29	0.23	0.26	0.18

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I add the extended set of geographic controls anticipated in the manuscript, namely locations' Absolute Latitude, Malaria Ecology, distance to Navigable Rivers, Major Lakes and Major Harbors and add the dummy non-EAC ≤ 100km {0,1} as well as the interaction of it with all period dummies to net out effects potentially stemming from a change (loss) in market access at non-EAC borders.

Table C.32: Robustness Check: Extended Set of Geographic Controls (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
Border [0,1] × EAC 1[2004]	-854.048	35.596	0.578*	0.075	0.102	0.314
	(1034.553)	(643.196)	(0.318)	(0.052)	(0.255)	(0.655)
Border [0,1] × CU 1[2010]	-1283.907		0.454		0.118	-0.188
	(1086.753)		(0.302)		(0.410)	(0.583)
Core {0,1} × EAC 1[2004]	172.061***	713.341***	-0.012	0.026***	0.143***	0.836***
	(39.808)	(69.683)	(0.024)	(0.009)	(0.025)	(0.030)
Core {0,1} × CU 1[2010]	269.142***		-0.036		0.080	0.846***
	(54.303)		(0.028)		(0.072)	(0.072)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,467	2,678	24,941	14,241	15,664	2,265
Observations - Fixed Effects	3,801	2,346	12,732	6,980	6,247	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.45
R-Squared -Within	0.14	0.10	0.04	0.01	0.14	0.42

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: I add the extended set of geographic controls anticipated in the manuscript, namely locations' Absolute Latitude, Malaria Ecology, distance to Navigable Rivers, Major Lakes and Major Harbors and add the dummy non-EAC ≤ 100km {0,1} as well as the interaction of it with all period dummies to net out effects potentially stemming from a change (loss) in market access at non-EAC borders.

Table C.33: Robustness Check: Extended Set of Geographic Controls (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
Border [0,1] × EAC 1[2004]	-174.383	-673.897	8162.184	-0.165	-86.860	5.782
	(249.275)	(1024.382)	(14988.404)	(0.370)	(144.913)	(9.909)
Border [0,1] × CU 1[2010]	-378.751***	-886.309***		-0.157***		5.605***
	(261.737)	(1058.009)		(0.329)		(9.731)
Core {0,1} × EAC 1[2004]	120.084***	37.350	3971.477***	-0.064	9.471	-1.209
	(35.210)	(36.610)	(577.551)	(0.041)	(25.328)	(1.489)
Core {0,1} × CU 1[2010]	127.440***	130.428***		-0.011***		-0.689***
	(37.672)	(35.080)		(0.047)		(1.486)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,469	5,499	2,678	16,303	1,775	5,380
Observations - Fixed Effects	3,802	3,815	2,346	11,581	1,186	3,754
R-Squared	0.89	0.83	0.69	0.86	0.95	0.89
R-Squared -Within	0.10	0.11	0.07	0.16	0.23	0.10

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: I add the extended set of geographic controls anticipated in the manuscript, namely locations' Absolute Latitude, Malaria Ecology, distance to Navigable Rivers, Major Lakes and Major Harbors and add the dummy non-EAC ≤ 100km {0,1} as well as the interaction of it with all period dummies to net out effects potentially stemming from a change (loss) in market access at non-EAC borders.

Table C.34: Robustness Check: Using Survey Weights (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.66]	[2.09]
Border [0,1] $\times$ EAC 1[2002-2004]	-0.032	0.078	0.093*	0.113	0.141	-0.288
	(0.315)	(0.059)	(0.057)	(0.075)	(0.125)	(0.300)
Border [0,1] $\times$ CU 1[2005-2009]	0.263	0.085	0.065	0.110	0.064	-0.179
	(0.293)	(0.068)	(0.077)	(0.076)	(0.129)	(0.319)
Border [0,1] $\times$ CM 1[t $\geq$ 2010]	0.152	0.103**	0.117	0.013	0.284***	-0.434
	(0.288)	(0.040)	(0.111)	(0.074)	(0.095)	(0.294)
Core {0,1} $\times$ EAC 1[2002-2004]	-0.245***	-0.016	-0.168**	0.016	-0.010	0.039
	(0.079)	(0.048)	(0.069)	(0.069)	(0.092)	(0.082)
Core {0,1} $\times$ CU 1[2005-2009]	-0.443***	0.001	0.910***	0.014	0.012	-0.187*
	(0.084)	(0.054)	(0.351)	(0.063)	(0.097)	(0.098)
Core {0,1} $\times$ CM 1[t $\geq$ 2010]	-0.340***	-0.089	0.916**	-0.268**	-0.002	0.032
	(0.067)	(0.078)	(0.388)	(0.114)	(0.096)	(0.082)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.12	0.13	0.27	0.24	0.25	0.10
R-Squared -Within	0.10	0.10	0.24	0.24	0.25	0.10

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I produce a weighted regression, employing the survey sampling weights provided by the Afrobarometer and accounting for the pooling across countries and years by standardizing the weights for each country-survey round pair.

Table C.35: Robustness Check: Using Survey Weights (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.18]	[0.00]	[0.76]	[1.51]	[0.52]
Border [0,1] × EAC 1[2002-2004]	-0.284	-0.129	0.104	0.114*	-0.184	-0.162
	(0.312)	(0.089)	(0.136)	(0.059)	(0.188)	(0.136)
Border [0,1] × CU 1[2005-2009]	-0.372*	-0.138**	0.084	0.118***	-0.291***	-0.271***
	(0.224)	(0.054)	(0.089)	(0.043)	(0.106)	(0.083)
Border [0,1] × CM 1[t ≥ 2010]	-0.547***	-0.034	0.060	0.081*	-0.149**	-0.141
	(0.202)	(0.040)	(0.095)	(0.046)	(0.075)	(0.092)
Core {0,1} × EAC 1[2002-2004]	0.433***	0.073*	0.146	0.067*	-0.046	-0.002
	(0.141)	(0.040)	(0.278)	(0.038)	(0.062)	(0.026)
Core {0,1} × CU 1[2005-2009]	0.314***	0.037	0.770***	0.014	0.058	0.053
	(0.105)	(0.039)	(0.177)	(0.043)	(0.111)	(0.033)
Core {0,1} × CM 1[t ≥ 2010]	0.361**	0.006	0.956***	0.061*	0.007	0.029
	(0.142)	(0.041)	(0.111)	(0.031)	(0.088)	(0.034)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,294	169,875	95,717	98,963
R-Squared	0.34	0.16	0.26	0.22	0.28	0.19
R-Squared -Within	0.34	0.10	0.24	0.22	0.28	0.19

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I produce a weighted regression, employing the survey sampling weights provided by the DHS and accounting for the pooling across countries and years by standardizing the weights for each country-survey round pair.

Table C.36: Robustness Check: Excluding Low-Precision Localities (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.09]	[0.22]	[0.00]	[0.63]	[1.75]	[2.03]
Border [0,1] × EAC 1[2002-2004]	0.288	0.088	0.137	0.139	0.006	0.127
	(0.448)	(0.091)	(0.092)	(0.123)	(0.202)	(0.399)
Border [0,1] × CU 1[2005-2009]	0.574	0.106	0.096	0.109	0.067	0.297
	(0.406)	(0.097)	(0.141)	(0.110)	(0.187)	(0.381)
Border [0,1] × CM 1[t ≥ 2010]	0.282	0.142***	0.170	0.063	0.337**	-0.102
	(0.373)	(0.052)	(0.117)	(0.083)	(0.145)	(0.380)
Core {0,1} × EAC 1[2002-2004]	-0.008	0.036	-0.139***	0.074	-0.011	0.336***
	(0.092)	(0.036)	(0.030)	(0.058)	(0.092)	(0.123)
Core {0,1} × CU 1[2005-2009]	-0.238**	0.054	0.757**	0.029	-0.067	0.142
	(0.108)	(0.043)	(0.366)	(0.070)	(0.147)	(0.134)
Core {0,1} × CM 1[t ≥ 2010]	-0.081	0.019	1.041***	-0.217**	-0.090	0.383***
	(0.078)	(0.050)	(0.213)	(0.089)	(0.107)	(0.110)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	15,507	12,116	1,926	13,074	10,345	15,479
R-Squared	0.13	0.14	0.31	0.23	0.27	0.09
R-Squared -Within	0.10	0.09	0.27	0.23	0.27	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I drop "low-precision" localities from the regression. This is implemented by dropping all observations for which the AidData precision code is above 2.

Table C.37: Robustness Check: Excluding Low-Precision Localities (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.09]	[0.16]	[0.00]	[0.77]	[1.48]	[0.45]
Border [0,1] $\times$ EAC 1[2002-2004]	-0.259	-0.103	0.519*	0.123*	-0.186	-0.174
	(0.314)	(0.084)	(0.306)	(0.063)	(0.227)	(0.146)
Border [0,1] $\times$ CU 1[2005-2009]	-0.402*	-0.126**	0.217	0.135***	-0.308***	-0.284***
	(0.226)	(0.054)	(0.138)	(0.046)	(0.113)	(0.081)
Border [0,1] $\times$ CM 1[t $\geq$ 2010]	-0.446**	-0.034	0.000	0.036	-0.143*	-0.142
	(0.192)	(0.039)	(0.077)	(0.043)	(0.074)	(0.093)
Core {0,1} $\times$ EAC 1[2002-2004]	0.439***	0.072*	-0.130	0.065*	-0.039	-0.003
	(0.147)	(0.040)	(0.303)	(0.039)	(0.069)	(0.030)
Core {0,1} $\times$ CU 1[2005-2009]	0.326***	0.038	0.734**	0.015	0.063	0.052
	(0.107)	(0.038)	(0.293)	(0.043)	(0.113)	(0.035)
Core {0,1} $\times$ CM 1[t $\geq$ 2010]	0.342**	0.006	0.950***	0.035*	0.058	0.042
	(0.157)	(0.043)	(0.103)	(0.019)	(0.075)	(0.041)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	149,062	61,998	3,224	140,076	75,568	78,587
R-Squared	0.33	0.17	0.31	0.21	0.28	0.17
R-Squared -Within	0.32	0.12	0.29	0.21	0.28	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I drop "low-precision" localities from the regression. This is implemented by dropping all observations for which coordinates are not generated from a GPS receiver used by the fieldworker.

Table C.38: Robustness Check: Agglomerations vs. Core (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.66]	[2.09]
Border [0,1] × EAC 1[2002-2004]	-0.121	0.061	0.098	0.084	0.029	-0.253
	(0.293)	(0.059)	(0.067)	(0.074)	(0.131)	(0.287)
Border [0,1] × CU 1[2005-2009]	0.170	0.062	0.027	0.093	0.044	-0.102
	(0.290)	(0.065)	(0.069)	(0.074)	(0.120)	(0.307)
Border [0,1] × CM 1[t ≥ 2010]	0.132***	0.093***	0.014***	0.016***	0.265***	-0.319***
	(0.280)	(0.042)	(0.087)	(0.077)	(0.090)	(0.286)
Agglomeration {0,1} × EAC 1[2002-2004]	0.108	-0.056**	0.022	-0.060*	-0.046	-0.199*
	(0.078)	(0.028)	(0.082)	(0.034)	(0.079)	(0.115)
Agglomeration {0,1} × CU 1[2005-2009]	-0.036	-0.006	0.129	-0.011	0.001	-0.350***
	(0.092)	(0.031)	(0.102)	(0.035)	(0.075)	(0.115)
Agglomeration {0,1} × CM 1[t ≥ 2010]	0.024***	-0.011***	0.242***	-0.063***	0.042***	-0.334***
	(0.080)	(0.027)	(0.150)	(0.045)	(0.071)	(0.117)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.12	0.13	0.29	0.23	0.26	0.10
R-Squared -Within	0.10	0.10	0.26	0.23	0.26	0.10

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: The Core {0,1} dummy is replaced with Agglomeration {0,1} which switches to 1 for households living within 50km of an "urban center" demarcated as such in the year 2000 (see manuscript).

Table C.39: Robustness Check: Agglomerations vs. Core (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.18]	[0.00]	[0.74]	[1.48]	[0.49]
Border [0,1] × EAC 1[2002-2004]	-0.571*	-0.171**	-0.063	0.125**	-0.219	-0.188
	(0.310)	(0.083)	(0.129)	(0.056)	(0.190)	(0.136)
Border [0,1] × CU 1[2005-2009]	-0.533**	-0.152***	-0.006	0.117***	-0.317***	-0.289***
	(0.237)	(0.045)	(0.106)	(0.042)	(0.106)	(0.084)
Border [0,1] × CM 1[t ≥ 2010]	-0.605***	-0.036***	-0.020***	0.092***	-0.168***	-0.155***
	(0.211)	(0.031)	(0.105)	(0.045)	(0.067)	(0.091)
Agglomeration {0,1} × EAC 1[2002-2004]	0.102	-0.044	0.256	0.015	-0.147**	-0.067
	(0.153)	(0.043)	(0.226)	(0.025)	(0.060)	(0.046)
Agglomeration {0,1} × CU 1[2005-2009]	-0.033	-0.052*	0.208*	0.034*	-0.117**	-0.029
	(0.108)	(0.027)	(0.111)	(0.019)	(0.050)	(0.041)
Agglomeration {0,1} × CM 1[t ≥ 2010]	-0.239***	-0.041***	0.258***	0.002***	-0.070***	-0.016***
	(0.090)	(0.017)	(0.169)	(0.022)	(0.033)	(0.031)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.34	0.16	0.21	0.23	0.26	0.17
R-Squared -Within	0.33	0.11	0.20	0.23	0.26	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: The Core {0,1} dummy is replaced with Agglomeration {0,1} which switches to 1 for households living within 50km of an "urban center" demarcated as such in the year 2000 (see manuscript).

Table C.40: Robustness Check: Agglomerations vs. Core (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
Border [0,1] × EAC 1[2004]	-1039.542	131.910	0.510	0.080	0.098	-0.980**
	(1026.698)	(700.554)	(0.324)	(0.052)	(0.251)	(0.465)
Border [0,1] × CU 1[2010]	-1160.586		0.347		0.050	-1.387**
	(1116.591)		(0.315)		(0.397)	(0.602)
Agglomeration {0,1} × EAC 1[2004]	191.119***	702.325***	-0.006	0.025***	0.134***	0.811***
	(65.909)	(125.914)	(0.023)	(0.009)	(0.021)	(0.027)
Agglomeration {0,1} × CU 1[2010]	259.145***		-0.036		0.076	0.754***
	(77.785)		(0.023)		(0.072)	(0.051)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,657	2,767	25,372	14,367	15,817	2,377
Observations - Fixed Effects	3,932	2,427	12,987	7,058	6,309	12
R-Squared	0.87	0.97	0.64	0.72	0.82	0.40
R-Squared -Within	0.12	0.09	0.04	0.01	0.14	0.38

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: The Core {0,1} dummy is replaced with Agglomeration {0,1} which switches to 1 for households living within 50km of an “urban center” demarcated as such in the year 2000 (see manuscript).

Table C.41: Robustness Check: Agglomerations vs. Core (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
Border [0,1] × EAC 1[2004]	-335.740	-720.571	7606.797	-0.108	-100.772	8.349
	(269.364)	(986.817)	(14133.842)	(0.344)	(137.806)	(8.808)
Border [0,1] × CU 1[2010]	-316.002	-838.609		-0.216		8.743
	(277.875)	(1070.563)		(0.329)		(8.815)
Agglomeration {0,1} × EAC 1[2004]	131.930***	48.610	4019.615***	-0.076	8.229	-1.135
	(47.167)	(36.410)	(1307.509)	(0.057)	(25.101)	(1.375)
Agglomeration {0,1} × CU 1[2010]	136.575***	113.736***		-0.005		-0.870
	(49.804)	(43.356)		(0.043)		(1.335)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,659	5,693	2,767	16,687	1,813	5,550
Observations - Fixed Effects	3,933	3,947	2,427	11,851	1,220	3,830
R-Squared	0.87	0.83	0.71	0.85	0.95	0.89
R-Squared -Within	0.11	0.09	0.07	0.15	0.18	0.08

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: The Core {0,1} dummy is replaced with Agglomeration {0,1} which switches to 1 for households living within 50km of an "urban center" demarcated as such in the year 2000 (see manuscript).

Table C.42: Robustness Check: Core ≤ 25km (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.76]	[2.09]
Border [0,1] $\times$ EAC 1[2002-2004]	-0.082	0.048	0.076	0.068	0.008	-0.341
	(0.295)	(0.059)	(0.057)	(0.077)	(0.135)	(0.295)
Border [0,1] $\times$ CU 1[2005-2009]	0.170	0.059	0.042	0.088	0.033	-0.231
	(0.292)	(0.065)	(0.070)	(0.073)	(0.118)	(0.313)
Border [0,1] $\times$ CM 1[t $\geq$ 2010]	0.139***	0.097***	0.077***	-0.002***	0.285***	-0.453***
	(0.282)	(0.039)	(0.079)	(0.075)	(0.086)	(0.292)
Core ≤ 25km {0,1} $\times$ EAC	-0.301***	-0.037	-0.208**	0.035	0.009	-0.021
	(0.076)	(0.061)	(0.087)	(0.076)	(0.079)	(0.075)
Core ≤ 25km {0,1} $\times$ CU	-0.520***	-0.051	0.728*	-0.040	-0.049	-0.244**
	(0.098)	(0.061)	(0.423)	(0.071)	(0.058)	(0.105)
Core ≤ 25km {0,1} $\times$ CM	-0.375***	-0.120***	1.195***	-0.342***	-0.101***	0.070***
	(0.062)	(0.090)	(0.440)	(0.126)	(0.093)	(0.080)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.12	0.13	0.34	0.23	0.26	0.10
R-Squared -Within	0.10	0.10	0.32	0.23	0.26	0.10

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: The spatial cut-off criteria for living in core agglomerations is reduced to 25km.

Table C.43: Robustness Check: Core ≤ 25km (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.11]	[0.18]	[0.00]	[0.78]	[1.51]	[0.52]
Border [0,1] × EAC 1[2002-2004]	-0.291	-0.166**	0.088	0.117**	-0.239	-0.194
	(0.315)	(0.082)	(0.123)	(0.059)	(0.167)	(0.129)
Border [0,1] × CU 1[2005-2009]	-0.363	-0.144***	0.078	0.119***	-0.293***	-0.274***
	(0.226)	(0.053)	(0.071)	(0.042)	(0.106)	(0.083)
Border [0,1] × CM 1[t ≥ 2010]	-0.546*	-0.036	0.048***	0.082**	-0.153	-0.142
	(0.209)	(0.039)	(0.076)	(0.045)	(0.073)	(0.090)
Core ≤ 25km {0,1} × EAC	0.411**	0.102**	-0.028	0.061	-0.016	0.020
	(0.172)	(0.047)	(0.293)	(0.041)	(0.076)	(0.035)
Core ≤ 25km {0,1} × CU	0.220**	0.060	0.759***	0.011	0.040	0.070**
	(0.108)	(0.049)	(0.160)	(0.044)	(0.132)	(0.036)
Core ≤ 25km {0,1} × CM	0.296*	-0.004	1.056***	0.061**	-0.023	0.012
	(0.162)	(0.049)	(0.180)	(0.030)	(0.084)	(0.040)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.34	0.17	0.37	0.23	0.26	0.17
R-Squared -Within	0.33	0.11	0.36	0.23	0.26	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: The spatial cut-off criteria for living in core agglomerations is reduced to 25km.

Table C.44: Robustness Check: Core ≤ 25km (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
Border [0,1] × EAC 1[2004]	-854.539	187.152	0.580*	0.070	0.107	-0.920**
	(1072.198)	(754.481)	(0.331)	(0.049)	(0.261)	(0.390)
Border [0,1] × CU 1[2010]	-1319.662		0.432		-0.002	-1.721***
	(1154.706)		(0.305)		(0.403)	(0.585)
Core ≤ 25km {0,1} × EAC 1[2004]	170.821***	706.607***	-0.013	0.026***	0.132***	0.930***
	(42.732)	(192.323)	(0.023)	(0.008)	(0.020)	(0.102)
Core ≤ 25km {0,1} × CU 1[2010]	276.569***		-0.037		0.074	0.987***
	(58.235)		(0.026)		(0.068)	(0.148)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,492	2,695	24,972	14,254	15,685	2,292
Observations - Fixed Effects	3,816	2,363	12,747	6,988	6,253	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.44
R-Squared -Within	0.12	0.09	0.04	0.01	0.14	0.42

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: The spatial cut-off criteria for living in core agglomerations is reduced to 25km.

Table C.45: Robustness Check: Core ≤ 25km (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
Border [0,1] × EAC 1[2004]	-177.656	-671.700	7592.679	-0.201	-96.031	5.806
	(237.098)	(1053.982)	(14152.494)	(0.377)	(137.206)	(9.285)
Border [0,1] × CU 1[2010]	-388.050	-913.067		-0.234		6.415
	(236.433)	(1116.349)		(0.344)		(9.148)
Core ≤ 25km {0,1} × EAC 1[2004]	122.246***	35.774	4013.722***	-0.058	8.109	-0.995
	(38.892)	(36.204)	(1314.084)	(0.047)	(25.258)	(1.439)
Core ≤ 25km {0,1} × CU 1[2010]	131.142***	135.227***		-0.007		-0.351
	(39.002)	(39.843)		(0.044)		(1.459)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,494	5,524	2,695	16,330	1,782	5,411
Observations - Fixed Effects	3,817	3,830	2,363	11,599	1,190	3,763
R-Squared	0.88	0.83	0.69	0.86	0.94	0.89
R-Squared -Within	0.10	0.09	0.07	0.16	0.18	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: The spatial cut-off criteria for living in core agglomerations is reduced to 25km.

Table C.46: Robustness Check: Core ≤ 10km (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.76]	[2.09]
Border [0,1] × EAC 1[2002-2004]	-0.089	0.051	0.091	0.075	0.017	-0.342
	(0.294)	(0.059)	(0.065)	(0.078)	(0.136)	(0.295)
Border [0,1] × CU 1[2005-2009]	0.151	0.060	0.061	0.090	0.039	-0.237
	(0.291)	(0.065)	(0.074)	(0.073)	(0.118)	(0.313)
Border [0,1] × CM 1[t ≥ 2010]	0.125	0.096**	0.147	-0.018	0.306***	-0.461
	(0.281)	(0.040)	(0.102)	(0.076)	(0.087)	(0.292)
Core ≤ 10km {0,1} × EAC	-0.309***	-0.090	0.098	0.007	-0.192***	-0.105
	(0.095)	(0.060)	(0.132)	(0.057)	(0.066)	(0.076)
Core ≤ 10km {0,1} × CU	-0.496***	-0.055	0.914*	-0.055	-0.209***	-0.341***
	(0.102)	(0.044)	(0.529)	(0.041)	(0.066)	(0.112)
Core ≤ 10km {0,1} × CM	-0.368***	-0.126	1.554***	-0.351***	-0.313***	0.091
	(0.068)	(0.079)	(0.597)	(0.097)	(0.114)	(0.082)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.12	0.13	0.36	0.23	0.25	0.09
R-Squared -Within	0.10	0.10	0.34	0.23	0.25	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: The spatial cut-off criteria for living in core agglomerations is reduced to 10km.

Table C.47: Robustness Check: Core ≤ 10km (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.11]	[0.18]	[0.00]	[0.78]	[1.51]	[0.52]
Border [0,1] × EAC 1[2002-2004]	-0.235	-0.138	0.134	0.116**	-0.212	-0.172
	(0.314)	(0.095)	(0.113)	(0.059)	(0.184)	(0.142)
Border [0,1] × CU 1[2005-2009]	-0.326	-0.132**	0.129	0.118***	-0.278**	-0.265***
	(0.232)	(0.055)	(0.081)	(0.041)	(0.113)	(0.087)
Border [0,1] × CM 1[t ≥ 2010]	-0.492**	-0.027	0.119*	0.082*	-0.132*	-0.129
	(0.212)	(0.039)	(0.071)	(0.044)	(0.073)	(0.091)
Core ≤ 10km {0,1} × EAC	0.420**	0.031	-0.110	0.062	-0.083	-0.019
	(0.175)	(0.043)	(0.330)	(0.041)	(0.058)	(0.062)
Core ≤ 10km {0,1} × CU	0.156	0.030	0.894***	-0.003	-0.008	0.046
	(0.108)	(0.029)	(0.249)	(0.035)	(0.125)	(0.062)
Core ≤ 10km {0,1} × CM	0.270*	-0.024	1.637***	0.052**	-0.087	-0.023
	(0.149)	(0.027)	(0.600)	(0.024)	(0.054)	(0.075)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.32	0.16	0.41	0.23	0.26	0.16
R-Squared -Within	0.31	0.11	0.41	0.23	0.26	0.16

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: The spatial cut-off criteria for living in core agglomerations is reduced to 10km.

Table C.48: Robustness Check: Core ≤ 10km (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
Border [0,1] × EAC 1[2004]	-885.490	187.211	0.575*	0.067	0.134	-1.111**
	(1142.344)	(767.899)	(0.321)	(0.051)	(0.280)	(0.474)
Border [0,1] × CU 1[2010]	-1299.896		0.440		0.083	-2.064***
	(1227.192)		(0.303)		(0.485)	(0.651)
Core ≤ 10km {0,1} × EAC 1[2004]	166.469***	694.026***	-0.010	0.027***	0.110***	1.004***
	(42.819)	(149.664)	(0.023)	(0.009)	(0.015)	(0.135)
Core ≤ 10km {0,1} × CU 1[2010]	228.369***		-0.042*		0.009	1.279***
	(58.904)		(0.025)		(0.080)	(0.279)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	NO
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,492	2,695	24,972	14,254	15,685	2,292
Observations - Fixed Effects	3,816	2,363	12,747	6,988	6,253	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.50
R-Squared -Within	0.11	0.09	0.04	0.01	0.13	0.48

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: The spatial cut-off criteria for living in core agglomerations is reduced to 10km.

Table C.49: Robustness Check: Core ≤ 10km (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[1885042.91]	[7.81]
Border [0,1] × EAC 1[2004]	-184.828	-696.169	8211.182	-0.215	-83.871	4.575
	(238.949)	(1110.470)	(15101.789)	(0.378)	(133.064)	(10.346)
Border [0,1] × CU 1[2010]	-375.958	-904.913		-0.275		5.905
	(259.850)	(1165.691)		(0.339)		(10.004)
Core ≤ 10km {0,1} × EAC 1[2004]	117.794***	36.303	3592.924	-0.058	7.206	-0.380
	(34.197)	(45.837)	(2577.006)	(0.042)	(19.772)	(1.427)
Core ≤ 10km {0,1} × CU 1[2010]	116.420***	101.922**		-0.001		-0.116
	(36.746)	(42.044)		(0.037)		(1.467)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,494	5,524	2,695	16,330	1,786	5,411
Observations - Fixed Effects	3,817	3,830	2,363	11,599	1,194	3,763
R-Squared	0.88	0.83	0.69	0.86	1.00	0.89
R-Squared -Within	0.09	0.08	0.06	0.15	0.17	0.08

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: The spatial cut-off criteria for living in core agglomerations is reduced to 10km.

Table C.50: Post-CM Estimate (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.66]	[2.09]
Border {0,1} × EAC	-0.080	0.049	0.098	0.072	0.017	-0.346
	(0.295)	(0.058)	(0.069)	(0.077)	(0.134)	(0.295)
Border {0,1} × CU	0.176	0.059	0.064	0.091	0.041	-0.229
	(0.292)	(0.065)	(0.081)	(0.073)	(0.119)	(0.313)
Border {0,1} × CM 1[2010-2014]	0.039	0.120**	0.084	-0.076	0.379***	-0.611**
	(0.293)	(0.051)	(0.118)	(0.103)	(0.111)	(0.274)
Border {0,1} × post-CM 1[t ≥ 2015]	0.277	0.087**	0.166	0.036	0.244***	-0.252
	(0.284)	(0.039)	(0.108)	(0.078)	(0.093)	(0.334)
Core {0,1} × EAC	-0.261***	-0.013	-0.189***	0.026	0.021	0.042
	(0.077)	(0.050)	(0.052)	(0.067)	(0.078)	(0.082)
Core {0,1} × CU	-0.470***	-0.006	0.731**	0.003	0.056	-0.233**
	(0.097)	(0.045)	(0.335)	(0.055)	(0.091)	(0.105)
Core {0,1} × CM 1[2010-2014]	-0.344***	-0.044	1.263***	-0.256**	-0.081	0.053
	(0.066)	(0.080)	(0.409)	(0.123)	(0.088)	(0.094)
Core {0,1} × post-CM 1[t ≥ 2015]	-0.332***	-0.097	0.897**	-0.259**	0.037	0.027
	(0.066)	(0.070)	(0.393)	(0.104)	(0.071)	(0.082)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.12	0.13	0.28	0.23	0.26	0.10
R-Squared -Within	0.10	0.10	0.25	0.23	0.26	0.10

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: The CM period dummy is split into a two dummies, i.e. timeframes, namely CM 1[2010-2014] and a post-CM time period post-CM 1[t ≥ 2015] to provide a test on the hypothesized transitory shift to a welfare-equalizing equilibrium.

Table C.51: Post-CM Estimate (DHS)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.18]	[0.00]	[0.76]	[1.51]	[0.52]
Border {0,1} × EAC	-0.271	-0.114	0.124	0.118**	-0.158	-0.154
	(0.304)	(0.086)	(0.133)	(0.060)	(0.186)	(0.131)
Border {0,1} × CU	-0.363*	-0.135**	0.095	0.121***	-0.284***	-0.270***
	(0.220)	(0.054)	(0.085)	(0.043)	(0.107)	(0.083)
Border {0,1} × CM 1[2010-2014]	-0.632***	-0.049	0.075	0.125**	-0.191**	-0.108
	(0.226)	(0.032)	(0.110)	(0.055)	(0.090)	(0.097)
Border {0,1} × post-CM 1[t ≥ 2015]	-0.443**	-0.023	0.066	0.026	-0.113	-0.158
	(0.210)	(0.047)	(0.082)	(0.056)	(0.096)	(0.104)
Core {0,1} × EAC	0.422***	0.066*	0.121	0.064*	-0.065	-0.004
	(0.131)	(0.036)	(0.257)	(0.038)	(0.049)	(0.023)
Core {0,1} × CU	0.310***	0.035	0.755***	0.012	0.047	0.052
	(0.104)	(0.038)	(0.184)	(0.044)	(0.109)	(0.032)
Core {0,1} × CM 1[2010-2014]	0.365***	0.009	1.275***	0.080	-0.061	0.062
	(0.114)	(0.037)	(0.217)	(0.050)	(0.072)	(0.054)
Core {0,1} × post-CM 1[t ≥ 2015]	0.349**	0.001	0.636***	0.041	0.034	0.006
	(0.168)	(0.042)	(0.094)	(0.025)	(0.081)	(0.036)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.34	0.16	0.30	0.23	0.26	0.17
R-Squared -Within	0.33	0.11	0.29	0.23	0.26	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: The CM period dummy is split into a two dummies, i.e. timeframes, namely CM 1[2010-2014] and a post-CM time period post-CM 1[t ≥ 2015] to provide a test on the hypothesized transitory shift to a welfare-equalizing equilibrium.

Table C.52: Robustness Check: Excluding "Post-EAC Migrants" (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.30]	[0.14]	[0.00]	[0.70]	[1.39]	[0.50]
Border [0,1] × EAC 1[2002-2004]	-1.316***	-0.017	0.530*	0.522***	-0.090	-0.250
	(0.402)	(0.073)	(0.293)	(0.126)	(0.222)	(0.184)
Border [0,1] × CU 1[2005-2009]	-0.694**	-0.155***	0.178	0.328***	-0.389***	-0.402***
	(0.296)	(0.044)	(0.141)	(0.081)	(0.133)	(0.084)
Border [0,1] × CM 1[t ≥ 2010]	-0.355	0.001	0.041	0.090	-0.159	-0.124
	(0.254)	(0.049)	(0.074)	(0.072)	(0.104)	(0.129)
Core {0,1} × EAC 1[2002-2004]	0.605***	0.066	-0.104	0.110**	-0.037	0.027
	(0.145)	(0.045)	(0.266)	(0.044)	(0.064)	(0.046)
Core {0,1} × CU 1[2005-2009]	0.375***	0.046	0.760***	0.037	0.034	0.137***
	(0.122)	(0.038)	(0.279)	(0.091)	(0.099)	(0.047)
Core {0,1} × CM 1[t ≥ 2010]	0.256**	-0.021	0.734***	0.062	0.023	0.116**
	(0.114)	(0.060)	(0.126)	(0.041)	(0.070)	(0.053)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	33,944	24,740	2,706	34,053	24,715	24,823
R-Squared	0.36	0.14	0.29	0.23	0.27	0.19
R-Squared -Within	0.34	0.10	0.27	0.23	0.27	0.19

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I exclude the individuals in the sample which did not live in the survey location at least three years before the establishment of the EAC (before 1999), i.e. excludes "post-EAC Migrants".

Table C.53: Robustness Check: Excluding “Post-EAC Migrants” (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[491.72]	[114.29]	[0.10]	[0.01]	[1.17]	[0.00]
Border [0,1] × EAC 1[2004]	20.962	195.460	0.492	-0.001	-0.036	-0.838**
	(341.477)	(644.348)	(0.312)	(0.076)	(0.213)	(0.375)
Border [0,1] × CU 1[2010]	-206.044		0.591**		-0.684	-1.762***
	(331.479)		(0.279)		(0.577)	(0.614)
Core {0,1} × EAC 1[2004]	119.373***	772.930***	0.009	0.017*	0.143***	0.930***
	(42.378)	(91.216)	(0.022)	(0.010)	(0.026)	(0.092)
Core {0,1} × CU 1[2010]	225.295***		-0.049*		0.141*	0.962***
	(46.662)		(0.026)		(0.074)	(0.102)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	4,480	2,392	22,531	13,600	14,418	2,010
Observations - Fixed Effects	3,155	2,086	12,318	6,718	5,990	12
R-Squared	0.90	0.97	0.68	0.74	0.84	0.45
R-Squared -Within	0.16	0.09	0.03	0.02	0.07	0.43

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: I exclude the individuals in the sample which did not live in the survey location at least three years before the establishment of the EAC (before 1999), i.e. excludes “post-EAC Migrants”.

Table C.54: Robustness Test: Excluding "Post-EAC Migrants" (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[316.66]	[174.90]	[670.91]	[0.27]	[26.19]	[10.84]
Border [0,1] × EAC 1[2004]	-132.988	157.347	7986.289	0.000	-141.718	-0.544
	(261.808)	(148.747)	(14898.392)	(0.352)	(187.157)	(18.589)
Border [0,1] × CU 1[2010]	-299.371	99.518		-0.074		1.058
	(259.523)	(163.821)		(0.337)		(19.004)
Core {0,1} × EAC 1[2004]	106.040***	0.311	4599.535***	-0.077	9.488	-0.456
	(36.417)	(16.952)	(1202.895)	(0.071)	(25.859)	(2.988)
Core {0,1} × CU 1[2010]	140.531***	74.285**		0.009		-0.171
	(35.606)	(36.188)		(0.047)		(2.939)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	4,482	4,495	2,392	14,166	1,586	2,748
Observations - Fixed Effects	3,157	3,162	2,086	10,659	1,037	2,277
R-Squared	0.88	0.88	0.68	0.89	0.94	0.96
R-Squared -Within	0.11	0.13	0.08	0.17	0.11	0.24

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: I exclude the individuals in the sample which did not live in the survey location at least three years before the establishment of the EAC (before 1999), i.e. excludes "post-EAC Migrants".

Table C.55: Robustness Test: DHS only (excluding AIS, MIS, KAP)(DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.19]	[0.00]	[0.75]	[1.51]	[0.53]
Border [0,1] × EAC 1[2002-2004]	-0.284	-0.129	0.365	0.413***	-0.184	-0.163
	(0.312)	(0.089)	(0.284)	(0.113)	(0.188)	(0.136)
Border [0,1] × CU 1[2005-2009]	-0.372*	-0.138**	0.140	0.161***	-0.291***	-0.269***
	(0.224)	(0.054)	(0.140)	(0.054)	(0.106)	(0.097)
Border [0,1] × CM 1[t ≥ 2010]	-0.547***	-0.034***	0.084***	0.098***	-0.149***	-0.141***
	(0.202)	(0.040)	(0.105)	(0.049)	(0.075)	(0.092)
Core {0,1} × EAC 1[2002-2004]	0.433***	0.073*	-0.133	0.096***	-0.046	0.000
	(0.141)	(0.040)	(0.270)	(0.027)	(0.062)	(0.027)
Core {0,1} × CU 1[2005-2009]	0.314***	0.037	0.665**	0.019	0.058	0.073*
	(0.105)	(0.039)	(0.267)	(0.059)	(0.111)	(0.042)
Core {0,1} × CM 1[t ≥ 2010]	0.361***	0.006***	1.045***	0.076***	0.007***	0.029***
	(0.142)	(0.041)	(0.112)	(0.038)	(0.088)	(0.035)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	4,843	125,539	95,717	96,277
R-Squared	0.34	0.16	0.29	0.22	0.26	0.17
R-Squared -Within	0.33	0.11	0.28	0.22	0.26	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: Excludes DHS' special survey rounds including the AIS, KAP and MIS surveys.

Table C.56: Robustness Check: Logged Dependent Variables (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Log(1+ Annual p.c. Consumption (TZS '000)	Log(1+ Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Log(1+ Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[6.06]	[112.23]	-	-	-	[0.60]
Border [0,1] * EAC 1[2004	0.008	-0.130	-	-	-	-1.313***
	(0.932)	(3.355)				(0.354)
Border [0,1] * CU 1[2010]	-0.835	-	-	-	-	-2.066***
	(0.869)					(0.576)
Core {0,1} * EAC 1[2004]	0.216*	0.318	-	-	-	1.185***
	(0.114)	(0.349)				(0.025)
Core {0,1} * CU 1[2010]	0.306***	-	-	-	-	1.052***
	(0.113)					(0.032)
Individual Controls	YES	YES	-	-	-	NO
Geographic Controls	YES	YES	-	-	-	YES
Individual Fixed Effects	NO	NO	-	-	-	NO
Household Fixed Effects	YES	YES	-	-	-	NO
Country-Year Fixed Effect	YES	YES	-	-	-	YES
Observations	5,492	2,695	-	-	-	2,292
Observations - Fixed Effec	3,816	2,363	-	-	-	12
R-Squared	0.90	0.95	-	-	-	0.97
R-Squared -Within	0.14	0.08	-	-	-	0.48

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: Relevant dependent variables are "logged", i.e. are transformed in the way log(1+Y).

Table C.57: Robustness Check: Logged Dependent Variables (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Log(1+ Ann. Food Consumption (TZS '000)	Log(1+ Ann. non-Food Consumption (TZS '000)	Log(1+ Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Log(1+ Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[5.62]	[4.89]	[4.18]	-	[2.79]	-
Border [0,1] × EAC 1[2004]	-0.179	0.293	1.416	-	-4.955	-
	(1.032)	(1.305)	(4.798)		(4.548)	
Border [0,1] × CU 1[2010]	-1.131	-0.099	-	-	-	-
	(0.978)	(1.314)				
Core {0,1} × EAC 1[2004]	0.271*	0.025	0.547	-	0.269	-
	(0.140)	(0.139)	(0.566)		(0.495)	
Core {0,1} × CU 1[2010]	0.300**	0.154	-	-	-	-
	(0.126)	(0.180)				
Individual Controls	YES	YES	YES	-	YES	-
Geographic Controls	YES	YES	YES	-	YES	-
Individual Fixed Effects	NO	NO	NO	-	YES	-
Household Fixed Effects	YES	YES	YES	-	NO	-
Country-Year Fixed Effects	YES	YES	YES	-	YES	-
Observations	5,494	5,524	2,695	-	1,782	-
Observations - Fixed Effects	3,817	3,830	2,363	-	1,190	-
R-Squared	0.86	0.90	0.96	-	0.89	-
R-Squared -Within	0.09	0.13	0.14	-	0.04	-

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: Relevant dependent variables are "logged", i.e. are transformed in the way log(1+Y).

Table C.58: Robustness Check: Beeline Distance to Border Crossings (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.76]	[2.09]
(Beeline to) Border [0,1] × EAC	0.096	0.022	0.080	0.012	-0.052	-0.421
	(0.295)	(0.054)	(0.062)	(0.068)	(0.131)	(0.325)
(Beeline to) Border [0,1] × CU	0.201	0.097	0.177	0.113	0.127	-0.211
	(0.327)	(0.063)	(0.124)	(0.082)	(0.107)	(0.346)
(Beeline to) Border [0,1] × CM	0.221	0.063	0.264	-0.051	0.256***	-0.433
	(0.259)	(0.041)	(0.168)	(0.075)	(0.088)	(0.288)
Core {0,1} × EAC 1[2002-2004]	-0.266***	-0.009	-0.173***	0.031	0.017	0.068
	(0.077)	(0.050)	(0.049)	(0.068)	(0.081)	(0.088)
Core {0,1} × CU 1[2005-2009]	-0.450***	0.006	0.750**	0.016	0.070	-0.178
	(0.107)	(0.050)	(0.321)	(0.062)	(0.097)	(0.110)
Core {0,1} × CM 1[t ≥ 2010]	-0.331***	-0.066	1.098***	-0.255**	0.016	0.062
	(0.066)	(0.073)	(0.357)	(0.108)	(0.095)	(0.079)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.12	0.13	0.28	0.23	0.25	0.10
R-Squared -Within	0.10	0.10	0.25	0.23	0.25	0.10

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I replace road distances to the border with "beeline" (as the crow flies) distances.

Table C.59: Robustness Test: Beeline Distance to Border Crossings (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.18]	[0.00]	[0.76]	[1.51]	[0.52]
(Beeline to) Border [0,1] × EAC	-0.160	-0.141	-0.042	0.076	-0.179	-0.131
	(0.312)	(0.091)	(0.139)	(0.054)	(0.194)	(0.135)
(Beeline to) Border [0,1] × CU	-0.277	-0.135***	0.067	0.121***	-0.207*	-0.141
	(0.265)	(0.045)	(0.143)	(0.043)	(0.118)	(0.101)
(Beeline to) Border [0,1] × CM	-0.550**	-0.066**	-0.027	0.083*	-0.178**	-0.182**
	(0.232)	(0.033)	(0.138)	(0.045)	(0.071)	(0.092)
Core {0,1} × EAC 1[2002-2004]	0.381**	0.058*	0.103	0.080**	-0.073	-0.024
	(0.155)	(0.031)	(0.266)	(0.040)	(0.047)	(0.029)
Core {0,1} × CU 1[2005-2009]	0.271**	0.019	0.767***	0.030	0.023	0.025
	(0.122)	(0.031)	(0.178)	(0.039)	(0.098)	(0.034)
Core {0,1} × CM 1[t ≥ 2010]	0.292**	0.000	0.934***	0.072**	-0.018	0.006
	(0.138)	(0.036)	(0.120)	(0.031)	(0.078)	(0.040)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,364	71,753	7,696	169,967	95,736	98,982
R-Squared	0.34	0.16	0.30	0.23	0.26	0.17
R-Squared -Within	0.33	0.11	0.29	0.23	0.26	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I replace road distances to the border with "beeline" (as the crow flies) distances.

Table C.60: Robustness Check: Beeline Distance to Border Crossings (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
(Beeline to) Border [0,1] × EAC	-816.913	69.400	0.563**	0.051	0.078	-0.710**
	(1112.207)	(715.743)	(0.279)	(0.044)	(0.238)	(0.312)
(Beeline to) Border [0,1] × CU	-1283.327		0.435		-0.033	-1.567***
	(1208.457)		(0.278)		(0.388)	(0.545)
Core {0,1} × EAC 1[2004]	154.777***	717.253***	-0.004	0.027***	0.137***	0.857***
	(19.851)	(132.421)	(0.020)	(0.008)	(0.018)	(0.058)
Core {0,1} × CU 1[2010]	254.967***		-0.032		0.074	0.864***
	(41.741)		(0.022)		(0.065)	(0.099)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,492	2,695	24,972	14,254	15,685	2,292
Observations - Fixed Effects	3,816	2,363	12,747	6,988	6,253	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.43
R-Squared -Within	0.12	0.09	0.04	0.01	0.14	0.41

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: I replace road distances to the border with "beeline" (as the crow flies) distances.

Table C.61: Robustness Check: Beeline Distance to Border Crossings (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
(Beeline to) Border [0,1] × EAC	-111.273	-697.103	4638.227	-0.180	-106.417	3.003
	(210.184)	(1104.755)	(12272.383)	(0.339)	(145.479)	(8.924)
(Beeline to) Border [0,1] × CU	-335.069	-924.540		-0.234		3.458
	(212.127)	(1172.534)		(0.319)		(8.728)
Core {0,1} × EAC 1[2004]	116.354***	25.677	4250.588***	-0.061	6.974	-0.861
	(35.359)	(17.989)	(1163.358)	(0.044)	(23.486)	(1.397)
Core {0,1} × CU 1[2010]	123.575***	121.374***		-0.010		-0.192
	(36.280)	(18.776)		(0.041)		(1.418)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,494	5,524	2,695	16,330	1,782	5,411
Observations - Fixed Effects	3,817	3,830	2,363	11,599	1,190	3,763
R-Squared	0.88	0.83	0.69	0.86	0.94	0.89
R-Squared -Within	0.10	0.10	0.07	0.16	0.18	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: I replace road distances to the border with "beeline" (as the crow flies) distances.

Table C.62: Robustness Check: Beeline Distance to Border Line (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.76]	[2.09]
(Beeline to) Borderline [0,1] × EAC	-0.045	0.062	0.098*	0.078	0.054	-0.402
	(0.282)	(0.058)	(0.060)	(0.077)	(0.129)	(0.297)
(Beeline to) Borderline [0,1] × CU	0.144	0.051	0.075	0.072	0.025	-0.235
	(0.287)	(0.059)	(0.079)	(0.071)	(0.108)	(0.309)
(Beeline to) Borderline [0,1] × CM	0.213	0.074**	0.131	-0.054	0.262***	-0.432
	(0.258)	(0.037)	(0.106)	(0.071)	(0.080)	(0.284)
Core {0,1} × EAC 1[2002-2004]	-0.270***	-0.011	-0.184***	0.029	0.022	0.064
	(0.077)	(0.051)	(0.051)	(0.068)	(0.080)	(0.083)
Core {0,1} × CU 1[2005-2009]	-0.466***	-0.003	0.733**	0.007	0.059	-0.204**
	(0.094)	(0.046)	(0.332)	(0.057)	(0.091)	(0.103)
Core {0,1} × CM 1[t ≥ 2010]	-0.340***	-0.071	1.090***	-0.256**	0.001	0.060
	(0.064)	(0.071)	(0.374)	(0.108)	(0.090)	(0.079)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.12	0.13	0.27	0.23	0.25	0.09
R-Squared -Within	0.10	0.10	0.25	0.23	0.25	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I replace road distances to the border with shortest "beeline" (as the crow flies) distance to the entire borderline.

Table C.63: Robustness Check: Beeline Distance to Border Line (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.18]	[0.00]	[0.76]	[1.51]	[0.52]
(Beeline to) Borderline [0,1] × EAC	-0.248	-0.152**	0.026	0.079	-0.168	-0.152
	(0.296)	(0.077)	(0.132)	(0.055)	(0.172)	(0.123)
(Beeline to) Borderline [0,1] × CU	-0.368	-0.106**	0.084	0.112**	-0.211*	-0.220***
	(0.235)	(0.049)	(0.107)	(0.044)	(0.108)	(0.084)
(Beeline to) Borderline [0,1] × CM	-0.559***	-0.024	-0.005	0.063	-0.102	-0.131
	(0.199)	(0.036)	(0.109)	(0.045)	(0.064)	(0.083)
Core {0,1} × EAC 1[2002-2004]	0.418***	0.069*	0.142	0.073*	-0.055	-0.009
	(0.142)	(0.036)	(0.272)	(0.039)	(0.056)	(0.025)
Core {0,1} × CU 1[2005-2009]	0.301***	0.033	0.778***	0.019	0.044	0.038
	(0.108)	(0.035)	(0.180)	(0.042)	(0.105)	(0.032)
Core {0,1} × CM 1[t ≥ 2010]	0.338**	0.007	0.958***	0.065**	0.001	0.023
	(0.134)	(0.041)	(0.116)	(0.031)	(0.086)	(0.035)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,364	71,753	7,696	169,967	95,736	98,982
R-Squared	0.34	0.16	0.30	0.23	0.26	0.17
R-Squared -Within	0.33	0.11	0.29	0.23	0.26	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I replace road distances to the border with shortest "beeline" (as the crow flies) distance to the entire borderline.

Table C.64: Robustness Check: Beeline Distance to Border Line (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
(Beeline to) Borderline [0,1] × EAC	-757.039	122.074	0.490*	0.041	0.062	-0.570*
	(998.849)	(577.684)	(0.261)	(0.044)	(0.234)	(0.318)
(Beeline to) Borderline [0,1] × CU	-1199.598		0.389		-0.043	-1.360***
	(1090.390)		(0.255)		(0.374)	(0.500)
Core {0,1} × EAC 1[2004]	162.273***	716.850***	-0.008	0.027***	0.137***	0.861***
	(33.694)	(120.817)	(0.022)	(0.008)	(0.019)	(0.062)
Core {0,1} × CU 1[2010]	267.119***		-0.036		0.076	0.872***
	(51.500)		(0.023)		(0.067)	(0.102)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,492	2,695	24,972	14,254	15,685	2,292
Observations - Fixed Effects	3,816	2,363	12,747	6,988	6,253	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.43
R-Squared -Within	0.12	0.09	0.04	0.01	0.14	0.41

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: I replace road distances to the border with shortest "beeline" (as the crow flies) distance to the entire borderline.

Table C.65: Robustness Check: Beeline Distance to Border Line (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
(Beeline to) Borderline [0,1] × EAC	-107.710	-638.292	4075.452	-0.135	-92.091	4.749
	(207.271)	(981.794)	(11014.623)	(0.327)	(133.728)	(8.546)
(Beeline to) Borderline [0,1] × CU	-323.756	-853.098		-0.203		5.328
	(206.881)	(1047.819)		(0.302)		(8.284)
Core {0,1} × EAC 1[2004]	117.191***	31.893	4207.051***	-0.061	8.250	-1.011
	(37.145)	(27.424)	(1186.314)	(0.048)	(24.091)	(1.419)
Core {0,1} × CU 1[2010]	126.877***	129.708***		-0.009		-0.351
	(38.103)	(29.727)		(0.043)		(1.441)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,494	5,524	2,695	16,330	1,782	5,411
Observations - Fixed Effects	3,817	3,830	2,363	11,599	1,190	3,763
R-Squared	0.88	0.83	0.69	0.86	0.94	0.89
R-Squared -Within	0.10	0.09	0.07	0.16	0.18	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: I replace road distances to the border with shortest "beeline" (as the crow flies) distance to the entire borderline.

Table C.66: Robustness Check: Logged Distance (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.66]	[2.09]
Log(BorderDistance) × EAC	0.029	-0.006	-0.018	-0.006	0.015	0.091
	(0.071)	(0.014)	(0.022)	(0.019)	(0.031)	(0.069)
Log(BorderDistance) × CU	-0.033	-0.023	-0.014	-0.027	-0.015	0.101
	(0.073)	(0.018)	(0.032)	(0.023)	(0.031)	(0.074)
Log(BorderDistance) × CM	-0.009	-0.027***	-0.032	0.007	-0.068***	0.143**
	(0.067)	(0.010)	(0.039)	(0.021)	(0.020)	(0.066)
Log(CoreDistance) × EAC	-0.022	0.026*	0.031	0.020	0.054*	-0.043
	(0.054)	(0.014)	(0.024)	(0.020)	(0.029)	(0.033)
Log(CoreDistance) × CU	0.067	0.016	-0.174**	0.022	0.037**	0.063
	(0.050)	(0.012)	(0.074)	(0.014)	(0.018)	(0.040)
Log(CoreDistance) × CM	0.009	0.030*	-0.302***	0.088***	0.034	-0.025
	(0.054)	(0.016)	(0.093)	(0.025)	(0.026)	(0.030)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.12	0.13	0.30	0.23	0.25	0.10
R-Squared -Within	0.10	0.10	0.28	0.23	0.25	0.10

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I log-transform the road distances to the border in the way (1+BorderDistance).

Table C.67: Robustness Check: Logged Distance (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.18]	[0.00]	[0.76]	[1.51]	[0.52]
Log(BorderDistance) × EAC	0.029	0.034**	-0.077*	-0.028*	0.018	0.046*
	(0.076)	(0.016)	(0.041)	(0.016)	(0.035)	(0.026)
Log(BorderDistance) × CU	0.083	0.040***	-0.053*	-0.024**	0.048	0.075***
	(0.057)	(0.013)	(0.032)	(0.011)	(0.030)	(0.023)
Log(BorderDistance) × CM	0.120**	0.016	-0.066*	-0.015	0.035**	0.048*
	(0.050)	(0.010)	(0.039)	(0.012)	(0.018)	(0.025)
Log(CoreDistance) × EAC	0.004	-0.004	0.057	-0.023**	0.038**	0.008
	(0.042)	(0.006)	(0.043)	(0.011)	(0.017)	(0.017)
Log(CoreDistance) × CU	-0.017	-0.001	-0.179***	-0.011	0.023	-0.010
	(0.027)	(0.008)	(0.055)	(0.011)	(0.025)	(0.011)
Log(CoreDistance) × CM	-0.011	0.007	-0.209***	-0.021**	0.027**	-0.005
	(0.027)	(0.007)	(0.057)	(0.009)	(0.013)	(0.018)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.34	0.16	0.31	0.23	0.25	0.17
R-Squared -Within	0.33	0.11	0.30	0.23	0.25	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I log-transform the road distances to the border in the way (1+BorderDistance).

Table C.68: Robustness Check: Logged Distance (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
Log(BorderDistance) × EAC	-14.571	-305.802	-0.051	-0.020*	-0.071	-0.404*
	(96.504)	(499.702)	(0.043)	(0.011)	(0.052)	(0.225)
Log(BorderDistance) × CU	24.227		-0.033		-0.040	-0.330
	(115.177)		(0.038)		(0.071)	(0.221)
Log(CoreDistance) × EAC	-2817.106	-3141.332	0.399	-0.074	-0.723	-4.660**
	(3095.358)	(6222.287)	(0.529)	(0.137)	(0.697)	(2.237)
Log(CoreDistance) × CU	-2850.940		0.390		-0.711	-4.669**
	(3096.469)		(0.527)		(0.696)	(2.230)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,475	2,678	24,726	14,018	15,433	2,288
Observations - Fixed Effects	3,800	2,347	12,660	6,899	6,161	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.49
R-Squared -Within	0.12	0.05	0.03	0.01	0.14	0.47

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: I log-transform the road distances to the border in the way (1+BorderDistance).

Table C.69: Robustness Check: Logged Distance (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
Log(BorderDistance) × EAC	-22.602	13.234	-3084.228	0.014	9.795	-1.344
	(56.181)	(62.018)	(3506.668)	(0.060)	(19.092)	(1.748)
Log(BorderDistance) × CU	17.695	8.645		0.025		-1.897
	(62.345)	(68.685)		(0.041)		(1.590)
Log(CoreDistance) × EAC	-519.714	-2241.741	-28757.050	-0.206	49.406	-22.758
	(727.313)	(2833.684)	(45558.582)	(0.658)	(300.652)	(21.510)
Log(CoreDistance) × CU	-483.284	-2312.587		-0.168		-23.282
	(727.883)	(2835.007)		(0.657)		(21.488)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,477	5,507	2,678	16,245	1,767	5,354
Observations - Fixed Effects	3,801	3,814	2,347	11,553	1,180	3,712
R-Squared	0.88	0.83	0.69	0.86	0.94	0.89
R-Squared -Within	0.10	0.10	0.06	0.16	0.18	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following amendments: I log-transform the road distances to the border in the way (1+BorderDistance).

Table C.70: Robustness Check: Border ≤ 100km (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.76]	[2.09]
Border ≤ 100km {0,1} × EAC	-0.137	0.002	-0.022	0.011	0.007	-0.125
	(0.135)	(0.023)	(0.041)	(0.039)	(0.066)	(0.165)
Border ≤ 100km {0,1} × CU	-0.019	0.034	-0.070	0.075	0.116*	-0.179
	(0.156)	(0.035)	(0.050)	(0.051)	(0.065)	(0.169)
Border ≤ 100km {0,1} × CM	-0.032	0.025	-0.035	-0.017	0.153***	-0.201
	(0.131)	(0.022)	(0.056)	(0.048)	(0.058)	(0.153)
Core {0,1} × EAC	-0.292***	-0.010	-0.189***	0.033	0.020	0.010
	(0.082)	(0.049)	(0.051)	(0.066)	(0.078)	(0.091)
Core {0,1} × CU	-0.457***	0.005	0.722**	0.024	0.085	-0.268**
	(0.110)	(0.044)	(0.332)	(0.055)	(0.089)	(0.117)
Core {0,1} × CM	-0.331***	-0.065	1.083***	-0.261**	0.037	-0.011
	(0.070)	(0.070)	(0.377)	(0.110)	(0.086)	(0.082)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.13	0.14	0.27	0.23	0.25	0.09
R-Squared -Within	0.11	0.10	0.25	0.23	0.25	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 100km to the border (using road distances).

Table C.71: Robustness Check: Border ≤ 100km (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.18]	[0.00]	[0.76]	[1.51]	[0.52]
Border ≤ 100km {0,1} × EAC	-0.421	-0.134***	-0.146*	0.023	-0.181***	-0.204***
	(0.261)	(0.035)	(0.084)	(0.038)	(0.063)	(0.025)
Border ≤ 100km {0,1} × CU	-0.345**	-0.099***	-0.051	0.038	-0.144***	-0.177***
	(0.167)	(0.023)	(0.058)	(0.023)	(0.048)	(0.034)
Border ≤ 100km {0,1} × CM	-0.379**	-0.065***	-0.056	0.024	-0.133***	-0.159***
	(0.150)	(0.015)	(0.056)	(0.025)	(0.034)	(0.049)
Core {0,1} × EAC	0.360**	0.041	0.138	0.075*	-0.090	-0.047**
	(0.157)	(0.041)	(0.275)	(0.039)	(0.060)	(0.020)
Core {0,1} × CU	0.249*	0.011	0.775***	0.026	0.017	0.006
	(0.128)	(0.040)	(0.181)	(0.045)	(0.113)	(0.032)
Core {0,1} × CM	0.272	-0.006	0.957***	0.068**	-0.024	-0.005
	(0.169)	(0.041)	(0.116)	(0.033)	(0.089)	(0.027)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.34	0.16	0.29	0.23	0.26	0.17
R-Squared -Within	0.33	0.11	0.28	0.23	0.26	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 100km to the border (using road distances).

Table C.72: Robustness Check: Border ≤ 100km (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
Border ≤ 100km {0,1} × EAC	-43.826	21.219	0.033	0.009	0.017	0.038
	(85.443)	(83.145)	(0.042)	(0.010)	(0.045)	(0.063)
Border ≤ 100km {0,1} × CU	-105.265		0.022		0.127	-0.085
	(108.035)		(0.041)		(0.089)	(0.102)
Core {0,1} × EAC	147.835***	688.652***	-0.001	0.024**	0.127***	0.809***
	(53.216)	(86.586)	(0.028)	(0.010)	(0.010)	(0.069)
Core {0,1} × CU	264.581***		-0.027		-0.002	0.844***
	(74.478)		(0.030)		(0.082)	(0.118)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,492	2,695	24,972	14,254	15,685	2,292
Observations - Fixed Effects	3,816	2,363	12,747	6,988	6,253	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.43
R-Squared -Within	0.12	0.09	0.03	0.01	0.14	0.40

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 100km to the border (using road distances).

Table C.73: Robustness Check: Border ≤ 100km (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
Border ≤ 100km {0,1} × EAC	-5.045	-38.379	1207.398	-0.019	-8.716	2.250
	(46.765)	(67.330)	(1921.629)	(0.068)	(26.978)	(1.438)
Border ≤ 100km {0,1} × CU	-36.354	-68.217		-0.040		2.278*
	(50.035)	(86.151)		(0.064)		(1.312)
Core {0,1} × EAC	114.786**	20.077	3679.881***	-0.059	10.154	-2.264
	(50.173)	(27.673)	(1389.983)	(0.066)	(21.925)	(1.573)
Core {0,1} × CU	131.199**	123.391***		0.002		-1.594
	(50.987)	(39.155)		(0.057)		(1.520)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,494	5,524	2,695	16,330	1,782	5,411
Observations - Fixed Effects	3,817	3,830	2,363	11,599	1,190	3,763
R-Squared	0.88	0.83	0.69	0.86	0.95	0.89
R-Squared -Within	0.10	0.09	0.07	0.15	0.20	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 100km to the border (using road distances).

Table C.74: Robustness Check: Border ≤ 50km (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.76]	[2.09]
Border ≤ 50km {0,1} × EAC	-0.190	0.029	0.047*	0.009	-0.093	0.078
	(0.149)	(0.024)	(0.026)	(0.032)	(0.061)	(0.116)
Border ≤ 50km {0,1} × CU	-0.035	0.095***	0.005	0.050	-0.019	0.039
	(0.122)	(0.032)	(0.039)	(0.058)	(0.096)	(0.146)
Border ≤ 50km {0,1} × CM	-0.177	0.077***	0.050*	0.012	0.091	-0.129
	(0.145)	(0.025)	(0.029)	(0.058)	(0.056)	(0.127)
Core {0,1} × EAC	-0.275***	-0.009	-0.185***	0.030	0.015	0.026
	(0.080)	(0.048)	(0.051)	(0.065)	(0.078)	(0.090)
Core {0,1} × CU	-0.457***	0.004	0.734**	0.010	0.059	-0.244**
	(0.103)	(0.044)	(0.332)	(0.054)	(0.088)	(0.113)
Core {0,1} × CM	-0.339***	-0.065	1.092***	-0.259**	0.017	0.005
	(0.068)	(0.068)	(0.376)	(0.109)	(0.087)	(0.085)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.13	0.14	0.27	0.23	0.25	0.09
R-Squared -Within	0.11	0.10	0.25	0.23	0.25	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 50km to the border (using road distances).

Table C.75: Robustness Check: Border ≤ 50km (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.18]	[0.00]	[0.76]	[1.51]	[0.52]
Border ≤ 50km {0,1} × EAC	0.040	-0.119**	-0.055	0.037	-0.086	-0.093***
	(0.214)	(0.050)	(0.078)	(0.041)	(0.061)	(0.033)
Border ≤ 50km {0,1} × CU	-0.048	-0.135***	0.038	0.010	-0.104**	-0.145***
	(0.154)	(0.034)	(0.064)	(0.028)	(0.047)	(0.043)
Border ≤ 50km {0,1} × CM	-0.092	-0.093***	0.013	0.020	-0.117***	-0.121*
	(0.172)	(0.024)	(0.045)	(0.026)	(0.023)	(0.063)
Core {0,1} × EAC	0.452***	0.063	0.163	0.073*	-0.057	-0.011
	(0.161)	(0.043)	(0.277)	(0.037)	(0.068)	(0.027)
Core {0,1} × CU	0.316**	0.022	0.786***	0.019	0.037	0.031
	(0.125)	(0.042)	(0.182)	(0.046)	(0.116)	(0.036)
Core {0,1} × CM	0.345**	-0.001	0.968***	0.064**	-0.006	0.016
	(0.164)	(0.043)	(0.115)	(0.032)	(0.093)	(0.032)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.34	0.16	0.29	0.23	0.26	0.17
R-Squared -Within	0.32	0.11	0.28	0.23	0.26	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 50km to the border (using road distances).

Table C.76: Robustness Check: Border ≤ 50km (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.00]
Border ≤ 50km {0,1} × EAC	49.969	-2.821	-0.004	0.035***	0.047	-0.143*
	(292.688)	(77.036)	(0.035)	(0.009)	(0.051)	(0.078)
Border ≤ 50km {0,1} × CU	-13.598		-0.033		-0.022	-0.164*
	(306.353)		(0.033)		(0.149)	(0.096)
Core {0,1} × EAC	118.770***	706.221***	0.020	0.032***	0.145***	0.808***
	(14.381)	(27.153)	(0.023)	(0.006)	(0.016)	(0.023)
Core {0,1} × CU	196.942***		-0.013		0.079	0.799***
	(59.110)		(0.025)		(0.063)	(0.117)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,492	2,695	24,972	14,254	15,685	2,292
Observations - Fixed Effects	3,816	2,363	12,747	6,988	6,253	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.43
R-Squared -Within	0.11	0.09	0.03	0.01	0.14	0.40

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 50km to the border (using road distances).

Table C.77: Robustness Test: Border $\leq$ 50km (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[28.83]	[7.81]
Border $\leq$ 50km {0,1} $\times$ EAC	0.159	46.895	1521.906	0.052	13.350	-2.628
	(186.972)	(107.532)	(1408.312)	(0.089)	(24.030)	(2.832)
Border $\leq$ 50km {0,1} $\times$ CU	-33.556	20.242		0.016		-2.064
	(215.623)	(97.817)		(0.058)		(2.874)
Core {0,1} $\times$ EAC	111.672***	-5.570	4526.138***	-0.069*	2.537	-0.820
	(30.388)	(28.461)	(968.869)	(0.038)	(24.289)	(1.357)
Core {0,1} $\times$ CU	108.651***	78.868**		-0.021		-0.138
	(37.305)	(32.295)		(0.037)		(1.365)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,494	5,524	2,695	16,330	1,782	5,411
Observations - Fixed Effects	3,817	3,830	2,363	11,599	1,190	3,763
R-Squared	0.88	0.83	0.69	0.86	0.94	0.89
R-Squared -Within	0.09	0.09	0.07	0.15	0.17	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 50km to the border (using road distances).

Table C.78: Robustness Check: Border ≤ 25km (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.76]	[2.09]
Border ≤ 25km {0,1} × EAC	-0.478**	-0.020	0.039	-0.034	-0.176	-0.381
	(0.225)	(0.048)	(0.039)	(0.054)	(0.151)	(0.288)
Border ≤ 25km {0,1} × CU	-0.029	0.150***	-0.002	0.093	0.007	-0.324
	(0.297)	(0.050)	(0.049)	(0.066)	(0.156)	(0.269)
Border ≤ 25km {0,1} × CM	-0.215	0.034	-0.019	-0.097	-0.083	-0.380***
	(0.255)	(0.043)	(0.055)	(0.079)	(0.107)	(0.146)
Core {0,1} × EAC	-0.275***	-0.011	-0.187***	0.030	0.019	0.010
	(0.078)	(0.048)	(0.051)	(0.065)	(0.078)	(0.087)
Core {0,1} × CU	-0.459***	0.000	0.734**	0.008	0.061	-0.255**
	(0.099)	(0.044)	(0.332)	(0.054)	(0.087)	(0.109)
Core {0,1} × CM	-0.333***	-0.069	1.088***	-0.261**	0.008	0.005
	(0.066)	(0.068)	(0.375)	(0.109)	(0.087)	(0.082)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.13	0.14	0.27	0.23	0.25	0.09
R-Squared -Within	0.11	0.10	0.25	0.23	0.25	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 25km to the border (using road distances).

Table C.79: Robustness Check: Border ≤ 25km (DHS)

	Dependent Variable					
	DHS					
	Consumpt.	Income				
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.14]	[0.18]	[0.00]	[0.76]	[1.51]	[0.52]
Border ≤ 25km {0,1} × EAC	0.548***	-0.051***	-0.083	0.090*	-0.010	-0.064
	(0.121)	(0.018)	(0.088)	(0.049)	(0.072)	(0.044)
Border ≤ 25km {0,1} × CU	0.261***	-0.068***	-0.004	0.039	-0.025	-0.158***
	(0.074)	(0.025)	(0.110)	(0.075)	(0.090)	(0.044)
Border ≤ 25km {0,1} × CM	0.244***	-0.031***	-0.036***	0.066***	-0.063***	-0.088***
	(0.167)	(0.026)	(0.092)	(0.043)	(0.062)	(0.038)
Core {0,1} × EAC	0.460***	0.070	0.166	0.072*	-0.051	-0.007
	(0.155)	(0.043)	(0.276)	(0.037)	(0.068)	(0.029)
Core {0,1} × CU	0.324***	0.029	0.784***	0.019	0.044	0.037
	(0.122)	(0.042)	(0.182)	(0.046)	(0.115)	(0.037)
Core {0,1} × CM	0.356***	0.005***	0.967***	0.064***	0.001***	0.023***
	(0.158)	(0.043)	(0.115)	(0.032)	(0.092)	(0.034)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.34	0.16	0.29	0.23	0.26	0.17
R-Squared -Within	0.32	0.11	0.28	0.23	0.26	0.17

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 25km to the border (using road distances).

Table C.80: Robustness Check: Border ≤ 25km (KHDS 1/2)

	Dependent Variable					
	KHDS					
	Consumpt.		Income			Agglom.
	Annual p.c. Consumption (TZS '000)	Value of dur. Assets (TZS '000)	Employed Work {0,1}	Salaried Work {0,1}	Occupation Level {1,3}	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[552.31]	[112.51]	[0.12]	[0.01]	[1.20]	[0.00]
Border ≤25km {0,1} × EAC	1046.331***	-242.913	-0.162	0.010	0.188	0.034
	(310.431)	(436.378)	(0.122)	(0.030)	(0.210)	(0.066)
Border ≤ 25km {0,1} × CU	1126.525***		-0.037		0.583*	-0.080
	(354.396)		(0.059)		(0.312)	(0.128)
Core {0,1} × EAC	115.804***	705.777***	0.020	0.030***	0.143***	0.816***
	(3.903)	(9.572)	(0.022)	(0.006)	(0.015)	(0.028)
Core {0,1} × CU	198.590***		-0.012		0.080	0.803***
	(31.553)		(0.024)		(0.060)	(0.118)
Individual Controls	YES	YES	YES	YES	YES	NO
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	NO
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	35,813	17,674	24,972	14,254	15,685	2,292
Observations - Fixed Effects	3,816	2,363	12,747	6,988	6,253	12
R-Squared	0.88	0.97	0.64	0.72	0.82	0.43
R-Squared -Within	0.12	0.09	0.03	0.01	0.14	0.40

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 25km to the border (using road distances).

Table C.81: Robustness Check: Border ≤ 25km (KHDS 2/2)

	Dependent Variable					
	KHDS					
	Consumpt.			Income		Agglom.
	Ann. Food Consumption (TZS '000)	Ann. non-Food Consumption (TZS '000)	Value of Dwelling (TZS '000)	Worked Last Year {0,1}	Monthly Salary (TZS '000)	Time lived at Location (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.53]	[203.27]	[660.74]	[0.26]	[28.83]	[7.81]
Border ≤25km {0,1} × EAC	536.903***	493.731***	3041.987	-0.033	-487.518	-0.755
	(166.385)	(181.678)	(3654.443)	(0.139)	(5036999.421)	(2.051)
Border ≤ 25km {0,1} × CU	663.445***	452.421***		-0.104		0.380
	(233.953)	(150.728)		(0.121)		(2.099)
Core {0,1} × EAC	107.507***	-4.390	4497.546***	-0.068*	1.907	-0.734
	(17.149)	(14.237)	(325.789)	(0.038)	(23.209)	(1.355)
Core {0,1} × CU	109.736***	79.415***		-0.020		-0.044
	(21.871)	(14.376)		(0.036)		(1.366)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	35,826	36,025	17,673	16,330	1,782	5,411
Observations - Fixed Effects	3,817	3,830	2,363	11,599	1,190	3,763
R-Squared	0.88	0.83	0.69	0.86	0.94	0.89
R-Squared -Within	0.10	0.09	0.07	0.15	0.17	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.3 such that most of the respective notes apply here. However, note the following important amendments: I replace the continuous distance variable with a dichotomous dummy switching to 1 for individuals within 25km to the border (using road distances).

Table C.82: Robustness Check: Flexible Distance Specification (Afrobarometer)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.	Income		
	Lived Poverty [0,4]	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Cash Inc. Deprivation {0,4}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[0.55]	[1.76]	[2.09]
Cont. Distance to Border × EAC (in '00 km)	0.010	0.013	-0.008	0.016	0.035	0.092
	(0.066)	(0.014)	(0.017)	(0.018)	(0.029)	(0.068)
Sq. Cont. Distance to Border × EAC (in '00	0.000	-0.002*	-0.001	-0.003*	-0.004*	-0.004
	(0.005)	(0.001)	(0.002)	(0.001)	(0.002)	(0.005)
Cont. Distance to Border × CM (in '00 km)	0.006	0.002	-0.015	-0.009	-0.014	0.182**
	(0.071)	(0.017)	(0.039)	(0.020)	(0.027)	(0.072)
Sq. Cont. Distance to Border × CU (in '00	-0.004	-0.001	0.002	0.000	0.001	-0.016***
	(0.006)	(0.001)	(0.003)	(0.002)	(0.002)	(0.006)
Cont. Distance to Border × CM (in '00 km)	-0.004	-0.013	-0.045	-0.001	-0.034	0.165***
	(0.061)	(0.010)	(0.055)	(0.019)	(0.022)	(0.062)
Sq. Cont. Distance to Border × CM (in '00	-0.001	0.000	0.004	0.000	0.000	-0.011**
	(0.005)	(0.001)	(0.004)	(0.001)	(0.002)	(0.005)
Cont. Distance to Core × EAC (in '00 km)	-0.048	0.026	0.084**	0.024	0.056	-0.065
	(0.101)	(0.018)	(0.036)	(0.027)	(0.046)	(0.065)
Sq. Cont. Distance to Core × EAC (in '00	0.001	0.000	-0.008**	0.001	0.001	0.004
	(0.010)	(0.002)	(0.003)	(0.003)	(0.005)	(0.007)
Cont. Distance to Core × CU (in '00 km)	0.126	0.013	-0.205***	0.019	0.015	0.135*
	(0.096)	(0.018)	(0.078)	(0.021)	(0.033)	(0.078)
Sq. Cont. Distance to Core × CU (in '00 km)	-0.012	0.000	0.015**	0.000	0.002	-0.011
	(0.009)	(0.002)	(0.007)	(0.002)	(0.003)	(0.007)
Cont. Distance to Core × CM (in '00 km)	0.031	0.030*	-0.318***	0.069**	0.034	0.005
	(0.096)	(0.018)	(0.110)	(0.031)	(0.038)	(0.054)
Sq. Cont. Distance to Core × CM (in '00 km)	-0.006	-0.001	0.023**	-0.003	-0.001	-0.002
	(0.008)	(0.001)	(0.009)	(0.002)	(0.003)	(0.005)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	4,156	26,563	21,232	35,975
R-Squared	0.13	0.14	0.20	0.23	0.25	0.09
R-Squared -Within	0.11	0.10	0.17	0.23	0.25	0.09

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I use the continuous (road) distances to both border crossings as well as core cities and also add their squared values as a separate regressor interacted with the treatment time periods.

Table C.83: Robustness Test: Flexible Distance Specification (DHS)

	Dependent Variable					
	Afrobarometer					
	Consumpt.	Income	Agglom.		Income	
	Wealth Index {1,5}	Employed Work {0,1}	Population Density (sdz.)	Worked last Year {0,1}	Occupation Level {1,3}	Paid in Cash {0,1}
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.11]	[0.18]	[0.00]	[0.78]	[1.51]	[0.52]
Cont. Distance to Border × EAC (in '00 km)	0.008	0.055**	-0.193***	-0.027*	-0.026	0.068*
	(0.080)	(0.026)	(0.066)	(0.016)	(0.049)	(0.040)
Sq. Cont. Distance to Border × EAC (in '00	-0.004	-0.011***	0.014***	0.002*	-0.001	-0.011**
	(0.006)	(0.003)	(0.005)	(0.001)	(0.005)	(0.005)
Cont. Distance to Border × CM (in '00 km)	0.060	0.015	-0.110**	-0.014	-0.038	0.053**
	(0.058)	(0.018)	(0.053)	(0.011)	(0.038)	(0.026)
Sq. Cont. Distance to Border × CU (in '00 km)	-0.006	-0.001	0.009**	0.001	0.007	-0.004
	(0.005)	(0.002)	(0.004)	(0.001)	(0.004)	(0.003)
Cont. Distance to Border × CM (in '00 km)	0.099*	-0.001	-0.149**	-0.007	0.014	0.050**
	(0.054)	(0.011)	(0.059)	(0.011)	(0.021)	(0.023)
Sq. Cont. Distance to Border × CM (in '00	-0.008*	0.000	0.012***	0.001	-0.001	-0.004**
	(0.005)	(0.001)	(0.004)	(0.001)	(0.002)	(0.002)
Cont. Distance to Core × EAC (in '00 km)	-0.015	-0.138***	-0.033	-0.050***	-0.078	-0.063
	(0.103)	(0.024)	(0.073)	(0.014)	(0.054)	(0.038)
Sq. Cont. Distance to Core × EAC (in '00 km)	0.003	0.035***	0.004	0.005***	0.027**	0.022**
	(0.009)	(0.005)	(0.007)	(0.001)	(0.011)	(0.009)
Cont. Distance to Core × CU (in '00 km)	-0.025	-0.003	-0.222**	-0.037*	0.013	0.009
	(0.054)	(0.022)	(0.096)	(0.020)	(0.050)	(0.028)
Sq. Cont. Distance to Core × CU (in '00 km)	0.003	0.003	0.018**	0.003*	0.007	0.002
	(0.006)	(0.003)	(0.008)	(0.002)	(0.008)	(0.005)
Cont. Distance to Core × CM (in '00 km)	0.020	0.014	-0.184**	-0.061***	0.061**	0.054***
	(0.051)	(0.012)	(0.082)	(0.015)	(0.025)	(0.020)
Sq. Cont. Distance to Core × CM (in '00 km)	0.001	-0.001	0.016**	0.006***	-0.005**	-0.007***
	(0.005)	(0.001)	(0.007)	(0.001)	(0.002)	(0.002)
Individual Controls	YES	YES	NO	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	183,250	71,738	7,692	169,875	95,717	98,963
R-Squared	0.32	0.15	0.18	0.23	0.24	0.16
R-Squared -Within	0.31	0.10	0.17	0.23	0.24	0.16

Notes: The results depicted in this table are estimated in the same way as those shown in Table 4.2 such that most of the respective notes apply here. However, note the following important amendments: I use the continuous (road) distances to both border crossings as well as core cities and also add their squared values as a separate regressor interacted with the treatment time periods.

Chapter 5

Concluding Remarks

This dissertation provided three distinct micro-level analyses of pertinent development challenges in sub-Saharan Africa, namely education, geography, and trade. Although these factors have been studied separately throughout the chapters, they contribute jointly to better explain underdevelopment, particularly in the context of sub-Saharan Africa: Whether it is geography per se that influences growth via low levels of (agricultural) productivity (Gallup et al. 1999), the hindrance to trade given its sizeable spatial extent and distance to (world) markets (Frankel and Romer 1999), or the diminished incentive to invest in institutions or human capital in face of a peculiar disease environment (Acemoglu et al. 2001; Sachs 2003), these three factors are interlinked (Rodrik 2002). As such, the thesis speaks to the broader discussion in the literature on factors related to development while taking a new perspective by focusing on individual- and household welfare.

Acknowledging the complex interplay between these forces, the dissertation aimed to contribute to the understanding of these developmental challenges by posing three specific questions for the continent: (i) How does coastal access explain individual economic well-being and what are the mechanisms behind? (ii) To what extent does education empower women, specifically in their choice on fertility outcomes and their freedom to access the labor market? And (iii) What are the spatial welfare consequences of market integration among three highly polarized economies?

Chapter 2 concerns itself with the first question and showed that physical geography remains a significant predictor of individual living standards within a comprehensive set of 28 African countries. In particular, individuals more distant to major harbors are significantly deprived across an array of wealth proxies such as consumer durables, the basic consumption items of food, water, or medical care, and lack access to cash employment. Importantly, these links remained when controlling for individual-level covariates, country-time specific influences via fixed-effects, as well as an extensive set of other established geographical influences of development. This underlines that there are most likely relevant development costs of remoteness that need to be addressed. Even though the results presented in the chapter cannot necessarily be interpreted as causal, they are informative and the focus on the individual level provides new insights in comparison to the existing literature. When exploring potential channels and mechanisms, we found two distinct factors which may provide starting points for policymakers in trying to rectify these and similar disparities across space. For one, we see that the presence of basic infrastructural endowments like paved roads, electricity grids, and healthcare dampens the negative effect across distance. And second, educated individuals fare significantly better even in these disadvantaged regions, which emphasizes the interlinked nature of the studied topics once more.

Chapter 3 picks up on education as a key instrument of development policy. In particular, the chapter studied the impact of a country-wide tuition abolishment policy on women's outcomes. What the chapter showed is that while educational attainment was positively impacted for individuals at all strata, the behavioral (downstream) changes induced by added schooling are almost exclusively identified for the poorer subgroup of the sample. Poor women benefit from added schooling in the form of increased literacy, remunerated employment opportunities as well as lowered desired and realized fertility. None of these effects of added schooling are documented for women from wealthier households. This confirms the notion that educating previously neglected groups provides significant marginal returns to schooling. These findings also provide clear considerations for policymakers, donor agencies, and governments: Interventions directly at budget-constrained individuals may be more cost-effective and ultimately more equity-enhancing than one-size-fits-all policies. Given that the external validity of our empirical strategy is limited, future research needs to investigate whether our findings can be confirmed in other settings.

Chapter 4 picks up on the recent concerns that specific trade arrangements may be inequality enhancing and investigated the distributional effect of regional market integration in the East African Community. Given that the founding members were already highly polarized economies before trade liberalization, a study of this case was deemed particularly promising. The theory constructed in the chapter made specific provisions for these spatial facts in an attempt to give the empirical investigation a prior on expected effects. Following the results of the theory, trade liberalization in the East African Community was predicted to draw economic activity to the border and thereby decrease previous spatial inequalities unless strong economies of scale present in agglomerations outweigh these effects. The empirical evidence drawn from individual-level data shows a relative welfare increase accompanied with accelerated agglomeration in the interior economic hubs. Given that these findings are in contrast to the theoretical predictions and also to what other recent empirical studies have shown, they are noteworthy and provoke further investigation. While the chapter provides some of these deliberations in a speculative way, the reduced-form empirics are ultimately not apt to inform on specific parameter coefficients, which would permit isolating a key factor driving these observed effects.

Additional Caveats. Next to the specific limitations and caveats mentioned within the scope of the chapters and the preceding remarks, there are some further overarching considerations that prompt future study on these topics. Despite its distinct benefits, the utilization of household surveys imposes specific constraints on the analysis. For instance, challenges such as the temporal limitations and the consistency of data, non-responses and the underreporting of sensitive topics, as well as practical concerns like the cost of conducting surveys and tracking long-term changes (in the same) households, prevail. However, given their usefulness, it is evident that future research must not only rely on complementary data but also make use of new approaches to data collection in developing contexts, such as high-frequency phone surveys recently introduced in many countries. Independent of the circumstances, exploring and combining a wide range of data sources will be crucial for advancing our understanding of the present and future developmental challenges across the world.