Economica

Economica (2022) **89**, 44–61
doi:10.1111/ecca.12394

Gravity Estimations with Interval Data: Revisiting the Impact of Free Trade Agreements

By PETER H. EGGER*,†, MARIO LARCH‡,§ and YOTO V. YOTOV¶,**

**ETH Zurich* *†KOF, CEPR, CESifo, WIFO* *‡University of Bayreuth* *§CEPII, CESifo, ifo, GEP* *¶Drexel University* ***CESifo, ifo*

Final version received 19 September 2021.

We challenge the common practice of estimating gravity equations with interval or averaged data in order to capture dynamic-adjustment effects to trade-policy changes. Instead, we point to a series of advantages of using consecutive-year data recognizing dynamic-adjustment effects. Our analysis reveals that, relative to interval or averaged data, the use of consecutive-year data avoids downward-biased effect estimates due to the distribution of trade-policy events during an event window as well as due to anticipation (pre-interval) and delayed (post-interval) effects, and it improves the efficiency of effect estimates due to the use of more data.

INTRODUCTION AND MOTIVATION

Highlights are as follows.

- Time-interval or time-averaged data may lead to biased estimates of the short- and long-run magnitudes of trade-policy effects.
- Time-interval or time-averaged data may lead to significant biases in the duration and timing of these effects.
- The effect of free trade agreements (FTAs) begins a few years prior to their entry into force.
- The total time to reach the full impact of FTAs on trade is about 10 years.
- The effects of FTAs on international trade are positive but non-monotonic.
- The strongest FTA effects take place a few years after their entry into force.
- We identify three phases that characterize the long-run impact of FTAs on trade.

Understanding the specifics of the dynamic adjustment of trade flows in response to trade policy is vital for academic work as well as policy-making for the following reasons. First, adjustment processes entail that the short- and long-run responses of outcomes of interest differ. Second, they bear the danger that the responses are not measured in full due to ignorance of anticipation effects or delayed effects that materialize after the sample period of analysis. Third, they may involve non-linear if not non-monotonic adjustment patterns of outcomes over time, which may be important to know and understand. In any case, the adjustment of economic outcomes to economic policy is typically not instantaneous, and a proper characterization of the adjustment process (including its beginning, form and duration) as well as the unbiased quantification of the short- and long-run effects may be challenging. All of these arguments are important in the context of the responses of outcomes such as trade flows to the enforcement of free-trade areas as one of the key bilateral policy measures in international trade.

The modern empirical analysis of trade-policy effects on trade flows may be characterized as follows.

First, FTAs feature prominently as a policy instrument, because details about them (including their inception and content) are made available to the public by the World

Trade Organization (WTO) for virtually all countries (and country pairs) in the world and the time period since the Second World War. In comparison, other policy instruments such as tariffs or non-tariff barriers are much harder to collect and are available for a much shorter sample period and smaller cross-section of parties by whom they are applied. Similarly, trade outcomes (values and quantities at the aggregate and product levels) are available also for large cross-sections and long sample periods, which enables the use of time series as well as cross-section data for economic analysis of FTA effects. Since both trade flows and FTAs can be measured at the country-pair level, the predominant approach to estimating direct responses makes use of the gravity equation of bilateral trade.

Second, the availability of time series and cross-section data on policy treatment and economic outcome together with the developments in econometric theory for such data permits the use of panel data methods that are now the leading tool to use for parameter (and direct-response) estimation of the effects of FTAs on trade.

However, the very responses of trade to policy instruments such as FTAs is likely dynamic for various reasons. First, FTAs are typically announced before they come into force, which may induce a boost in (anticipated) firm value and associated investments (see Moser and Rose 2014). Second, FTAs often involve phase-in periods for tariff-barrier and non-tariff-barrier adjustments, which means that their full trade-facilitation effect comes into play only years after their inception. Third, firms need time to adjust—their investments, their business network, etc.—so there is a delay of their response to the materialized facilitation of trade flows. All of those aspects require a proper treatment in the analysis in order to obtain unbiased estimates of the (average) pattern of the adjustment process (its beginning and duration as well as its pattern) so that short-, medium- and long-run effects of FTAs can be properly measured.

With regard to the use of panel data methods to estimate gravity models in contexts where dynamic adjustment is important, Cheng and Wall (2005, p. 52, fn 8) note that '[f] ixed-effects estimation is sometimes criticized when applied to data pooled over consecutive years on the grounds that dependent and independent variables cannot fully adjust in a single year's time'. Clearly, the use of fixed (country-pair) effects is motivated by the opportunity to control for all measured (e.g. geographical distance) and unmeasured time-invariant obstacles to and facilitations of bilateral trade. However, with dynamic adjustment processes present, there is a danger of relying on fixed-effects estimates where policy responses are measured (allowed to materialize) only contemporaneously with the inception of associated treatments; whereas pooling the data over longer time periods will lead to a convergence of the parameter estimates to the long-run responses, using the fixed-effects parameter estimates will lie somewhere between the short-run and long-run responses (see Egger and Pfaffermayr 2005).

To address such critiques and potentially avoid the associated biases, researchers have used panel data with time intervals instead of ones based on consecutive years to estimate the direct responses (net of or before general-equilibrium adjustments) of trade flows to FTA membership. For example, Cheng and Wall (2005) and Baier and Bergstrand (2007) use 5-year intervals, Olivero and Yotov (2012) experiment with 3- and 5-year interval data, and Masood (2021) use 3-year, 4-year and 5-year intervals and different time periods.[1]

This paper challenges the practice of using interval data to estimate the impact of trade policy effects in gravity equations. Instead, we argue in favour of specifications that employ all available data—that is, data pooled over consecutive years—but pay attention to dynamic adjustment processes. The motivation for our argument is threefold whenever adjustment processes are present. First, time-interval data of trade and FTAs may lead to biased estimates of the short-run as well as long-run effects during the predefined intervals due to the unequal spacing of FTAs during the considered time windows. With positive FTA effects and an accumulated adjustment, there will be a systematic downward bias in the long-run effect attributable to the time during the window. Second, the interval approach may suffer from averaged-out anticipation (pre-window) and delayed (post-window) effects. Either one of those will lead to a downward bias of both the short-run and long-run responses. A third disadvantage from using time-interval data relative to annual data is that data are unnecessarily discarded so that parameters are measured less precisely than possible. For example, the sample size is reduced by 80% with 5-year intervals, by construction. Moreover, the aforementioned variation of the inception times of FTAs between the chosen interval boundaries is itself a source of parameter uncertainty and will show in inflated standard errors on the FTA parameters of interest.

We propose an approach that relies on annual data and pays attention to the non-linear response process. We illustrate the merits of this approach in a framework that respects the latest developments in the literature on estimating gravity models, focusing on the direct effects of FTAs on trade flows (net of general-equilibrium responses). The comparison between the results that we obtain with interval or time-averaged versus consecutive-year estimating samples reveals the following.

i. While the cumulative effects of FTAs are similar, the use of interval or time-averaged data leads to biases of the responses relative to the short-run and long-run (direct) effects as discussed above, and these biases pertain to during-interval when FTAs are newly implemented as well as prior to and after such phases associated with anticipation and delay effects.
ii. The pattern of the response surface in time is not identifiable with larger-interval/time-averaged data, and important phases of the process cannot be discerned with such an approach.
iii. The direct trade response to FTAs is highly non-linear, and the duration of the response adjustment takes altogether about 10 years on average.

Note that while we see many merits of using annual data, and many leads and lags to identify the dynamic adjustments of FTAs, we also want to emphasize that the correlations among the leads and lags of FTAs are potentially high, specifically for consecutive lags, which may make identification of individual FTA parameters challenging.

We uncover several systematic patterns that characterize the impact of FTAs on trade among member countries. Consistent with the existing literature, we find that the cumulative effects of FTAs on international trade are positive and statistically significant. However, our estimates suggest that the evolution of these effects is non-linear in time. The analysis reveals that the impact of FTAs begins about 3 years prior to their entry into force, possibly at the time when they are announced or signed. According to our estimates, the main part of the positive impact of FTAs takes place between 3 and 6 years after their implementation, following a period of gradual initial adjustment. There are still positive and statistically significant effects of FTAs in our sample between 7 and 8 years after their entry into force. In combination with the estimated anticipation effects, these estimates imply that the full impact of FTAs on trade is reached about 10 years after they start mattering in the anticipation period.

Using the full, annual, pooled estimating sample enables us to identify three distinct and intuitive phases that characterize the evolution of the (direct) impact of FTAs on trade in our sample akin to the lifecycle for products. The first phase, labelled '*Pre-FTA and Anticipation Phase*', covers the period of up to 4 years prior to the implementation of FTAs. We obtain some positive and significant FTA estimates during this period, but the cumulative effects up until entry into force are small. The second phase, labelled '*Growth Phase*', covers the years following the entry into force of an FTA. We obtain positive and statistically significant effects in this phase. The final phase, labelled '*Maturity Phase*', begins about 8 years after an FTA's entry into force, and from this phase onwards there are no additional associated trade effects—that is, the FTA may be said to have reached its full potential.

The remainder of the paper is organized as follows. Section II introduces our estimating equation and describes the data. Section III presents and analyses our main findings. In Section IV, we summarize and attempt to generalize our results, and we discuss the broader implications of our findings. Section V concludes. An Online Appendix offers details on the countries and the free trade agreements in our estimating sample, and provides additional results.

I. Estimating Equation and Data

We rely on the latest developments in the theoretical and empirical gravity literatures to specify our econometric model as follows:[2]

$$X_{ij,t} = \exp\left[\alpha FTA_{ij,t} + \sum_s \alpha_s FTA_{ij,t+s} + \sum_k \beta_k\, FTA_{ij,t-k}\right] \times \exp\left[\sum_t \beta_t\, BRDR_{ij,t} + \gamma_{ij} + \psi_{i,t} + \phi_{j,t}\right] \times \varepsilon_{ij,t}, \qquad \text{for all } i,\ j,\ t.$$

Here, $X_{ij,t}$ denotes valued bilateral trade flows from country i to country j at time t in levels, and—following the recommendation of Yotov *et al.* (2016) for the use of theory-consistent trade flows—$X_{ij,t}$ includes both international and intranational trade flow observations. Following the recommendations of Santos Silva and Tenreyro (2006, 2011), one should abstain from log-transforming the model in (1) but estimate it with the Poisson pseudo maximum likelihood (PPML) estimator, which accounts for heteroscedasticity in trade data and takes advantage of the information that is contained in zero trade flows. The parameters α, α_s and β_k are of particular interest here, as they measure the direct contemporaneous, leading (anticipation) and delayed (phasing-in or sluggish-adjustment) responses, respectively, of country pairs' trade flows to the inception of an average FTA.[3] $FTA_{ij,t}$ is an indicator variable that takes value 1 when i and j are members of the same FTA in force at time t, and it is equal to 0 otherwise.[4] Following Wooldridge (2010) and Baier and Bergstrand (2007), the motivation for the inclusion of the first FTA lead from the estimating sample is that its estimate can be used to test for strict exogeneity, for example, due to reverse causality. Our empirical analysis reveals that some of the FTA lead effects are actually significant, and we argue that they should be accounted for explicitly in econometric specifications.

Equation (1) includes four sets of fixed effects. As is standard in the literature, we use exporter-time fixed effects ($\psi_{i,t}$) and importer-time fixed effects ($\phi_{j,t}$) to control *inter alia* for the unobservable exporter and importer multilateral resistances established by Anderson and van Wincoop (2003). These fixed effects will also absorb/control for any other country-time-specific characteristics that may impact bilateral trade on the exporter and importer sides. In addition, following the recommendations of Baier and Bergstrand (2007) and Egger and Nigai (2015), we also employ symmetric country-pair fixed effects (γ_{ij}), which will absorb/control for all time-invariant bilateral trade costs and will mitigate endogeneity concerns with respect to our key policy variable of interest, $FTA_{ij,t}$. Finally, we follow Bergstrand *et al.* (2015) to account for common globalization effects with a set of time-varying border dummy variables $BRDR_{ij}$, for which a parameter β_t is estimated for each year t separately. $BRDR_{ij,t}$ is created on the basis of an indicator for external (foreign) sales as opposed to domestic sales, $BRDR_{ij}$, which is interacted with a binary indicator for each year t to obtain $BRDR_{ij,t}$.

As we are dealing with a three-way fixed effects gravity model, we adjust for the (asymptotic) bias in the coefficient estimates as well as the standard errors by using the bias correction developed recently by Weidner and Zylkin (2021) in all our specifications.

We will use the model in equation (1) and estimate it on the dataset of Baier *et al.* (2019). This dataset covers total bilateral trade of manufactures among 69 countries over the period 1986–2006, and it has several advantages for our purposes.[5] First, the dataset includes domestic/intranational trade flows, which are needed for theory-consistent gravity estimations.[6] As demonstrated by Baier *et al.* (2019), the estimates of FTAs are affected by the inclusion of domestic trade flows, since the presence of the latter allows for capturing of trade diversion effects from domestic sales. Second, the time span of the dataset is relatively long, and it covers a period of intense globalization efforts with a number of new FTAs. Third, consistent with our focus, the dataset has already been used to analyse the impact of FTAs. Data on FTAs come from the NSF-Kellogg Database on Economic Integration Agreements of Jeff Bergstrand. A list of the FTAs that are used in the estimating sample, as well as a list of the countries in the dataset, appears in the Online Data Appendix. For further description of the data, its sources and construction, we refer the reader to Baier *et al.* (2019).

In order to obtain estimates of the parameters $\alpha, \quad \alpha_s, \quad \beta_k$, we make use of the dataset in three conceptually different ways. In one type of analysis, we use only data from every τth year, where we choose $\tau = \{5,4\}$ to form intervals. Hence when $\tau = 5$, we use only 20% of the years, and when $\tau = 4$, we use 25% of the years. Alternatively, we use averages of trade flow data over 5-year and 4-year intervals. This leads to the same number of observations as in the first case, but trade data will contain information from the whole interval due to averaging. Finally, we use data for all years, that is, consecutive-year data. With the 5-year interval approach, this means that we use data from only the years 1986, 1991, 1996, 2001, 2006 for estimating the parameters in the columns of Table 1 (below) labelled 'Interval'. With the averaging approach over 5 years, we calculate trade flow averages in the periods 1986–90, 1991–2000, 2001–5, 2006, and use for the independent variables the years 1986, 1991, 1996, 2001, 2006 in the columns labelled 'Average'.[7] Finally, with the consecutive-year approach, we estimate the parameters pertaining to the years of the interval or the averaging, but using data for all consecutive years 1986, 1987, . . ., 2005, 2006 in the columns labelled 'Consecutive'.

II. Estimation Results and Analysis

Our main findings are presented in Table 1. Following Cheng and Wall (2005), and Baier and Bergstrand (2007), our first specification is based on 5-year interval data.

TABLE 1
ON THE IMPACT OF FTAS ON TRADE

	5-year lags			4-year lags			2-year lags	1-year lags
	Interval (1)	Average (2)	Consecutive (3)	Interval (4)	Average (5)	Consecutive (6)	Consecutive (7)	Consecutive (8)
FTA_{t+5}	−0.018 (0.079)	0.168 (0.155)	−0.045 (0.068)					−0.093 $(0.055)^+$
FTA_{t+4}				−0.062 (0.062)	−0.018 (0.054)	−0.013 (0.062)	−0.061 (0.055)	−0.026 (0.017)
FTA_{t+3}								0.073 (0.021)**
FTA_{t+2}							0.093 (0.038)*	0.025 (0.039)
FTA_{t+1}								0.069 (0.012)**
FTA_t	0.121 $(0.066)^+$	0.181 (0.138)	0.108 $(0.055)^+$	0.057 (0.044)	0.086 (0.044)*	0.072 (0.048)	−0.005 (0.028)	−0.059 (0.018)**
FTA_{t-1}								0.040 $(0.024)^+$
FTA_{t-2}							0.070 (0.022)**	0.022 (0.021)
FTA_{t-3}								0.060 $(0.032)^+$
FTA_{t-4}				0.196 (0.052)**	0.163 (0.037)**	0.190 (0.034)**	0.127 (0.029)**	0.071 (0.020)**
FTA_{t-5}	0.189 (0.043)**	0.120 (0.102)	0.190 (0.030)**					0.050 (0.012)**
FTA_{t-6}							0.058 (0.020)**	0.002 (0.021)
FTA_{t-7}								0.072 (0.032)*

TABLE 1
CONTINUED

	5-year lags			4-year lags			2-year lags	1-year lags
	Interval (1)	Average (2)	Consecutive (3)	Interval (4)	Average (5)	Consecutive (6)	Consecutive (7)	Consecutive (8)
FTA_{t-8}				0.116	0.080	0.065	0.039	−0.009
				(0.057)*	(0.047)$^{+}$	(0.044)	(0.044)	(0.017)
FTA_{t-9}								−0.015
								(0.015)
FTA_{t-10}	0.069	0.061	0.044					0.034
	(0.055)	(0.137)	(0.039)					(0.022)
Total FTA	0.362	0.529	0.297	0.307	0.311	0.314	0.321	0.315
	(0.118)**	(0.221)*	(0.110)**	(0.107)**	(0.090)**	(0.103)**	(0.105)**	(0.112)**
No. of obs.	14,045	14,045	58,989	16,854	16,854	58,989	58,989	58,989

Notes
This table reports estimates of the effects of FTAs on international trade. The dependent variable is always trade in levels, and all estimates are obtained with the PPML estimator with (asymptotic) bias-corrected coefficient estimates and standard errors (Weidner and Zylkin 2021) and exporter-time, importer-time, country-pair fixed effects, and time-varying border dummies. The estimates of all fixed effects are omitted for brevity. Column (1) reports estimates that are obtained with 5-year interval data and 5-year FTA lags and leads. Column (2) reports estimates that are obtained with 5-year time-averages and 5-year lags and leads. Column (3) reproduces the results from columns (1) and (2) but with a sample that uses all years. Columns (4), (5) and (6) correspond to columns (1), (2) and (3), but instead of using 5-year intervals/time-averages and 5-year FTA lags and leads, they employ 4-year intervals/time-averages and 4-year FTA lags and leads. Column (7) employs the whole dataset but with 2-year FTA lags and leads. Finally, column (8) also uses all years in the sample but 1-year FTA lags and leads. Standard errors are clustered by country pair. See text for further details.
$^{+}$, *, ** indicate $p<0.10$, $p<0.05$, $p<0.01$, respectively.

Specifically, we use only data for the years 1986, 1991, 1996, 2001 and 2006. Several findings stand out from column (1). First, the positive and significant estimate of the contemporaneous effect of FTAs suggests that FTAs promote trade between their members (relatively) immediately, as expected. Second, the positive and significant estimate on FTA_{t-5} captures the phasing-in effects of FTAs, suggesting that FTAs need time to expand their full effect on trade between member countries. Third, the estimate on FTA_{t-10} is economically small and statistically insignificant, implying that FTAs have reached their full potential 10 years after their entry into force. Fourth, the insignificant estimate on FTA_{t+5} is an indicator that there is a lack of anticipation and the FTA indicator passes the strict exogeneity test; cf. Baier and Bergstrand (2007). Finally, the 'Total FTA' effect that appears at the bottom of Table 1—which we obtain as the sum of all FTA parameters above, that is, the anticipation FTA_{t+5}, the contemporaneous FTA_t, the 5-year FTA_{t-5} and 10-year FTA_{t-10} phasing-in FTA estimates—implies that, all else equal, the FTAs that entered into force during the period of investigation have led to an average increase of 43.6% (with standard error 17.924) in bilateral trade between the FTA members relative to the non-members.

In column (2) of Table 1, we use averages over the respective 5-year intervals for the trade flows instead of information of only every fifth year. When using time averages, we have to make a choice how to treat the independent variables. Say we consider the interval from 1986 to 1990. If an FTA is concluded in 1988, for example, we could set it to either 0 or 1, or use the average also for the FTA variable. In the latter case, we will no longer end up with a binary indicator variable, which is consistent with the rest of the specifications. We decided to use the values for the FTA corresponding to the beginning of the interval. As can be seen by comparing columns (1) and (2), using averages over the intervals instead of data for one point in time for each interval leads to qualitatively similar results, while we lose significance.

The estimates in column (3) of Table 1 replicate the results from columns (1) and (2), but using data for all years. A comparison between the new estimates and the results from column (1), which are based on interval data, reveals that they are not statistically different from each other.

In columns (4), (5) and (6) of Table 1, we conduct the same analysis as in columns (1), (2) and (3) except that we use every fourth (rather than every fifth) year of data. Specifically, the years that we cover include 1986, 1990, 1994, 1998, 2002 and 2006, and we use 4-year intervals/time-averages with corresponding leads and lags. The main takeaway from the comparisons of the estimates in columns (1)–(3) and (4)–(6), respectively, is that estimations with 4- and 5-year interval/time-averaged data are only marginally different, and the qualitative insights gained are identical. Moreover, estimations with interval/time-averaged data and those with data that are pooled over consecutive years deliver very similar estimates of the effects of FTAs.

The results in columns (7) and (8) of Table 1 take advantage of the additional information in the consecutive-year sample to obtain a full set of 2-year and year-on-year responses (including leads and lags) to the inception of FTAs.[8] These estimates generate several new insights in relation to the results from the previous columns. First, we note that the estimates of the FTA anticipation effects for 3 years prior to the agreement entering into force are positive, and those of the 3- and 1-year leads are also statistically significant. Hence FTAs lead to an increase in trade between partners even before entering into force. We offer two explanations for this result. One is that once an agreement is announced, some firms start to adjust in anticipation of the implementation of the agreement (see Breinlich 2014; Moser and Rose 2014). In addition, it is possible that the potential member countries are already relaxing some administrative measures to reduce trade costs even before the agreement enters into force.[9] The pattern of anticipation effects that we document in column (8) are not (and cannot be) captured with the interval/time-averaged samples in columns (1)–(6). Based on these estimates, we believe that it is important for econometric models to explicitly allow for anticipation effects of FTAs with data that are measured at a sufficiently fine granularity in the time dimension.

The second interesting result from the estimates in column (5) of Table 1 is that we obtain a negative estimate of the immediate/contemporaneous effect of FTAs (FTA_t) and relatively small and not (or barely) statistically significant estimates of the FTA effects for the first 2 years after entry into force. The small initial response of trade flows to the formation of FTAs may be due to various reasons, including gradual removal of policy barriers, adjustments on behalf of the firms, and so on; that is, this is a period of adjustment to the new trade rules and conditions. This leads to our third observation, namely that the increase of the FTA effects is the strongest between 3 (t+3, associated with FTA_{t-3}) and 5 years (t+5, associated with FTA_{t-5}) from their entry into force. The effects of FTAs during this period are all positive and statistically significant, increasing at first, then peaking in the middle and tapering off towards the end of that phase. The final result from column (5) is that the cumulative impact of FTAs in our sample reaches its full potential after about 7 years (t+7, associated with FTA_{t-7}) from their implementation. This, in combination with the positive anticipation effects that we estimate, implies that the dynamic-adjustment process takes about 10 years until FTAs unfold their full effect.

It is still common to perform gravity estimations without the use of domestic trade flows.[10] Hence we also estimated our specification on international trade flows data only. Note that in this case we cannot identify the globalization effects. We present these results in Online Appendix B. Overall, the FTA effects are positive but no longer significant when not including intranational trade flows. This is independent of whether we use intervals, time averages or annual data. The finding that intranational trade data are important for estimating the effects of FTAs is in line with several recent contributions in the literature, for example, Dai *et al.* (2014) and Baier *et al.* (2019). More recently, Campos *et al.* (2021) study the implications of using different methods to construct domestic trade flows (e.g. using input–output data versus using gross output data) for the gravity FTA estimates, and they confirm the result that the estimates of FTAs increase once domestic trade flows are explicitly taken into account.

III. Discussion and Implications

A comparison of the cumulative effects at the bottom of Table 1 indicates that they are not starkly statistically different from each other, but the point estimates in columns (1)–(7) differ in magnitude from the ones in column (8) in a way that we try to summarize and generalize in Figure 1. In the figure, we draw the dynamic response locus of bilateral trade for a hypothetical FTA that came into force in year t=1. In the illustration, we assume that a researcher ‘bins’ the data in (non-overlapping) 5-year intervals between which the parameters are identified. Of those intervals, we focus on the one between t=1 and t=5. We display three alternative hypothetical density functions of FTA enforcements during this interval. With a true dynamic response as displayed by the solid curve, note that FTAs behind each one of the three density functions would have their own respective response, and the density functions would suggest how to weight (average) them.

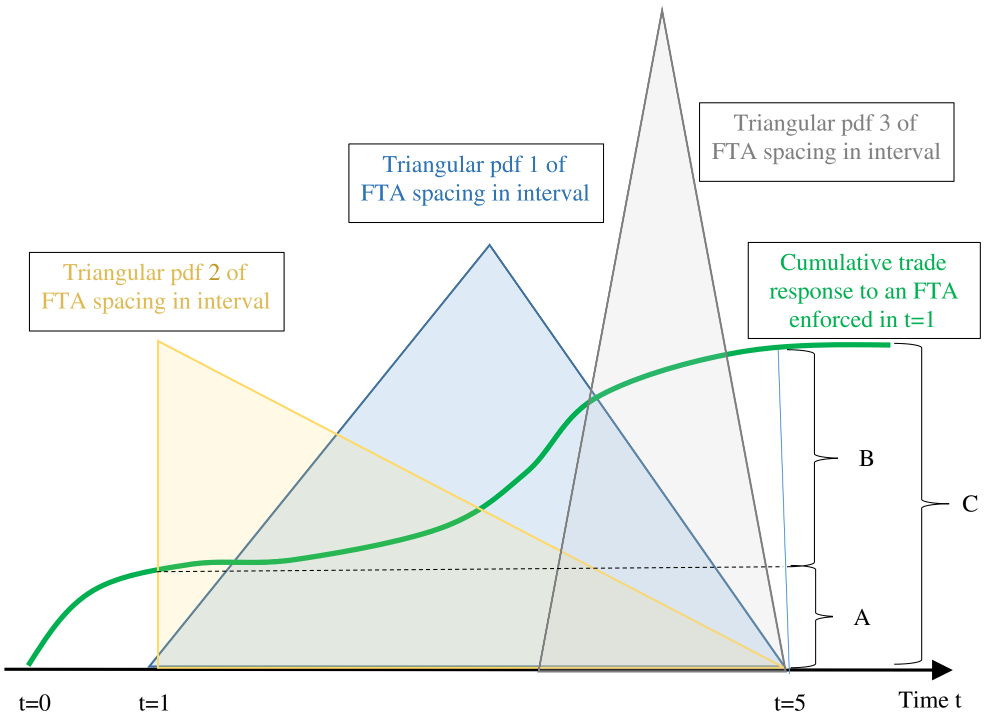

FIGURE 1. Spacing of FTA events over an arbitrary time window of 5 years.*Notes*: The solid curve displays a hypothetical cumulative effect of an FTA enforced in $t = 1$ on bilateral trade. The three shaded triangles display hypothetical event densities of FTA enforcements across country pairs in the data. In the portrayed situation, a 5-year time-interval data approach that uses an interval from $t = 1$ to $t = 5$ would bias results due to mixing density-weighted delay-, anticipation- and during-window effects. A indicates the magnitude of the anticipation effect of an FTA implemented at $t = 1$. B indicates the accumulated associated effect between t=1 and $t = 5$. The cumulative effect at $t = 5$ is $A + B$. The long-run effect is $C > A + B$. A bias of the effect of FTAs may emerge from a mix of the enforcement of FTAs in time between the interval endpoints.

Figure 1 illustrates that the problem with intervals and unequally spaced FTAs is that they cannot possibly do justice to the adjustment process, because the response is measured not relative to the true enforcement time but to the arbitrarily placed time window.[11] With the example at hand (if all FTAs in the data would materialize within the chosen window), the parameter on an FTA indicator would be a biased estimate of the short-run (within-first-year) and long-run (accumulated-process) response.[12]

The event-time pattern (associated with the densities) in Figure 1, together with the effect schedule, entails that the effects estimated from the window relative to before are mixes of density-weighted delay-, anticipation- and during-window effects. Note that depending on when an FTA is ‘born’, the solid curve starts one period ahead of that. Hence we obtain a density-weighted average of curves that are horizontally (in t) shifted. That means that the interval estimator may lead to biased short- and long-run effects. One source of the bias is that some anticipation happens before the window and delay effects happen after it. Note that there are even anticipation effects during the window (for late-coming FTAs), and there the cumulated delay effects will be large. Overall, the figure illustrates that it would be necessary to measure the response intensity relative to the true inception time of FTAs rather than anchored in a fixed way on the time array. In other words, the misalignment and the variation of the phases of the adjustment process to an FTA inception together relative to the spacing of the interval boundaries in time are a source of bias of the response estimates.

To demonstrate this point with our data, Table 2 illustrates the distribution of new FTA memberships across the years 1986–2006 and country pairs in the data. In this table, we provide the specific years, the number of new country pairs in FTAs in a year, and two columns pertaining to the distribution of new FTA memberships within a 5-year interval approach by way of an example. The table demonstrates that during the study period, FTA memberships were distributed quite unevenly (in other words, they were unequally spaced in time).

The 5-year interval approach starting in 1986 leads to a situation where there is a concentration of new memberships in the centre of some intervals (e.g. in 1986–90), while there is one towards the end in other intervals (e.g. in 1991-5).

The numbers column (4) of Table 2 are interval-length averaged durations of new memberships within 5-year intervals. A lower number suggests that new memberships are skewed towards the end (with a theoretical minimum value of 0.2 if all FTAs within the

TABLE 2
DISTRIBUTION OF NEW FTA MEMBERSHIPS WITHIN 5-YEAR INTERVALS

Year (1)	Number of FTAs (2)	Share of FTAs (3)	Interval-length-averaged durations (4)
1986	0	0.00	
1987	0	0.00	
1988	26	0.93	
1989	2	0.07	
1990	0	0.00	0.59
1991	0	0.00	
1992	12	0.05	
1993	68	0.28	
1994	48	0.20	
1995	114	0.47	0.38
1996	42	0.17	
1997	8	0.03	
1998	72	0.28	
1999	2	0.01	
2000	130	0.51	0.47
2001	2	0.01	
2002	6	0.03	
2003	38	0.18	
2004	142	0.68	
2005	20	0.10	0.43
2006	10	1.00	

Notes
This table reports the specific years in column (1), the number of new-FTAs country pairs in a year in column (2), the share of FTAs concluded in a specific year within 5-year intervals (distinguished by spaces) of all FTAs concluded in the 5-year interval in column (3), and the interval-length-averaged durations of new memberships within 5-year intervals in column (4). We normalized the latter interval to unity, so that the number in the last column is the fraction of years for which the average new FTA in an interval existed.

5-year interval would enter into force in the last year of the interval), while a higher number suggests that they are skewed towards the beginning of an interval (with a theoretical maximum value of 1 if all FTAs within the 5-year interval would enter into force in the first year of the interval). With underlying dynamics of adjustment of outcome to new memberships as in Figure 1, this means that the parameters measured on 'contemporaneous' FTAs measured only at interval boundaries (1986, 1991, ...) are prone to a bias that comes from the underlying aggregation of dynamic patterns from within intervals. In any case, the time-interval aggregation hinders the identification of the form of the response process of trade flows to FTAs, which is reflected in the estimates in Table 1.

Turning back to the estimates from column (8) of our main results in Table 1, we identify three distinct phases of the long-run impact of FTAs on international trade. We visualize these phases in Figure 2 by plotting the FTA estimates from column (8), and we discuss each phase in turn.

- *The 'Pre-FTA and Anticipation Phase'*. Prior to the entry into force of the agreements, we see some positive estimates up to 3 years before entry into force. We attribute such positive FTA effects to firm adjustment in anticipation of the agreement and/or to leading changes in trade costs in preparation of the agreement between the member

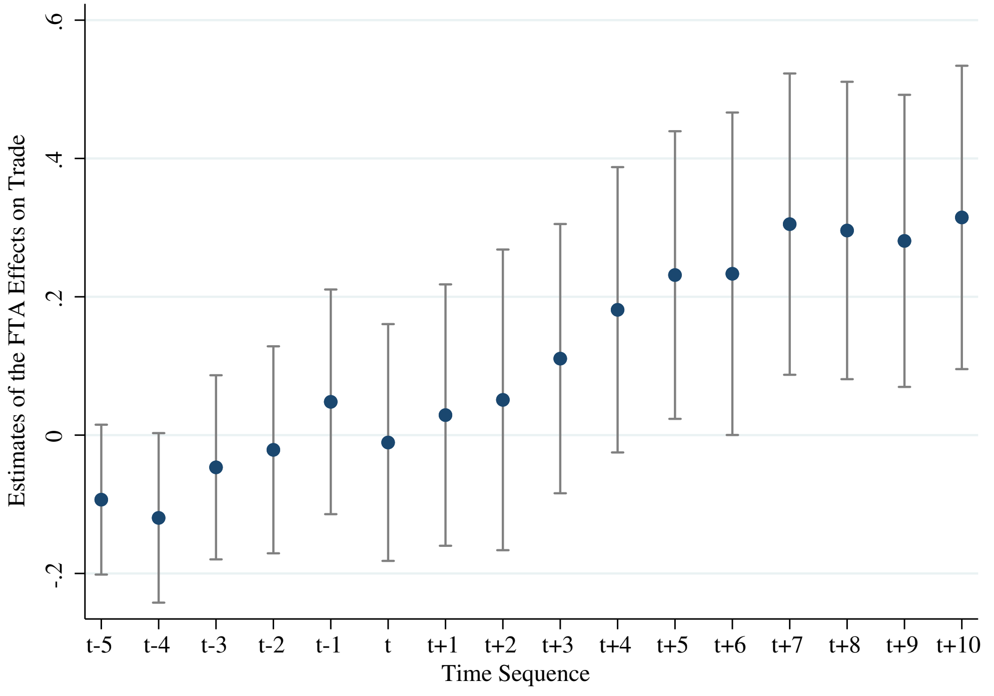

FIGURE 2. On the impact of FTAs on trade.*Notes*: This figure visualizes the estimates from column (8) of Table 1. Specifically, the estimates in the figure are constructed by adding all the leads, lags and contemporaneous FTA estimates from column (8). Thus, for example, by construction '$t-5$' (associated with FTA_{t+5}), 't–4' (associated with FTA_{t+5} plus FTA_{t+4}) and 't+10' (associated with the sum of effects under FTA_{t+5}, FTA_{t+4}, FTA_{t+3}, FTA_{t+2}, FTA_{t+1}, FTA_{t}, FTA_{t-1}, FTA_{t-2}, FTA_{t-3},FTA_{t-4}, FTA_{t-5}, FTA_{t-6}, FTA_{t-7}, FTA_{t-8}, FTA_{t-9} and FTA_{t-10})—i.e. the solid dots in the figure—represent the total average FTA effect for the FTA members in our sample. The whiskers give the 95% confidence intervals.

countries. In terms of timing this phase, intuitively, we expect that such lead effects may start showing up when negotiations of an agreement start, which is often around 2–4 years before entry into force (see Breinlich 2014; Moser and Rose 2014). Cumulatively, as seen in Figure 2, the effects are not statistically significant. However, for proper timing of the duration of this phase, our practical estimation recommendations are (i) to use country-pair fixed effects as in Baier and Bergstrand (2007), and (ii) to allow for sufficiently many lead effects of FTAs.

- *The 'Growth Phase'*. This phase covers the year of entry into force and the first few following years. We see positive but barely significant FTA estimates in the first 2 years after entry into force, which, as discussed earlier, we attribute to the fact that this is a period of adjustment to the new trade rules and conditions. Moreover, most FTAs come with a phase-in period where policy barriers are slowly reduced at the beginning. Finally, some firms may have had exaggerated expectations about the immediate consequences of the inception of an FTA or underestimated the competition of firms in the markets with which they are integrating. After the slow initial response, we see positive and statistically significant FTA estimates. The natural explanation for such positive estimates is that the impact of the FTAs is starting to kick and the continued reductions in policy barriers are gaining bite. Firms are now able to take full advantage of the new trade opportunities and conditions. Also, any uncertainty about the time of entry into force of the FTA and the specifics of its design is resolved, and the trade costs between the member countries have been lowered to a more or less full extent, on average.[13]
- *The 'Maturity Phase'*. According to our estimates, the impact of the FTAs in our sample has reached its full effect after about 8 years from their implementation, or about 10–11 years after the initial (anticipation) impact of FTAs on trade among member countries. We believe that in order to obtain proper estimates of the impact of FTAs in the 'Maturity Phase', one must use country-pair fixed effects in combination with pooled data to ensure a proper identification of its timing. In combination with our findings regarding anticipation effects, these results imply that it takes about 10–11 years until the FTAs unfold their full potential. This is consistent with estimates from the existing literature (e.g. Baier and Bergstrand 2007). Importantly, however, our analysis reveals details on the phases of adjustment processes to FTA effects on trade that are beyond the reach of estimates that do not use panel data with a sufficiently high frequency.

Stimulated by the findings of Baier *et al.* (2019), who argue that the impact of FTAs varies significantly across agreements, we conclude our analysis with an experiment that evaluates the performance of using annual data to identify FTA effects when we allow for heterogeneous FTA effects. To this end, we isolate the impact of two specific agreements: the FTA between Canada and Chile from 1997, and the FTA between Canada and Israel also from 1997. We chose these agreements because they are bilateral, both were signed in 1996 and entered into force in 1997 (early enough so that we are able to identify nine lags), and negotiations started also similarly in 1994 and 1995 (see Moser and Rose 2012, 2014), but they differ substantially in terms of depth measured by the number of provisions using information from the World Bank (the Canada–Chile FTA has 307 provisions, while the Canada–Israel FTA has 161).[14]

We present our findings with graphs similar to Figure 2. The corresponding table with the regression results is provided in Online Appendix C. Note that to ease the interpretation, we isolate the FTA indicators for the specific FTA under consideration from the ones referring to other FTAs. Thus the estimates of the effects of the Canada–Chile FTA and the Canada–Israel FTA can be interpreted in levels. Before we analyse the results, we note that we cannot obtain an estimate on the 10-year lag, because it is outside the span of our sample.

Figure 3 plots the evolution of the FTA estimates for the deep Canada–Chile FTA. As for the overall FTAs, we see that before entering into force, there are some positive, but overall not significant effects. After entry into force, we see again a growth phase, which seems to end about 7 years after the entry into force. This is remarkably similar to the overall FTA pattern that we found before. Note, however, that while the point estimates are quite sizeable, the agreement-specific estimates are not precisely estimated. A possible explanation for the large standard errors is the small number of observations from which we can identify the impact of individual agreements.

The cumulative estimates for the more shallow agreement between Canada and Israel are plotted in Figure 4. Again, we see the three phases: the 'Pre-FTA and Anticipation Phase' with slight positive, but not statistically significant effects, the 'Growth Phase' up until about 7 years after entry into force, followed by the 'Maturity Phase'. In comparison to the effects for the deeper Canada–Chile agreement, two things are noteworthy. First, the point estimates are substantially larger for the deep Canada–Chile

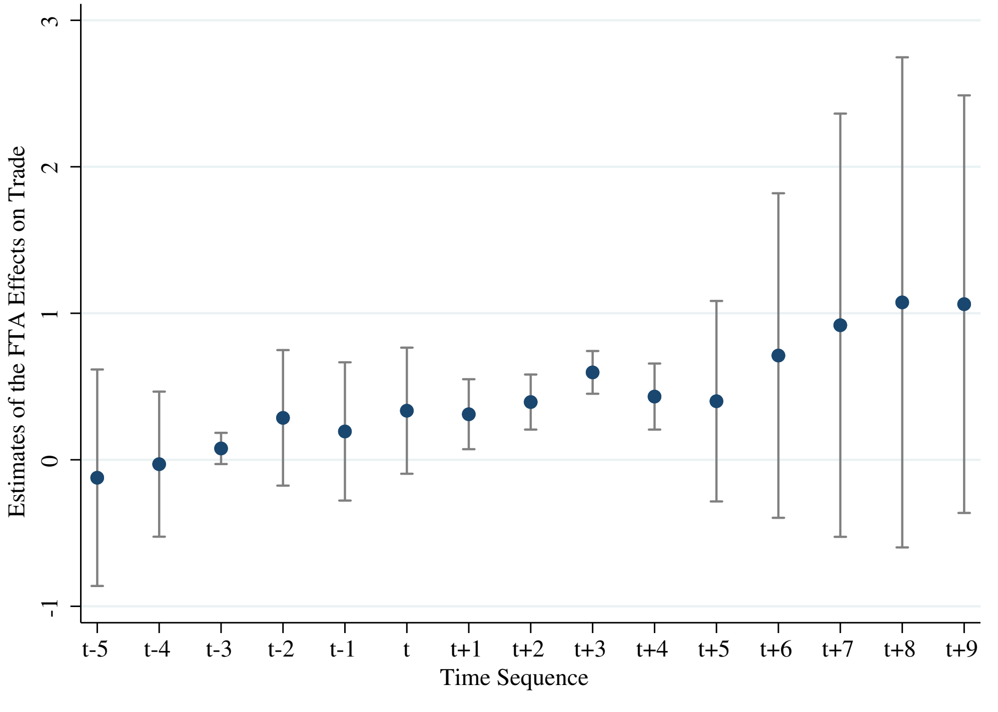

FIGURE 3. On the impact of the Canada–Chile FTA on trade.*Notes*: This figure visualizes the estimates from the Canada–Chile FTA from column (4) of Table A-4 of the Online Appendix. Specifically, the estimates in the figure are constructed by adding all the leads, lags and contemporaneous FTA estimates. Thus, for example, by construction '$t-5$' (associated with FTA_{t+5}), '$t-4$' (associated with FTA_{t+5} plus FTA_{t+4}) and '$t+10$' (associated with the sum of effects under FTA_{t+5}, FTA_{t+4}, FTA_{t+3}, FTA_{t+2}, FTA_{t+1}, FTA_t, FTA_{t-1}, FTA_{t-2}, FTA_{t-3}, FTA_{t-4}, FTA_{t-5}, FTA_{t-6}, FTA_{t-7}, FTA_{t-8}, FTA_{t-9} and FTA_{t-10})—i.e. the solid dots in the figure—represent the total average FTA effect for the FTA members in our sample. The whiskers give the 95% confidence intervals.

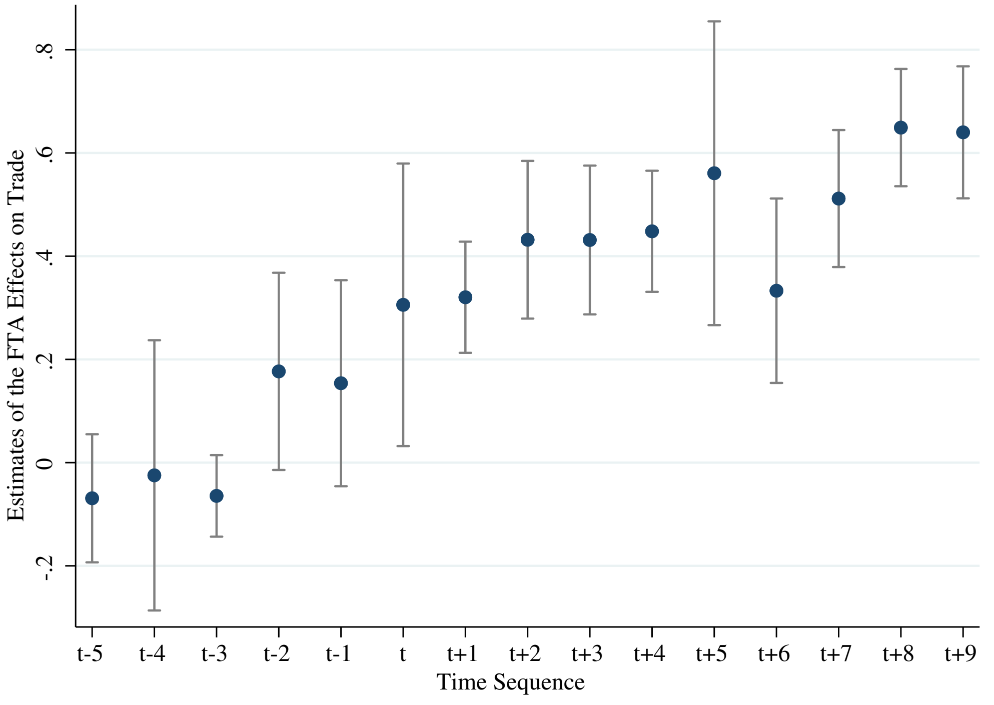

FIGURE 4. On the impact of the Canada–Israel FTA on trade.*Notes*: This figure visualizes the estimates from the Canada–Israel FTA from column (8) of Table A-4 of the Online Appendix. Specifically, the estimates in the figure are constructed by adding all the leads, lags and contemporaneous FTA estimates. Thus, for example, by construction '$t-5$' (associated with FTA_{t+5}), '$t-4$' (associated with FTA_{t+5} plus FTA_{t+4}) and 't+10' (associated with the sum of effects under FTA_{t+5}, FTA_{t+4}, FTA_{t+3}, FTA_{t+2}, FTA_{t+1}, FTA_{t}, FTA_{t-1}, FTA_{t-2}, FTA_{t-3},FTA_{t-4}, FTA_{t-5}, FTA_{t-6}, FTA_{t-7}, FTA_{t-8}, FTA_{t-9} and FTA_{t-10})—i.e. the solid dots in the figure—represent the total average FTA effect for the FTA members in our sample. The whiskers give the 95% confidence intervals.

agreement compared to the Canada–Israel agreement, suggesting that the deeper agreement has larger effects on trade. Second, the estimates for the Canada–Israel agreement are estimated with sharper precision. Inspecting the data, the main reason seems to be that exports and imports between Canada and Israel are more balanced than trade between Canada and Chile.

Based on the results in Table A-4 of the Online Appendix, we conclude that the agreement-specific estimates of the effects of FTAs can be very different when obtained with interval versus consecutive-year data. This reinforces our main argument that estimates that are obtained with annual/consecutive-year data—that is, the full sample—should be preferred over interval-data estimates, which may be biased.

IV. CONCLUSION

Using time-interval or time-averaged data of bilateral trade flows and trade-policy variables has become a widely used practice, in particular, when estimating the direct effects of free trade agreements (FTAs) on bilateral trade flows. Such direct effects are important structural parameters that are used to quantify multicountry general equilibrium models of open economies. By using time-interval or time-averaged data rather than annual data, it is argued that adjustments and short-run fluctuations are taken into account. We challenge this practice.

One problem with interval or time-averaged data is that they can lead to biases of the total (accumulated) direct effects of FTAs or other trade-policy variables. Moreover, they are prone to misrepresent—in particular, with longer intervals—the pattern of the response function. Using annual instead of longer-interval data, we find similar cumulative effects but are able to identify at least three phases of FTA effects on trade flows during the dynamic adjustment process for the average country pair. Distinguishing these phases and more detailed dynamic adjustments is possible only when using the information of trade flows and FTAs at a sufficiently high frequency. Hence we argue, for efficiency and identification reasons, to use all data available rather than using intervals.

While we provide empirical evidence for three different phases of FTA effects on trade flows, we think that it would be fruitful to inform theoretical models based on these stylized facts. Such theorizing might be the basis of richer quantitative dynamic models that are currently not available.

V. Funding

Open Access funding enabled and organized by Projekt DEAL. WOA Institution: UNIVERSITAET BAYREUTH. Blended DEAL: Projekt DEAL.

NOTES

1. An alternative is to average the trade flow data over years within intervals. While averaging leads to smoother data, it reduces the number of observations and, as we demonstrate below, this approach is subject to caveats that are similar to those when using interval data.
2. As demonstrated by Arkolakis *et al.* (2012), the structural estimating gravity equation can be derived from, and is therefore representative of, a very wide class of underlying micro-foundations. We refer the reader to Anderson (2011) and Costinot and Rodríguez-Clare (2014) for recent surveys of the theoretical gravity literature, and to Head and Mayer (2014) and Yotov *et al.* (2016) for recent surveys of the empirical gravity literature. What is elemental is that with time-indexed aggregate bilateral trade data, exporter-time and importer-time fixed effects should be included, as they control for endogenous price and income as well as other country-specific variables.
3. Following the related literature, we use leads to capture the potential impact of FTAs prior to their entry into force, and we use lags to capture possible phasing-in effects, that is, effects after entry into force. The reason for this conventional notation is that this is exactly how these variables are constructed. For example, to construct the FTA variable that will capture the FTA impact one year after entry into force, we take the lag of the FTA variable.
4. We focus exclusively on the impact of FTAs for clarity and simplicity. This is consistent with the FTA analysis of Baier and Bergstrand (2007). The implications of our analysis most likely extend to other policy variables, such as WTO membership, EU membership, and so on.
5. For computational ease and given the main goal of their project to estimate the impact of FTAs, Baier *et al.* (2019) group the 17 countries in their sample that had not signed any agreements during the period of investigation into a 'Rest of the World' (ROW) region. Treating these countries individually does not alter the main conclusions. A list of the countries in the estimating sample appears in the Online Data Appendix.
6. An intranational trade flow is constructed as the difference between the gross value of total production and total exports. The original international trade data come from the United Nations COMTRADE database, accessed via the World Integrated Trade Solution (WITS). The data on total gross production come from the CEPII TradeProd and UNIDO IndStat databases.
7. Using the data for the year 2006 without averaging it over a longer period ensures that we have the same number of observations as for the results reported in the columns labelled 'Interval'.
8. The year-to-year identification of FTA leads and lags in column (8) of Table 1 is a demanding specification, given all our fixed effects. Hence it is not surprising that some of the leads and lags are not significant. We therefore also present results using 2-year lags and leads in column (7).
9. While we see scope for a rigorous theoretical investigation, this is beyond the scope of this paper. However, recent papers investigating the dynamics of trade flows, such as Eaton *et al.* (2016) or Anderson *et al.* (2020), may be fruitful starting points to build a theoretical framework able to capture the phases of FTAs.

For example, in a dynamic framework, anticipation of the conclusion of FTAs in the future may lead to more consumption and trade today. If these anticipation effects work through general-equilibrium effects, then they should be accounted for by our fixed effects (see Anderson *et al*. 2020). But the anticipation of trade-cost changes may change the allocation of bilateral trade transactions in time. Additionally, sluggish adjustments of prices could potentially motivate lags of FTA effects.

10. Yotov (2021) offers 15 reasons why it may be beneficial to estimate gravity with domestic trade flows in addition to international trade flows.
11. The aggregation of data in bracketed time intervals is a well-studied problem in the macroeconometrics literature, and some of the problems and biases that we mention here are the same as those addressed in the respective literatures (see Hassler 2013).
12. Note that the event densities could be used to generate a fractional variable from a binary FTA indicator, which would reflect the fraction of years for which the average FTA was in place during the window. Alternatively, one could use a binary indicator that is unity whenever an FTA was in place any time during the interval. Clearly, the former is preferable over the latter, yet it matters only for the magnitude and not the direction of the biases in the example.
13. While in Figure 2, the cumulative effect of FTAs from t until t+5 seems to increase pretty linearly, column (8) of Table 1 shows that the first 3 years do not lead to significant FTA effects, and only after that do FTAs significantly increase trade. Based on this, an alternative split of this phase could be into an 'Adjustment Phase' followed by a 'Growth Phase'.
14. See Hofmann *et al*. (2017), available online at https://datacatalog.worldbank.org/search/dataset/0039575/Content-of-Deep-Trade-Agreements (accessed 30 September 2021).

REFERENCES

ANDERSON, J. (2011). The gravity model. *Annual Review of Economics*, **3**, 133–60.

ANDERSON, J., LARCH, M. and YOTOV, Y. (2020). Transitional growth and trade with frictions: a structural estimation framework. *Economic Journal*, **130**(630), 1583–607.

ANDERSON, J. and VAN WINCOOP, E. (2003). Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, **93**(1), 170–92.

ARKOLAKIS, C., COSTINOT, A. and RODRÍGUEZ-CLARE, A. (2012). New trade models, same old gains? *American Economic Review*, **102**(1), 94–130.

BAIER, S. and BERGSTRAND, J. (2007). Do free trade agreements actually increase members' international trade? *Journal of International Economics*, **71**(1), 72–95.

BAIER, S., YOTOV, Y. V. and ZYLKIN, T. (2019). On the widely differing effects of free trade agreements: lessons from twenty years of trade integration. *Journal of International Economics*, **116**, 206–26.

BERGSTRAND, J., LARCH, M. and YOTOV, Y. (2015). Economic integration agreements, border effects, and distance elasticities in the gravity equation. *European Economic Review*, **78**, 307–27.

BREINLICH, H. (2014). Heterogeneous firm-level responses to trade liberalization: a test using stock price reactions. *Journal of International Economics*, **93**(2), 270–85.

CAMPOS, R., TIMINI, J. and VIDAL, E. (2021). Structural gravity and trade agreements: does the measurement of domestic trade matter? Banco de Espana Working Paper no. 2117; available online at https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/ DocumentosTrabajo/21/Files/dt2117e.pdf (accessed 1 October 2021).

CHENG, I. and WALL, H. (2005). Controlling for heterogeneity in gravity models of trade and integration. *Federal Reserve Bank of St. Louis Review*, **87**(1), 49–63.

COSTINOT, A. and RODRÍGUEZ-CLARE, A. (2014). Trade theory with numbers: quantifying the consequences of globalization. In G. Gopinath, E. Helpman and K. S. Rogoff (eds), *Handbook of International Economics*, Vol. **4**. Oxford: Elsevier.

DAI, M., YOTOV, Y. V. and ZYLKIN, T. (2014). On the trade-diversion effects of free trade agreements. *Economics Letters*, **122**(2), 321–5.

EATON, J., KORTUM, S., NEIMAN, B. and ROMALIS, J. (2016). Trade and the global recession. *American Economic Review*, **106**(11), 3401–38.

EGGER, P. and NIGAI, S. (2015). Structural gravity with dummies only: constrained ANOVA-type estimation of gravity models. *Journal of International Economics*, **97**, 86–99.

EGGER, P. and PFAFFERMAYR, M. (2005). Estimating long and short run effects in static panel models. *Econometric Reviews*, **23**(3), 199–214.

HASSLER, U. (2013). Effect of temporal aggregation on multiple time series in the frequency domain. *Journal of Time Series Analysis*, **34**(5), 562–73.

HEAD, K. and MAYER, T. (2014). Gravity equations: workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman and K. S. Rogoff(eds), *Handbook of International Economics*, Vol. **4**. Oxford: Elsevier.

HOFMANN, C., OSNAGO, A. and RUTA, M. (2017). Horizontal depth: a new database on the content of preferential trade agreements. Policy Research Working Paper no. WPS 7981, Washington, DC, World Bank Group.

MASOOD, A. (2021). Anatomy of the trade-effect of regional trade agreements and agenda for future research. Munich Personal RePEc Archive Working Paper; available online at https://mpra.ub.uni-muenchen.de/108284 (accessed 1 October 2021).

MOSER, C. and ROSE, A. (2012). Why do trade negotiations take so long? *Journal of Economic Integration*, **27** (2), 280–90.

MOSER, C. and ROSE, A. (2014). Who benefits from regional trade agreements? The view from the stock market. *European Economic Review*, **68**, 31–47.

OLIVERO, M. and YOTOV, Y. (2012). Dynamic gravity: endogenous country size and asset accumulation. *Canadian Journal of Economics*, **45**(1), 64–92.

SANTOS SILVA, J. and TENREYRO, S. (2006). The log of gravity. *Review of Economics and Statistics*, **88**(4), 641–58.

SANTOS SILVA, J. and TENREYRO, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, **112**(2), 220–2.

WEIDNER, M. and ZYLKIN, T. (2021). Bias and consistency in three-way gravity models. *Journal of International Economics*, **132**, 103513.

WOOLDRIDGE, J. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd edn. Cambridge, MA: MIT Press.

YOTOV, Y. (2021). The variation of gravity within countries. School of Economics Working Paper Series 2021-12, LeBow College of Business, Drexel University.

YOTOV, Y., PIERMARTINI, R., MONTEIRO, J.-A. and LARCH, M.(2016). *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*. Geneva: UNCTAD and WTO.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

A Data Appendix
B Results without Intra-National Trade
C Heterogeneous FTA results