How Useful is Statistical Skewness of Financial Data in Decision Making?

Michael Heinrich Baumann
University of Bayreuth

April 13, 2023

Abstract

Statistical skewness is an important concept in the analysis of gambling and financial investment opportunities. Possibly, investors take the skewness of returns' distributions into account and, classically, search for highly skewed financial products. The concept of skewness can be used in some cases to explain or, perhaps, even predict decisions of agents. However, it is known from the literature that there are not only discrepancies between the formal definition of skewness via the third moment and the usual idea of skewness (as Pearson skewness) but also a mismatch between preference structures built on skewness and those built on expected utility. We contribute to the literature by showing via easy-to-understand, exemplary random variables that preference structures built on specific risk indicators—namely loss probabilities, worst-case losses, or value at risk—are, in general, inconsistent with those built on skewness. We illustrate the connection of risk and skewness on a basic level where we can explain the problem of mismatched preference structures avoiding unnecessarily complex mathematical models. Finally, we investigate the relationship between skewness respectively Pearson skewness and probabilities and prove mathematically that it is possible to make statements about probabilities in one special case, namely for random variables whose Pearson skewness values have different signs.

Keywords: skewness, preferences, risk measure

Author:
Michael Heinrich Baumann
Department of Mathematics
University of Bayreuth
Universitätsstraße 30, D-95447 Bayreuth, Germany
E-mail: `michael.baumann@uni-bayreuth.de`
URL: `num.math.uni-bayreuth.de/en/team/michael-baumann`

1 Introduction

In decision and game theory the principle of maximizing the expected value of some target variable—let it be utility [see von Neumann and Morgenstern, 1947] or gain [cf. Bellman, 1954]—is one of the most widespread principles for solving games or calculating equilibria. Yet, this maxim does not seem sufficient for questions of finance, to which we also include lotteries and betting. Presumably, most people care whether they get \$1,000 as a fix amount or whether they can flip a coin that pays them \$2,000 if it's heads and \$0 if it's tails.

We can model these decision alternatives via two random variables A and B, which are, as all other random variables to appear, real-valued and assumed to be defined on some adequate probability space. Consequently, we define A and B with $A = +1,000$ (w/ prob. 1) and $B = \begin{cases} \text{w/ prob. 0.5:} & 0 \\ \text{w/ prob. 0.5:} & +2{,}000 \end{cases}$.

Because most people might care whether they take part in lottery A or B, since the advent of modern portfolio theory [see Markowitz, 1952] [and cf. Markowitz and Todd, 2000], so-called mean-variance analyses have become popular in finance. The basic assumption is that variance (or dispersion in general) is a risk that should be avoided by the trader. In other words, variance can be modeled as a cost that has to be balanced against the expected profit. Thus, the trader does not exclusively maximize $\mathbb{E}[X]$ of a random variable X that models a payoff but either seeks any Pareto-optimal choice that simultaneously promises a high expected payoff and a low variance, or maximizes $\mathbb{E}[X] - \lambda\sqrt{\mathrm{Var}(X)}$ with a trader-specific parameter $\lambda > 0$ that models the trader's idiosyncratic risk appetite.

When two payoff profiles X and Y are equally preferable to a trader but X has a larger variance than Y, it follows directly that X must in turn also have a higher expected value than Y. This difference between the expected values is referred to as the risk premium (cf. Steindl [1941] and Footnote 2 on Page 43 therein for classical references and Fernandez [2020] for a recent review). Conversely, this means that investment (or gambling) opportunities that have no variance (and lie on the Pareto frontier) can only have the smallest possible Pareto-optimal expected value. If we follow the idea of identifying variance and risk, this can explain statements like "no-risk no-gain" [cf. Campi and Garatti, 2018, Section 4.1]. Thus, if one wants to expect (excess) gains, one has to take risks. In modern portfolio theory there are two maxims. Either one looks for the investment alternative with the lowest variance for a given minimal desired expected payout, or one looks for the investment alternative with the highest expected payout value for a given maximal risk (respectively dispersion) tolerance.

What we have not yet taken into account is the distinction between downside and upside risk. Many traders are likely to be at least indifferent to the 'risk' of getting an unexpectedly large payout, if they do not even aspire to it. However, the risk of an unexpectedly high loss (or, equivalently, an unexpectedly low payout) is the crucial deterrent. Monetary risk measures can help distinguishing between variance as a chance and variance as a risk [see Föllmer and Schied, 2011, Chapter 4]. In principle following Föllmer and Schied [2011, Definition 4.1], we define:

Definition 1. *A function ρ that assigns real numbers (i.e. the risks) to financial products is a monetary risk measure if for all two financial positions X and Y it holds true that I) $X \leq Y \Rightarrow \rho(X) \geq \rho(Y)$ (monotonicity) and II) $\rho(X+m) = \rho(X) - m$ for all real m (cash invariance; translation property).*

Monetary risk measures (which are sometimes shortly just called "risk measures") may have further properties like "normalization," "convexity," "quasi convexity," "positive homogeneity," or "subadditivity" [see Föllmer and Schied, 2011, Chapter 4]. However, these monetary risk measures are often very specific or can only be calculated under restrictive assumptions. This is why i.a. statistical skewness is used as an alternative to distinguish between downside and upside risk [see also Kraus and Litzenberger, 1976]. Skewness is not only considered by researches in theoretical works on preferences and analyses of gambling [see Garrett and Sobel, 1999] but may also be taken into account by practitioners [cf. Kim and Park, 2020, Swedroe, 2015, Yang and Nguyen, 2019].

In this paper, we are going to show that the power of skewness to assess risk (and, thus, risk-based preferences) is very limited. In doing so, we state risk indicator and skewness measure definitions in Section 2, in which we also mention advantages of skewness considerations. Section 3 briefly summarizes some related work. In Section 4 we show with the help of easy-to-understand examples that skewness is not always a meaningful target variable for financial decisions. In Section 5 we look at theoretical connections between (Pearson) skewness and probabilities. Finally, Section 6 presents possible future research directions and concludes the paper.

2 Skewness and its Explanatory Potential

We begin this section with an example. We consider the two alternatives C and D. Here, C is the alternative where we do not invest anything, i.e., C has a payoff of zero (w/ prob. 1). In contrast, D has the following payoff profile:

$$D = \begin{cases} \text{w/ prob. } 0.99751 & -10 \\ \text{w/ prob. } 0.0024938: & +1,995 \end{cases}$$

Please note that here and so in the whole paper we round all values to five significant digits. An exception are the plots in Figures 1 and 2 where we round to two digits to enhance readability.

We can think of D as a possible lottery ticket which costs \$10 and offers the chance to win \$2,005 with a rather small probability. It is easy to calculate that $\mathbb{E}[C] = 0$, $\text{sd}(C) = \sqrt{\text{Var}(C)} = 0$, $\mathbb{E}[D] = -5$, and $\text{sd}(D) = 100$ hold. That is, with a pure mean-variance analysis, every player/investor should prefer C over D. However, the payout profile of D as a possible lottery ticket does not seem to be very unrealistic. Thus, one can easily imagine that there are people who would buy a lottery ticket D. How can this be explained? Would there also be any person who would invest in E with

$$E = \begin{cases} \text{w/ prob. } 0.022005: & -671.67 \\ \text{w/ prob. } 0.97800: & +10 \end{cases}$$

where also $\mathbb{E}[E] = -5$ and $\mathrm{sd}(E) = 100$ hold? It may be hard to imagine that there are many people that prefer E over C. And even if one adds 10 to E, i.e., $F = E + 10$ (from which $\mathbb{E}[F] = +5$ and $\mathrm{sd}(F) = 100$ follow), most people might still prefer C (and D) over F. Concerning the concept of 'preferences,' we shall note that preferences are subjective and arise, among others, from past experience, intuition, or instinct. We call a mathematical concept (such as skewness or some specific risk measure) an explanation for a preference if this concept can be used to calculate some indicator that is strictly monotonic in the preference structure. Conversely, we call a preference structure based on a mathematical concept if it can be explained by this concept.

One way to explain why E and also F are unlikely to be preferred over C or D is the Worst-Case loss (WC), which is for discrete bounded below random variables simply the smallest possible payoff that happens with positive probability. In general, one can express the worst-case loss in mathematical terms as follows.

Definition 2. *The worst-case loss* $\mathrm{WC}(X)$ *is the infimum of the support of the distribution of the real-valued random variable* X *that models the gain.*

Thus, the mathematical number "worst-case loss" can according to the above definition in general also be minus infinity, leading to:

Theorem 1. *The worst-case loss is no monetary risk measure.*

That means that in this case the realized loss is not bounded below. In our examples, $\mathrm{WC}(C) = 0$, $\mathrm{WC}(D) = -10$, $\mathrm{WC}(E) = -671.67$, and $\mathrm{WC}(F) = -661.67$ hold. With the help of the worst-case loss, however, one still cannot explain why someone should prefer D over C. One possible answer to this is the concept of skewness [cf. Golec and Tamarkin, 1998, Garrett and Sobel, 1999].

Remark 1 (Verbal)**.** *For example, one may think that distributions typically have a large (positive) skewness if agents cannot make a large loss, have a very large probability for making a small loss, and if there is the chance, albeit small, of making a large gain. Conversely and also exemplarily, one may think that distributions typically have a small skewness (i.e. a negative skewness with a large absolute value) when traders have a high probability of making a small profit but cannot make a large profit and face the, albeit small, risk of making a large loss.*

Remark 2 (Graphical)**.** *Viewed graphically, see Figures 1 and 2, one may think of a positively skewed distribution as one that is skewed (i.e. flat) on the right and steep on the left, likewise one may think of a negatively skewed one as one that is skewed (i.e. flat) on the left and steep on the right.*

Such a graphical explanation is not mathematically precise and, in fact, graphical 'definitions' of skewness might, just as the verbal examples ('... if agents cannot make a large loss, have a very large probability for making a small loss, and ...') be missleading [cf. von Hippel, 2005].

Hence, the question arises how to determine skewness formally. One possible skewness definition is Fisher's moment coefficient of skewness, which we simply call skewness.

Remark 3. *"Fisher's moment coefficient of skewness" is sometimes also called "moment coefficient of skewness" or—confusingly—"Pearson's moment coefficient of skewness."*

The skewness $v(X)$ of a random variable X can be written down similarly to expected value, variance, and standard deviation by means of moments.

Definition 3. *For a sufficiently often integrable random variable X with mean, i.e. expected value, $\mu(X) = \mathbb{E}[X]$, we denote its variance by $\sigma^2(X) = \mathrm{Var}(X) = \mathbb{E}[(X - \mu(X))^2]$, its standard deviation by $\sigma(X) = \mathrm{sd}(X) = \sqrt{\sigma^2(X)}$, and its skewness (i.e. Fisher's moment coefficient of skewness) by $v(X) = \mathbb{E}\left[\left(\frac{X-\mu(X)}{\sigma(X)}\right)^3\right]$ if $\sigma(X) \neq 0$ and $v(X) = 0$ if $\sigma(X) = 0$.*

Remark 4. *Note that $\sigma^2(X) = \mathbb{E}[X^2] - \mu(X)^2$ and $v(X) = \frac{\mathbb{E}[X^3] - 3\mathbb{E}[X^2]\mu(X) + 2\mu(X)^3}{\sigma(X)^3}$ if $\sigma(X) \neq 0$ hold.*

Remark 5. *Definition 3 expands the standard definition of Fisher's moment coefficient of skewness by the zero-variance case. That way, we do not have to exclude this case if necessary. When the variance is positive, this skewness measure is the third standardized moment.*

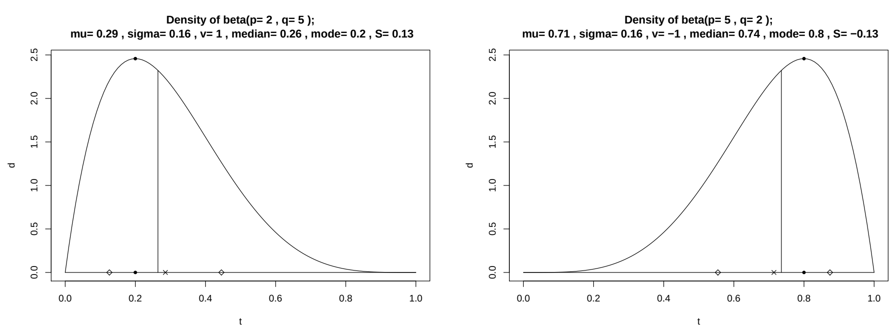

Figure 1: Plots of the density functions of a $\beta(2, 5)$ distribution (right-skewed) on the left and a $\beta(5, 2)$ distribution (left-skewed) on the right.

Returning to the lottery tickets from the beginning of this section, we see that for C, D, E, and F the skewness evaluates to: $v(C) = 0$, $v(D) = +19.95$, and

$v(E) = v(F) = -6.5167$. This may indicate the explanatory power of skewness [cf. Golec and Tamarkin, 1998, Garrett and Sobel, 1999] since it might explain why someone may prefer D over C when assuming that agents prefer highly skewed gains. Loosely spoken, we can think of a trader/agent, who prefers D over C even though the worst-case loss is an argument against D, because D (in the sense of skewness) makes much higher profits possible than C, but not much worse losses.

Although Fisher's moment coefficient of skewness is likely to be the most widely used measure of skewness, we note that there are many alternative skewness measures, some of which use statistical location parameters.

Definition 4. *'A' median is a value such that at least 50% of the probability mass lies left of or on this median and at least 50% of the probability mass lies right of or on this median. When there is exactly one median, we denote it by* $\mathrm{med}(X)$*. When there is more than one median, we define—as commonly done—'the' median* $\mathrm{med}(X)$ *as the midpoint of the closed interval of all medians.*

Definition 5. *The mode* $\mathrm{mod}(X)$ *of a random variable is the unique argument of the global maximum of its probability mass/density function respectively the set of all arguments of the non-unique global maxima of its probability mass/density function.*

Remark 6. *The mode does not have to exist. But we assume all random variables in this work to have an unique mode if we use this mode (directly or indirectly).*

Remark 7. *Via the last sentence of Definition 4, for any random variable* X *and any* $c \in \mathbb{R}$ *the equality* $\mathrm{med}(X + c) = \mathrm{med}(X) + c$ *holds. This is a property also the mean and—when it exists and when it is unique—the mode have.*

If we look at Figure 1 again, we see that for the distribution with positive skewness, the mean lies on the right of the respective median $\mathrm{med}(X)$, for the distribution with negative skewness, it lies on the other side. This is the motivation for our next definition.

Definition 6. *We define the Pearson skewness* $S(X)$ *of an adequate random variable* X *as* $S(X) = \frac{\mu(X) - \mathrm{med}(X)}{\sigma(X)}$ *if* $\sigma(X) \neq 0$ *and* $S(X) = 0$ *if* $\sigma(X) = 0$*.*

Remark 8. *The Pearson skewness used in the work at hand is also known as "nonparametric skew." We note that there are other definitions of Pearson skewness like* $\frac{\mu(X) - \mathrm{mod}(X)}{\sigma(X)}$ *("mode skewness" or "Pearson's first skewness coefficient") and* $\frac{3 \cdot (\mu(X) - \mathrm{med}(X))}{\sigma(X)}$ *("median skewness" or "Pearson's second skewness coefficient"). Since we are only interested in comparisons of Pearson skewness values or in their signs, the factor three is irrelevant for this work.*

Note that for any random variable X and any $c \in \mathbb{R}$ the equalities $v(X+c) = v(X)$ and $S(X + c) = S(X)$ hold, thus, neither skewness nor Pearson skewness nor "minus skewness" nor "minus Pearson skewness" is cash invariant, which leads to:

Theorem 2. *Neither skewness nor Pearson skewness nor "minus skewness" nor "minus Pearson skewness" is a monetary risk measure.*

Additionally, neither skewness nor Pearson skewness nor "minus skewness" nor "minus Pearson skewness" is monotone: take any left/right (Pearson) skewed distribution S (i.e. "$v(X) < 0$" resp. "$S(X) < 0$" or "$v(X) > 0$" resp. "$S(X) > 0$") that is bounded above by zero and have a look at $-X$.

When we again come back to the lottery tickets example, we observe that $S(C) = 0$, $S(D) = +0.05$, and $S(E) = S(F) = -0.15$. Regarding Figure 1, we observe that the $\beta(2,5)$ distribution is positively skewed, positively Pearson skewed, and $\text{mod}(\beta(2,5)) < \text{med}(\beta(2,5)) < \mu(\beta(2,5))$ holds. Further, the $\beta(5,2)$ distribution is negatively skewed, negatively Pearson skewed, and $\text{mod}(\beta(2,5)) > \text{med}(\beta(2,5)) > \mu(\beta(2,5))$ holds. This should be kept in mind and is particularly interesting for the discussion which follows in the next paragraph.

Von Hippel (2005) reports that many textbooks mention rules of the following type: a random variable is positively/negatively skewed—presumably thought of in the sense of Fisher's moment coefficient of skewness—if and only if its mean lies on the right-/left-hand side of its median, i.e., if it has a positive/negative Pearson skewness. Some textbooks—again according to von Hippel [2005]—extend this connection of mean, median, and skewness to a connection of mean, median, mode (in this order), and skewness. However, von Hippel [2005] demonstrats with many examples (mainly discrete but also continuous ones) that this 'textbook rule' does not have to hold in general. Also we give a simple example where this rule does not hold. We define G as follows:

$$G = \begin{cases} \text{w/ prob. } 1 \cdot 10^{-9}: & -10{,}000 \\ \text{w/ prob. } 0.99999: & -0.0029899 \\ \text{w/ prob. } 0.9999 \cdot 10^{-5}: & +300.01 \end{cases}$$

Then $\mu(G) = 0$, $\sigma(G) = 1$, $v(G) = -729.99$, and $S(G) = +0.0029899$ hold. Thus, the desired counterexample $\text{sgn}(v(G)) \neq \text{sgn}(S(G))$ follows.

Remark 9. *Additionally, von Hippel [2005] shows that for discrete distributions some standard interpretation of the median does not have to hold.*

Although skewness (and also Pearson skewness) may explain why agents might prefer D over C, its explanatory power for preferences seems to be limited. This becomes clear when having a look at H with

$$H = \begin{cases} \text{w/ prob. } 1 \cdot 10^{-12}: & -100 \\ \text{w/ prob. } 0.8: & -50 \\ \text{w/ prob. } 0.19999: & +174.31 \\ \text{w/ prob. } 1 \cdot 10^{-5} - 1 \cdot 10^{-12}: & +13{,}959 \end{cases}.$$

Which alternative, D or H, would an agent prefer? Since $\mu(D) = -5 = \mu(H)$ and $\sigma(D) = 100 = \sigma(H)$ hold, mean-variance considerations do not help answering this question. So, we have a look at (Pearson) skewness and calculate $v(H) = 28.307$ and $S(H) = 0.45$, thus, H is positively skewed and positively Pearson skewed as well as higher skewed than D and higher Pearson skewed than D (recall: $v(D) = 19.95$ and $S(D) = 0.05$). Nonetheless, it is rather unclear whether someone would prefer H over D. When we compare D and H, we see that the probability for winning in D is 0.0024938 while the probability for winning in H is nearly 0.2. Thus, the probability for winning in H is around 80 times the respective probability in D. The amount to be maximally won is much higher in H, however, also the worst-case loss is with $\mathrm{WC}(H) = -100$ much higher in H than in D (recall: $\mathrm{WC}(D) = -10$). But is it really the worst-case loss that determines agents' preferences? We might doubt this since the worst-case loss may be of limited explanatory power, which becomes obvious when considering distributions with a support without lower bound (see, e.g., Figure 2 right-hand side).

Remark 10. *In detail, in Figure 2 right-hand side, a Gumbel distribution is depicted whose parameters are chosen in such way that* $\mu(\mathit{Gumbel}) = 0$ *and* $\sigma(\mathit{Gumbel}) = 1$ *hold, i.e. loc* $= -0.45005$ *and scale* $= 0.77970$. *Both the Gumbel and the* $\chi^2(6)$ *distribution (which is depicted on the left-hand side of Figure 2) are positively skewed, positively Pearson skewed, and* $\mathrm{mod}(\chi^2(6)) < \mathrm{med}(\chi^2(6)) < \mu(\chi^2(6))$ *as well as* $\mathrm{mod}(\mathit{Gumbel}) < \mathrm{med}(\mathit{Gumbel}) < \mu(\mathit{Gumbel})$ *hold. The support of the* $\chi^2(6)$ *distribution is* $\mathbb{R}_0^+$ *while the support of the Gumbel distribution is* $\mathbb{R}$, *in particular,* $\mathrm{WC}(\mathit{Gumbel}) = -\infty$ *holds.*

Yet, even if the worst-case loss is a real number, as in the case of H, its explanatory power for preferences is questionable. We observe that the probability for loosing 100 is only $1 \cdot 10^{-12}$. We can imagine that one might say that the event of losing 100 in H is 'practically impossible' (despite this statement is mathematical vague). Thus, can the "-100" be ignored?

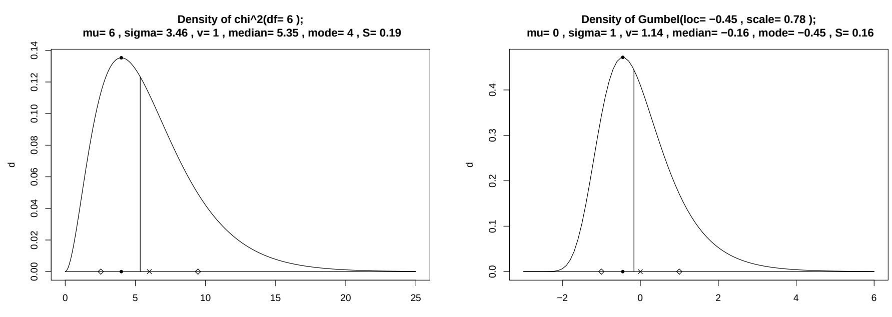

Figure 2: Plots of the density functions of a $\chi^2(6)$ distribution on the left and a (shifted and scaled) Gumbel distribution on the right.

The monetary risk measure named Value at Risk (V@R) can be motivated via this idea of neglecting very rare events such as the "-100" in H. For a random variable modeling a profit, the value at risk can be defined as follows [see Föllmer and Schied, 2011, Example 4.11, Equation (4.41), and Section 4.4 in general].

Definition 7. *Let X be a real-valued random variable modeling a profit and $\alpha \in (0,1)$. Then*

$$\mathrm{V@R}_{\alpha}(X) = \inf\{m \in \mathbb{R} \mid \mathbb{P}(X + m < 0) \leq \alpha\}$$

*is called the value at risk of X to the level α. The Loss Probability (*LP*) of X is*

$$\mathrm{LP}(X) = \mathbb{P}(X < 0).$$

Theorem 3. *The loss probability is no monetary risk measure since it is not cash invariant. However, it is monotone.*

Proof. 1.) Let $X \sim \mathcal{U}([0,1])$. Then, $\mathrm{LP}(X) = 0 = \mathrm{LP}(X + m)$ for all nonnegative m. 2.) Let $X \leq Y$. Then, $\mathrm{LP}(X) = \mathbb{P}(X < 0) \geq \mathbb{P}(Y < 0) = \mathrm{LP}(Y)$. □

Loosely spoken, the value at risk is the amount of money one has to add to an investment opportunity such that the probability of loosing becomes smaller or equal to α. The level α can be interpreted as the "level of risk" an agent resp. trader is willing to take or as a probability that is so small that the trader does not care about. Here we note that the value at risk has some disadvantages, esp. it is not convex [see Föllmer and Schied, 2011, Section 4.4] and—although $\mathrm{V@R}_{\alpha}(X + \mathrm{V@R}_{\alpha}(X)) = 0$ holds—it is possible that $\mathrm{WC}(X + \mathrm{V@R}_{\alpha}(X)) = -\infty$ (cf. Figure 2 right-hand side), i.e., the value at risk can be small while at the same time the worst-case loss may be enormous.

Nonetheless, the value at risk is widespread in practice since it has some advantages, too: it can be convex (under specific assumptions concerning distributions, portfolio types, and parameters), it is often easily computable, and there is no smallest value-at-risk dominating convex risk measure [Föllmer and Schied, 2011, Theorem 4.47]. When we come back to the question whether someone should prefer D or H, we might consider that $\mathrm{LP}(D) = 0.99751 > \mathrm{LP}(H)$ (which is an argument for H) but $\mathrm{V@R}_{1\%}(H) = 50 > \mathrm{V@R}_{1\%}(D) = 10$ (which is an argument for D). That means, what speaks against H is not so much the worst-case loss of -100 but that the probability for "-50" is 80% and, hence, not neglectable (at a 1% level).

The next steps in this work are as follows: The succeeding section discusses some of the related literature. In Section 4, we investigate connections between skewness and loss probability, worst-case loss, as well as value at risk. Further, in Section 5, we analyze connections between skewness as well as Pearson skewness and probabilities.

3 Literature Review

In Golec and Tamarkin [1998], the behavior of bettors at horse tracks is investigated empirically. For this reason, an adequate skewness measure is discussed.

It is reported that some literature concludes that bettors are risk loving (in the sense of high variance), however, the conclusion in Golec and Tamarkin [1998] is that bettors prefer high variance if it is connected to high skewness. In Garrett and Sobel [1999] it is theoretically and empirically explained how skewness can be used to explain why risk averse individuals take unfair gambles.

But not only in the field of games, gambles, and lotteries skewness has been used to explain agents' behavior. In finance, i.e. in portfolio selection, not only mean-variance considerations are of importance but also skewness (or the gain's third moment; see Alderfer and Bierman Jr. [1970] and the references therein, esp. Footnote 2). Although a large share of the literature cited up to now is quite old (more than 40 years), the fundamental problem for financial decision makers has not changed much and is also present in rather new branches of finance such as 'feedback trading' [see Malekpour and Barmish, 2012, Baumann, 2022] or 'momentum investing' [cf. Barroso and Santa-Clara, 2015, Martin, 2012]. A trader i who uses a specific strategy $X_{\mathbf{p}}$ which depends on some parameter vector $\mathbf{p}$ has to choose these parameters in an optimal way. When staying on the mean-variance Pareto frontier, the parameters can be changed in such a way that the mean becomes larger/smaller while at the same time the variance becomes also larger/smaller. The question is which optimality criterion should be used. For this end, 'classical' mean-variance considerations can help, e.g., $\mathbf{p}^* = \text{argmax}_{\mathbf{p}}\{\mu(X_{\mathbf{p}}) - \lambda_i \sigma(X_{\mathbf{p}})\}$. This reduces the possibly high dimensional problem of choosing $\mathbf{p}$ to a one dimensional problem of choosing λ_i. However, it is not clear how λ_i should be chosen because the basic tradeoff between mean and variance stays the same. The problem might become even worse when keeping in mind that λ_i is rather abstract, i.e. hard to interpret for the trader. Here, skewness might help since traders can (or at least: they think they can) imagine what skewness is (maybe they have some plots like those in Figure 1 in mind). Hence, traders may specify a desired (idiosyncratic) skewness level and choose λ_i (and so $\mathbf{p}$) to match the desired skewness at the mean-variance Pareto frontier. Such approaches are used in finance [cf., e.g., Briec et al., 2007, Li et al., 2010]. In Szpiro [1991], the variance and the skewness of the risk are used to model risk aversion. Mills [1995] describes the mathematical modeling of skewness and kurtosis (i.e. of higher moments).

In Brockett and Kahane [1992] it is shown that preferences based on expected utility and preferences based on moments (mean, variance, skewness) do not have to coincide. This is a strong hint that preferences based on risks and preferences based on moments (mean, variance, skewness) might also be in conflict in some cases. This is exactly what is shown in the next section. In Section 5, one specific exception is proven; in detail: a specific case where certain risk preferences may be deduced from Pearson skewness (see Theorem 8).

4 Examples

In this section, we show for three preference models that are based on risk (namely based on loss probability, worst-case loss, and value at risk) that they can be in conflict with preferences based on skewness (i.e. on Fisher's moment coefficient of skewness).

Definition 8. *Let X and Y be real-valued random variables and $\alpha \in (0,1)$. We define*

$$X \begin{Bmatrix} \succ_\xi \\ \sim_\xi \\ \prec_\xi \end{Bmatrix} Y \;\Leftrightarrow\; \xi(X) \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} \xi(Y) \text{ and } X \begin{Bmatrix} \succ_\zeta \\ \sim_\zeta \\ \prec_\zeta \end{Bmatrix} Y \;\Leftrightarrow\; \zeta(X) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} \zeta(Y)$$

with $\xi \in \{v, S, WC\}$ and $\zeta \in \{LP, \mathrm{V@R}_\alpha\}$ as five possible preference models.

Remark 11. *Note that for skewness v and Pearson skewness S we assumed that traders prefer high values [cf. Golec and Tamarkin, 1998]. However, since we are going to show non-monotonicity of these preferences, this assumption is without loss of generality.*

To demonstrate that preferences based on skewness and those based on risk (in detail: on the risk indicators loss probability, worst-case loss, and value at risk) do not necessarily coincide we look at different simple counterexamples. In detail, we consider four-point distributed random variables, i.e., random variables that can take at most four different real values with positive probability. To ensure comparability for all random variables considered in both the present and the next section, $\mu(\cdot) = 0$ and $\sigma^2(\cdot) = 1$ hold, i.e., they are standardized. This is an important point since otherwise it would be very easy to construct random variables with arbitrary skewness values—but these random variables would not necessarily be counterexamples in a mean-variance-skewness sense.

Remark 12. *Note that for a four-point distribution when fixing the four probabilities adequately, two values, the mean, and the variance, there are exactly zero, one, or two possibilities for the remaining two of values.*

4.1 Loss Probability

Let us have a look at the random variables I, J, K with

$$I = \begin{cases} \text{w/ prob. } 0.1: & -2 \\ \text{w/ prob. } 0.25: & -1 \\ \text{w/ prob. } 0.25: & +0.38462 \\ \text{w/ prob. } 0.4: & +0.88462 \end{cases},$$

$$J = \begin{cases} \text{w/ prob. } 0.1: & -1.4 \\ \text{w/ prob. } 0.25: & -1 \\ \text{w/ prob. } 0.25: & -0.28752 \\ \text{w/ prob. } 0.4: & +1.1547 \end{cases},$$

$$K = \begin{cases} \text{w/ prob. } 0.1: & -1.5 \\ \text{w/ prob. } 0.24: & -1 \\ \text{w/ prob. } 0.4: & +0.043249 \\ \text{w/ prob. } 0.26: & +1.4335 \end{cases}.$$

On the one hand $v(I) = -0.75888$, $v(J) = +0.085498$, $v(K) = +0.18836$ hold but on the other hand we observe $\mathrm{LP}(I) = 0.35$, $\mathrm{LP}(J) = 0.6$, $\mathrm{LP}(K) = 0.34$. Thus,

$$I \prec_v J \prec_v K \quad \text{but} \quad J \prec_{\mathrm{LP}} I \prec_{\mathrm{LP}} K.$$

That means:

Theorem 4. *The risk measured in terms of loss probability is not monotone in skewness (neither increasing nor decreasing).*

4.2 Worst Case Risk

When we have a look at I, J, K from Section 4.1, we observe that $\mathrm{WC}(I) = -2$, $\mathrm{WC}(J) = -1.4$, and $\mathrm{WC}(K) = -1.5$. Thus

$$I \prec_v J \prec_v K \quad \text{but} \quad I \prec_{\mathrm{WC}} K \prec_{\mathrm{WC}} J.$$

That means:

Theorem 5. *The risk measured in terms of the worst-case loss is not monotone in skewness (neither increasing nor decreasing).*

4.3 Value at Risk

As explained in Section 2, it might be preferable to investigate the value at risk rather than loss probabilities or worst-case losses. Here, we exemplarily use the levels 5% and 10%. Hence, let us have a look at the random variables L, M, N with

$$L = \begin{cases} \text{w/ prob. } 0.1: & -2 \\ \text{w/ prob. } 0.1: & -1 \\ \text{w/ prob. } 0.7: & +0.11195 \\ \text{w/ prob. } 0.1: & +2.2164 \end{cases},$$

$$M = \begin{cases} \text{w/ prob. } 0.1: & -1.9 \\ \text{w/ prob. } 0.1: & -1.1 \\ \text{w/ prob. } 0.7: & +0.10591 \\ \text{w/ prob. } 0.1: & +2.2586 \end{cases},$$

$$N = \begin{cases} \text{w/ prob. } 0.1: & -2.1 \\ \text{w/ prob. } 0.1: & -1.05 \\ \text{w/ prob. } 0.79: & +0.32207 \\ \text{w/ prob. } 0.01: & +6.0564 \end{cases}.$$

On the one side $v(L) = +0.18972$, $v(M) = +0.33408$, $v(N) = +1.2061$ hold but on the other side we calculate $\text{V@R}_{5\%}(L) = +2$, $\text{V@R}_{5\%}(M) = +1.9$, $\text{V@R}_{5\%}(N) = +2.1$. Thus,

$$L \prec_v M \prec_v N \quad \text{but} \quad N \prec_{\text{V@R}_{5\%}} L \prec_{\text{V@R}_{5\%}} M.$$

That means:

Theorem 6. *The risk measured in terms of the value at risk is in general not monotone in skewness (neither increasing nor decreasing).*

We showed this for the 5% level. Things become even worse when considering additionally the 10% value at risk: $\text{V@R}_{10\%}(L) = +1$, $\text{V@R}_{10\%}(M) = +1.1$, $\text{V@R}_{10\%}(N) = +1.05$, thus,

$$L \prec_v M \prec_v N \quad \text{but} \quad M \prec_{\text{V@R}_{10\%}} N \prec_{\text{V@R}_{10\%}} L.$$

Further, the examples I, J, K from Section 4.1 as well as L, M, N from the present section demonstrate that risk preferences built upon worst case loss are generally inconsistent to those built on loss probabilities as well as that preferences built on value at risk are inconsistent among different levels. This is true because $J \prec_{\text{LP}} I \prec_{\text{LP}} K$ but $I \prec_{\text{WC}} K \prec_{\text{WC}} J$ and $N \prec_{\text{V@R}_{5\%}} L \prec_{\text{V@R}_{5\%}} M$ but $M \prec_{\text{V@R}_{10\%}} N \prec_{\text{V@R}_{10\%}} L$ hold.

5 Connections of Probabilities and (Pearson) Skewness

In this section, we analyze the connections between skewness as well as Pearson skewness to (loss) probabilities in greater detail. When thinking about Markov's inequality [see, e.g., Georgii, 2012, Proposition 5.4 and esp. its proof] or Chebyshev's inequality, it might be quite natural to guess that one can conclude statements about (loss) probabilities from information about skewness resp. third moments. Thus, when again having a look at Figure 1 one might guess that, e.g., by means of Markov's inequality a connection between the skewness of two random variables and their probabilities to be above their respective means could be proven. However, the following theorem shows that this is not true.

Theorem 7. *Let X and Y be two real-valued random variables with $v(X) < v(Y)$. In general, one can conclude neither $\mathbb{P}(X < \mathbb{E}[X]) < \mathbb{P}(Y < \mathbb{E}[Y])$ nor $\mathbb{P}(X < \mathbb{E}[X]) \leq \mathbb{P}(Y < \mathbb{E}[Y])$ nor $\mathbb{P}(X < \mathbb{E}[X]) > \mathbb{P}(Y < \mathbb{E}[Y])$ nor $\mathbb{P}(X < \mathbb{E}[X]) \geq \mathbb{P}(Y < \mathbb{E}[Y])$.*

Proof. Proof of Theorem 7. The statement can either be derived from the results provided by von Hippel [2005] or be easily seen by recalling I, J, K from Section 4.1. It holds $v(I) < v(J)$ but $\text{LP}(I) < \text{LP}(J)$, and $v(J) < v(K)$ but $\text{LP}(J) > \text{LP}(K)$, from which the statement follows when keeping in mind that $\mu(I) = \mu(J) = \mu(K) = 0$. □

We notice that $\sigma(I) = \sigma(J) = \sigma(K) = 1$ holds, i.e., the counterexample does not rely on differing variances.

Now, the reader may ask for the intuition behind this result. Maybe, the answer is that it is rather hard to imagine a third moment (upon which Fisher's moment coefficient of skewness is built). Rather, we might have plots like Figure 1 in mind when thinking about skewness. Via such visualizations Pearson skewness can be motivated, especially the properties 'positively Pearson skewed' and 'negatively Pearson skewed' hold if and only if the mean is on the right resp. left of the median. Indeed it turns out that positively/negatively Pearson skewed is an adequate concept for making statements about probabilities, as shown next.

Theorem 8. *Let X and Y be two real-valued random variables with $S(X) < 0 < S(Y)$. Then the inequality $\mathbb{P}(X < \mathbb{E}[X]) \leq \mathbb{P}(Y < \mathbb{E}[Y])$ holds.*

Proof. Proof of Theorem 8. We note that $S(X) < 0 \;\Leftrightarrow\; \mu(X) < \mathrm{med}(X)$, $S(Y) > 0 \;\Leftrightarrow\; \mathrm{med}(Y) < \mu(Y)$, and $\mathbb{P}(X \geq \mathrm{med}(X)) \geq 0.5$. Now, we calculate:

$$\begin{aligned}\mathbb{P}(X < \mu(X)) &\leq \mathbb{P}(X \leq \mu(X)) \leq \mathbb{P}(X < \mathrm{med}(X)) \\ &\leq 0.5 \leq \mathbb{P}(Y \leq \mathrm{med}(Y)) \leq \mathbb{P}(Y < \mu(Y)) \leq \mathbb{P}(Y \leq \mu(Y)).\end{aligned}$$

□

The last step in the proof is not necessary but may be of interest to the reader, too. When comparing Theorems 7 and 8 we highlight three aspects: In Theorem 8, i) the inequality is not strict, ii) the Pearson skewness instead of Fisher's moment coefficient of skewness is used, and iii) the Pearson skewness values are separated by zero, i.e. the values have different signs. In the three theorems to follow, we show that all these aspects are essential. First, a strict inequality in Theorem 8 cannot be achieved, as the following theorem shows.

Theorem 9. *Let X and Y be two real-valued random variables with $S(X) < 0 < S(Y)$. In general, $\mathbb{P}(X < \mathbb{E}[X]) < \mathbb{P}(Y < \mathbb{E}[Y])$ does not hold.*

Proof. Proof of Theorem 9. We consider O and P with

$$O = \begin{cases} \text{w/ prob. } 0.2: & -1.7944 \\ \text{w/ prob. } 0.3: & -0.0037133 \\ \text{w/ prob. } 0.4: & +0.5 \\ \text{w/ prob. } 0.1: & +1.6 \end{cases},$$

$$P = \begin{cases} \text{w/ prob. } 0.1: & -1.6 \\ \text{w/ prob. } 0.4: & -0.5 \\ \text{w/ prob. } 0.3: & +0.0037133 \\ \text{w/ prob. } 0.2: & +1.7944 \end{cases}.$$

We compute $S(O) = -0.24814 < 0 < S(P) = +0.24814$ and remark $\mathrm{LP}(O) = \mathrm{LP}(P) = 0.5$ (with $\mu(O) = \mu(P) = 0$ and $\sigma(O) = \sigma(P) = 1$). □

Clearly, this counterexample is based on the fact that there is more than one median in each case. Further, it is not possible to replace the Pearson skewness by the skewness in Theorem 8.

Theorem 10. *Let X and Y be two real-valued random variables with $v(X) < 0 < v(Y)$. In general, one can conclude neither $\mathbb{P}(X < \mathbb{E}[X]) < \mathbb{P}(Y < \mathbb{E}[Y])$ nor $\mathbb{P}(X < \mathbb{E}[X]) \leq \mathbb{P}(Y < \mathbb{E}[Y])$ nor $\mathbb{P}(X < \mathbb{E}[X]) > \mathbb{P}(Y < \mathbb{E}[Y])$ nor $\mathbb{P}(X < \mathbb{E}[X]) \geq \mathbb{P}(Y < \mathbb{E}[Y])$.*

Proof. Proof of Theorem 10. Again we recall I, J, K from Section 4.1. It holds $v(I) < 0 < v(J)$ but $\mathrm{LP}(I) < \mathrm{LP}(J)$ and $v(I) < 0 < v(K)$ but $\mathrm{LP}(I) > \mathrm{LP}(K)$, from which the statement follows when keeping in mind that $\mu(I) = \mu(J) = \mu(K) = 0$. □

As the last statement in this paper we formulate that in Theorem 8 the separation by zero is essential.

Theorem 11. *Let X and Y be two real-valued random variables with $S(X) < S(Y)$. In general, one can conclude neither $\mathbb{P}(X < \mathbb{E}[X]) < \mathbb{P}(Y < \mathbb{E}[Y])$ nor $\mathbb{P}(X < \mathbb{E}[X]) \leq \mathbb{P}(Y < \mathbb{E}[Y])$ nor $\mathbb{P}(X < \mathbb{E}[X]) > \mathbb{P}(Y < \mathbb{E}[Y])$ nor $\mathbb{P}(X < \mathbb{E}[X]) \geq \mathbb{P}(Y < \mathbb{E}[Y])$.*

Proof. Proof of Theorem 11. We consider I and K from Section 4.1 and Q with

$$Q = \begin{cases} \text{w/ prob. } 0.1: & -1.5 \\ \text{w/ prob. } 0.26: & -1 \\ \text{w/ prob. } 0.4: & +0.15424 \\ \text{w/ prob. } 0.24: & +1.4513 \end{cases}.$$

It holds $\mu(Q) = \mu(I) = \mu(K) = 0$, $\sigma(Q) = 1$, $S(I) = -0.38462$, $S(K) = -0.043250$, $S(Q) = -0.15424$, and $\mathrm{LP}(Q) = 0.36$, thus,

$$S(I) < S(Q) < S(K) \quad \text{but} \quad \mathrm{LP}(K) < \mathrm{LP}(I) < \mathrm{LP}(Q).$$

□

6 Conclusion and Future Research

In this paper, we discussed statistical skewness measures (Fisher's moment coefficient of skewness as well as Pearson skewness) and elaborated with the help of examples and the literature that preferences of agents, traders, players may be explained with the help of skewness. However, we were able to show with easy-to-understand examples that preference structures based on loss probability, worst-case loss, or value at risk do not have to match those based on Fisher's moment coefficient of skewness. Furthermore, we analytically investigated the relationship of Fisher's moment coefficient of skewness as well as Pearson skewness and probabilities. We proved that for two random variables, one of which is positively Pearson skewed and one of which is negatively Pearson skewed, the probability of the positively Pearson skewed being below its expected value is larger than those of the negatively Pearson skewed (below its respective expected value), see Theorem 8.

From a practical point of view, the results in this paper might be noteworthy insofar as traders might want to carefully consider whether to use skewness as a proxy for a risk indicator or rather use a monetary risk measure that could better suit their idiosyncratic needs. Also, traders should always be aware of the fact that skewness and Pearson skewness can differ not only quantitatively but also qualitatively [see von Hippel, 2005]. Theorem 8 may suggest that traders might think about incorporating Pearson skewness in their decision making process for portfolio selection.

For further research, the question arises whether one can show the mismatch of preference structures based on skewness and those based on risk for general (i.e. all) monetary risk measures. Possibly, there are classes of risk measures that match to skewness while other do not. Other measures of skewness could also be interesting, just like assumptions on distributions (e.g., three-point distributions or absolutely continuous ones) or parameters. These points are particularly critical when observing the mismatches of preferences based on different risk indicators as shown in this work.

Acknowledgements

The author thanks Melanie Birke, Bernhard Herz, and Lars Grüne (all with Universtiy of Bayreuth, Bayreuth, Germany) and Michaela Baumann (with NÜRNBERGER Lebensversicherung AG, Nuremberg, Germany).

All pictures were created with `R` using the packages `evd: Functions for Extreme Value Distributions` and `VGAM: Vector Generalized Linear and Additive Models.` All calculations were made by hand, in `R`, or in "Microsoft Excel," partially with "Visual Basic for Applications." This paper was set in LaTeX using the packages `amssymb, xcolor, amsmath, amsthm, graphicx, natbib.`

References

C. P. Alderfer and H. Bierman Jr. Choices with Risk: Beyond the Mean and Variance. *The Journal of Business*, 43(3):341–353, 1970.

P. Barroso and P. Santa-Clara. Momentum has its Moments. *Journal of Financial Economics*, 116(1):111–120, 2015.

M. H. Baumann. Beating the Market? A Mathematical Puzzle for Market Efficiency. *Decisions in Economics and Finance*, 45(1):279–325, 2022.

R. Bellman. Decision Making in the Face of Uncertainty? I. *Naval Research Logistics Quarterly*, 1(3):230–232, 1954.

W. Briec, K. Kerstens, and O. Jokung. Mean-Variance-Skewness Portfolio Performance Gauging: A General Shortage Function and Dual Approach. *Management Science*, 53(1):135–149, 2007.

P. L. Brockett and Y. Kahane. Risk, Return, Skewness and Preference. *Management Science*, 38(6):851–866, 1992.

M. C. Campi and S. Garatti. *Introduction to the Scenario Approach*. SIAM, 2018.

P. Fernandez. The Equity Premium in 100 Textbooks. *Journal of New Finance*, 1(3/3), 2020.

H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*, volume Third Edition. De Gruyter Textbook, 2011.

T. A. Garrett and R. S. Sobel. Gamblers Favor Skewness, Not Risk: Further Evidence from United States' Lottery Games. *Economics Letters*, 63(1):85–90, 1999.

H.-O. Georgii. *Stochastics*. De Gruyter, 2012.

J. Golec and M. Tamarkin. Bettors Love Skewness, Not Risk, at the Horse Track. *Journal of Political Economy*, 106(1):205–225, 1998.

J. Kim and Y. J. Park. Individual Investors, Average Skewness, and Market Returns. *Sustainability*, 12(20):8357, 2020.

A. Kraus and R. H. Litzenberger. Skewness Preference and the Valuation of Risk Assets. *The Journal of Finance*, 31(4):1085–1100, 1976.

X. Li, Z. Qin, and S. Kar. Mean-Variance-Skewness Model for Portfolio Selection with Fuzzy Returns. *European Journal of Operational Research*, 202(1):239–247, 2010.

S. Malekpour and B. R. Barmish. How Useful are Mean-Variance Considerations in Stock Trading via Feedback Control? In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 2110–2115. IEEE, 2012.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

H. M. Markowitz and G. P. Todd. *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, volume 66. John Wiley & Sons, 2000.

R. Martin. Momentum Trading:'skews me. *Risk*, 25(8):40, 2012.

T. C. Mills. Modelling Skewness and Kurtosis in the London Stock Exchange FT-SE Index Return Distributions. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 44(3):323–332, 1995.

J. Steindl. On Risk. *Oxford Economic Papers*, (5):43–53, 1941.

L. Swedroe. Swedroe: Keep Skewness in Perspective, 2015. URL www.finance.yahoo.com/news/swedroe-keep-skewness-perspective-163703392.html. Retrieved April 21, 2022.

G. G. Szpiro. Risk Aversion as a Function of Variance and Skewness. In *Progress in Decision, Utility and Risk Theory*, pages 355–363. Springer, 1991.

P. T. von Hippel. Mean, Median, and Skew: Correcting a Textbook Rule. *Journal of Statistics Education*, 13(2), 2005.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*, volume Second Rev. Edition. Princeton University Press, 1947.

S.-P. Yang and T. Nguyen. Skewness Preference and Asset Pricing: Evidence from the Japanese Stock Market. *Journal of Risk and Financial Management*, 12(3):149, 2019.