Modeling Skewness Determinants in Accounting Research

Sudipta Basu

Dmitri Byzalov

Temple University

Abstract

Skewness-based proxies are widely used in accounting and finance research. To study how the skewness of a dependent variable Y varies with conditioning variables X, researchers typically compute firm-specific skewness measures over a short window and regress them on X. However, we show that this standard approach can cause severe biases and produce false findings of both conditional skewness on average and systematic variation in conditional skewness. We develop alternative methods that address these biases by directly modeling the conditional skewness of Y for each observation as a function of X. Simulations confirm that our methods have good type-I errors and test power even in scenarios in which the standard method is severely biased. Our methods are transparent, robust, and can be implemented in a few lines of code. Use of our methods changes major prior findings in the literature.

1. Introduction

Researchers in accounting, finance, and other fields often study how the conditional skewness of a Y variable is affected by a vector of X variables. For example, accrual skewness and stock return skewness are widely used as proxies for conditional conservatism (Basu, 1995; Givoly and Hayn, 2000; Ahmed and Duellman 2013; and many others) and crash risk (e.g., Chen et al., 2001), respectively, and researchers examine various determinants of these phenomena. We show that the standard empirical approach in the literature for studying skewness determinants can cause severe biases. We propose alternative methods that resolve these biases and are robust and easy to use.

Skewness is pervasive.[1] It arises from the economics of superstars (Rosen, 1981), asymmetric reporting rules (e.g., Basu, 1995), real-option effects (e.g., Burgstahler and Dichev, 1997), and other phenomena, which researchers can learn about by studying skewness determinants.

To study any variation in skewness, a researcher must use the appropriate *conditional skewness* definition. For example, to study variation across subsamples, she should use the definition $Skew(Y|Subsample)$, i.e., the skewness of the conditional distribution of Y within a given subsample. To study how skewness varies with an X vector (e.g., firm characteristics), she should use $Skew(Y|X)$, i.e., the skewness of the conditional distribution of Y for a given value of X.

To estimate how an X vector affects the conditional skewness of Y, researchers typically use a rolling-window approach. They first compute the skewness of Y over a firm-specific rolling window, implicitly assuming that skewness "conditional on the window" is an acceptable proxy

---

[1] For example, among research faculty in 30 leading US economics departments, a top-1% researcher has 4.9 times more citations than a top-10% researcher and 49.4 times more citations than a median researcher (Hamermesh, 2018). Among savanna baboons, the top-ranked male in a troop has 3.6 times more reproductive success than ranks 2–4 combined (Bulger, 1993). The top 3% of YouTube channels account for 90% of all views (Bartl, 2018). The top 3% of super-spreaders caused 61% of all infections during the West African Ebola outbreak of 2014 (Wong et al., 2015). The top 3% of street segments in Boston generate more than half of all gun violence (Braga et al., 2010). The sickest 5% of the population consume more than half of all US health care spending (Pearl and Madvig, 2020). The top 10 trading days account for two-thirds of the total returns on the Dow Jones Industrial Average from 1900 to 2006 (Estrada, 2009).

for skewness "conditional on X" for the observations in the window, and then regress this proxy on X. However, this standard method can cause spurious inferences. Suppose that X for a given firm varies over time, and that X affects the conditional mean $E(Y|X)$ and standard deviation $SD(Y|X)$. The assumed distribution of Y around the conditional mean is symmetric, i.e., the true conditional skewness $Skew(Y|X)$ is zero for any X (Panel A of Figure 1). Because X varies over time, the window for the computation of skewness combines the conditional distributions of Y for different values of X. Although each underlying distribution is symmetric, the combined distribution within the window is asymmetric due to uncontrolled variation in X (Panel B of Figure 1). This leads to spurious findings of conditional skewness in the standard method. Further, the spurious conditional skewness varies systematically with the values of X in the window (Panel C of Figure 1). This leads to spurious findings that X affects the conditional skewness of Y, even when the true effect is zero. Figures 2 and 3 illustrate two additional biases in the standard method that arise respectively when there is persistent time-series variation in Y and when the number of observations varies across the windows. In all three scenarios, the faulty inferences arise because skewness "conditional on the window" is an invalid proxy for skewness "conditional on X" for the observations in the window.

To address these biases, we develop alternative methods that estimate the conditional skewness of Y as a function of X. Instead of using an *ad hoc* window to approximate skewness conditional on X, we directly implement the statistical definition of conditional skewness. We estimate the individual components of this definition, such as the mean and standard deviation of Y conditional on X, and then combine them to empirically model the skewness of Y for each *individual observation* as a function of its X. The different model components can be estimated jointly using the Generalized Method of Moments (GMM) or sequentially using off-the-shelf regressions (the

Appendix shows how both methods can be implemented in a few lines of code). By directly linking the data to the statistical definitions, we avoid the biases from Figures 1–3 and get valid estimates of the conditional skewness of Y as a function of X. Simulations confirm that our methods have good type-I errors and test power even in scenarios where the standard method is severely biased.

Researchers typically use Pearson's moment coefficient of skewness, which is based on the normalized cubic moment of Y. However, for many fat-tailed distributions, Pearson's skewness is theoretically undefined because the cubic moment explodes to infinity (Jondeau and Rockinger, 2003). The empirical skewness estimate is then uninterpretable. Even when Pearson's skewness is interpretable (i.e., the cubic moment is finite), it is extremely sensitive to outliers (Kim and White, 2004). Therefore, following Kim and White (2004), we examine an alternative skewness measure that is based on asymmetry in the conditional quantiles of Y (e.g., Q10 and Q90) relative to the conditional median. We show how a researcher can estimate this skewness measure as a function of X using off-the-shelf quantile regressions. In simulations, the estimates for quantile-based skewness are more robust to outliers than are the estimates for Pearson's skewness, as expected. Therefore, we recommend that researchers apply our methods to quantile-based skewness because it is more reliable than standard Pearson's moment coefficient of skewness.[2]

Our statistical modeling approach generalizes beyond skewness. For example, Dutta and Patatoukas (2017) revive a conservatism measure that is based on asymmetry in accrual variances between good-news and bad-news observations (Basu, 1995). If a researcher uses the standard

---

[2] Quantile-based skewness has not been used in accounting research because it cannot be computed on a typical short rolling window. For example, Q10 and Q90 cannot be measured at all on a rolling window of 5 observations, whereas Pearson's moment coefficient of skewness can be measured (albeit very unreliably) on just 3 observations. Because we directly model the conditional quantiles of Y as a function of X for each observation, instead of loosely approximating this function using a short window, we avoid this problem. Thus, while the standard approach forces a researcher to use Pearson's moment coefficient of skewness, our approach offers more flexibility in the choice of skewness metric. Some research contexts might suggest alternative skewness metrics, such as the scaled difference between the conditional mean and the conditional median (Section 2.5).

window-based method to study this measure's determinants, then she risks false inferences through the mechanisms shown in Figures 1–3. However, a straightforward modification of our methods resolves this problem in a few lines of code (Section 2.5).

Ahmed and Duellman (2013) find that managerial overconfidence is significantly negatively associated with skewness-based conservatism proxies, computed on a five-year rolling window. We successfully replicate their results. However, these results are susceptible to the biases from Figures 1–2. When we use our methods to remove these biases, the relation between managerial overconfidence and skewness-based proxies becomes insignificant in almost all cases (Section 4).

Section 2 explains the biases in the standard method and presents our new methods for estimating the determinants of conditional skewness. Section 3 provides simulation evidence. Section 4 revisits prior findings. Section 5 concludes.

2. Estimation of skewness determinants

2.1. *Unconditional versus conditional skewness*

Pearson's unconditional moment coefficient of skewness for a dependent variable Y is

$$Skew(Y) = E\left[\left(\frac{Y - E(Y)}{SD(Y)}\right)^3\right] \tag{1}$$

where $E(Y)$ and $SD(Y)$ are the unconditional mean and standard deviation of Y, respectively. Its empirical counterpart is the sample mean of $\left(\frac{Y-m}{s}\right)^3$, where m and s are the sample mean and standard deviation of Y.

Unconditional skewness is meaningful only for the full sample. To study variation in skewness, a researcher must use *conditional* skewness defined for specific conditioning variables. In other words, because any variation is *across something* (e.g., across subsamples or across different values of X), the analysis must be clearly conditioned on that something.

For example, to study variation in skewness across subsamples, a researcher must measure conditional skewness $Skew(Y|Subsample)$ for each subsample. $Skew(Y|Subsample)$ focuses on the conditional Y distribution for an individual subsample, which has a subsample-specific conditional mean $E(Y|Subsample)$ and standard deviation $SD(Y|Subsample)$. By only using within-subsample variation, this measure separates the skewness of Y within a given subsample from the variation in the conditional mean and standard deviation of Y across the subsamples.

Similarly, to study how skewness varies with explanatory variables $X = (X_1, ..., X_K)$, a researcher must measure skewness conditional on X, i.e., $Skew(Y|X)$ as a function of X. Formally,

$$Skew(Y|X) = E\left[\left(\frac{Y - E(Y|X)}{SD(Y|X)}\right)^3 |X\right] \qquad (2)$$

For a given value of the X vector, Y is a random variable that fluctuates around the conditional mean $E(Y|X)$ with conditional standard deviation $SD(Y|X)$, where $E(Y|X)$, $SD(Y|X)$, and the shape of the conditional Y distribution can all vary with X. The rescaling $\frac{Y-E(Y|X)}{SD(Y|X)}$ in (2) normalizes the mean to zero and standard deviation to one for all values of X while preserving the shape of the conditional Y distribution, including its skewness (i.e., asymmetry in the shape of the distribution). Therefore, (2) isolates the net effect of X on the conditional skewness of Y, separating it from potential confounding effects of X on the conditional mean and standard deviation of Y.

The conditioning vector X in the formal skewness definition (2) must be consistent with the empirical variables that a researcher uses to explain variation in skewness. For example, if a researcher wants to study the impact of X_2 on skewness after controlling for X_1, then skewness (2) must be defined conditional on these specific variables, i.e., $Skew(Y|X_1, X_2)$. It would be *a priori* unreasonable to study how the empirical explanatory variables X_1, X_2 affect $Skew(Y|X_1)$, or $Skew(Y|X_1, X_2, X_3)$, or $Skew(Y|X_3, X_4)$. Because by definition $Skew(Y|X_1)$ is a function of X_1

only, it is meaningless to ask how it varies with X_2 after controlling for X_1. Because $Skew(Y|X_1, X_2, X_3)$ is a function of X_1, X_2, and X_3, any analysis of how it varies with X_1 and X_2 only is confounded by self-inflicted correlated omitted variable bias for X_3.[3] Because the conditioning variables in $Skew(Y|X_3, X_4)$ differ from the empirical X_1 and X_2, any analysis of how $Skew(Y|X_3, X_4)$ varies with X_1 and X_2 is uninterpretable. Notably, this last issue undermines typical regression analyses for subsample-based skewness measures, where the subsample identifier variables (e.g., $X_3 = industry$ and $X_4 = year$) differ from the regression variables.

Which conditioning variables should and should not be included in the X vector depends on a researcher's economic interpretation of skewness. For example, if accrual skewness is used as a proxy for conditional conservatism, then the measurement should focus on the distribution of accrual fluctuations that are driven by concurrent news. In other words, the researcher should interpret news for the firm-year as a *random variable* that is the source of relevant fluctuations in accruals, instead of using the *observed value* of news for the firm-year as a control to remove these news-driven fluctuations. Therefore, the X vector should *not* include concurrent news.[4] The X vector should include the test variables (i.e., new skewness determinants) and relevant control

---

[3] Because conditional skewness (2) is a non-linear function of several conditional expectations, the law of iterated expectations does not fully apply. In other words, while $E\{E(Y|X_1, X_2, X_3)|X_1, X_2\}$ simplifies to $E(Y|X_1, X_2)$ through iterated expectations, $E\{Skew(Y|X_1, X_2, X_3)|X_1, X_2\}$ does *not* simplify to $Skew(Y|X_1, X_2)$. Instead, it becomes $E\left[\left(\frac{Y-E(Y|X_1,X_2,X_3)}{SD(Y|X_1,X_2,X_3)}\right)^3 |X_1, X_2\right]$, which is an uninterpretable mess of mutually inconsistent conditioning vectors. It captures how the skewness of the residual variation in Y after controlling for X_1, X_2, and X_3 varies with X_1 and X_2 without controlling for X_3. We are not aware of any potential applications for such a measure.

[4] Suppose that the true accrual process is $Accr = \beta_1 News + \beta_2 I_{bad} News + \varepsilon$, where $News$ is any appropriate news proxy, I_{bad} is a bad-news dummy, ε is a random shock, and β_2 is a positive asymmetric timeliness coefficient. Conservatism leads to accrual skewness because $\beta_1 News + \beta_2 I_{bad} News$ converts any given news distribution into a more left-skewed accrual distribution. Accrual skewness conditional on news would miss this effect, because $\beta_1 News + \beta_2 I_{bad} News$ conditional on $News$ is a constant, and the only remaining source of fluctuations in accruals is the random shock ε. Thus, $Skew(Accr|News)$ only captures the skewness of ε. In contrast, accrual skewness without conditioning on news captures the distribution of $\beta_1 News + \beta_2 I_{bad} News + \varepsilon$, treating both $News$ and ε as random variables. This captures the impact of conservatism on the accrual distribution. Similarly, if the true process is $Accr = (\beta_1 + \beta_{1X}X)News + (\beta_2 + \beta_{2X}X)I_{bad} News + \gamma_X X + \varepsilon$ with moderating factors X, then $Skew(Accr|X)$ captures the impact of conservatism on the accrual distribution at a given value of X, treating $News$ and ε as random variables.

variables that are unrelated to news and that likely affect the conditional mean, variance, or skewness of accruals (e.g., beginning-of-period size, book-to-market ratio, and leverage[5]).

2.2. *Biases in the standard method for measuring variation in conditional skewness*

When the X vector takes on just a few discrete values, a researcher can directly estimate conditional skewness (2) using subsamples for each possible value of X. However, when X includes continuous variables (or many discrete variables), this direct approach is impractical because the sample likely contains only a few observations for each distinct value of X.

To address this problem, researchers typically compute skewness on a firm-specific rolling window, implicitly assuming that this computation yields an acceptable empirical proxy for skewness conditional on X. In other words, they use unstructured variation across firms and rolling-window intervals to approximate variation conditional on X. For example, Gu and Wu (2003), Beatty et al. (2008), and Ahmed and Duellman (2013) use rolling windows of 8 quarters, 20 quarters, and 5 years, respectively.[6] After computing these skewness proxies, researchers use them either as a dependent variable in an analysis of skewness determinants (e.g., Peek et al., 2010; Ahmed and Duellman, 2013) or as an explanatory variable in an analysis of skewness consequences (e.g., Gu and Wu, 2003; Beatty et al., 2008). However, the standard rolling-window approximation of conditional skewness is likely to cause both conceptual and empirical problems.

---

[5] End-of-period controls can embed concurrent news relevant for conservatism. Therefore, it is better to use beginning-of-period values to avoid "included variable bias" (e.g., Killingsworth, 1993) due to overcontrolling.

[6] For ease of exposition, we focus on firm-specific measures. Givoly and Hayn (2000) compute cross-sectional skewness based on year-specific subsamples of firms. Peek et al. (2010) compute skewness in subsamples based on country, year, and listing status. Because these computations use much larger subsamples than a typical firm-specific computation, they are even more susceptible to the bias illustrated in Figure 1. Our methods generalize to these alternative subsamples. In the finance literature, Harvey and Siddique (1999), White et al. (2008), Engle (2011), and others estimate complex GARCH-type time-series models for conditional skewness. However, these models are not designed to easily incorporate multiple explanatory variables and the estimation is prohibitively complex.

Conceptually, the rolling-window computation produces estimates of $E(Y|window\ FE)$, $SD(Y|window\ FE)$, and $Skew(Y|window\ FE)$, where $window\ FE$ can be interpreted as a vector of dummies for all rolling windows in the data at the firm × time interval level. To study skewness determinants, a researcher then regresses the estimate of $Skew(Y|window\ FE)$ on explanatory variables X, which produces an estimate of $E\{Skew(Y|window\ FE)|X\}$. This estimate is uninterpretable because it uses two mutually inconsistent sets of conditioning variables.[7]

Empirically, the standard method causes three systematic biases. The first bias (Figure 1) arises when there is time-series variation in explanatory variables. It can generate spurious findings of both conditional skewness on average and systematic variation in conditional skewness, even when the true conditional skewness is zero. To show this bias in a simple example, we assume that Y follows a Normal distribution with mean $E(Y|X) = X$ and standard deviation $SD(Y|X) = 1.5 + 0.25 \times X$ (Panel A of Figure 1). For simplicity, X follows a symmetric three-point distribution $\{-2, 0, 2\}$ with probability 1/3 each.[8]

Because the Normal distribution is symmetric, the true conditional skewness of Y is zero for all values of X. However, the standard method will falsely detect conditional skewness. When X varies over time for a given firm, the window for the standard computation of skewness combines the conditional distributions of Y for several different values of X. For example, because X takes on

---

[7] For example, because $window\ FE$ is constant for the duration of the window, $Skew(Y|window\ FE)$ cannot adequately control for the effect of time-series variation in X on the conditional mean and standard deviation of Y. Because $window\ FE$ allows any pattern of variation across firms, it is likely to overcontrol for persistent cross-sectional differences in Y that are unrelated to X. Thus, the conditioning on $window\ FE$ is both too inflexible and too flexible. Additionally, the standard estimate of $Skew(Y|window\ FE)$ is biased and inconsistent, even for an infinite sample, because estimation on a short rolling window causes the incidental parameter problem (Neyman and Scott, 1948) for all non-linear estimators, including the skewness estimator. It is unclear how this bias affects inferences, even when the confounding effects from Figures 1–3 do not arise.

[8] A skewed X distribution and/or a non-linear effect of X on Y could be additional, more mechanical sources of spurious findings of conditional skewness in the standard method. Our example in Figure 1 does not rely on such mechanical asymmetries. Instead, it shows that spurious conditional skewness arises even when all of the model components are perfectly linear and/or symmetric.

the values {−2, 0, 2} with equal probabilities in our example, a representative rolling window contains observations with $X=-2$, $X=0$, and $X=2$ in equal proportions. The probability density function of Y for this representative window is an equal-weighted average of the three conditional probability density functions from Panel A of Figure 1. Although each of these conditional density functions is symmetric, their equal-weighted average is asymmetric, as shown in Panel B of Figure 1. Thus, the Y distribution in a representative rolling window is skewed due to uncontrolled within-window variation in X.[9] Therefore, the standard method will falsely detect conditional skewness on average, even though the true skewness conditional on X is zero by construction.

This spurious conditional skewness varies systematically with the values of X for the observations in the rolling window. Panel C of Figure 1 presents the probability density function of Y for a window with $X=-2$ and $X=0$ versus a window with $X=0$ and $X=2$, computed by averaging the conditional densities from Panel A for the relevant X values. The Y distribution is considerably more skewed in the first window than in the second window (Panel C). Therefore, the standard method will falsely detect systematic variation in conditional skewness and will falsely attribute it to X, even though the true conditional skewness is zero by construction and is unaffected by X.

The second bias (Figure 2) arises when Y contains a persistent time-series component that is unrelated to X, such as a firm random effect or an autoregressive shock. The standard method filters out persistent variation in Y, whether relevant or irrelevant, by subtracting the firm-specific mean of Y for the rolling window. Thus, it can exclude relevant variation from the computation of skewness conditional on X. For example, let $Y = \beta X + \varepsilon_{firm} + \varepsilon_{firm-year}$, where X only varies

---

[9] After computing skewness using a rolling window, a researcher can in principle regress it on X for all observations in the window. However, this would not adequately control for within-window variation in X. To eliminate the confounding effect from Figure 1, such controls must be implemented *inside* the computation of skewness. Our estimation methods in Sections 2.3–2.5 directly implement such controls.

cross-sectionally (such that the bias from Figure 1 does not arise), ε_{firm} is a firm effect drawn from a skewed distribution, and $\varepsilon_{firm-year}$ is a temporary shock drawn from a symmetric distribution. Skewness conditional on X captures all sources of variation in Y that are unrelated to X, i.e., the firm effects ε_{firm} and the temporary shocks $\varepsilon_{firm-year}$ (Panel A of Figure 2). In contrast, the rolling-window skewness computation only focuses on the within-window variation in Y. Because the firm effect ε_{firm} is fixed within the window, it drops out. As a result, measured skewness in the window only captures the distribution of the temporary shocks $\varepsilon_{firm-year}$ and ignores the distribution of the firm effects ε_{firm} (Panel B). Thus, because the standard method overcontrols for firm effects (and other forms of persistent variation in Y), it misses a major potential source of skewness conditional on X.[10]

The third bias (Figure 3) arises in the standard method when the number of observations varies across rolling windows. For example, Gu and Wu (2003) use a rolling window of 8 periods and require at least 5 valid observations, so the actual number of observations in the window varies from 5 to 8 depending on the available data. Because empirical Pearson's skewness is a non-linear function of noisy empirical mean and standard deviation, it is biased in a short rolling window, and the severity of the bias varies with the window size (Cox, 2010). Therefore, variation in the number of observations per window is likely to be mistaken for variation in conditional skewness. Beatty et al. (2008) require only 5 non-missing observations out of 20 total observations, so the bias is likely to be greater than in Gu and Wu (2003).

---

[10] Firm effects are particularly important in strategy research (e.g., Alcacer et al., 2018). This research emphasizes a firm's unique strategic positioning, which by definition is incremental to any generic firm characteristics X. Strategic positioning has long-term performance consequences, which can be modeled as systematic firm effects, and these consequences are likely to be highly right skewed, especially in winner-take-all markets.

To illustrate this bias, we generate 10,000 firm-specific windows of N observations each, where N = 5, 10, 20, 100, and 1000. For each observation, we draw Y from non-central Student's t distribution (Harvey and Siddique, 1999) with 5 degrees of freedom and non-centrality parameter δ=3. This distribution has constant true Pearson's skewness of approximately 2.5. We compute empirical skewness of Y in each window and then compute the mean of these estimates across the 10,000 windows of size N.

Figure 3 summarizes the results. For N=5, the mean skewness estimate is 0.34, less than one-sixth of the true value (2.5). Researchers sometimes adjust skewness for degrees of freedom to mitigate the small-sample bias.[11] The adjusted mean estimate is 0.70, which is less than one-third of the true value. For N=20, the mean estimate (both unadjusted and adjusted) is less than one-half of the true value. Even for N=100, the mean estimate is only two-thirds of the true value.

The bias in Figure 3 varies systematically with the window size N. Thus, if the number of valid observations varies across rolling windows (e.g., because a researcher allows a few missing observations per window), then this variation in bias will be mistaken for variation in conditional skewness. This can confound inferences about any variable that is correlated with a firm's number of observations (e.g., any proxy for a firm's age or the probability of exiting the sample due to bankruptcy, delisting, acquisition, or any other reason).[12]

---

[11] The standard Pearson's moment coefficient of skewness formula $\frac{1}{N}\sum\left(\frac{Y_i-m}{s}\right)^3$ does not adjust for degrees of freedom. The adjusted formula is $\frac{N}{(N-1)(N-2)}\sum\left(\frac{Y_i-m}{s}\right)^3$ (e.g., Gu and Wu, 2003). Stata uses the unadjusted formula, Excel uses the adjusted formula, and SAS uses the adjusted formula by default but allows the unadjusted option.

[12] A researcher could try to address this bias by requiring 100% data availability within a fixed-length rolling window. However, this approach is practical only for a relatively short window (e.g., requiring 20 valid observations in a 20-period window would massively reduce sample size and could cause severe survivorship bias). A researcher could also try to control for the number of valid observations N in the rolling window; however, N would need to be interacted with all of the X variables because the bias for a given N depends on the true level of skewness. Our methods in Sections 2.3–2.5 provide a direct solution.

The biases from Figures 1–3 are likely to distort standard inferences about both the determinants and the consequences of conditional skewness. In the remainder of this section, we develop methods that address these biases in the analysis of skewness determinants. Technically, our methods can also be used to study skewness consequences, as we explain in Section 2.6. However, this analysis is appropriate only if a researcher has good instruments for conditional skewness, which is unusual in practice.

2.3. *Our method for standard Pearson's moment coefficient of skewness*

Instead of using an *ad hoc* rolling-window approximation for skewness conditional on X, we directly estimate conditional skewness (2) from the data. To implement (2) empirically for a general X vector that can incorporate any number of variables, we specify the functional forms

$$E(Y|X) = \mu_0 + \mu_X X \qquad (3a)$$

$$SD(Y|X) = exp(\pi_0 + \pi_X X) \qquad (3b)$$

$$Skew(Y|X) = \beta_0 + \beta_X X \qquad (3c)$$

where μ_0, π_0, and β_0 are scalar intercepts, and μ_X, π_X, and β_X are coefficient vectors conformable with the X vector. To capture non-linear effects, a researcher can include in the X vector any relevant transformations and interactions of the original explanatory variables. The exponential specification for the conditional standard deviation in (3b) ensures that $SD(Y|X)$ is strictly positive and monotonic with respect to X.[13] Notably, although conditional skewness (2) incorporates a cubic function of $E(Y|X) = \mu_0 + \mu_X X$ scaled by $SD(Y|X) = exp(\pi_0 + \pi_X X)$, and thus might

---

[13] A simpler specification $SD(Y|X) = |\pi_0 + \pi_X X|$ has two major drawbacks. First, when $\pi_0 + \pi_X X = 0$, (2) explodes due to division by zero. This causes convergence problems in GMM estimation when $\pi_0 + \pi_X X$ gets near zero even for a single observation during the numerical search. Second, it is non-monotonic with respect to X (i.e., it first decreases and then increases with X, or vice versa). Our exponential specification avoids these problems.

seem to require a complex functional form, these non-linearities cancel out in (2).[14] Therefore, we

can use a linear functional form for the empirical specification of conditional skewness (3c).

Generalized Method of Moments (GMM) estimation

GMM is the most direct way to estimate our model. Equations (2) and (3a)–(3c) imply the

following conditional moment restrictions:

$$E[Y - \mu_0 - \mu_X X | X] = 0 \tag{4a}$$

$$E[(Y - \mu_0 - \mu_X X)^2 - exp(\pi_0 + \pi_X X)^2 | X] = 0 \tag{4b}$$

$$E\left[\left(\frac{Y - \mu_0 - \mu_X X}{exp(\pi_0 + \pi_X X)}\right)^3 - \beta_0 - \beta_X X | X\right] = 0 \tag{4c}$$

These conditional moment restrictions have the form $E(\xi | X) = 0$, where the random variable

ξ is the difference between the actual value in the data and the predicted value from equations

(3a)–(3c). For example, in (4a), ξ is the difference between the observed Y and the predicted

conditional mean $\mu_0 + \mu_X X$ from (3a). In (4b), ξ is the difference between the observed squared

error $\varepsilon^2 = (Y - \mu_0 - \mu_X X)^2$ and the predicted conditional variance $E(\varepsilon^2 | X) = SD(Y|X)^2 = exp(\pi_0 + \pi_X X)^2$.

We estimate (4a)–(4c) using the *gmm* command in Stata, as illustrated in Appendix A. This

command can handle any number of moment conditions with any functional forms, and the

standard errors can be clustered and/or bootstrapped as needed.

---

[14] Any random variable Y can be rewritten as $Y = E(Y|X) + SD(Y|X) \times \epsilon$, where ϵ is a normalized random variable with $E(\epsilon|X) = 0$ and $SD(\epsilon|X) = 1$. After this notation change, conditional skewness (2) simplifies to $Skew(Y|X) = E\left[\left(\frac{E(Y|X)+SD(Y|X)\times\epsilon-E(Y|X)}{SD(Y|X)}\right)^3 |X\right] = E(\epsilon^3|X)$. Thus, both $E(Y|X)$ and $SD(Y|X)$ cancel out, and their functional forms do not restrict the range of acceptable functional forms for $Skew(Y|X) = E(\epsilon^3|X)$.

A simpler three-stage estimation method

Because GMM uses numerical search, a researcher can encounter convergence problems in models with a large number of X variables.[15] As a more resilient (but less efficient) alternative, we break down the estimation into three stages that only use off-the-shelf regression tools.

Stage 1 estimates the conditional mean $E(Y|X)$ using an OLS regression

$$Y = \mu_0 + \mu_X X + \varepsilon \tag{5a}$$

Stage 2 estimates the conditional variance $E(\varepsilon^2|X) = SD(Y|X)^2$ using a non-linear regression with squared residual $\hat{\varepsilon}^2$ from stage 1 as the dependent variable

$$\hat{\varepsilon}^2 = exp(\pi_0 + \pi_X X)^2 + \nu \tag{5b}$$

We use the *nl* command in Stata for this stage.[16]

In stage 3, we define the dependent variable as

$$\tilde{Y}^{skew} \equiv \left(\frac{Y - \hat{E}(Y|X)}{\widehat{SD}(Y|X)}\right)^3 \tag{5c}$$

where $\hat{E}(Y|X) \equiv \hat{\mu}_0 + \hat{\mu}_X X$ is the predicted conditional mean from stage 1, and $\widehat{SD}(Y|X) \equiv exp(\hat{\pi}_0 + \hat{\pi}_X X)$ is the predicted conditional standard deviation from stage 2. We then estimate conditional skewness $Skew(Y|X)$ using an OLS regression

---

[15] For example, if there are 25 X variables, then GMM estimation of (3a)–(3c) involves numerical search in a 78-dimensional parameter space [=3×(25+1), where +1 is for the intercept]. We note that many convergence problems are unrelated to the size of the parameter space and are caused by sloppy scaling of the X variables. For example, if a variable X_1 varies in the range ±100,000 and X_2 varies in the range ±0.0003, then any numerical optimization algorithm will likely have convergence problems because a given step size $\Delta\beta$ for the coefficients affects $\beta_1 X_1$ and $\beta_2 X_2$ very differently. A researcher can avoid these problems by using consistent scaling for all X variables (e.g., normalize all variables to have zero mean and unit standard deviation).

[16] This stage is formulated for variance (rather than standard deviation) because ε^2 is an unbiased proxy for $Var(Y|X)$, but $\sqrt{\varepsilon^2}$ would be a biased proxy for $SD(Y|X)$. Stata has a standard command *hetregress* that estimates the variance of Y conditional on X. This command regresses $\log(\hat{\varepsilon}^2)$ on X to get a consistent estimate of $E(\log(\varepsilon^2)|X)$ and then converts it into an estimate of $E(\varepsilon^2|X)$ following Harvey (1976). However, this conversion is valid only for a Normally distributed ε, for which there is a simple algebraic relation between $E(\log(\varepsilon^2)|X)$ and $E(\varepsilon^2|X)$. For a general non-Normal distribution, this algebraic relation does not hold, and *hetregress* yields an inconsistent estimate of $E(\varepsilon^2|X)$. In contrast, our stage 2 is valid for any symmetric or asymmetric distribution with finite variance.

$$\tilde{Y}^{skew} = \beta_0 + \beta_X X + v_i \qquad (5d)$$

This regression produces an estimate $\hat{E}\left(\tilde{Y}^{skew}|X\right) = \hat{E}\left[\left(\frac{Y-\hat{E}(Y|X)}{\widehat{SD}(Y|X)}\right)^3 |X\right]$, which parallels the

theoretical definition of conditional skewness (2). We use bootstrapping (wrapped around all three

stages) to adjust the standard errors for the estimation noise from stages 1 and 2.[17] The coefficients

β_X capture the impact of X on the skewness of Y. They can be used in hypothesis tests and any

other standard post-estimation analysis, just like regular OLS regression coefficients. We provide

a Stata code example for implementing this method in Appendix B.

In principle, a researcher could add firm fixed effects to regressions (5a)–(5c). However, that

would break the consistency of the estimates due to the incidental parameter problem (Neyman

and Scott, 1948), just as the use of a short firm-level rolling window destroys the consistency of

all empirical analyses for standard firm-specific skewness measures.

2.4. *Our method for robust quantile-based skewness*

While Pearson's moment coefficient of skewness (1) is widely used, it has major flaws. First,

for many fat-tailed distributions (e.g., Student's t with 3 degrees of freedom), Pearson's skewness

does not exist because the true theoretical value of the cubic moment $E\left[\left(\frac{Y-E(Y)}{SD(Y)}\right)^3\right]$ explodes to

infinity (Jondeau and Rockinger, 2003).[18] Second, even when Pearson's skewness exists

---

[17] In untabulated simulations, clustered standard errors without this adjustment are biased considerably, as expected, while bootstrapping yields good standard errors. When researchers use multi-stage estimation (e.g., in any regression with an estimated *C-score*, *F-score*, or discretionary accrual), they typically report off-the-shelf standard errors. However, these standard errors for the last stage are wrong (in hundreds of published articles) because they ignore the estimation noise from the earlier stages. Chen, Hribar, and Melessa (2018) show that the popular two-stage estimation procedure that uses stage 1 residuals (e.g., discretionary accruals) as the dependent variable in stage 2 yields inconsistent estimates because the two stages use different vectors of explanatory variables, causing correlated omitted variable bias. This bias does not arise in our method because we use the same explanatory variables in all stages based on the logic of the statistical model.

[18] In a finite sample, a researcher can always get an estimate of Pearson's skewness. However, this estimate is meaningless when the underlying theoretical measure does not exist.

theoretically, its empirical estimation is very sensitive to outliers because of the cubic moment (Kim and White, 2004). These issues are caused by the basic structure of Pearson's skewness rather than any estimation method. Therefore, Kim and White (2004) recommend that researchers replace Pearson's skewness with an alternative quantile-based skewness measure of Hinkley (1975):

$$SkewQ(Y) = \frac{[Q_{top}(Y) - Q_{median}(Y)] - [Q_{median}(Y) - Q_{bottom}(Y)]}{Q_{top}(Y) - Q_{bottom}(Y)}$$
$$= \frac{Q_{top}(Y) + Q_{bottom}(Y) - 2Q_{median}(Y)}{Q_{top}(Y) - Q_{bottom}(Y)} \tag{6}$$

where $Q_{bottom}(Y)$ and $Q_{top}(Y)$ are the bottom and top quantiles of size α (i.e., percentiles α and $1-\alpha$, respectively) in the distribution of Y, and $Q_{median}(Y)$ is the median of Y. Unlike Pearson's skewness (1), quantile-based skewness (6) exists even for very fat-tailed distributions, and it can be estimated reliably even when there are large outliers (Kim and White, 2004).[19]

This skewness measure captures the asymmetry of the bottom and top quantiles relative to the median. For a symmetric distribution, the bottom and top quantiles are equidistant from the median (i.e., $Q_{median}(Y) - Q_{bottom}(Y)$ and $Q_{top}(Y) - Q_{median}(Y)$ in (6) are equal and fully cancel out), and skewness is zero (Panel A of Figure 4). For a left-skewed distribution (i.e., a longer left tail), the distance from the median to the top quantile is smaller than the distance from the median to the bottom quantile, and skewness is negative (Panel B). Conversely, for a right-skewed distribution, skewness is positive. Quantile-based skewness has a range of ± 1 (Pearson's moment coefficient of skewness is scaled differently and has a range of $\pm\infty$).

---

[19] For example, suppose that 99% of observations have moderate Y values of less than ± 1, 1% of observations have more extreme Y values of ± 10, and the mean of Y is zero. In Pearson's skewness (1), the cubic term $(Y - E(Y))^3$ is less than 1 in absolute value for the moderate observations (99% of the sample) versus $10^3 = 1,000$ in absolute value for the extreme observations (1% of the sample). Therefore, the computation is dominated 10 to 1 (1,000×1% vs 1×99%) by the extreme observations. In contrast, the computation of quantile-based skewness (6) for a typical choice of α (e.g., 0.25, 0.1, or 0.05) is unaffected by the extreme 1% of the sample.

The quantile size α is chosen by a researcher. For example, in the context of operating cash flow skewness attributed to non-linear production technology (e.g., diminishing marginal product of variable inputs), a researcher can use a moderate α such as 0.25 (i.e., Q25 and Q75) or 0.1 (Q10 and Q90). For accrual skewness caused by large and rare goodwill impairments, a lower α such as 0.05 is likely more appropriate. For severe stock market crashes and other rare events, the economic context can dictate an even lower α.[20] A researcher can also examine a range of economically plausible values for α.

Conditional on explanatory variables X, quantile-based skewness is

$$SkewQ(Y|X) = \frac{Q_{top}(Y|X) + Q_{bottom}(Y|X) - 2Q_{median}(Y|X)}{Q_{top}(Y|X) - Q_{bottom}(Y|X)} \tag{7}$$

where $Q_{bottom}(Y|X)$, $Q_{top}(Y|X)$, and $Q_{median}(Y|X)$ are the quantiles of the Y distribution conditional on X. We estimate conditional skewness in two stages.[21]

In stage 1, we estimate the bottom and top quantiles $Q_{bottom}(Y|X)$ and $Q_{top}(Y|X)$. We use a quantile regression of Y on X for percentiles α and $1 - \alpha$, respectively (Stata command *qreg* with option *quantile*)

$$Q_{bottom}(Y|X) = \gamma_0 + \gamma_X X \tag{8a}$$

$$Q_{top}(Y|X) = \lambda_0 + \lambda_X X \tag{8b}$$

where γ and λ are the quantile regression coefficients. This stage yields the predicted bottom and top quantiles $\hat{Q}_{bottom}(Y|X) \equiv \hat{\gamma}_0 + \hat{\gamma}_X X$ and $\hat{Q}_{top}(Y|X) \equiv \hat{\lambda}_0 + \hat{\lambda}_X X$ for each observation.

---

[20] For extreme "black swan" events (Taleb, 2007), the economic context might dictate an α very close to zero. However, the researcher's data are unlikely to contain enough information to draw statistical inferences about truly rare "black swan" events, regardless of the empirical method used.

[21] Although in theory the model components can be estimated jointly using GMM, there is a practical complication. Conditional quantile estimation minimizes absolute deviations, which are non-differentiable at zero. Therefore, standard gradient-based numerical optimization methods cannot be used. Quantile regression uses a custom linear programming algorithm that does not require differentiability. However, this algorithm cannot be combined with the general-purpose GMM estimation command in Stata.

In stage 2, we define the dependent variable as

$$\tilde{Y}^{skewQ} \equiv \frac{\hat{Q}_{top}(Y|X) + \hat{Q}_{bottom}(Y|X) - 2Y}{\hat{Q}_{top}(Y|X) - \hat{Q}_{bottom}(Y|X)} \tag{9a}$$

where the predicted values $\hat{Q}_{bottom}(Y|X)$ and $\hat{Q}_{top}(Y|X)$ for each observation are from stage 1, and the actual value of Y for each observation in (9a) provides the information for the estimation of the conditional median of Y as a function of the observation's X.

Then, we run a quantile regression of $\tilde{Y}^{skewQ}$ on X for the conditional median

$$Q_{median}(\tilde{Y}^{skewQ}|X) = \beta_0 + \beta_X X \tag{9b}$$

Because $\hat{Q}_{bottom}$ and $\hat{Q}_{top}$ in (9a) are known constants conditional on X, and the only remaining source of variation in $\tilde{Y}^{skewQ}$ is the conditional distribution of Y, this median regression replaces the realized Y in (9a) with an estimate of its conditional median $\hat{Q}_{median}(Y|X)$

$$\hat{Q}_{median}(\tilde{Y}^{skewQ}|X) = \frac{\hat{Q}_{top}(Y|X) + \hat{Q}_{bottom}(Y|X) - 2\hat{Q}_{median}(Y|X)}{\hat{Q}_{top}(Y|X) - \hat{Q}_{bottom}(Y|X)} \tag{9c}$$

This empirical expression parallels the theoretical conditional skewness (7). Therefore, the median regression for $\tilde{Y}^{skewQ}$ characterizes the skewness of Y conditional on X, and the coefficients β_X in (9b) directly reflect the impact of X on skewness.[22,23]

---

[22] Alternatively, a researcher could separately estimate the conditional median $\hat{Q}_{median}(Y|X)$ and then compute conditional skewness through equation (7) by combining the estimates $\hat{Q}_{median}(Y|X)$, $\hat{Q}_{bottom}(Y|X)$, and $\hat{Q}_{top}(Y|X)$. However, to compute the marginal effects of the X variables on skewness, the researcher would need to use the delta method or simulation to trace how each X variable flows through the different estimated components of equation (7). In contrast, by defining the dependent variable per equation (9a), we directly combine the effects that flow through the different components of equation (7), such that the second-stage coefficients are immediately interpretable as the marginal effect of X on conditional skewness.

[23] Although conditional skewness (7) is a non-linear function of the conditional quantiles, these non-linearities cancel out. Any random variable Y can be rewritten as $Y = v_0(X) + v_1(X) \times \epsilon$, where ϵ is a normalized random variable with $Q_{top}(\epsilon|X) = 1$ and $Q_{bottom}(\epsilon|X) = -1$, and $v_0(X)$, $v_1(X)$ are deterministic functions. After this notation change, conditional skewness (7) simplifies to $SkewQ(Y|X) = \frac{[v_0(X)+v_1(X)]+[v_0(X)-v_1(X)]-2[v_0(X)+v_1(X)Q_{median}(\epsilon|X)]}{[v_0(X)+v_1(X)]-[v_0(X)-v_1(X)]} = -Q_{median}(\epsilon|X)$. Thus, the functional forms of the conditional quantiles $Q_{top}(Y|X)$ and $Q_{bottom}(Y|X)$ in stage 1 do not restrict the range of acceptable functional forms for $SkewQ(Y|X) = -Q_{median}(\epsilon|X)$ in stage 2. Therefore, we can use a linear functional form for $SkewQ(Y|X)$ in (9b).

We use bootstrapping (wrapped around both estimation stages) to adjust the second-stage standard errors for the first-stage estimation noise. We provide a code example in Appendix C.

2.5. *Examples of extensions of our method for additional asymmetry metrics*

Skewness based on the difference between the mean and the median conditional on X

If analysts forecast the conditional median of EPS rather than the conditional mean, then their forecasts have a systematic bias on average that varies with EPS skewness (Gu and Wu, 2003; Basu and Markov, 2004). In this research context, it is more natural to use an alternative skewness definition that is based on the difference between the conditional mean and the conditional median, known as non-parametric skew or Pearson's median skewness:

$$SkewM(Y|X) = \frac{E(Y|X) - Q_{median}(Y|X)}{SD(Y|X)} \tag{10}$$

A straightforward modification of our multi-stage approach from Sections 2.3–2.4 can estimate $SkewM(Y|X)$ as a function of X. Stages 1 and 2 estimate the conditional mean $E(Y|X)$ and standard deviation $SD(Y|X)$ using regressions (5a) and (5b), respectively. Stage 3 defines the dependent variable as

$$\tilde{Y}^{skewM} \equiv \frac{\hat{E}(Y|X) - Y}{\widehat{SD}(Y|X)} \tag{11}$$

where $\hat{E}(Y|X)$ and $\widehat{SD}(Y|X)$ are the predicted values from stages 1 and 2. After that, a median regression of $\tilde{Y}^{skewM}$ on X produces an estimate $\hat{Q}_{median}(\tilde{Y}^{skewM}|X) = \frac{\hat{E}(Y|X) - \hat{Q}_{median}(Y|X)}{\widehat{SD}(Y|X)}$, which parallels the theoretical definition (10). The standard errors can be obtained through bootstrapping (wrapped around all three stages).

Asymmetric accrual variance measure of conservatism conditional on X

Dutta and Patatoukas (2017) revive an alternative conservatism measure that is based on the difference in accrual variance between good-news and bad-news observations (Basu, 1995):

$$AVar\ conservatism(Y) = Var(Y|r < 0) - Var(Y|r \geq 0) \qquad (12a)$$

where Y is scaled accrual and r is unexpected stock return (i.e., a news proxy). Although this measure is distinct from skewness, it is amenable to similar analysis. Dutta and Patatoukas measure the average level of conservatism (12a) in their sample, and they do not provide a way to incorporate explanatory variables. However, researchers typically study variation in conservatism, rather than just its average level. Further, a standard rolling-window approach to studying variation in (12a) would be susceptible to the biases from Figures 1–3. To provide a more viable solution, we extend (12a) to incorporate explanatory variables X

$$AVar\ conservatism(Y|X) = Var(Y|r < 0, X) - Var(Y|r \geq 0, X) \qquad (12b)$$

To implement (12b) empirically, we specify the functional forms

$$E(Y|r \geq 0, X) = \mu_0^+ + \mu_X^+ X \qquad (13a)$$

$$E(Y|r < 0, X) = \mu_0^- + \mu_X^- X \qquad (13b)$$

$$Var(Y|r \geq 0, X) = exp(\pi_0 + \pi_X X)^2 \qquad (13c)$$

$$Var(Y|r < 0, X) = \beta_0 + \beta_X X + exp(\pi_0 + \pi_X X)^2 \qquad (13d)$$

where the conditional means (13a)–(13b) are used in the computation of conditional variances, the exponential functional form (13c) ensures a non-negative variance, and $\beta_0 + \beta_X X$ in (13d) directly measures $AVar\ conservatism(Y|X) = Var(Y|r < 0, X) - Var(Y|r \geq 0, X)$.[24]

---

[24] For consistency with Basu (1995) and Dutta and Patatoukas (2017), we formulate the model in terms of variances rather than standard deviations. If conservatism has a multiplicative effect on accrual variance for bad news (and accrual variance for both good and bad news also varies with X through operating volatility), then it might be better to define (13d) as $Var(Y|r < 0, X) = exp(\beta_0 + \beta_X X) \times exp(\pi_0 + \pi_0 X)^2$, such that $\beta_0 + \beta_X X$ captures the percentage difference between accrual variances for bad and good news.

Like conditional Pearson's moment coefficient of skewness in Section 2.3, this model can be estimated using either GMM or multi-stage estimation with bootstrapped standard errors. The conditional moment restrictions for equations (13a)–(13d) are

$$E\big((Y - \mu_0^+ - \mu_X^+ X) \times I_{good}|X\big) = 0 \tag{14a}$$

$$E\big((Y - \mu_0^- - \mu_X^- X) \times I_{bad}|X\big) = 0 \tag{14b}$$

$$E\big([(Y - \mu_0^+ - \mu_X^+ X)^2 - exp(\pi_0 + \pi_X X)^2] \times I_{good}|X\big) = 0 \tag{14c}$$

$$E\big([(Y - \mu_0^- - \mu_X^- X)^2 - exp(\pi_0 + \pi_X X)^2 - \beta_0 - \beta_X X] \times I_{bad}|X\big) = 0 \tag{14d}$$

where the good-news dummy $I_{good} \equiv I\{r \geq 0\}$ and the bad-news dummy $I_{bad} \equiv I\{r < 0\}$ restrict each moment condition to the relevant subsample. These GMM moment conditions can be implemented by slightly adapting the Stata code example from Appendix A.

To implement multi-stage estimation, stage 1 regresses Y on X separately for good-news and bad-news subsamples to obtain $\hat{E}(Y|r \geq 0, X)$ and $\hat{E}(Y|r < 0, X)$, respectively. Stage 2 is a non-linear regression (*nl* in Stata) of the squared residuals $\hat{\varepsilon}^2$ from stage 1 on $exp(\pi_0 + \pi_X X)^2$ for the good-news subsample. Stage 3 defines the dependent variable for the bad-news subsample as $\hat{\varepsilon}^2 - exp(\hat{\pi}_0 + \hat{\pi}_X X)^2$ (i.e., excess accrual variance for bad news relative to the variance predicted by the coefficients $\hat{\pi}$ from stage 2) and regresses it on X to obtain an estimate of conservatism $\beta_0 + \beta_X X$ as a function of X. Bootstrapping provides valid standard errors.

2.6. *Using predicted conditional skewness as an explanatory variable*

Our methods are primarily designed to study the determinants of conditional skewness. However, predicted conditional skewness from our methods can also be used as an explanatory variable, similar to how Altman's (1968) *Z-score*, Khan and Watts' (2009) *C-score*, or Dechow et al.'s (2011) *F-score* are often used as explanatory variables. We provide an implementation example in Appendix D. However, the effect of skewness in this analysis is identified reliably *only*

if a researcher has a valid instrument for skewness (i.e., a variable *Z* that affects conditional skewness, affects the outcome of interest only through conditional skewness, and is uncorrelated with unobservables for the outcome). This major caveat applies to all measures that are based on a first-stage predictive regression, including *Z-score*, *C-score*, and *F-score*.

3. Simulation evidence on test performance of our methods versus the standard method

3.1. Test performance when there is time series variation in X, which causes the bias from Figure 1 in the standard method

Simulation protocol

In each scenario, we generate 250 simulated samples. Each sample is a panel of 1,000 firms with 10 annual observations per firm. For each firm-year, we first draw an explanatory variable

$$X = \sqrt{1 - \rho} \times \chi_{firm} + \sqrt{\rho} \times \chi_{firm-year} \qquad (15a)$$

where χ_{firm} and $\chi_{firm-year}$ are firm-level and firm-year-level random variables, respectively, drawn from independent *Uniform*[-1,1] distributions. The mean of *X* is zero by construction. The parameter ρ determines the proportion of time-series variation in *X*.[25] We then generate the dependent variable

$$Y = \mu_0 + \mu_X X + exp(\pi_0 + \pi_X X) \times \varepsilon \qquad (15b)$$

where the random shock ε is the source of variation (and skewness) in *Y* when *X* is held constant. ε is independent across firm-years and has zero mean and unit standard deviation, such that the

---

[25] From (15a), $Var(X) = (1 - \rho)Var(\chi_{firm}) + \rho Var(\chi_{firm-year})$, where $Var(\chi_{firm}) = Var(\chi_{firm-year})$. Thus, the time-series variation $\sqrt{\rho} \times \chi_{firm-year}$ in (15a) accounts for a fraction ρ of the total variance of *X*.

conditional mean of Y is $\mu_0 + \mu_X X$ and the conditional standard deviation is $exp(\pi_0 + \pi_X X)$. The parameters μ and π in (15b) do not affect the skewness of Y conditional on X.[26]

To model conditional skewness, we draw ε from a normalized non-central Student's t distribution (Harvey and Siddique 1999) with a firm-year-specific non-centrality parameter

$$\delta = \delta_0 + \delta_X X \tag{15c}$$

When the non-centrality parameter δ is zero, ε follows a symmetric Student's t distribution with zero skewness. When δ is positive, the distribution is right-skewed (i.e., the right tail is larger than the left tail) and skewness is positive, and vice versa.[27] The coefficients δ_0 and δ_X for the non-central Student's t distribution in (15c) are scaled differently from the skewness parameters β_0 and β_X in (5d) and (9b) in estimation, but they have consistent predicted signs.

After generating each simulated sample, we estimate the conditional skewness of Y as a function of X using the following methods: (1) the standard method from Section 2.2, which computes firm-level skewness based on a window of 10 observations,[28] (2) our method for Pearson's skewness from Section 2.3, with either GMM or three-stage estimation, (3) our method for quantile-based skewness from Section 2.4, with quantile size α set to 0.25, 0.1, and 0.05. We then test whether conditional skewness differs from zero on average, and whether conditional skewness varies with X, using two-tailed tests with a 5% nominal significance level.

---

[26] Because skewness is scale-independent, it is fully determined by the shape of the ε distribution. Any viable estimation method should reliably distinguish variation in conditional mean and standard deviation through μ and π from variation in conditional skewness. We use an exponential specification for the standard deviation $exp(\pi_0 + \pi_X X)$ of the simulated Y for consistency with the estimation model (3a)–(3c); the results are robust to using a linear specification of the standard deviation in untabulated sensitivity checks.

[27] Non-central Student's t distribution is typically defined in unnormalized form, in which changes in δ affect not only the skewness but also the mean and the standard deviation. We prefer all sources of variation to be clearly visible in the model. Therefore, we normalize ε to have zero mean and unit standard deviation for each firm-year, such that δ only affects conditional skewness, and we directly model variation in the mean and standard deviation through the parameters μ and π in (15b).

[28] By using constant window length in the standard method, we remove the confounding effect from Figure 3. Our methods from Sections 2.3–2.5 do not use a window and therefore cannot be confounded by window length, as we confirm in simulations in Section 3.3.

Type-I error (Table 1)

We set the non-centrality parameter δ to zero for all firm-years, such that Y conditional on X follows a symmetric Student's t distribution with zero conditional skewness (i.e., true $\beta_0 = \beta_X = 0$ in the empirical skewness equation). Because the null hypotheses of zero conditional skewness on average and zero variation in conditional skewness are true, the rejection rates in Table 1 represent type-I error. They should be near 5%.

Panel A of Table 1 presents our main type-I error simulations. In column 1, X only varies across firms ($\rho = 0$). Because there is no time-series variation in X in this special case, the bias from Figure 1 for the standard method does not arise. As expected, the standard method and all our methods have acceptable type-I errors near 5%.

In columns 2–4, we add time-series variation in X. Consistent with Figure 1, the standard method breaks down. Even when the time-series variation is just 25% of the total ($\rho = 0.25$ in column 2), the standard method falsely detects significant conditional skewness on average in 100.0% of the simulated samples and falsely detects a significant effect of X on skewness in 79.2% of the simulated samples. The scenarios with 50% and 75% time-series variation in columns 3 and 4 yield similarly excessive type-I errors in the standard method.

In contrast, our methods continue to perform well, as expected, with type-I errors near 5% in all columns.[29] Because we explicitly model the conditional mean, conditional standard deviation, and conditional skewness of Y for each firm-year observation as a function of X, by construction our methods can accommodate any variation in X, yielding good type-I errors.

---

[29] The three-stage version of our method for Pearson's skewness tends to have abnormally low type-I errors in the test of $\beta_X = 0$. The reason is that the estimates for Pearson's skewness are often dominated by relatively few extreme observations, and therefore the standard asymptotic Normal approximation is relatively inaccurate in a sample of 10,000 firm-years.

To directly show why it is important to fully control for X, in additional tests we remove the controls for time-series variation in X in our methods by using a firm-specific mean of X (rather than the actual X for the firm-year) as an explanatory variable in estimation. We retain all other features of our methods. As expected, this modification breaks type-I errors, yielding rejection rates of 77.2–100.0% in the test of conditional skewness on average and 20.8–83.2% in the test of variation in conditional skewness (untabulated). Thus, the good type-I error performance of our methods in Table 1 is directly related to their ability to control for time-series variation in X.

Panel B presents sensitivity checks for the simulation with 50% time-series variation in X. As expected, the standard method continues to generate false findings of significant conditional skewness on average and significant variation in conditional skewness when we use (1) a shorter window of 5 years per firm, (2) a longer window of 20 years per firm, (3) a less fat-tailed distribution with df=10, and (4) a more fat-tailed distribution with df=3. In contrast, our methods perform well in all of these sensitivity checks.[30]

Test power (Table 2)

In Table 2, we set the non-centrality parameter for each firm-year to $\delta = 1 + 0.5 \times X$, where X has mean zero. Thus, Y conditional on X follows a right-skewed non-central Student's t distribution whose skewness increases with X. The rejection rates reflect test power.

---

[30] For the fat-tailed distribution with df=3 in column 4, Pearson's skewness is theoretically undefined (Jondeau and Rockinger, 2003), which invalidates all inferences based on this skewness measure (both in the standard method and in our method). Therefore, even though our method for Pearson's skewness appears to have reasonable type-I errors in column 4, it should not be used in this scenario (many of the coefficient estimates for this scenario are in the billions, which confirms that the method operates in abnormal conditions). As expected, our quantile-based method remains valid for df=3, yielding good type-I errors and reasonable point estimates. In additional tests for the standard method, we regress firm-specific skewness on the firm-specific mean of X, which is better aligned with the timing of the dependent variable than the original firm-year-level X. As expected, this modification does not improve type-I errors (untabulated).

Panel A of Table 2 presents our main test power simulations. In column 1, X does not vary over time ($\rho = 0$), and therefore the bias from Figure 1 for the standard method does not arise. As expected, the standard method performs well in this special case, detecting significant conditional skewness on average in 100% of the simulated samples and a significant effect of X on conditional skewness in 93.2% of the simulated samples.

However, even a small amount of time-series variation in X disrupts the standard method, as predicted by Figure 1. When time-series variation in X increases to just 25% ($\rho = 0.25$ in column 2), the rejection rate for the effect of X on conditional skewness in the standard method decreases from 93.2% to 29.2%. The rejection rate deteriorates further to just 20.4% and 8.4% for the scenarios with 50% and 75% time-series variation in X in columns 3 and 4, respectively.[31] All of these rejection rates are lower than the parallel type-I errors in Table 1, which indicates that the standard method lacks any statistical power in these scenarios.

In contrast, our methods are largely unaffected by time-series variation in X. In our method for Pearson's skewness, GMM estimation detects a significant effect of X on conditional skewness in 39.2–45.6% of the simulated samples across all four columns. Three-stage estimation is less efficient, detecting a significant effect of X in 23.2–30.8% of the simulated samples, but it performs consistently across the different scenarios. Similarly, our method for quantile-based skewness exhibits consistent performance across all four columns.

The test power of our quantile-based method increases considerably for smaller values of α (i.e., quantiles that reach further into the tails of the distribution). For example, for $\alpha=0.25$ (i.e., skewness based on the asymmetry of Q25 and Q75 relative to the median), the test detects a

---

[31] In the test of whether conditional skewness exists on average (the top half of the table), the standard method has a rejection rate of 100%. However, this does not indicate high test power. For example, in the parallel scenarios in Table 1, where true conditional skewness is zero, the standard method has an identical rejection rate of 100.0%.

significant effect of X on skewness in 18.8–26.8% of the simulated samples. For $\alpha=0.1$ (i.e., Q10 and Q90 relative to the median), test power improves to 70.4–77.2%. For $\alpha=0.05$ (i.e., Q5 and Q95 relative to the median), it improves further to 88.8–92.8%.

Figure 5 shows why. We plot how the quantiles of a normalized non-central Student's t distribution shift when the non-centrality parameter δ increases from 0 (the solid lines) to 1 (the dashed lines). For ease of visual comparison, the medians of both distributions are normalized to zero. When the non-centrality parameter δ increases, there is only a slight shift in Q25 and Q75, a larger shift in Q10 and Q90, and a much larger shift in Q5 and Q95, as indicated by the arrows underneath the X axis. Thus, there is only a slight change in measured skewness for quantile size $\alpha=0.25$, which is based on Q25 and Q75, a larger change in skewness for $\alpha=0.1$ (Q10 and Q90), and a much larger change in skewness for $\alpha=0.05$ (Q5 and Q95). These larger changes in measured skewness for small α improve test power. However, if α is pushed too close to zero (e.g., $\alpha=0.01$ in untabulated tests), then this larger effect size is offset by insufficient data density around the quantiles α and $1-\alpha$, which reduces test power.

Notably, our method yields considerably higher test power when it is combined with quantile-based skewness than when it is combined with Pearson's moment coefficient of skewness. For example, our quantile-based tests with $\alpha=0.1$ and $\alpha=0.05$ detect a significant effect of X on skewness in 70.4–92.8% of the simulated samples, versus 23.2–45.6% for the parallel Pearson-based tests. The reason is that quantile-based skewness is less sensitive to extreme values of Y than Pearson's skewness, and therefore the quantile-based estimates incorporate less noise.

Panel B of Table 2 presents sensitivity checks for the simulation with 50% time-series variation in X. Although the standard method detect a significant effect of X on conditional skewness in 28.8–52.0% of the simulated samples in columns 2 and 3, these rejection rates are lower than the

type-I errors in the parallel scenarios in Table 1, i.e., the standard method lacks any statistical power.[32] In contrast, our methods continue to have good test power across the different scenarios.

3.2. Test performance when the skewness of Y is driven by firm effects, which cause the bias from Figure 2 in the standard method

We generate 250 simulated samples of 1,000 firms × 10 years. For each firm, we first draw an explanatory variable $X = \chi_{firm}$ from a $Uniform[-1,1]$ distribution. X varies across firms but not over time, such that the bias from Figure 1 does not arise. We then generate the dependent variable $Y = \mu_0 + \mu_X X + exp(\pi_0 + \pi_X X) \times \varepsilon$, where $\varepsilon = \varepsilon_{firm} + \varepsilon_{firm-year}$. The firm effect ε_{firm} is drawn from a skewed non-central Student's t distribution with non-centrality parameter δ, and the firm-year shock $\varepsilon_{firm-year}$ is drawn from a symmetric Normal distribution. Thus, the Y distribution conditional on X is skewed, and the skewness is concentrated in the firm effect ε_{firm}. For each sample, we estimate skewness as a function of X and conduct hypothesis tests.

Table 3 presents the simulation results. In columns 1–2, skewness conditional on X is constant (δ=0.5 and δ=3, respectively), and in columns 3–4, it increases with X (δ=1+0.5×X and δ=1+X, respectively). Consistent with our argument in Figure 2, the standard method cannot detect the existence of conditional skewness (the rejection rate in the test for β_0 in all columns is just 3.2%), and it cannot detect the effect of X on conditional skewness (the rejection rate in the test for β_X in columns 3–4 is just 4.0%).[33] In contrast, our methods successfully detect both the existence of skewness in all columns and the effect of X on skewness in columns 3–4.

---

[32] In the fat-tailed scenario with df=3 in column 4, only quantile-based inferences are valid, as we explain for the parallel analysis in Table 1.

[33] To confirm that this bad test performance arises specifically because skewness is embedded in the firm effects, in additional simulations we (1) replace the firm-level ε_{firm} with a firm-year-level shock drawn from same non-central Student's t distribution, or (2) swap the distributions of ε_{firm} and $\varepsilon_{firm-year}$. In both cases, the true skewness of Y conditional on X is identical to the original simulation, and the only change is in the persistence of the time-series variation in Y. As expected, the standard method performs much better in these modified simulations (untabulated).

3.3. Test performance when there is variation in the number of observations across rolling windows, which causes the bias from Figure 3 in the standard method

We generate 250 simulated samples. Each sample comprises 10,000 firm-year observations with variation in the time series length per firm. Each firm exits the sample with probability 0.1 at the end of each period, such that the number of observations per firm is a random variable with mean 10 [$=1/0.1$] and variance 90 [$=(1-0.1)/0.1^2$]. To directly model the bias from Figure 3, we use a firm's total number of observations N_{firm} as the explanatory variable. We draw the dependent variable Y from a non-central Student's t distribution with a constant non-centrality parameter δ. Thus, by construction, the conditional skewness of Y does not vary with N_{firm}. Because N_{firm} does not incorporate any time-series variation and the draws of Y are independent over time, the results are not confounded by the additional biases from Figures 1 and 2, respectively.

After generating each simulated sample, we estimate the conditional skewness of Y as a function of N_{firm} using the methods from Section 2. In the standard method, we compute firm-specific skewness using all available observations for the firm, such that the number of observations in the computation window varies across firms. In each method, we then test whether N_{firm} affects measured conditional skewness. Because true conditional skewness is unaffected by N_{firm}, a valid test should yield a rejection rate near the nominal significance level of 5%.[34]

Table 4 presents type-I errors. Even for a moderately skewed distribution with $\delta=0.5$ in column 1, the standard method falsely detects a significant effect of N_{firm} on conditional skewness in 53.6% of the simulated samples. The degrees-of-freedom adjustment reduces this type-I error to 27.6%,

---

[34] Unlike the bias from Figure 1, the bias from Figure 3 arises only for a skewed distribution. We do not conduct type-I error analysis for the test of whether skewness exists on average, because it is meaningful only when true skewness is zero. We do not conduct test power simulations because they do not align with the logic of the bias from Figure 3. The *df*-adjusted version of the standard method is reported in Table 4 but not in Tables 1–3, because only the former has variation in the number of observations across firms.

i.e., it partly mitigates but does not eliminate the bias. We then restrict the estimation sample to firms with at least 5 valid observations, thus removing the most biased skewness measures. The unadjusted and adjusted type-I errors improve slightly but remain excessive (40.0% and 25.2%, respectively). For the more skewed distribution with $\delta=3$ in column 2, type-I errors in the standard method (both unadjusted and adjusted) are 98.0–100%. Thus, if a researcher allows variation in the number of observations per window (e.g., by allowing a few missing observations in a rolling window), then she risks false findings due to spurious variation in conditional skewness. This spurious variation can confound inferences about any explanatory variable that is correlated with the number of observations N_{firm} (e.g., any correlate of a younger firm, whose time series starts later, or a good acquisition target, whose time series as an independent firm is likely to end earlier).

In contrast, our methods yield acceptable type-I errors near 5%, as expected. Because we directly model conditional skewness for each firm-year instead of using a window, by construction our methods are immune from the bias from Figure 3.

3.4. Summary

Our simulations confirm that the biases in the standard rolling-window method can have a major impact on inferences. First, when X varies over time, the standard method is likely to falsely detect conditional skewness on average and systematic variation in conditional skewness, because the rolling window is too inflexible and cannot control for time-series variation in X (Figure 1). Second, when Y contains skewed firm effects (or other skewed persistent variation), the standard method is likely to miss this source of conditional skewness, because the rolling window is too flexible and overcontrols for persistent variation in Y that is unrelated to X (Figure 2). Third, when the number of observations varies across rolling windows (e.g., because a researcher allows a few missing observations in a window), the standard method is likely to falsely detect systematic

variation in conditional skewness, because the small-sample bias in the rolling-window skewness computation varies with the window size (Figure 3).

Our methods directly resolve all three biases, yielding good type-I errors. They also enable a researcher to replace Pearson's moment coefficient of skewness with the more robust quantile-based skewness, which improves test power considerably.

4. Application: The relation between managerial overconfidence and conservatism

Ahmed and Duellman (2013) use the difference between operating cash flow skewness and earnings skewness (i.e., *Skew*(*CFO*)−*Skew*(*Earn*)), computed on a five-year rolling window, as a conditional conservatism proxy. They predict and find a significant negative relation between managerial overconfidence and this proxy. However, their rolling-window skewness computation can distort inferences by adding spurious variation (Figure 1) and by removing relevant variation (Figure 2). Therefore, we re-examine these findings using our methods.

Our sample and variable definitions follow Ahmed and Duellman (2013). We use the intersection of Execucomp, Compustat, and CRSP data during 1993–2009, and exclude financial firms (SIC codes 6000–6999), firm-years with missing data, and firm-years with CEO turnover. The final sample comprises 17,467 firm-year observations.[35]

We use three managerial overconfidence indicators from Ahmed and Duellman (2013): *Holder67* reflects delayed exercise of deep-in-the-money options by the CEO, *CAPEX* reflects excess capital expenditures relative to the industry median, and *OverInvest* reflects excess asset growth relative to sales growth. All variables are defined in the notes to Table 5.[36]

---

[35] Untabulated descriptive statistics in our sample are consistent with Table 1 in Ahmed and Duellman (2013). Our sample is approximately 20% larger, which could be due to retroactive inclusion of additional firms in standard datasets or unreported additional screening criteria in Ahmed and Duellman (2013).

[36] We do not use Ahmed and Duellman's fourth indicator, *Purchase*, because it requires additional data from Thomson Reuters that we do not have access to.

Columns 1–3 of Panel A in Table 5 replicate Ahmed and Duellman's (2013) analysis, in which rolling-window $Skew(CFO)–Skew(Earn)$ is regressed on the overconfidence indicators and control variables. As expected, we find a significant negative relation between all three overconfidence indicators and $Skew(CFO)–Skew(Earn)$. Because conditional conservatism manifests most directly in accruals, in columns 4–6 we use rolling-window accrual skewness $Skew(Accr)$ as an alternative proxy.[37] The results continue to hold. Thus, when we use the standard method, we successfully reproduce Ahmed and Duellman's (2013) results.

Next, we re-run the analysis using our methods. Panel B models conditional accrual skewness $Skew(Accr|X)$, where the X vector comprises an overconfidence indicator plus all of the controls. Columns 1–3 present GMM estimates for conditional Pearson's moment coefficient of skewness, implemented using the code from Appendix A (the inferences are robust to using three-stage estimation from Appendix B in untabulated tests). For all three overconfidence indicators, the relation with conditional accrual skewness is insignificant even at the 10% level, contrary to Ahmed and Duellman's (2013) results. Additionally, all three indicators are significantly associated with the conditional mean and/or standard deviation of accruals, which can confound the standard method as shown in Figure 1.

Columns 4–6 of Panel B present two-stage estimates for quantile-based conditional accrual skewness with α=0.05 (i.e., asymmetry of Q5 and Q95 relative to Q50 conditional on X). For two overconfidence indicators out of three, *Holder67* and *CAPEX*, the relation with this proxy is insignificant at the 10% level (these non-results are robust to using α=0.25 and α=0.1 in untabulated tests). Only one overconfidence indicator, *OverInvest*, yields significant results.

---

[37] Greater conservatism corresponds to a more positive $Skew(CFO)–Skew(Earn)$ and a more negative $Skew(Accr)$. Therefore, we flip the sign of all accrual-based proxies in Panels A and B, such that the coefficient signs have consistent interpretation across all estimates in Table 5.

Panel C models the difference in conditional skewness between operating cash flows and earnings, i.e., *Skew(CFO|X)–Skew(Earn|X)*.[38] We estimate $Skew(CFO|X) = \beta_0^{CFO} + \beta_X^{CFO}X$ and $Skew(Earn|X) = \beta_0^{EARN} + \beta_X^{Earn}X$ and report the coefficients for the difference $(\beta_0^{CFO} - \beta_0^{EARN}) + (\beta_X^{CFO} - \beta_X^{Earn})X$. We slightly adapt our methods to compute the standard errors for this difference as described in the notes to Table 5. Similar to Panel B, all three overconfidence indicators are insignificant at the 10% level in the analysis for Pearson's moment coefficient of skewness in columns 1–3.[39] Likewise, two out of three indicators are insignificant in the analysis for quantile-based skewness in columns 4–6.

In summary, use of our methods leads to qualitatively different inferences about the relation between managerial overconfidence indicators and skewness-based conservatism proxies.

5. Conclusion

Researchers in many areas analyze skewness determinants. We show that the standard window-based approximation for conditional skewness in the literature is likely to cause false findings of conditional skewness on average and false findings of systematic variation in conditional skewness, even when the true conditional skewness is zero. These biases arise because skewness

---

[38] This differential skewness proxy is the conditional counterpart of Ahmed and Duellman's (2013) proxy from Panel A. A potential drawback of differential skewness is that it uses mutually inconsistent scaling for *CFO* and *Earn* (i.e., each variable is scaled by its respective standard deviation in the computation of skewness). Our accrual skewness proxies in Panels A and B avoid this drawback. We exclude CFO from the list of control variables in Panel C because *Skew(CFO|CFO)* is undefined (i.e., the distribution of CFO conditional on itself collapses to a single point, and a single point does not have skewness).

[39] The biases from Figures 1 and 2 can confound the standard rolling-window estimates in Panel A but cannot confound our estimates in Panels B and C. To directly show that these two biases change the inferences, we artificially inject them in our method. We replace the conditional mean and standard deviation equations (4a)-(4b) in our GMM estimation with the moment-condition equivalent of the standard rolling-window computation, while keeping all other features of our method. This modification removes the controls for within-window variation in *X* and adds excessive controls for persistent variation in *Y*, thus activating the biases from Figures 1 and 2 (we cannot activate each bias separately in estimation but can measure their combined impact). As expected, when we add these biases to our method, we find a significant negative relation between all three overconfidence indicators and differential skewness (untabulated). The bias from Figure 3 is irrelevant in this application because the rolling window has the same length (5 valid observations) for all firm-years.

"conditional on the window" is an invalid proxy for skewness "conditional on X" for the observations in the window.

To resolve these biases, we develop new methods that directly model the conditional skewness of a Y variable as a function of X variables. Simulations confirm that our methods successfully resolve the biases without sacrificing test power. Our methods are transparent, robust, and easy to use. A researcher can implement them in a few lines of code following the code examples that we provide in Appendices A–D.

Our methods change the interpretation of a major prior finding in the accounting literature. Ahmed and Duellman (2013) report a significant negative relation between indicators of managerial overconfidence and skewness-based conservatism proxies. However, these results for standard rolling-window skewness measures are likely confounded by the biases from Figures 1 and 2. When we remove these confounding effects using our methods, the relation between overconfidence indicators and conditional skewness becomes insignificant in almost all cases.

While we emphasize accounting applications such as accrual skewness as a proxy for conservatism, skewed outcomes are pervasive across domains. For example, labor income skewness in individual-level data can be used as proxy for income inequality. Researchers in labor economics can apply our methods to income skewness to study the determinants of labor income inequality at the individual level (in contrast, traditional Gini-type inequality measures can only be analyzed across large subsamples). Researchers in health economics can use our methods to study the skewness of health care spending across patients. Researchers in strategy can use our methods to study winner-take-all dynamics that manifest in skewed firm-level outcomes. Our methods could benefit researchers in these and many other areas.

References

Ahmed, A., Duellman, S., 2013. Managerial overconfidence and accounting conservatism. Journal of Accounting Research 51(1), 1–30.

Alcacer, J., Chung, W., Hawk, A., Pacheco-de-Almeida, G., 2018. Applying random coefficient models to strategy research: Identifying and exploring firm heterogeneous effects. Strategy Science 3(3), 481–553.

Altman, E., 1968. Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. Journal of Finance 23(4), 189–209.

Bartl, M., 2018. YouTube channels, uploads, and views: A statistical analysis of the past 10 years. Convergence: The International Journal of Research into New Media Technologies 24(1), 16–32.

Basu, S., 1995. Conservatism and the Asymmetric Timeliness of Earnings. Ph.D. thesis, University of Rochester.

Basu, S., Markov, S., 2004. Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts. Journal of Accounting and Economics 38, 171–203

Beatty, A., Weber, J., Yu, J., 2008. Conservatism and debt. Journal of Accounting and Economics 45(2-3), 154–174

Braga, A., Papachristos, V., Hureau, D., 2010. The concentration and stability of gun violence at micro places in Boston. Journal of Quantitative Criminology 26(1), 33–53.

Bulger, J., 1993. Dominance rank and access to estrous females in male savanna baboons. Behaviour 127(1/2), 67–103.

Burgstahler, D., Dichev, I., 1997. Earnings, adaptation, and equity value. The Accounting Review 72(2), 185–215.

Chen, J., Hong, H., Stein, J., 2001. Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. Journal of Financial Economics 61, 345–381.

Chen, W., Hribar, P., Melessa, S., 2018. Incorrect inferences when using residuals as dependent variables. Journal of Accounting Research 56(3), 751–796.

Cox, N., 2010. Speaking Stata: The limits of sample skewness and kurtosis. The Stata Journal 10(3), 482–495.

Dechow, P., Ge, W., Larson, C., Sloan, R., 2011. Predicting material accounting misstatements. Contemporary Accounting Research 28(1), 17–82.

Dutta, S., Patatoukas, P., 2017. Identifying conditional conservatism in financial accounting data: Theory and evidence. The Accounting Review 92(4), 191–216.

Engle, R., 2011. Long-term skewness and systemic risk. Journal of Financial Econometrics 9(3), 437–468.

Estrada, J., 2009. Black swans, market timing, and the Dow. Applied Economics Letters 16, 1117–1121.

Givoly, D., Hayn, C., 2000. The changing time-series properties of earnings, cash flows, and accruals: Has financial reporting become more conservative? Journal of Accounting and Economics 29(3), 287–320

Gu, Z., Wu., J., 2003. Earnings skewness and analyst forecast bias. Journal of Accounting and Economics 35(1), 5–29.

Hamermesh, D., 2018. Citations in economics: Measurement, uses, and impacts. Journal of Economic Literature 56(1), 115–156.

Harvey, A., 1976. Estimating regression models with multiplicative heteroskedasticity. Econometrica 44(3), 461–465.

Harvey, C., Siddique, A., 1999. Autoregressive conditional skewness. The Journal of Financial and Quantitative Analysis, 34(4), 465–487.

Hinkley, D., 1975. On power transformations to symmetry. Biometrica 62, 101–111.

Jondeau, E., Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. Journal of Economic Dynamics and Control 27, 1699–1737.

Khan, M., Watts, R., 2009. Estimation and empirical properties of a firm-year measure of accounting conservatism. Journal of Accounting and Economics 48(2-3), 132–150.

Kim, T., White, H., 2004. On more robust estimation of skewness and kurtosis. Finance Research Letters 1, 56–73.

Killingsworth, M., 1993. Analyzing employment discrimination: From the seminar room to the courtroom. American Economic Review 83(2), 67–72.

Neyman, J., Scott, E, 1948. Consistent estimation from partially consistent observations. Econometrica 16, 1–32.

Pearl, R., Madvig, P., 2020. Managing the most expensive patients: A new primary-care model can lower costs and improve outcomes. Harvard Business Review, January-February, 68–75.

Peek, E., Cuijpers, R., Buijink, W., 2010. Creditors' and shareholders' reporting demands in public versus private firms: Evidence from Europe. Contemporary Accounting Research 27(1), 49–91.

Rosen, S., 1981. The economics of superstars. American Economic Review 71(5), 845–858.

Taleb, N., 2007. The Black Swan: The Impact of Highly Improbable Events, first ed. Random House, New York.

White, H., Kim, T., Manganelli, S., 2008. Modeling autoregressive conditional skewness and kurtosis with multi-quantile CAViaR. Working paper, European Central Bank.

Wong, G., Liu, W., Liu, Y., Zhou, B., Bi, Y., Gao, G., 2015. MERS, SARS, and Ebola: The role of super-spreaders in infectious disease. Cell Host & Microbe 18(4), 398–401.
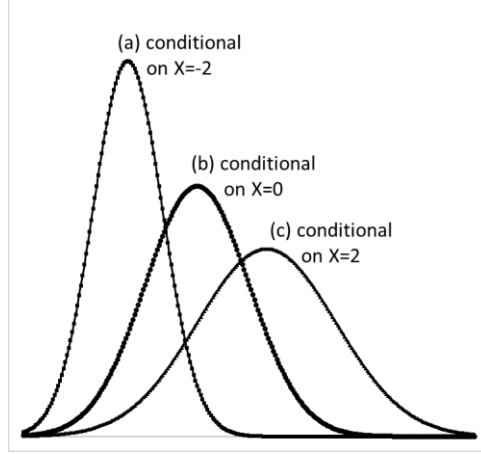
Panel A: The distribution of *Y* conditional on *X*
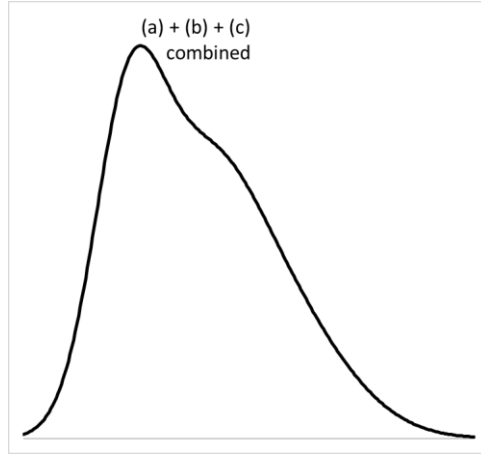
Panel B: The distribution of *Y* for a rolling window that contains observations
with *X*=-2, *X*=0, and *X*=2 (i.e., distributions (a) + (b) + (c) combined)
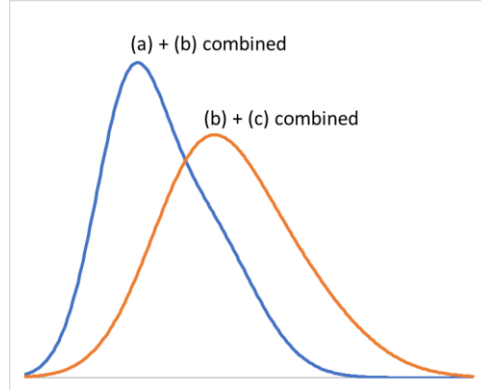
Panel C: The distribution of *Y* for a rolling window with *X*=-2 and *X*=0 (i.e., (a)+(b) combined)
versus a rolling window with *X*=0 and *X*=2 (i.e., (b)+(c) combined)

**Figure 1. Within-window variation in *X* leads to spurious findings of skewness and
variation in skewness in the standard method**

The figure presents the probability density function of *Y* for different scenarios on the conditioning variable *X*. The
distribution of *Y* conditional on *X* in Panel A is Normal with *mean* = *X* and *SD* = 1.5+(0.25×*X*). The combined
distributions in Panels B and C are obtained by adding up the relevant density functions from Panel A and rescaling.
The Pearson skewness is 0 in Panel A, 0.52 in Panel B, and 0.45 for (a)+(b) and 0.30 for (b)+(c) in Panel C.
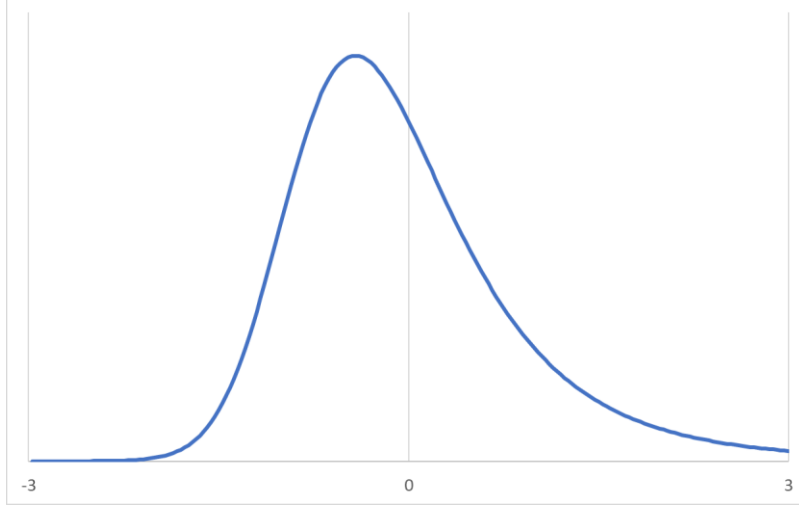
Panel A: The distribution of Y conditional on X (the variation reflects $\varepsilon_{firm} + \varepsilon_{firm-year}$)
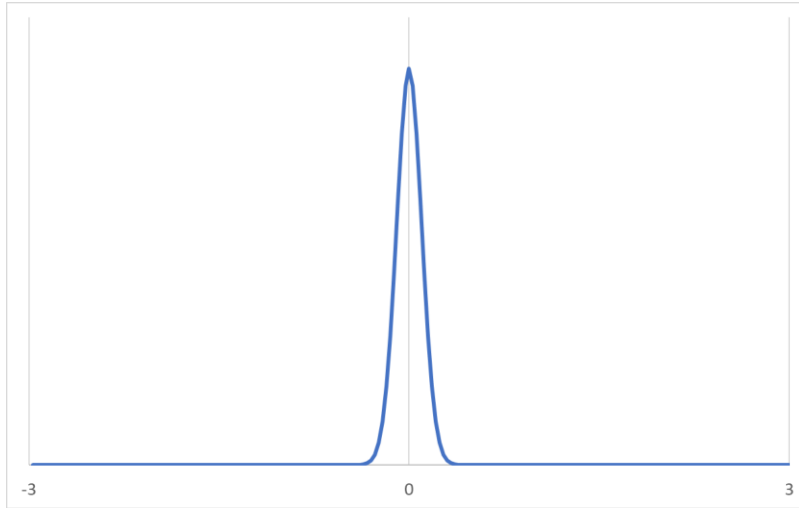
Panel B: The distribution of Y within a firm-specific rolling window (the variation only reflects $\varepsilon_{firm-year}$, while ε_{firm} is absorbed by the firm-specific mean for the window)

Figure 2. The standard method cannot detect skewness embedded in the firm effects because the rolling window overcontrols for persistent variation in Y

The figure presents the probability density function of Y conditional on X in Panel A, and the probability density function of Y within a firm-specific rolling window in Panel B. We generate $Y = \beta X + \varepsilon_{firm} + \varepsilon_{firm-year}$, where X only varies cross-sectionally (such that the bias from Figure 1 does not arise), the firm effect ε_{firm} is drawn from a normalized non-central Student's t distribution with 5 degrees of freedom and non-centrality parameter $\delta=3$, and the firm-year shock $\varepsilon_{firm-year}$ is drawn from a Normal distribution with mean zero and standard deviation 0.1. Thus, the firm effects ε_{firm} are skewed while the firm-year shocks $\varepsilon_{firm-year}$ are symmetric. The distribution of Y conditional on X in Panel A reflects variation in $\varepsilon_{firm} + \varepsilon_{firm-year}$, while βX is held constant (we normalize $\beta X = 0$ without loss of generality because it is subtracted from Y during the computation of conditional skewness). This distribution is skewed because it includes the firm effects ε_{firm}. The distribution of Y within a firm-specific rolling window in Panel B reflects variation in $\varepsilon_{firm-year}$ only, while βX and ε_{firm} are constant within the window (and are absorbed by the window-specific mean during the computation of conditional skewness). This distribution is symmetric because it excludes the skewed firm effects ε_{firm}.
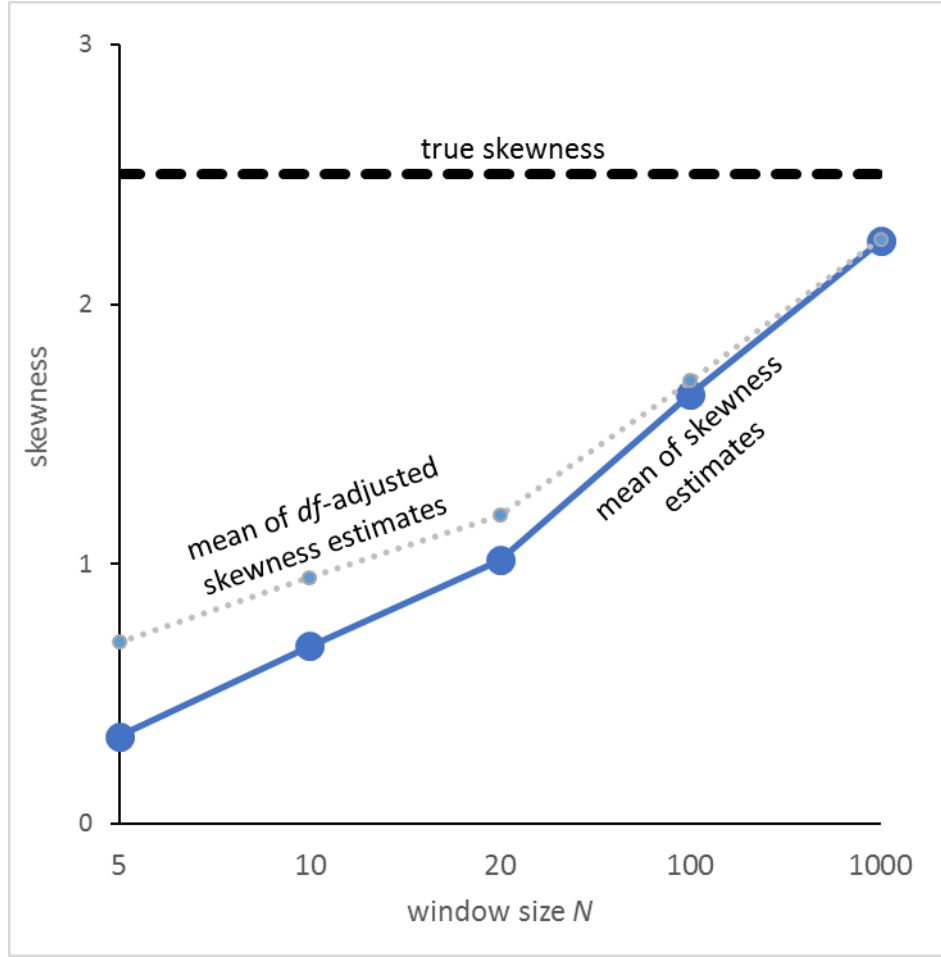
Figure 3. Variation in the number of observations across rolling windows leads to spurious findings of variation in skewness in the standard method

Y follows a non-central Student's t distribution with 5 degrees of freedom and non-centrality parameter δ=3. We generate Y for 10,000 firm-specific windows of N observations each (for N = 5, 10, 20, 100, and 1,000), compute Pearson's moment coefficient of skewness of Y in each window, and report the average skewness estimate across the 10,000 windows (the standard error of this average estimate is negligible relative to the scale of the vertical axis, and therefore we omit the confidence intervals). Standard Stata implementation of Pearson's skewness uses the formula $\frac{1}{N}\Sigma\left(\frac{Y_i-m}{s}\right)^3$, where m is the sample mean and s is the sample standard deviation. The df-adjusted formula $\frac{N}{(N-1)(N-2)}\Sigma\left(\frac{Y_i-m}{s}\right)^3$ incorporates a degrees-of-freedom adjustment for small samples (e.g., Gu and Wu, 2003) and is the default implementation in SAS.
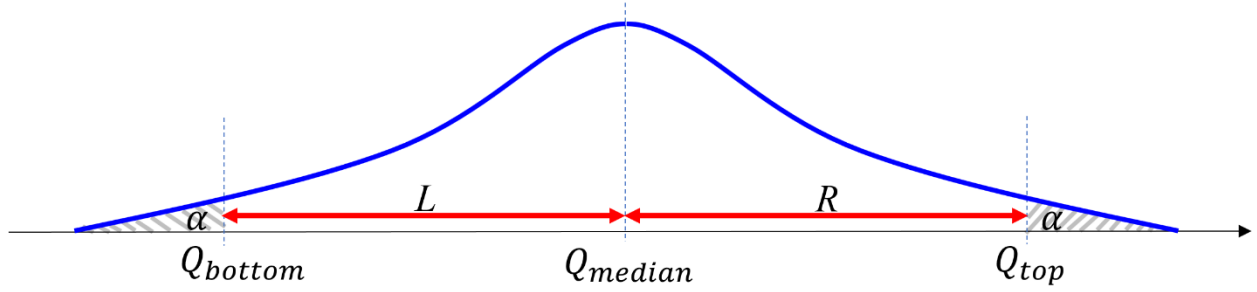
Panel A: Symmetric distribution (R=L, zero skewness)
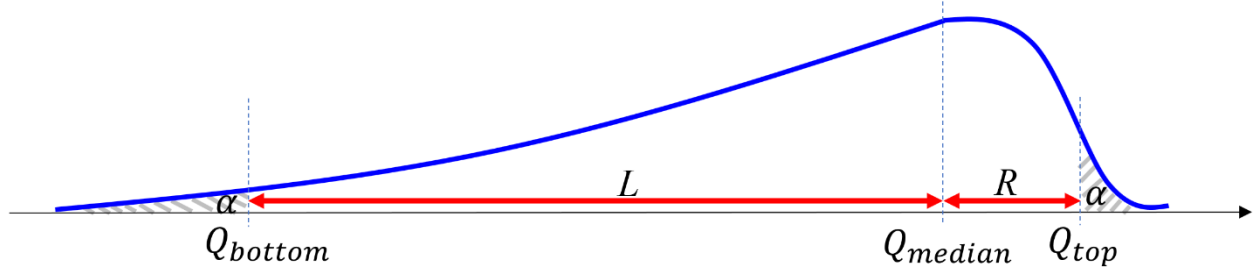
Panel B: Left-skewed distribution (R<L, negative skewness)

Figure 4. Quantile-based skewness measure

L and R in the figure are the inter-quantile distances $L \equiv Q_{median} - Q_{bottom}$ and $R \equiv Q_{top} - Q_{median}$. Quantile-based skewness (6) can be rewritten as $SkewQ(Y) = \frac{R-L}{R+L}$. In Panel A, the distribution is symmetric, and $R=L$. Therefore, $SkewQ(Y) = 0$. In Panel B, the distribution is left-skewed, and $R<L$. Therefore, $SkewQ(Y)$ is negative.
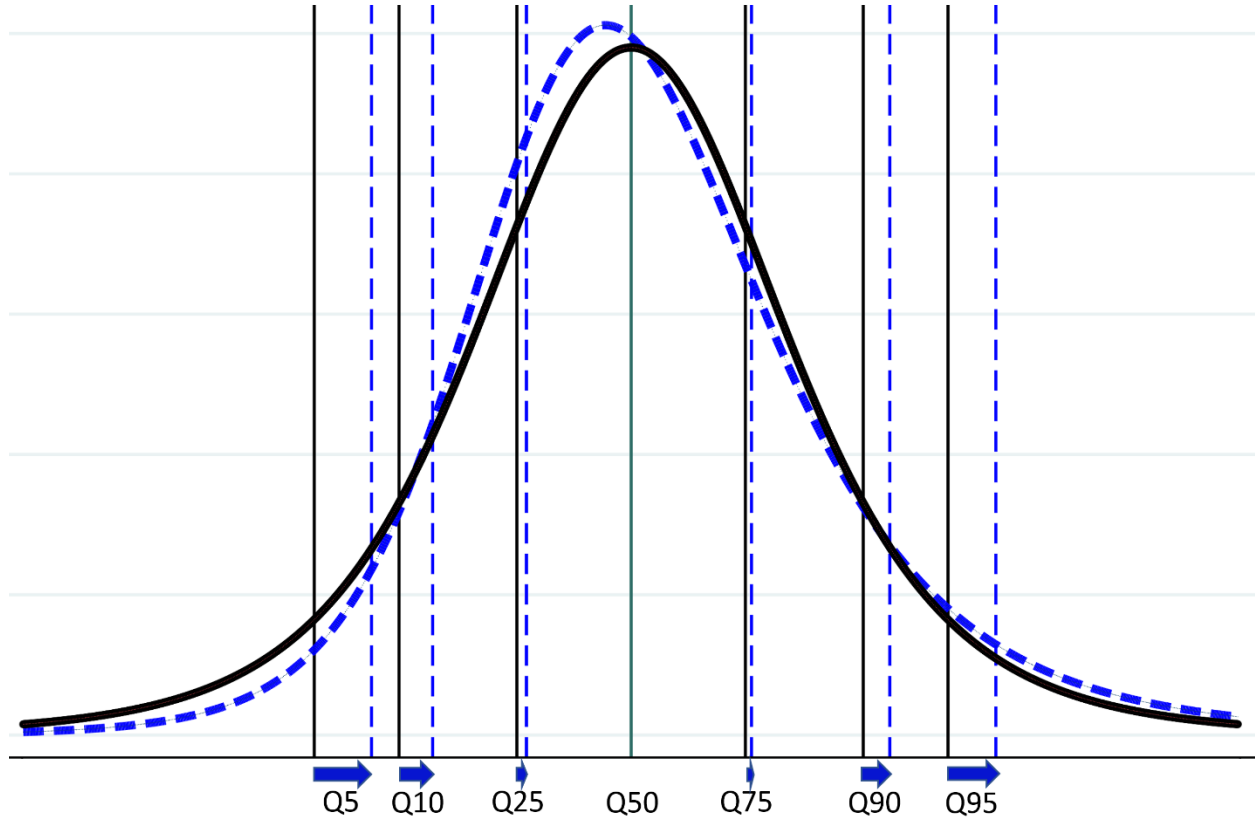
Figure 5. Impact of a given change in skewness on different distribution quantiles

The figure presents the probability density function of a normalized non-central Student's t distribution with non-centrality parameter $\delta=0$ (the solid curve that represents a symmetric distribution) and $\delta=1$ (the dashed curve that represents a right-skewed distribution). For ease of visual comparison, the medians of both distributions are normalized to zero. The vertical lines represent the distribution quantiles (solid lines for $\delta=0$ and dashed lines for $\delta=1$). The arrows represent the shift in quantiles when the non-centrality parameter increases from $\delta=0$ to $\delta=1$.

Table 1. Type-I error when *X* varies over time, which can bias the standard estimates of both skewness on average and variation in skewness (Figure 1)

Panel A: Simulation for a symmetric Student's t distribution with df=5

	Percentage of time-series variation in X (ρ)			
	0%	25%	50%	75%
Is there skewness on average? (H_0: $\beta_0 = 0$)				
Standard method	7.6	100.0	100.0	100.0
Our method for Pearson's skewness				
GMM estimation	5.2	4.8	4.8	4.8
Three-stage estimation	4.0	3.6	3.6	4.0
Our method for quantile-based skewness				
α=0.25	5.2	4.8	5.2	4.0
α=0.1	4.8	4.4	5.2	5.6
α=0.05	6.8	5.6	6.8	7.6
Does *X* affect skewness? (H_0: $\beta_X = 0$)				
Standard method	7.6	79.2	76.4	78.4
Our method for Pearson's skewness				
GMM estimation	5.6	4.8	2.4	2.8
Three-stage estimation	2.8	2.0	0.8	0.4
Our method for quantile-based skewness				
α=0.25	2.8	6.4	8.0	6.4
α=0.1	5.6	4.4	3.2	4.0
α=0.05	4.4	2.8	5.6	4.4

Panel B: Sensitivity checks for the simulation with 50% time-series variation (ρ=0.5)

	Shorter window 5 years×2,000 firms	Longer window 20 years×500 firms	Less fat tails *df*=10	Fatter tails *df*=3
Is there skewness on average? (H_0: $\beta_0 = 0$)				
Standard method	100.0	100.0	100.0	100.0[a]
Our method for Pearson's skewness				
GMM estimation	4.8	4.8	5.2	9.2[a]
Three-stage estimation	3.6	4.0	3.6	5.2[a]
Our method for quantile-based skewness				
α=0.25	5.2	3.6	4.8	5.2
α=0.1	5.6	4.8	4.0	5.2
α=0.05	6.0	8.0	6.4	6.4
Does *X* affect skewness? (H_0: $\beta_X = 0$)				
Standard method	87.6	54.4	90.0	37.6[a]
Our method for Pearson's skewness				
GMM estimation	4.4	3.6	4.8	8.0[a]
Three-stage estimation	2.8	1.2	0.8	2.8[a]
Our method for quantile-based skewness				
α=0.25	4.4	4.4	7.2	7.6
α=0.1	5.6	5.2	3.6	3.6
α=0.05	5.2	4.8	5.2	5.2

The table presents rejection rates across 250 simulated samples. In Panel A, each sample comprises 1,000 firms × 10 annual observations per firm. For each firm-year, we draw $X = \sqrt{1-\rho} \times \chi_{firm} + \sqrt{\rho} \times \chi_{firm-year}$, where χ_{firm} and $\chi_{firm-year}$ are *Uniform*[-1,1] random variables at the firm and firm-year levels, respectively. We then generate the dependent variable $Y = \mu_0 + \mu_X X + exp(\pi_0 + \pi_X X) \times \varepsilon$, where $\mu_0 = 2$, $\mu_X = 1.5$, $\pi_0 = 0$, $\pi_X = 1$, and the random variable ε is drawn from a symmetric Student's *t* distribution with non-centrality parameter $\delta = 0$, such that

the true conditional skewness is zero for all X by construction. The sensitivity checks in Panel B follow the definitions from Panel A, except for the specific change noted in the column title.

After generating each sample, we estimate the conditional skewness of Y as a function of X (i.e., $skewness = \beta_0 + \beta_X X$) using the methods described in Section 2. Firm-level Pearson's moment coefficient of skewness in the standard method is computed on a window of 10 observations per firm. The GMM version of our method uses off-the-shelf standard errors, while the multi-stage version uses bootstrapped standard errors with 100 replications. In all methods, the standard errors are clustered by firm.

We use two-tailed tests with a 5% nominal significance level. The test of whether there is conditional skewness on average only focuses on the coefficient β_0 because the mean of X is zero. Because the null hypotheses for the tests are true by construction, a well-performing test should have rejection rates near 5%. The 95% confidence interval for the rejection rates across the 250 simulations is 2.3–7.7%.

[a] In the fat-tailed scenario with df=3 in Panel B, inferences for the standard method and for the Pearson-based version of our method are *a priori* invalid because Pearson's moment coefficient of skewness is theoretically undefined for df=3 (Jondeau and Rockinger, 2003). Inferences in the quantile-based method remain valid.

Table 2. Test power when X varies over time, which can bias the standard estimates of both skewness on average and variation in skewness (Figure 1)

Panel A: Simulation for a right-skewed non-central Student's t distribution with non-centrality parameter $\delta = 1 + 0.5 \times X$ and df=5

	Percentage of time-series variation in X (ρ)			
	0%	25%	50%	75%
Is there skewness on average? ($H_0: \beta_0 = 0$)				
Standard method	100.0	100.0	100.0	100.0
Our method for Pearson's skewness				
GMM estimation	99.6	99.6	99.6	99.6
Three-stage estimation	86.8	74.8	76.4	77.2
Our method for quantile-based skewness				
α=0.25	100.0	100.0	98.4	99.6
α=0.1	100.0	100.0	100.0	100.0
α=0.05	100.0	100.0	100.0	100.0
Does X affect skewness? ($H_0: \beta_X = 0$)				
Standard method	93.2	29.2	20.4	8.4
Our method for Pearson's skewness				
GMM estimation	42.0	39.2	40.8	45.6
Three-stage estimation	26.0	23.2	28.4	30.8
Our method for quantile-based skewness				
α=0.25	18.8	21.2	26.8	24.4
α=0.1	77.2	73.6	70.4	72.0
α=0.05	92.8	90.0	88.8	90.8

Panel B: Sensitivity checks for the simulation with 50% time-series variation (ρ=0.5)

	Shorter window 5 years×2,000 firms	Longer window 20 years×500 firms	Less fat tails df=10	Fatter tails df=3
Is there skewness on average? ($H_0: \beta_0 = 0$)				
Standard method	100.0	100.0	100.0	100.0[a]
Our method for Pearson's skewness				
GMM estimation	99.6	99.6	100.0	92.4[a]
Three-stage estimation	75.2	81.2	90.8	54.4[a]
Our method for quantile-based skewness				
α=0.25	99.6	98.8	62.8	100.0
α=0.1	100.0	100.0	100.0	100.0
α=0.05	100.0	100.0	100.0	100.0
Does X affect skewness? ($H_0: \beta_X = 0$)				
Standard method	6.8	52.0	28.8	99.2[a]
Our method for Pearson's skewness				
GMM estimation	39.6	40.4	36.0	26.8[a]
Three-stage estimation	30.4	29.2	30.8	22.0[a]
Our method for quantile-based skewness				
α=0.25	25.6	30.0	10.0	58.8
α=0.1	68.0	69.6	22.8	97.6
α=0.05	87.6	88.0	34.8	100.0

The table presents rejection rates across 250 simulated samples. In Panel A, each sample comprises 1,000 firms × 10 annual observations per firm. For each firm-year, we draw $X = \sqrt{1-\rho} \times \chi_{firm} + \sqrt{\rho} \times \chi_{firm-year}$, where χ_{firm} and $\chi_{firm-year}$ are *Uniform*[-1,1] random variables at the firm and firm-year levels, respectively. We then generate the dependent variable $Y = \mu_0 + \mu_X X + exp(\pi_0 + \pi_X X) \times \varepsilon$, where $\mu_0 = 2$, $\mu_X = 1.5$, $\pi_0 = 0$, $\pi_X = 1$. The random

variable ε is drawn from a right-skewed non-central Student's t distribution with non-centrality parameter $\delta = 1 + 0.5 \times X$, such that the true conditional skewness is positive and increasing with X. The sensitivity checks in Panel B follow the definitions from Panel A, except for the specific change noted in the column title.

After generating each sample, we estimate the conditional skewness of Y as a function of X (i.e., $skewness = \beta_0 + \beta_X X$) using the methods described in Section 2. Firm-level Pearson's moment coefficient of skewness in the standard method is computed on a window of 10 observations per firm. The GMM version of our method uses off-the-shelf standard errors, while the multi-stage version uses bootstrapped standard errors with 100 replications. In all methods, the standard errors are clustered by firm.

We use two-tailed tests with a 5% nominal significance level. The test of whether there is conditional skewness on average only focuses on the coefficient β_0 because the mean of X is zero.

[a] In the fat-tailed scenario with df=3 in Panel B, inferences for the standard method and for the Pearson-based version of our method are *a priori* invalid because Pearson's moment coefficient of skewness is theoretically undefined for df=3 (Jondeau and Rockinger, 2003).

Table 3. Test power when Y includes a skewed firm random effect unrelated to X, which can bias the standard estimates due to overcontrolling (Figure 2)

	Constant skewness		Skewness increases with X	
	$\delta=0.5$	$\delta=3$	$\delta=1+0.5\times X$	$\delta=1+1\times X$
Is there skewness on average? (H_0: $\beta_0 = 0$)				
Standard method	3.2	3.2	3.2	3.2
Our method for Pearson's skewness				
GMM estimation	59.2	96.8	93.6	90.4
Three-stage estimation	22.0	41.2	44.8	48.8
Our method for quantile-based skewness				
$\alpha=0.25$	17.6	99.2	42.4	39.6
$\alpha=0.1$	56.0	100.0	97.2	97.2
$\alpha=0.05$	62.8	100.0	99.2	99.2
Does X affect skewness? (H_0: $\beta_X = 0$)				
Standard method	4.0	4.0	4.0	4.0
Our method for Pearson's skewness				
GMM estimation	7.2	9.6	19.6	42.6
Three-stage estimation	4.0	7.6	13.2	26.4
Our method for quantile-based skewness				
$\alpha=0.25$	4.8	5.6	8.0	12.8
$\alpha=0.1$	5.2	5.2	16.8	45.2
$\alpha=0.05$	4.0	5.6	23.2	59.2

The table presents rejection rates across 250 simulated samples. Each sample comprises 1,000 firms × 10 annual observations per firm. For each firm-year, we draw $X = \chi_{firm}$, where χ_{firm} is a $Uniform[-1,1]$ random variable at the firm level without any variation across years (such that the bias from Figure 1 does not arise in the standard method). We then generate the dependent variable $Y = \mu_0 + \mu_X X + exp(\pi_0 + \pi_X X) \times \varepsilon$, where $\mu_0 = 2$, $\mu_X = 1.5$, $\pi_0 = 0$, $\pi_X = 1$. The random shock $\varepsilon = \varepsilon_{firm} + \varepsilon_{firm-year}$, where ε_{firm} is a firm-level random effect drawn from a right-skewed non-central Student's t distribution with non-centrality parameter δ that is defined in the column heading, and $\varepsilon_{firm-year}$ is a firm-year-level shock drawn from a Normal distribution with mean 0 and standard deviation 0.1.

After generating each sample, we estimate the conditional skewness of Y as a function of X (i.e., $skewness = \beta_0 + \beta_X X$) using the methods described in Section 2. Firm-level Pearson's moment coefficient of skewness in the standard method is computed on a window of 10 observations per firm. The GMM version of our method uses off-the-shelf standard errors, while the multi-stage version uses bootstrapped standard errors with 100 replications. In all methods, the standard errors are clustered by firm.

We use two-tailed tests with a 5% nominal significance level. The test of whether there is conditional skewness on average only focuses on the coefficient β_0 because the mean of X is zero. Because the true skewness is positive on average in all four scenarios, the rejection rates in the test for β_0 in the top half of the table reflect test power. The rejection rates in the test for β_X in the bottom half of the table reflect type-I error in the first two scenarios (the true skewness does not vary with X) and reflect test power in the last two scenarios (true skewness increases with X).

Table 4. Type-I error when the number of observations varies across windows, which can bias the standard estimates of variation in skewness (Figure 3)

	Constant skewness δ=0.5	Constant skewness δ=3
Does the number of observations N_{firm} affect skewness? ($H_0: \beta_X = 0$)		
Standard method		
without *df* adjustment	53.6	100.0
with *df* adjustment	27.6	99.2
Standard method after discarding firms with fewer than 5 observations		
without *df* adjustment	40.0	99.6
with *df* adjustment	25.2	98.0
Our method for Pearson's skewness		
GMM estimation	6.8	6.8
Three-stage estimation	6.4	7.2
Our method for quantile-based skewness		
α=0.25	7.6	7.2
α=0.1	7.2	6.0
α=0.05	5.2	4.8

The table presents rejection rates across 250 simulated samples. Each sample comprises 10,000 firm-years with variation in the time series length per firm. At the end of each period, a firm exits the sample with probability 0.1, such that the number of observations per firm is a random variable with mean 10 [=1/0.1] and variance 90 [=(1−0.1)/0.1²]. We define the explanatory variable N_{firm} as the number of observations for the firm in the simulated data. For each firm-year, we then generate the dependent variable Y from a normalized non-central Student's *t* distribution with *df*=5 and a constant non-centrality parameter δ. Thus, the true skewness of Y does not vary with any explanatory variables.

After generating each simulated sample, we estimate the skewness of Y as a function of the number of observations N_{firm} (i.e., $skewness = \beta_0 + \beta_X N_{firm}$) using the methods described in Section 2. Firm-level Pearson's moment coefficient of skewness in the standard method is computed using all available observations for the firm. The *df*-adjusted version of the standard method uses the degrees-of-freedom adjustment $\frac{N_{firm}}{(N_{firm}-1)(N_{firm}-2)}$ in the computation of firm-level skewness (e.g., Gu and Wu, 2003). This adjustment affects inferences only when the number of observations varies across windows, and therefore it is irrelevant for Tables 1–3. The GMM version of our method uses off-the-shelf standard errors, while the multi-stage version uses bootstrapped standard errors with 100 replications. In all methods, the standard errors are clustered by firm.

We test whether N_{firm} affects skewness using a two-tailed test with a 5% nominal significance level. Because the null hypothesis of zero effect is true by construction, a well-performing test should have rejection rates near 5%. The 95% confidence interval for the rejection rates across the 250 simulations is 2.3–7.7%.

Table 5. The relation between managerial overconfidence and skewness-based conservatism proxies

Panel A: Standard analysis for rolling-window skewness measures

	Y = Skew(CFO) – Skew(Earn)			Y = –Skew(Accr)		
Holder 67	-0.287***			-0.208***		
	(-3.25)			(-2.96)		
CAPEX		-0.319***			-0.374***	
		(-5.00)			(-6.96)	
OverInvest			-0.284***			-0.461***
			(-6.13)			(-11.92)
Own	1.188	1.171	1.154	0.805	0.802	0.783
	(1.30)	(1.30)	(1.28)	(1.15)	(1.15)	(1.12)
MTB	-0.102***	-0.105***	-0.105***	-0.085***	-0.086***	-0.086***
	(-8.31)	(-8.67)	(-8.60)	(-7.85)	(-8.09)	(-8.00)
Lev	2.927***	2.929***	2.947***	2.706***	2.689***	2.700***
	(10.53)	(10.63)	(10.66)	(10.67)	(10.76)	(10.78)
Size	0.069	0.052	0.110	-0.076	-0.089	0.004
	(0.88)	(0.68)	(1.41)	(-1.15)	(-1.36)	(0.05)
Litigation	18.762***	18.817***	17.423***	15.906***	16.009***	13.810***
	(9.06)	(9.11)	(8.48)	(8.63)	(8.71)	(7.59)
SalesGrowth	-1.360***	-1.306***	-1.421***	-1.221***	-1.150***	-1.306***
	(-11.16)	(-10.66)	(-11.57)	(-12.04)	(-11.24)	(-12.72)
R&D AD	6.448***	6.448***	6.390***	8.079***	8.105***	8.038***
	(10.11)	(10.10)	(10.00)	(12.39)	(12.34)	(12.31)
CFO	17.253***	17.258***	17.203***	16.637***	16.685***	16.639***
	(33.53)	(33.74)	(33.63)	(35.27)	(35.57)	(35.71)
σRevenue	1.051***	1.011***	1.080***	1.242***	1.182***	1.262***
	(4.23)	(4.09)	(4.37)	(5.91)	(5.68)	(6.09)
Year fixed effects	included	included	included	included	included	included
Firm fixed effects	included	included	included	included	included	included
R^2	0.165	0.166	0.166	0.202	0.205	0.210

Panel B: Our methods, estimates for skewness of Y = –Accr conditional on X

	Pearson's skewness, GMM			Quantile-based skewness with α=0.05		
	Coefficients for conditional skewness			Coefficients for conditional skewness		
Holder 67	**0.147**			**0.006**		
	(0.85)			**(0.22)**		
CAPEX		**0.240**			**-0.004**	
		(1.34)			**(-0.17)**	
OverInvest			**0.152**			**-0.156*** **
			(0.72)			**(-7.35)**
Own	-0.199	-0.189	-0.138	-0.034***	-0.033***	-0.029***
	(-1.57)	(-1.57)	(-1.04)	(-2.97)	(-3.15)	(-2.77)
MTB	-0.179	-0.157	-0.211	0.059***	0.062***	0.053***
	(-1.03)	(-0.93)	(-1.15)	(4.59)	(4.67)	(4.27)
Lev	-0.245**	-0.256**	-0.209*	-0.023*	-0.021*	-0.018
	(-2.26)	(-2.32)	(-1.78)	(-1.83)	(-1.73)	(-1.33)
Size	0.323	0.347*	0.485**	-0.065***	-0.070***	-0.057***
	(1.59)	(1.73)	(2.32)	(-4.01)	(-4.10)	(-3.43)
Litigation	-0.002	-0.014	-0.077	0.091***	0.094***	0.078***
	(-0.01)	(-0.11)	(-0.60)	(6.42)	(6.50)	(5.32)
SalesGrowth	0.255*	0.272*	0.444***	-0.069***	-0.070***	-0.073***
	(1.78)	(1.85)	(2.77)	(-7.35)	(-6.85)	(-7.34)
R&D AD	0.413**	0.417**	0.502**	-0.049***	-0.050***	-0.055***
	(2.10)	(2.12)	(2.39)	(-4.42)	(-4.76)	(-5.26)
CFO	-0.171	-0.192	-0.182	-0.077***	-0.076***	-0.073***
	(-1.23)	(-1.34)	(-1.22)	(-6.36)	(-6.46)	(-6.95)
σRevenue	0.084	0.095	0.072	0.064***	0.068***	0.069***
	(0.84)	(0.94)	(0.64)	(5.93)	(6.30)	(6.89)
Year fixed effects	included	included	included	included	included	included
	Coefficients for conditional mean			Coefficients for conditional median		
Holder 67	-0.002			-0.003**		
	(-1.43)			(-2.57)		
CAPEX		0.003**			0.003***	
		(2.01)			(4.05)	
OverInvest			-0.018***			-0.009***
			(-18.33)			(-13.72)
Controls *Own… σRevenue*	included	included	included	included	included	included
Year fixed effects	included	included	included	included	included	included
	Coefficients for conditional standard deviation			Coefficients for interquantile range (Q95–Q5)		
Holder 67	0.009			-0.002		
	(0.33)			(-0.62)		
CAPEX		-0.045*			-0.004	
		(-1.96)			(-1.08)	
OverInvest			-0.118***			-0.007**
			(-5.09)			(-1.98)
Controls *Own… σRevenue*	included	included	included	included	included	included
Year fixed effects	included	included	included	included	included	included

Panel C: Our methods, estimates for the difference between the skewness of Y1 = Earn and Y2 = CFO conditional on X

	Pearson's skewness, GMM			Quantile-based skewness with α=0.05		
	Coefficients for conditional skewness			Coefficients for conditional skewness		
Holder 67	**0.100**			**0.009**		
	(0.50)			**(0.29)**		
CAPEX		**0.109**			**-0.043**	
		(0.59)			**(-1.55)**	
OverInvest			**-0.185**			**-0.164*****
			(-0.83)			**(-6.45)**
Own	-0.070	-0.052	-0.021	-0.038**	-0.038**	-0.040**
	(-0.94)	(-0.68)	(-0.24)	(-2.13)	(-2.18)	(-2.48)
MTB	0.020	0.002	0.004	0.027*	0.032**	0.018
	(0.18)	(0.02)	(0.04)	(1.91)	(2.22)	(1.22)
Lev	-0.420**	-0.383**	-0.454**	0.013	0.001	0.004
	(-2.25)	(-2.14)	(-2.53)	(0.83)	(0.08)	(0.26)
Size	0.786	0.834*	1.179**	-0.017	-0.014	0.005
	(1.51)	(1.67)	(2.27)	(-0.77)	(-0.66)	(0.19)
Litigation	0.042	0.021	-0.008	0.048**	0.056***	0.032
	(0.30)	(0.15)	(-0.06)	(2.45)	(2.97)	(1.59)
SalesGrowth	1.184***	1.162***	1.445***	-0.045***	-0.048***	-0.054***
	(3.22)	(3.02)	(3.78)	(-3.83)	(-4.06)	(-4.31)
R&D AD	0.283**	0.297**	0.299**	-0.022	-0.013	-0.031
	(2.27)	(2.35)	(2.33)	(-1.21)	(-0.70)	(-1.62)
σRevenue	0.136	0.138	0.134	0.063***	0.066***	0.065***
	(1.48)	(1.49)	(1.44)	(4.42)	(4.15)	(4.23)
Year fixed effects	included	included	included	included	included	included
	Coefficients for conditional mean			Coefficients for conditional median		
Holder 67	0.004**			0.006***		
	(2.25)			(3.63)		
CAPEX		0.015***			0.018***	
		(9.86)			(13.08)	
OverInvest			-0.017***			-0.006***
			(-13.47)			(-5.66)
Controls Own… σRevenue	included	included	included	included	included	included
Year fixed effects	included	included	included	included	included	included
	Coefficients for conditional standard deviation			Coefficients for interquantile range (Q95–Q5)		
Holder 67	0.045			0.012**		
	(1.62)			(2.21)		
CAPEX		0.082***			0.007	
		(3.32)			(1.16)	
OverInvest			0.213***			0.025***
			(9.55)			(6.08)
Controls Own… σRevenue	included	included	included	included	included	included
Year fixed effects	included	included	included	included	included	included

The sample comprises 17,476 firm-year observations during 1993–2009. The *t*-statistics in parentheses are clustered by firm. The coefficients highlighted in boldface capture the relation between overconfidence indicators and skewness-based conservatism proxies. For consistent interpretation of the coefficient signs across all panels, we flip the sign of the dependent variable where relevant (e.g., $Y=-Skew(Accr)$ in Panel A and $Y=-Accr$ in Panel B), such that a positive coefficient in all cases indicates an increase in conservatism.

The overconfidence indicators follow Ahmed and Duellman (2013). *Holder67* is activated if the CEO's ratio of the value of options in the money to the strike price is above 0.67 at least twice during the sample period. When this

condition holds, *Holder67* is set to one from the first fiscal year in which the ratio exceeds 0.67 and until the end of the sample, and is set to zero otherwise. The value per option is computed as Execucomp items OPT_UNEX_EXER_EST_VAL / OPT_UNEX_EXER_NUM, and the strike price is computed as Compustat item PRCC_F minus the value per option. *CAPEX* equals one if capital expenditure (Compustat item CAPX) scaled by lagged total assets (AT) is above the median for the year in the firm's Fama-French industry, and zero otherwise. *OverInvest* equals one if the residual in a regression of total asset growth (AT) on sales growth (SALE), run within the firm's Fama-French industry-year, is positive, and zero otherwise.

The control variables follow Ahmed and Duellman (2013). *Own* is the percentage of the firm's outstanding shares (CSHO) that are owned by the CEO (SHROWN_EXCL_OPTS). *MTB* is the market-to-book ratio (items PRCC_F×CSHO and CEQ). *Lev* is total liabilities (LT) divided by total assets (AT). *Size* is the natural logarithm of total assets (AT). *Litigation* is the predicted probability of litigation for the firm-year based on the estimates from model (2) in Table 7 of Kim and Skinner (2012). *SalesGrowth* is the percentage change in sales (SALE). *R&D AD* is research and development expense plus advertising expense (XRD+XAD) scaled by sales (SALE). *CFO* is operating cash flow (OANCF−XIDOC) scaled by average total assets (AT). *σRevenue* is the standard deviation of log-changes in sales (SALE) over a five-year rolling window. *Earn* is earnings before extraordinary items and discontinued operations (IB) scaled by average total assets (AT). *Accr* is accruals (*Earn – CFO*).

Panel A presents standard skewness regressions. Following Ahmed and Duellman (2013), the skewness of a variable y is computed as $(y−\mu)^3/\sigma^3$, where μ and σ are the mean and standard deviation of y over a five-year rolling window. Ahmed and Duellman's computation differs from typical rolling-window skewness measures in the literature: it uses $(y−\mu)^3$ for the current year only, rather than the average of $(y−\mu)^3$ over the rolling window. When we re-run the models from Panel A for the more typical rolling-window computation in untabulated tests, the coefficients on the overconfidence proxies become insignificant for $Y = Skew(CFO) – Skew(Earn)$ and are weakened considerably for $Y = −Skew(Accr)$.

Panel B presents our methods' estimates for conditional accrual skewness, defined as $Skew(−Accr|X) = −Skew(Accr|X)$, where the sign is flipped for interpretation consistency across all panels. The GMM estimation for Pearson's moment coefficient of skewness in the first three columns uses the code from Appendix A, and the two-stage estimation for quantile-based skewness in the last three columns uses the code from Appendix C with 100 bootstrap replications. For brevity, we tabulate the coefficients on the control variables only in the skewness part of the model, but all other parts of the model include the same full list of controls. In GMM estimation, the coefficients in Panel B represent the conditional skewness equation (4c), the conditional mean equation (4a), and the conditional standard deviation equation (4c), respectively. In quantile-based estimation, the coefficients represent the conditional skewness equation (9b) and its individual components that parallel the conditional mean and standard deviation (i.e., the conditional median $Q_{median}(Y|X)$ and the interquantile range $Q_{top}(Y|X) − Q_{bottom}(Y|X)$, respectively).

Panel C presents our methods' estimates for the difference in conditional skewness between operating cash flow and earnings (*Skew(CFO|X) – Skew(Earn|X)*). The GMM estimation in the first three columns uses stacked regression design. We first create two copies of each firm-year observation in the original sample. We define *Ystack=Earn* for copy 1 and *Ystack=CFO* for copy 2. We define the vector of explanatory variables *Xstack={X, 0}* for copy 1 and *Xstack={X, X}* for copy 2, where *X* is the original vector of explanatory variables, and 0 is a conformable vector of zeros. We then run our regular GMM code from Appendix A for *Ystack* and *Xstack* in the full stacked sample. The estimates for the first half of *Xstack* in this stacked design capture the coefficients on *X* for *Skew(Earn|X)*, and the estimates for the second half of *Xstack* capture the difference between the coefficients on *X* for *Skew(CFO|X)* and the coefficients on *X* for *Skew(Earn|X)*. In two-stage estimation for quantile-based skewness in the last three columns, we separately estimate the model for *Y=Earn* and *Y=CFO*, compute the difference between the skewness coefficients on *X* for these two dependent variables, and then bootstrap this difference to obtain the *t*-statistics. Because we directly bootstrap the difference, we do not need to estimate this model jointly for *Y=Earn* and *Y=CFO* using a stacked design. We exclude *CFO* from the list of control variables in Panel C because *Skew(CFO|CFO)* is undefined. We do not use firm fixed effects in Panels B and C, because they would destroy the consistency of the estimates due to the incidental parameter problem (Neyman and Scott, 1948).

To speed up the numerical search in Panels B and C, we normalize the control variables in these panels to have zero mean and unit standard deviation. This does not change the statistical significance of the estimates or their signs. For better comparability with Ahmed and Duellman (2013), we do not normalize the variables in Panel A.

Appendix A. Stata code example for the GMM version of our method for Pearson's moment coefficient of skewness

The following Stata command implements GMM estimation with moment conditions (4a)–(4c) from Section 2.3. The dependent variable is Y, the explanatory variables are X1–X20, and the standard errors are clustered by firmID.

```
gmm  ( Y - {mu: _cons X1-X20} )                              ///    // note A below
     ( (Y - {mu:})^2 - exp(2*{pi: _cons X1-X20}) )           ///
     ( ((Y - {mu:})^3)/exp(3*{pi:}) - {beta: _cons X1-X20} ),  ///
     inst(X1-X20) vce(cluster firmID) onestep winitial(identity)  ///    // note B below
     derivative( 1/mu = -1 )                                 ///    // note C below
     derivative( 2/mu = -2*(Y-{mu:}) )                       ///
     derivative( 2/pi = -2*exp(2*{pi:}) )                    ///
     derivative( 3/mu = -3*((Y-{mu:})^2)/exp(3*{pi:}) )      ///
     derivative( 3/pi = -3*((Y-{mu:})^3)/exp(3*{pi:}) )      ///
     derivative( 3/beta  = -1 )
```

This command uses moment conditions of the form $E(\xi|X) = 0$. For example, condition (4a) defines $\xi = Y - \mu_0 - \mu_X X$. The first line of code implements this ξ for $X = (X_1, ..., X_{20})$, and the option *inst(X1-X20)* in the fourth line defines the list of conditioning variables X.

Note A: The notation *{mu: _cons X1-X20}* defines a linear combination $\mu_0 + \mu_1 X_1 + \cdots + \mu_{20} X_{20}$.[40] To re-use this linear combination in another moment condition, just write *{mu:}*.

Note B: Because the model is just-identified (i.e., the number of moment conditions is equal to the number of coefficients), we use one-step GMM estimation, defined using options *onestep* and *winitial(identity)*. Two-step estimation and alternative settings for *winitial()* are relevant only for overidentified models.
The option *vce(cluster firmID)* defines clustered standard errors based on the firm index *firmID*. A researcher can use these off-the-shelf clustered standard errors for GMM (bootstrapping is required only in the multi-stage estimators, such as those in Appendices B and C).

Note C: The option *derivative* defines the analytical derivative of each moment condition (indexed by 1–3) with respect to the linear combinations *{mu:}*, *{pi:}*, and *{beta:}*. These analytical derivatives speed up the estimation considerably. When this option is omitted, the estimation uses slower numerical derivatives.

The following examples illustrate the syntax for accessing coefficients in hypothesis tests:
test [pi]_cons = [pi]X1 = 0   // jointly test the intercept π_0 and the slope coefficient π_1 on X1
                                   // in the linear combination *{pi: _cons X1-X20}*
lincom [pi]_cons+0.5[pi]X1* // test the linear combination $\pi_0 + 0.5\pi_1$

---

[40] To incorporate a long unstructured list of explanatory variables, a researcher can define a local macro, e.g.., *local Xvars = "X1 X2 X_A X_B BM AT ABC KFC"*, and then define the linear combination as *{mu: _cons `Xvars'}*.

Appendix B. Stata code example for the three-stage version of our method for Pearson's moment coefficient of skewness

The three estimation stages (without bootstrap) are packaged as a stand-alone estimation command *skewXpearson.ado* in a separate file. We wrap bootstrap around this command in the main file.

Code in skewXpearson.ado:

```
program define skewXpearson, eclass byable(recall) sortpreserve
        version 15
        syntax varlist [if]
        marksample touse

        tokenize `varlist'
        local Y "`1'"          // dependent variable `Y'
        macro shift 1
        local X "`*'"          // list of explanatory variables `X'

        tempvar eps eps2 predVar Y_tilde

        // Stage 1 – conditional mean as a function of `X'
        reg `Y' `X' if `touse'
        qui predict `eps', resid
        qui gen `eps2'=`eps'^2

        // Stage 2 – conditional variance as a function of X
        nl (`eps2'=exp({b0}+{xb: `X'})^2) if `touse', nolog
        qui predict `predVar'

        // Stage 3 – conditional skewness as a function of X
        qui gen `Y_tilde' = (`eps'^3)/(`predVar'^1.5)  // ^1.5 b/c predVar is Var not SD
        reg `Y_tilde' `X' if `touse'     // s.e. are handled by bootstrap outside of this file

        tempname b V e_N
        matrix `b' = e(b)
        matrix `V' = e(V) // these s.e. do not matter, but I need to return a matrix V here
        sca `e_N'=e(N)

        ereturn clear
        ereturn post `b' `V', esample(`touse')
        ereturn scalar N = `e_N'
end
```

Code in the main file:

```
bootstrap, reps(100) cluster(firmID): skewXpearson Y X1-X20
```

The option *reps()* defines the number of bootstrap replications. The option *cluster(firmID)* specifies that entire clusters (and not individual observations) are resampled during bootstrapping. This produces the bootstrap equivalent of clustered standard errors.

Appendix C. Stata code example for our method for quantile-based skewness

The two estimation stages (without bootstrap) are packaged as a stand-alone estimation command *skewXquantile.ado* in a separate file. We wrap bootstrap around this command in the main file.

Code in skewXquantile.ado:
```
program define skewXquantile, eclass byable(recall) sortpreserve
        version 11     // when version=11, qreg uses a different, much faster s.e. computation
        syntax varlist [if], Quantile(real)  // the option quantile sets the value of α∈(0,0.5)

        marksample touse

        tokenize `varlist'
        local Y "`1'"            // dependent variable `Y'
        macro shift 1
        local X "`*'"            // list of explanatory variables `X'

        // stage 1 –bottom and top quantiles as a function of X
        tempvar qBottom qTop rescaledY
        qreg `Y' `X' if `touse', quantile(`quantile')
        predict `qBottom', xb

        local tmpTop = 1-`quantile'
        qreg `Y' `X' if `touse', quantile(`tmpTop')
        predict `qTop', xb

        // stage 2 – conditional skewness as a function of X
        gen `rescaledY' = (`qTop'+`qBottom'-2*`Y')/(`qTop'-`qBottom') if `touse'
        qreg `rescaledY' `X' if `touse'  // s.e. are handled by bootstrap outside of this file

        tempname b V e_N
        matrix `b' = e(b)
        matrix `V' = e(V) // these s.e. do not matter, but I need to return a V matrix here
        sca `e_N'=e(N)

        ereturn clear
        ereturn post `b' `V', esample(`touse')
        ereturn scalar N = `e_N'
end
```

Code in the main file:
bootstrap, reps(100) cluster(firmID): skewXquantile Y X1-X20, quantile(0.1)
The option *reps()* defines the number of bootstrap replications. The option *cluster(firmID)* specifies that entire clusters (and not individual observations) are resampled during bootstrapping. This produces the bootstrap equivalent of clustered standard errors. The required option *quantile()* for the estimation command sets the quantile size α, which should be between 0 and 0.5.

Appendix D. An example of how a researcher can use conditional skewness as an explanatory variable when she has a valid instrument for skewness

Suppose that an intermediate outcome W has conditional mean $E(W|X,Z) = \mu_0 + \mu_X X + \mu_Z Z$, conditional standard deviation $SD(W|X,Z) = \exp(\pi_0 + \pi_X X + \pi_Z Z)$, and conditional Pearson's moment coefficient of skewness $Skew(W|X,Z) = \beta_0 + \beta_X X + \beta_Z Z$, where X and Z are explanatory variables, and $\beta_Z \neq 0$.

The final outcome $Y = \delta_0 + \delta_X X + \delta_{skew} Skew(W|X,Z) + \nu$, where $E(\nu|X,Z) = 0$.

A researcher wants to estimate the coefficient δ_{skew} on conditional skewness $Skew(W|X,Z)$. For example, W could be accruals, $Skew(W|X,Z)$ could be a proxy for conditional conservatism, and a researcher could want to measure the effect of this proxy on Y = *creditor recovery rate*. Because true $Skew(W|X,Z)$ is unobserved, she cannot simply regress Y on X and $Skew(W|X,Z)$.[41]

Crucially, we assume that Z affects Y only through $Skew(Y|X,Z)$, i.e., Z is a valid instrument for the effect of $Skew(Y|X,Z)$ on Y. Under this assumption, the effect of $Skew(Y|X,Z)$ on Y can be identified using the standard two-stage estimation logic for instrumental variables. For example, in stage 1, a researcher could estimate $Skew(Y|X,Z)$ as a function of X and Z using the code from Appendix A or B and compute predicted $\widehat{Skew}(Y|X,Z)$ for each observation. In stage 2, she could then regress Y on X and $\widehat{Skew}(Y|X,Z)$ to obtain consistent estimates of δ_X and δ_{skew}. Off-the-shelf OLS standard errors for this stage 2 are invalid because they ignore the estimation noise from stage 1, but valid standard errors can be obtained easily through bootstrapping (wrapped around both stages).

Alternatively, a researcher could jointly estimate the full model using GMM. For example:

```
gmm (  W - {mu: _cons X Z} )                              ///
     ( (W - {mu:})^2 - exp(2*{pi: _cons X Z}) )           ///
     ( ((W - {mu:})^3)/exp(3*{pi:}) - {beta: _cons X Z} ) ///
     ( Y - {delta: _cons X} - {deltaSkew}*{beta:} ),      /// // NB: {delta:} does NOT include Z
     inst(X Z) vce(cluster firmID) winitial(identity) onestep
```

The first three lines in this code define the moment conditions for the estimation of the conditional skewness of W and are similar to the code in Appendix A. The fourth line is the moment condition that links Y to X and the predicted conditional skewness of W (represented by the linear combination *{beta:}* in the code). Because all of the model components are estimated jointly, a researcher can use off-the-shelf clustered standard errors (option *vce(cluster firmID)*).

We caution that this approach requires a valid instrument Z for conditional skewness, and a researcher must carefully justify this instrument. Good instruments rarely exist in practice. We are not aware of a "silver bullet" solution that could yield valid estimates in this setting without a good instrument.

---

[41] Researchers often use rolling-window skewness as an empirical explanatory variable that approximates $Skew(W|X,Z)$. However, untabulated exploratory simulations show that this approach causes false inferences because this explanatory variable is susceptible to the biases from Figures 1–3.