# Which parametric model for conditional skewness?*

Bruno Feunou† Mohammad R. Jahan-Parvar‡

*Bank of Canada* *Federal Reserve Board*

Roméo Tédongap§

*Stockholm School of Economics and Swedish House of Finance*

First Draft: December 2011

This Draft: November 2013

**Abstract**

This paper addresses an existing gap in the developing literature on conditional skewness. We develop a simple procedure to evaluate parametric conditional skewness models. This procedure is based on regressing the realized skewness measures on model-implied conditional skewness values. We find that an asymmetric GARCH-type specification on shape parameters with a skewed generalized error distribution provides the best in-sample fit for the data, as well as reasonable predictions of the realized skewness measure. Our empirical findings imply significant asymmetry with respect to positive and negative news in both conditional asymmetry and kurtosis processes.

**Keywords:** Conditional skewness; Downside risk; Mincer-Zarnowitz regression; Realized skewness.

**JEL Classification:** C22, C51, G12, G15.

---

*We thank two anonymous referees, Chris Adcock (the editor), Torben Andersen, Bruce Hansen, Scott Hendry, Glen Keenleyside, Stanislav Khrapov, Nicola Loperfido, Dilip K. Patro, Kazuhiko Shinki, Akhtar Siddique, seminar participants at Joint Statistical Meetings 2011, Midwest Econometric Group meeting 2011, the Bank of Canada, Wayne State University (mathematics department), Office of the Comptroller of the Currency, and the MAF 2012 conference for many useful comments. An earlier version of this paper has been circulated and presented at various seminars and conferences under the title "Regime Switching in the Conditional Skewness of S&P500 Returns." The remaining errors are ours. The analysis and the conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Federal Reserve Board of Governors or the Bank of Canada.

†Bank of Canada, 234 Wellington St., Ottawa, Ontario, Canada K1A 0G9. e-mail: feun@bankofcanada.ca, Telephone No: (613) 782-8302.

‡Corresponding author: Office of Financial Stability Policy and Research, Federal Reserve Board of Governors, 20th St and Constitution Ave NW, Washington, DC 20551 USA. e-mail: Mohammad.Jahan-Parvar@frb.gov, Telephone No: (202) 475-6631.

§Department of Finance, Stockholm School of Economics and Swedish House of Finance, Drottninggatan 98, Stockholm SE-111 60, Sweden. e-mail: Romeo.Tedongap@hhs.se, Telephone No: +46-8-736 9143.

# 1 Introduction

Traditional modelling of financial time series critically relies on the assumption of conditional normality of returns. This assumption implies that conditional skewness and excess kurtosis should be equal to zero. However, empirical evidence is in sharp contrast to this assertion. Unconditionally, these moments prove not to be zero. Moreover, similar to the first two conditional moments, higher moments demonstrate considerable time variation as noted by Bekaert et al. (1998) and Ghysels et al. (2011). Thus, explicit modelling of conditional higher moments that allows for time variation, is necessary to avoid model misspecification.

Since the pioneering work of Hansen (1994), a number of researchers have proposed parametric models for conditional skewness. Examples of studies on the economic importance of conditional skewness in financial asset returns, its econometric modelling and its empirical applications include Harvey and Siddique (1999, 2000), Chen et al. (2001), Brännäs and Nordman (2003), Jondeau and Rockinger (2003), Patton (2004), León et al. (2005), Lanne and Saikkonen (2007), Grigoletto and Lisi (2009), Wilhelmsson (2009), Ghysels et al. (2011), Durham and Park (2013), Conrad et al. (2013), and Feunou et al. (2013). A number of studies investigate conditional kurtosis, among them Brooks et al. (2005) and Guidolin and Timmermann (2008). In this paper we focus on conditional skewness, or conditional asymmetry.

The existing research does not tell us which parametric conditional skewness model provides a better fit for the data. As noted by Kim and White (2004), this is partially due to the extreme sensitivity of traditional skewness and kurtosis measures to outliers.[1] They propose several robust measures for skewness and kurtosis. In previous work (Feunou et al. 2013), we found that conditional asymmetry in returns is related to the "relative semi-variance," defined as the upside variance minus the downside variance.[2] We showed that modelling downside risk is possible when a measure of skewness is explicitly incorporated in the model. We used Pearson's (1895) "mode skewness" as the measure of choice. Pearson's mode skewness is more robust to outliers than traditional skewness measures. Building upon and expanding on suggestions in Kim and White (2004) and Feunou et al. (2013), this paper fills the existing gap in the literature regarding model adequacy for parametric conditional skewness models.

Our objective is threefold. First, we establish through a proposition that the relative semi-variance,

[1] By traditional, we mean standardized third and fourth moments of a random variable.

[2] We define the "upside variance" as the variance of the returns conditional upon their realization above a certain threshold. Their variance conditional upon their realization below the same threshold is called "downside variance." Based on these two definitions, we define the difference between upside variance and downside variance as the "relative semi-variance."

divided by the total variance, is a measure of skewness that satisfies the properties proposed by Groeneveld and Meeden (1984) for any reasonable skewness measure. Second, we develop an intuitive and easy-to-implement method for non-parametric measurement of realized asymmetry. Third, based on this measure of realized asymmetry, we can test any parametric model for conditional skewness and provide a method of how to model the conditional skewness. We test a number of parametric models of conditional asymmetry with various functional and distributional assumptions. The testing procedure is based on Mincer and Zarnowitz (1969) regressions and is similar in spirit to the methodology developed by Chernov (2007) to close the realized-implied volatility predictive regression gap. We find that in addition to allowing time variation in the conditional asymmetry, we need to allow for a "leverage effect," but also for asymmetry-in-asymmetry to obtain the best characterization of the conditional skewness dynamics.[3] The most successful characterization of the conditional asymmetry shares several functional features with the celebrated exponential GARCH (EGARCH) model of Nelson (1991).

Ghysels et al. (2011) propose a methodology for modelling and estimating the conditional skewness based on a mixed data sampling (MIDAS) method of volatility estimation introduced and extensively studied by Ghysels et al. (2005, 2006, 2007), and Bowley's (1920) measure of skewness. Our work differs from Ghysels et al. in two important dimensions. First, we are interested in assessing the adequacy of different models of conditional asymmetry, while Ghysels et al. focus on a single model. Second, Ghysels et al. build their model of conditional asymmetry based on Bowley's (1920) robust coefficient of skewness. This measure is constructed using the inter-quantile ranges of the series investigated, while our measure is based on the difference between upside and downside semi-variances.

Durham and Park (2013) study the contribution of conditional skewness in a continuous-time framework. For tractability, they assume simple dynamics based on a single Lévy process for the conditional skewness in their estimated models. We model the conditional asymmetry in discrete time and assume much richer dynamics. Durham and Park establish the cost of ignoring conditional higher moments in modelling returns dynamics. Thus, their simple model is adequate to motivate their work. The economic relevance of conditional asymmetry has been established in several asset pricing studies. As pointed out by Christoffersen et al. (2006), conditionally non-symmetric return innovations are critically important, since in option pricing, for example, heteroskedasticity and the leverage effect alone do not suffice to explain the option smirk. In this study, we try to find the model that best characterizes the dynamics of the conditional asymmetry in S&P 500 returns in a rich set of models.

Jondeau and Rockinger (2003) characterize the maximal range of skewness and kurtosis for which

[3]We find that the evidence for asymmetry-in-asymmetry itself is relatively weak. But the flexibility offered by separating the contributions of positive and negative shocks improves the model's performance significantly.

a density exists. They claim that the generalized Student-$t$ distribution spans a large domain in this maximal set, and use this distribution to model innovations of a GARCH-type model with conditional parameters. They find time dependency of the asymmetry parameter, but a constant degree-of-freedom parameter in the series they study. They provide evidence that skewness is strongly persistent, but kurtosis is much less so. While influenced by Jondeau and Rockinger (2003), our study differs from their work in two important dimensions. First, we study a larger number of models and distributions than Jondeau and Rockinger (2003), and we therefore consider our study to be more comprehensive. Second, since we compare non-nested models, we rely on Mincer and Zarnowitz's (1969) methodology to investigate the adequacy of models.

The rest of the paper proceeds as follows. In section 2, we provide the theoretical background for our study. We discuss the implications of various distributional assumptions in section 3. Section 4 describes the different parametric model specifications for the conditional skewness that we test in our empirical analysis. We report our empirical findings in section 5. Section 6 concludes.

## 2 Theoretical Background

Conventional asymptotic theory in econometrics typically leads to limiting distributions for economic variables that are conditionally Gaussian as sample size increases. Examples of such work include Bollerslev et al. (1994) and Davidson (1994). Thus, conditional skewness should converge to zero as sample size increases. However, as Ghysels et al. (2011), Brooks et al. (2005), and Jondeau and Rockinger (2003) show, conditional skewness for many financial time series does not vanish in large samples or through sampling at higher frequencies. In what follows, we extend their findings and develop a testing framework to compare different parametric models of conditional skewness.

In general, common parametric distributions considered in empirical work to characterize the distribution of logarithmic returns are unimodal and satisfy the following conditions:

$$\begin{aligned} Var\left[r \mid r \geq m\right] > Var\left[r \mid r < m\right] &\Leftrightarrow Skew\left[r\right] > 0 \\ Var\left[r \mid r \geq m\right] = Var\left[r \mid r < m\right] &\Leftrightarrow Skew\left[r\right] = 0 \\ Var\left[r \mid r \geq m\right] < Var\left[r \mid r < m\right] &\Leftrightarrow Skew\left[r\right] < 0, \end{aligned} \tag{1}$$

where $m$ is a suitably chosen threshold. A few studies in the literature use distributions that explicitly allow for skewness in returns; in particular, the skewed generalized Student-$t$ distribution popularized by Hansen (1994), the skewed generalized error distribution of Nelson (1991), and the binormal distribution applied to financial data in our previous work (Feunou et al. 2013) all satisfy equation (1). We introduce a new measure of asymmetry, called the relative semi-variance (RSV), defined by the difference between the upside variance and the downside variance, where upside and downside are relative to a cut-off point equal to the mode.

**Proposition 2.1** *Let the random variable $x$ follow a unimodal distribution with mode $m$. Denote the upside variance as $\sigma_u^2 = Var[x|x \geq m]$ and the downside variance as $\sigma_d^2 = Var[x|x < m]$. We standardize the relative semi-variance $RSV \equiv \sigma_u^2 - \sigma_d^2$ by dividing it by the total variance to obtain a scale-invariant and dimensionless measure for skewness defined as*

$$\gamma(x) = \frac{\sigma_u^2 - \sigma_d^2}{\sigma^2}, \tag{2}$$

*where $\sigma^2 = Var[x]$ is the total variance. The distribution is right-skewed if $\sigma_u^2 > \sigma_d^2$, and left-skewed if $\sigma_u^2 < \sigma_d^2$. The proposed skewness measure is coherent; that is, it satisfies the three properties proposed by Groeneveld and Meeden (1984) that any reasonable skewness measure should satisfy.*[4] *These properties are:*

- *(P1) for any $a > 0$ and $b$, $\gamma(x) = \gamma(ax + b)$;*
- *(P2) if $x$ is symmetrically distributed, then $\gamma(x) = 0$;*
- *(P3) $\gamma(-x) = -\gamma(x)$.*

**Proof:** Note that, for any $a > 0$ and $b$, the mode of $ax + b$ is equal to $am + b$. Besides the total variance, the upside variance and the downside variance of $ax + b$ are given by

$$Var[ax + b] = Var[ax] = a^2 Var[x] = a^2\sigma^2,$$

$$Var[ax + b \mid ax + b \geq am + b] = Var[ax + b \mid x \geq m] = Var[ax \mid x \geq m] = a^2 Var[x \mid x \geq m] = a^2\sigma_u^2$$

$$Var[ax + b \mid ax + b < am + b] = Var[ax + b \mid x < m] = Var[ax \mid z < m] = a^2 Var[x \mid x < m] = a^2\sigma_d^2.$$

Thus, the skewness of $ax + b$ is given by

$$\begin{aligned}\gamma(ax + b) &= \frac{Var[ax + b \mid ax + b \geq am + b] - Var[ax + b \mid ax + b < am + b]}{Var[ax + b]} \\ &= \frac{a^2\sigma_u^2 - a^2\sigma_d^2}{a^2\sigma^2} = \frac{\sigma_u^2 - \sigma_d^2}{\sigma^2} = \gamma(x).\end{aligned}$$

The $\gamma(x)$ skewness measure thus satisfies (P1).[5]

To demonstrate that $\gamma(x)$ satisfies the second property, suppose that $x$ is symmetric and unimodal; then we know that the mode is equal to the mean. As a result, $x - m$ is symmetric and unimodal with mean zero. Consequently, $x - m$ and its opposite, $m - x$, have the same distribution. The upside variance of $x - m$ is equal to $\sigma_u^2$ and the downside variance of $x - m$ is equal to $\sigma_d^2$. However,

$$\sigma_d^2 = Var[x - m \mid x - m < 0] = Var[m - x \mid x - m < 0] = Var[m - x \mid m - x > 0].$$

[4]Suitability of $\gamma(x)$ as a skewness measure critically depends on the measure of volatility used in modelling the returns process. Later in the paper, we show that our results are based on the Engle and Ng (1993) NGARCH volatility model. Based on empirical results, we argue that NGARCH is a perfectly adequate volatility measure and, thus, our conditional skewness measures are well specified.

[5]This result means that relative semi-variance, $\sigma_u^2 - \sigma_d^2$, satisfies (P1) up to a multiplicative constant.

So, $\sigma_d^2$ is also the upside variance of $m - x$. Since $x - m$ and $m - x$ have the same distribution, then $\sigma_u^2 = \sigma_d^2$ and in consequence $\gamma(x) = 0$. This shows that our measure of skewness satisfies (P2).

To demonstrate that $\gamma(x)$ satisfies (P3), note that the mode of $-x$ is simply $-m$. The upside variance of $-x$ is thus the downside variance of $x$:

$$Var[-x \mid -x \geq -m] = Var[x \mid -x \geq -m] = Var[x \mid x \leq m] = \sigma_d^2.$$

Similarly, we can show that the downside variance of $-x$ is equal to $\sigma_u^2$, the upside variance of $x$. On the other hand, $-x$ and $x$ have the same total variance $\sigma^2$. Consequently, we have

$$\begin{aligned}\gamma(-x) &= \frac{Var[-x \mid -x \geq -m] - Var[-x \mid -x < -m]}{Var[-x]} \\ &= \frac{\sigma_d^2 - \sigma_u^2}{\sigma^2} = -\frac{\sigma_u^2 - \sigma_d^2}{\sigma^2} = -\gamma(x).\end{aligned}$$

Our skewness measure thus satisfies (P3).

## 2.1 Building a realized skewness measure

Based on our discussion in section 2, we posit that modelling conditional skewness or asymmetry is equivalent to modelling relative semi-variance, $RSV_t$. The literature on modelling and measuring volatility in finance and economics is vast. It suffices to say that, for years now, using realized variance following the methodology of Andersen et al. (2001, 2003) is the standard method for measuring volatility in financial time series. As an example, Chernov's (2007) study on the adequacy of option-implied volatility in forecasting future volatility crucially depends on this methodology. We modify this standard methodology in the literature to build a non-parametric and distribution-free measure for conditional asymmetry in returns. In a recent paper, Neuberger (2012) discusses a somewhat similar measure of realized skewness.

We construct our measures following the common practice in the realized variance literature by summing up finely sampled squared-return realizations over a fixed time interval, $RV_t = \sum_{j=1}^{n_t} r_{j,t}^2$, where there are $n_t$ high-frequency returns in period $t$, $r_{j,t}$ is the $j$th high-frequency return in period $t$.

We then construct the realized downside and upside variance series as

$$RV_t^d = \frac{n_t}{2n_t^d} \sum_{j=1}^{n_t} r_{j,t}^2 I(r_{j,t} < m_t) \quad \text{and} \quad RV_t^u = \frac{n_t}{2n_t^u} \sum_{j=1}^{n_t} r_{j,t}^2 I(r_{j,t} \geq m_t), \tag{3}$$

where $n_t^d$ and $n_t^u$ are, respectively, the number of high-frequency returns below and above the conditional mode of return $m_t$ in period $t$, and where $I(\cdot)$ denotes an indicator function. Thus, the measure for realized relative semi-variance is simply defined as

$$RRSV_t = RV_t^u - RV_t^d, \tag{4}$$

which, divided by realized variance, will define realized skewness according to our proposed measure for skewness introduced in Proposition 2.1. Realized volatility will refer to the square root of realized variance.

It is a well-known fact that $E_t\left[RV_{t+1}\right] = \sigma_t^2$, where $\sigma_t^2 = Var_t\left[r_{t+1}\right]$ is the conditional variance, and where $r_t = \sum_{j=1}^{n_t} r_{j,t}$ is the return of period $t$. This is a well-established result based on Corollary 1 in Andersen et al. (2003). We establish that $E_t\left[RRSV_{t+1}\right] \approx \sigma_{u,t}^2 - \sigma_{d,t}^2$, and the following proposition and its proof show the veracity of our assertion.

**Proposition 2.2** *Let the unidimensional continuous-price process $\{P_t\}_{t=0}^{T}$, where $T > 0$, be defined on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Let $\{\mathcal{F}\}_{t\in[0,T]} \subseteq \mathcal{F}$ be an information filtration, defined as a family of increasing $\mathbf{P}$-complete and right-continuous $\sigma$-fields. Information set $\mathcal{F}_t$ includes asset prices and relevant state variables through time $t$. Let $RV_{t+1}^u$ and $RV_{t+1}^d$ be defined as above for this price process. Then $E_t\left[RV_{t+1}^u - RV_{t+1}^d\right] \approx \sigma_{u,t}^2 - \sigma_{d,t}^2$.*

**Proof:** See the appendix.

Thus, a simple testing procedure consists of regressing $RV_{t+1}^u - RV_{t+1}^d$ on $\sigma_{u,t}^2 - \sigma_{d,t}^2$, or

$$RV_{t+1}^u - RV_{t+1}^d = \beta_0 + \beta_1\left(\sigma_{u,t}^2 - \sigma_{d,t}^2\right) + \varepsilon_{t+1}. \quad (5)$$

Following the standard Mincer and Zarnowitz (1969) methodology, we view the model with the $\beta_0$ closest to zero, the $\beta_1$ closest to one and the highest regression $R^2$ as the better model for conditional skewness.[6]

We have followed Barndorff-Nielsen et al. (2010) closely in our treatment of sources of conditional skewness. This implies that in the proof of the above proposition, and following Barndorff-Nielsen et al. (2010), we have shut down the “instantaneous” or “high-frequency leverage effect” to derive the desired results. In practice, we have shown that our realized relative semi-variance shares many features with the realized semi-variances studied by Barndorff-Nielsen et al. (2010).

We have observed, but have not studied this issue in depth, that in data sampled at high enough frequency, conditional skewness is driven purely by jumps. This is a reasonable assumption at frequencies such as 5-minute, 15-minute, half-hour, hourly, or even daily sampled data, since the instantaneous leverage effect is clearly weak at such high frequencies. However, as the sampling frequency is lowered to, for example, monthly or quarterly periods, this assertion loses power. The leverage effect is an important contributor to conditional skewness in lower sampling frequencies.

[6] In practice, we put less weight on the first condition, with $\beta_0$ statistically indistinguishable from 0. As Chernov (2007) documents this issue in a parallel literature, pinning down the correct functional form of the statistical relationship between latent variables is difficult. Thus, we focus on the more robust and theoretically more important relationship between our skewness measures through slope parameters. That said, we report results for joint tests for $\beta_0 = 0, \beta_1 = 1$ in our discussion of empirical findings.

Figure 1 shows the paths for daily realized relative semi-variance and realized variance for S&P 500 returns in the 1980–2010 period. Both series are constructed using 15-minute returns. In order to provide a more tractable picture of the behavior of realized relative semi-variance, we thin out the plot and use every 22nd data point in the figure. We study the correlation between realized relative semi-variance and realized variance and volatility. We find that the correlation between realized variance and relative semi-variance is -0.5076, the correlation between realized variance and skewness is -0.0206, the correlation between realized volatility and relative semi-variance is -0.4475, and finally, the correlation between realized volatility and skewness is -0.0538. Thus, RRSV and RV are negatively correlated. It is immediately clear that spikes in realized relative semi-variance typically lead to significant jumps in realized variance. It is also clear that there is evidence of clustering visible in relative semi-variance, particularly in the first half of the sampling period. The range of realized relative semi-variance is comparable to that reported in our previous work (Feunou et al. 2013).

# 3 Model Specification for the Conditional Skewness

In modeling the first two conditional moments, it is not necessary to take a stance on the parametric distribution of the returns. Unlike the first two conditional moments where no assumption on the parametric distribution is required, modelling higher moments requires a specification of a parametric distribution. Two flexible families of distributions attract a lot of attention in the literature. They are the skewed generalized Student-$t$ (GST) and the skewed generalized error distribution (SGED). We also study the binormal distribution, which we recently introduced to the finance literature (Feunou et al. 2013). Without loss of generality, we standardize these distributions by fixing their mean to be equal to zero, and their variance to be equal to one. Standardized returns are denoted by $z$.

## 3.1 The skewed generalized Student-$t$ distribution

Hansen (1994) popularized the skewed GST distribution. Its density is defined by

$$f^{GST}(z) = \begin{cases} bc\left(1+\frac{1}{\eta-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} & if \quad z < -a/b \\ bc\left(1+\frac{1}{\eta-2}\left(\frac{bz+a}{1+\lambda}\right)^2\right)^{-(\eta+1)/2} & if \quad z \geq -a/b \end{cases}$$

where $\lambda$ is the skewness parameter, $\eta$ represents degrees of freedom and

$$a \equiv 4\lambda c\frac{\eta-2}{\eta-1}, \quad b^2 \equiv 1+3\lambda^2-a^2, \quad c \equiv \frac{\Gamma\left(\left(\eta+1\right)/2\right)}{\sqrt{\pi\left(\eta-2\right)}\Gamma\left(\eta/2\right)}.$$

This density is defined for $2 < \eta < \infty$ and $-1 < \lambda < 1$. GST density nests a large set of conventional densities. For example, if $\lambda = 0$, Hansen's GST distribution reduces to the traditional Student-$t$

distribution. We recall that the traditional Student-$t$ distribution is not skewed. In addition, if $\eta = \infty$, the Student-$t$ distribution collapses to a normal density.

Since $\lambda$ controls skewness, if $\lambda$ is positive, the probability mass concentrates in the right tail. If it is negative, the probability mass is in the left tail. It is well known that the traditional Student-$t$ distribution with $\eta$ degrees of freedom admits all moments up to the $\eta$th. Therefore, given the restriction $\eta > 2$, Hansen's distribution is well defined and its second moment exists. The third and fourth moments of this distribution are defined as

$$\begin{aligned} E\left[z^3\right] &= \left(m_3 - 3am_2 + 2a^3\right)/b^3, \\ E\left[z^4\right] &= \left(m_4 - 4am_3 + 6a^2m_2 - 3a^4\right)/b^4, \end{aligned}$$

with $m_2 = 1 + 3\lambda^2$, $m_3 = 16c\lambda\left(1+\lambda^2\right)\frac{(\eta-2)^2}{(\eta-1)(\eta-3)}$ if $\eta > 3$, and $m_4 = 3\frac{\eta-2}{\eta-4}(1 + 10\lambda^2 + 5\lambda^4)$ if $\eta > 4$. The mode of the distribution is $-a/b$. Thus, the relative semi-variance is

$$Var\left[z|z > -a/b\right] - Var\left[z|z < -a/b\right] = \frac{4\lambda}{b^2}\left[1 - 4c^2\frac{(\eta-2)^2}{(\eta-1)^2}\right]. \tag{6}$$

We find that using relative semi-variance as a measure of skewness in the context of generalized Student-$t$ distribution adds flexibility to the analysis. Note that for the third moment to exist for a random variable with skewed GST distribution, $\eta$ must be greater than 3. However, we need only $\eta > 2$ for relative semi-variance to exist, which is the same condition for the existence of the second moment. Thus, it is possible to study asymmetry even when the third moment does not exist.

### 3.2 The skewed generalized error distribution

The probability density function for the SGED is

$$f^{SGED}(z) = C\exp\left(-\frac{|z+\delta|^\eta}{\left[1 + sign(z+\delta)\lambda\right]\theta^\eta}\right).$$

We define $C = \frac{\eta}{2\theta}\Gamma\left(\frac{1}{\eta}\right)^{-1}$, $\theta = \Gamma\left(\frac{1}{\eta}\right)^{\frac{1}{2}}\Gamma\left(\frac{3}{\eta}\right)^{-\frac{1}{2}}S(\lambda)^{-1}$, $\delta = 2\lambda AS(\lambda)^{-1}$, $S(\lambda) = \sqrt{1 + 3\lambda^2 - 4A^2\lambda^2}$, and $A = \Gamma\left(\frac{2}{\eta}\right)\Gamma\left(\frac{1}{\eta}\right)^{-\frac{1}{2}}\Gamma\left(\frac{3}{\eta}\right)^{-\frac{1}{2}}$, where $\Gamma(\cdot)$ is the gamma function. Scaling parameters $\eta$ and $\lambda$ are subject to $\eta > 0$ and $-1 < \lambda < 1$.

This density function nests a large set of conventional densities. For example, when $\lambda = 0$, we have the generalized error distribution, as in Nelson (1991). When $\lambda = 0$ and $\eta = 2$, we have the standard normal distribution; when $\lambda = 0$ and $\eta = 1$, we have the double exponential distribution; and when $\lambda = 0$ and $\eta = \infty$, we have the uniform distribution on the interval $\left[-\sqrt{3}, \sqrt{3}\right]$.

The parameter $\eta$ controls the height and the tails of the density function, and the skewness parameter $\lambda$ controls the rate of descent of the density around the mode $(-\delta)$. The third and the fourth moment are defined as

$$E\left[z^3\right] = A_3 - 3\delta - \delta^3,$$
$$E\left[z^4\right] = A_4 - 4A_3\delta + 6\delta^2 + 3\delta^4,$$

where $A_3 = 4\lambda\left(1+\lambda^2\right)\Gamma\left(4/\eta\right)\Gamma\left(1/\eta\right)^{-1}\theta^3$ and $A_4 = \left(1+10\lambda^2+5\lambda^4\right)\Gamma\left(5/\eta\right)\Gamma\left(1/\eta\right)^{-1}\theta^4$. The mode of this distribution is $-\delta$.

As a result, we find that the relative semi-variance is

$$Var\left[z|z > -\delta\right] - Var\left[z|z < -\delta\right] = \frac{4\lambda\left(1-A^2\right)}{S\left(\lambda\right)^2}. \tag{7}$$

### 3.3 The binormal distribution

The binormal distribution was introduced by Gibbons and Mylroie (1973). It is an analytically tractable distribution that accommodates empirically plausible values of skewness and kurtosis, and nests the familiar normal distribution.[7] In our previous work (Feunou et al. 2013), we showed that a GARCH model based on the binormal distribution, which we call BiN-GARCH, is quite successful in characterizing the elusive risk-return trade-off in the U.S. and international index returns. Our BiN-GARCH model explicitly links the market price of risk to conditional skewness.

The conditional density function of a standardized binormal distribution (SBin), or binormal distribution with zero mean unit variance, and Pearson mode skewness $\lambda$, is given by

$$f^{SBin}\left(z\right) = A\exp\left(-\frac{1}{2}\left(\frac{z+\lambda}{\nu_d}\right)^2\right)I\left(z < -\lambda\right) + A\exp\left(-\frac{1}{2}\left(\frac{z+\lambda}{\nu_u}\right)^2\right)I\left(z \geq -\lambda\right),$$

where $\nu_d = -\sqrt{\pi/8}\lambda + \sqrt{1-\left(3\pi/8-1\right)\lambda^2}$ and $\nu_u = \sqrt{\pi/8}\lambda + \sqrt{1-\left(3\pi/8-1\right)\lambda^2}$, and where $A = \sqrt{2/\pi}/\left(\nu_d+\nu_u\right)$. If $\lambda = 0$, then $\nu_d = \nu_u = 1$, and this distribution collapses to the familiar standard normal distribution.

We find that $-\lambda$ is the conditional mode, and up to a multiplicative constant, $\nu_d^2$ and $\nu_u^2$ are interpreted as downside variance and upside variance with respect to the mode, respectively. Specifically,

$$Var\left[z \mid z < -\lambda\right] = \left(1-\frac{2}{\pi}\right)\nu_d^2 \text{ and } Var\left[z \mid z \geq -\lambda\right] = \left(1-\frac{2}{\pi}\right)\nu_u^2. \tag{8}$$

We consider this property to be the most important characteristic of the binormal distribution, given our objectives. The existence and positivity of the quantities $\nu_d$ and $\nu_u$ impose a bound on the parameter $\lambda$, given by $|\lambda| \leq 1\Big/\sqrt{\pi/2-1} \approx 1.3236$. Finally, it is trivial to show that the relative semi-variance

[7] See Bangert et al. (1986), Kimber and Jeynes (1987), and Toth and Szentimrey (1990), among others, for examples of using the binormal distribution in data modelling, statistical analysis and robustness studies.

for the standardized binormal distribution is

$$Var\left[z|z \geq -\lambda\right] - Var\left[z|z < -\lambda\right] = \sqrt{2\pi}\left(1 - \frac{2}{\pi}\right)\lambda\sqrt{1 - \left(3\pi/8 - 1\right)\lambda^2}. \tag{9}$$

### 3.4 The skewness-kurtosis boundary

Let $\mu_3 = E(z_t^3)$ and $\mu_4 = E(z_t^4)$ denote the non-centered third and fourth moments of a random variable $\{z_t\}_{t=0}^{\infty}$. For any distribution on $z_t$ with $(-\infty, \infty)$ support, we have $\mu_3^2 < \mu_4 - 1$ with $\mu_4 > 0$ (see Widder 1946, p. 134, Theorem 12.a; Jondeau and Rockinger 2003).

This relation confirms that, for a given level of kurtosis, only a finite range of skewness may be spanned. This is known as the skewness-kurtosis boundary, and it ensures the existence of a density. Thus, the real challenge while modelling the distribution of $z_t$ is to get close enough to this skewness-kurtosis boundary.

Figure 2 shows the skewness-kurtosis boundary for Hansen's skewed generalized Student-$t$, SGED, and binormal distributions against the theoretical boundary discussed in Widder (1946) and Jondeau and Rockinger (2003). It is clear from this figure that the SGED spans a larger area of the theoretical skewness-kurtosis boundary than does the skewed generalized Student-$t$ distribution. Thus, we expect models based on the SGED to outperform models based on the GST distribution. The sharp limit on permissible skewness levels in the binormal distribution seriously limits its ability to span the theoretical skewness-kurtosis boundary.

Figures 3 and 4 demonstrate the contribution of skewness and peakedness parameters, $\lambda$ and $\eta$, respectively, to generating skewness and kurtosis in skewed generalized Student-$t$ distribution and SGED, respectively. Note that the patterns for the skewness surface in both distributions are very similar. The difference lies in the permissible values for $\eta$ and the level of skewness generated by comparable combinations of skewness and peakedness parameters. That is, the skewed generalized Student-$t$ seems to generate larger values for skewness in comparison with SGED. The pattern of the kurtosis surface, however, is different for these distributions. The skewed generalized Student-$t$ distributions demonstrate a more explosive pattern as we move toward the corners of the admissible set for $\lambda$. The behavior of the kurtosis surface for all admissible values of $\lambda$ is more subdued for SGED. Both distributions show mild evidence of asymmetry in kurtosis for lower values of $\eta$.

## 4 Model Specification

In this section, we introduce the functional forms of the models that we fit to the data for testing purposes. We first discuss how we make our models comparable. We want to estimate the parameters of interest without imposing restrictions on our estimation procedure; however, we also want to preserve the theoretical bounds imposed on the shape parameters. We use sign-preserving transformations. In particular, following Hansen (1994) and Nelson (1991), because of the different restrictions on the shape parameters, we map the transformed parameters to be estimated into the true parameters, using a logistic mapping for the skewness and an exponential mapping for the peakedness. This step allows us to estimate the transformed parameters $\tilde{\lambda}$ and $\tilde{\eta}$ as free values, and then recover the original parameters.

For the generalized Student-$t$ distribution, we use the mappings

$$\lambda = -1 + \frac{2}{1 + \exp\left(-\tilde{\lambda}\right)} \quad \text{and} \quad \eta = 2 + \exp\left(\tilde{\eta}\right). \tag{10}$$

These transformations are intuitive. Recall that skewed generalized Student-$t$ distribution requires that $|\lambda| < 1$. Transforming $\lambda$ following equation (10) ensures that these bounds are preserved, regardless of the estimated value of $\tilde{\lambda}$. Similarly, GST requires that $2 < \eta < \infty$. Equation (10) preserves these limits for $\eta$, regardless of the estimated value of $\tilde{\eta}$. For the skewed generalized error distribution, we have

$$\lambda = -1 + \frac{2}{1 + \exp\left(-\tilde{\lambda}\right)} \quad \text{and} \quad \eta = \exp\left(\tilde{\eta}\right), \tag{11}$$

so as to maintain the restrictions $|\lambda| < 1$ and $\eta > 0$.

Finally, for the standardized binormal distribution, we have

$$\lambda = \frac{1}{\sqrt{\frac{\pi}{2} - 1}} \left[-1 + \frac{2}{1 + \exp\left(-\tilde{\lambda}\right)}\right]. \tag{12}$$

Note that the binormal distribution does not have a distinct peakedness parameter. The transformation in equation (12) is to maintain the bound on Pearson mode skewness, $|\lambda| \leq 1 \Big/ \sqrt{\pi/2 - 1} \approx 1.3236$, for the binormal distribution discussed immediately after equation (8).

We estimate and compare results from up to nine specifications for skewness and peakedness factors, across the three distributions discussed in section 3. A total of 23 models fit to the data. In the most basic model, both of these factors are constant parameters to be estimated. We relax this specification and allow for time variation and functional complexity in these processes. We assume that the conditional variance process for returns follows an Engle and Ng (1993) NGARCH specification. Thus, the conditional variance process follows

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 \sigma_t^2 \left(z_{t+1} - \theta\right)^2 + \beta_2 \sigma_t^2. \tag{13}$$

This choice of functional form for the conditional variance allows for a "leverage effect" in returns. We assume that returns follow $r_{t+1} = \mu + \sigma_t z_{t+1}$, where $z_{t+1} \mid \mathcal{I}_t \sim GST\left(\lambda_t, \eta_t\right)$, $z_{t+1} \mid \mathcal{I}_t \sim SGED\left(\lambda_t, \eta_t\right)$, or $z_{t+1} \mid \mathcal{I}_t \sim SBiN(\lambda_t)$, and where $\mathcal{I}_t$ denotes the information set up to time $t$.[8]

The most basic model that we study assumes constant $\tilde{\lambda}$ and $\tilde{\eta}$. We call this model M0. We relax the assumption of time invariance for the skewness process in Model 1, but maintain that the peakedness process is still a parameter to be estimated. In model M1, we assume that the skewness process follows a symmetric ARCH(1) process,

$$\tilde{\lambda}_{t+1} = \kappa_0 + \kappa_1 z_{t+1}. \tag{14}$$

The assumption of symmetry here means that the arrival of good or bad news impacts the skewness process with the same magnitude.

Asymmetry in volatility is a well-documented feature of financial data. Several studies in the (G)ARCH literature address this issue, which leads to the leverage effect in financial data. Among these studies, we note Nelson (1991), Glosten et al. (1993), and Engle and Ng (1993). Jondeau and Rockinger (2003) argue that asymmetry in ARCH for skewness and kurtosis requires investigation. Hence, we study this issue in model M2, where we assume an Asym-ARCH structure in the skewness process, $\tilde{\lambda}$, but assume constant $\tilde{\eta}$. In this model, $\tilde{\lambda}$ follows

$$\tilde{\lambda}_{t+1} = \kappa_0 + \kappa_{1,+} z_{t+1} I(z_{t+1} > 0) + \kappa_{1,-} z_{t+1} I(z_{t+1} < 0), \tag{15}$$

where $I(\cdot)$ denotes an indicator function.

We allow for the richer GARCH(1,1) dynamics in the skewness process in model M3. In this model, we still assume constant $\tilde{\eta}$. $\tilde{\lambda}$ follows

$$\tilde{\lambda}_{t+1} = \kappa_0 + \kappa_1 z_{t+1} + \kappa_2 \tilde{\lambda}_t. \tag{16}$$

This model, except for distributional assumptions and normalization, is the same model studied by Harvey and Siddique (1999).

To study the potential existence of an asymmetry-in-asymmetry or leverage effect in the conditional skewness, we assume that the skewness process follows an asymmetric GARCH form in model M4. In this model, $\tilde{\eta}$ is still assumed to be constant. Asym-GARCH in $\tilde{\lambda}$ implies the following functional form:

$$\tilde{\lambda}_{t+1} = \kappa_0 + \kappa_{1,+} z_{t+1} I(z_{t+1} > 0) + \kappa_{1,-} z_{t+1} I(z_{t+1} < 0) + \kappa_2 \tilde{\lambda}_t. \tag{17}$$

In the remaining four models, we relax the assumption of constant $\tilde{\eta}$. In model M5, we assume a

[8] To keep our notation consistent with the realized volatility literature, our timing convention differs slightly from the familiar GARCH notation. Throughout the paper, the subscript $t$ on any variable means that it is observed exactly at time $t$. In the traditional GARCH notation, the subscript $t$ in the conditional variance means that it is the variance of the time $t$ returns. Hence, the variance is observed at time $t-1$.

symmetric ARCH structure for both $\tilde{\lambda}$ and $\tilde{\eta}$:

$$\tilde{\lambda}_{t+1} = \kappa_0 + \kappa_1 z_{t+1}, \tag{18}$$
$$\tilde{\eta}_{t+1} = \gamma_0 + \gamma_1 z_{t+1}.$$

In model M6, we study the impact of good and bad news on both skewness and peakedness processes by assuming an asymmetric ARCH in both $\tilde{\lambda}$ and $\tilde{\eta}$:

$$\tilde{\lambda}_{t+1} = \kappa_0 + \kappa_{1,+} z_{t+1} I(z_{t+1} > 0) + \kappa_{1,-} z_{t+1} I(z_{t+1} < 0), \tag{19}$$
$$\tilde{\eta}_{t+1} = \gamma_0 + \gamma_{1,+} z_{t+1} I(z_{t+1} > 0) + \gamma_{1,-} z_{t+1} I(z_{t+1} < 0).$$

Model M7 investigates the implications of assuming a GARCH specification for both skewness, $\tilde{\lambda}$, and peakedness, $\tilde{\eta}$, processes:

$$\tilde{\lambda}_{t+1} = \kappa_0 + \kappa_1 z_{t+1} + \kappa_2 \tilde{\lambda}_t, \tag{20}$$
$$\tilde{\eta}_{t+1} = \gamma_0 + \gamma_1 z_{t+1} + \gamma_2 \tilde{\eta}_t.$$

As with M3, this formulation is very similar to the model in Harvey and Siddique (1999).

Finally, we study the implications of assuming an asymmetric GARCH functional form for both $\tilde{\lambda}$ and $\tilde{\eta}$ in model M8:

$$\tilde{\lambda}_{t+1} = \kappa_0 + \kappa_{1,+} z_{t+1} I(z_{t+1} > 0) + \kappa_{1,-} z_{t+1} I(z_{t+1} < 0) + \kappa_2 \tilde{\lambda}_t \tag{21}$$
$$\tilde{\eta}_{t+1} = \gamma_0 + \gamma_{1,+} z_{t+1} I(z_{t+1} > 0) + \gamma_{1,-} z_{t+1} I(z_{t+1} < 0) + \gamma_2 \tilde{\eta}_t.$$

# 5 Estimation Results

## 5.1 Data

We use daily Standard and Poor's 500 (S&P 500) index excess returns from Thomson Reuters Datastream. The data series starts in January 1980 and ends in September 2010.

Table 1 reports summary statistics of the data. In Panel A, we report annualized return means and standard deviations, in percentages, in columns two and three. We report unconditional skewness in column four. We observe negative unconditional skewness for market returns. The value of unconditional skewness is not small relative to the average daily returns. Our data seem to be highly leptokurtotic, since the series demonstrates significant unconditional excess kurtosis, as seen in column five. The reported $p$-values of Jarque and Bera's (1980) normality test imply a significant departure from normality in the data. Our proxy for the risk-free rate is the yield of the 3-month constant-maturity U.S. Treasury bill, which we obtained from the Federal Reserve Bank of St. Louis FRED II data bank. The crash of

October 1987, the Asian crisis of 1997, the Russian default of 1998 and the 2007–09 Great Recession are represented in the data.

Our intraday data series comes from Olsen Financial Technologies and is their longest available one-minute close level S&P 500 index price series. The data span the period from February 1986 to September 2010. To reduce the market microstructure effect in our empirical results, we construct intraday returns at a 15-minute frequency. These results appear in Panel B of Table 1.

### 5.2 Discussion of the results

We report two sets of results for each distribution. First, we report estimation results for all models studied for each distribution. We then report the results from running the Mincer and Zarnowitz (1969) regressions of realized skewness measure on parametric skewness results. We report estimation results for the full sample (1980–2009) and three subsamples, spanning 1980–89, 1990–99, and 2000-09. For model selection, we mainly rely on empirical findings based on our full-sample estimates. Estimation results based on subsamples are generally for demonstration of time variation or robustness and play a secondary role.

In each case, for each distribution and for each sample period, we first identify the model that best characterizes the returns. Our strategy to identify such a model is to compute likelihood ratio (LR) test statistics. A model is considered viable when (i) the LR test rejects equal goodness of fit between a baseline normally distributed model and the model with conditional skewness, and (ii) the LR test rejects equal goodness of fit between each model and the model preceding it.[9] That is, we control for overparameterization of models by comparing models sequentially. The LR test is not applicable for non-nested models. In such cases, we do not compute LR test statistics. If the LR test does not differentiate between two models' fit or it is inapplicable, then we first look at the Bayesian information criteria (BIC) of models. If this is still not helpful, we prefer the less parameterized model over the more parameterized model with the same goodness of fit and similar BIC. Based on the estimated parameters of the preferred parametric model, we filter out the model-implied relative semi-variance process. Note that, in our study, model M8 nests all other models.[10]

[9] This means that, in our tables, the reported LR test statistics are computed by using log-likelihood values from model M$x$ and model M$x-1$. That is, the LR statistic for model M2 is based on log-likelihood values for models M1 and M2. When necessary, we indicate that we have used log-likelihood values from non-sequential but nested models to construct a test statistic.

[10] Our testing procedure is similar in spirit to the forward selection method used in econometrics for model selection through sequential significance testing. Thus, our procedure does not account for global significance level of tests. For nested models, this procedure is set up to reject restricted (null) against the unrestricted (alternative) models. Our goal is to provide a simple rule for model selection. One may consider a different rule which leads to a different model choice. For example, backward elimination, forward selection and stepwise

We then regress realized relative semi-variance onto model-implied relative semi-variance. Next, we turn to model evaluation. Our criteria for the success of a parametric model of skewness (relative semi-variance) here – in descending order of importance – are: (i) a slope parameter that is statistically indistinguishable from unity at 5 or 10% confidence levels, (ii) the highest possible $R^2$ given that the previous condition is met, and finally (iii) whether the slope and intercept coefficients of the Mincer-Zarnowitz regression in question are jointly equal to one and zero, respectively, given that the previous two conditions are met. The last condition indicates the difficulty of pinning down the correct functional form of a latent variable, in our case conditional skewness. Thus, we put less weight on intercept estimates that are statistically not different from zero. Chernov (2007) documents this issue for realized and implied volatility literature.

Tables 2 and 3 report the estimation results for models in section 4 for the GST distribution. Note that the baseline model (represented as $N$ in both tables) is rejected in favor of all alternative models entertained in our study, based on likelihood ratio tests that are not reported to save space. Models M0–M8 deliver a better fit for the data. Thus, the first criterion is met.

For model M0, we find that estimated peakedness parameters are statistically different from zero at the 5 percent level across all samples studied. However, the skewness parameter is statistically significant in the full sample and the 2000–09 subsample. We conclude that there is very strong evidence in favor of the existence of excess kurtosis and convincing evidence supporting skewness, which gets stronger as we use the 21st-century data.[11]

Models M1–M4 show a somewhat similar pattern. While the size of the estimated parameters may differ across these models and across samples, they are all significantly different from zero at the 5 percent level. Estimated autoregressive parameters, whether symmetric or asymmetric ($\kappa_{1+}$ or $\kappa_{1-}$), and the GARCH-like parameter $\kappa_2$ are generally statistically different from zero. In addition, we observe evidence of asymmetry-in-asymmetry in our results. That is, positive returns increase the skewness – or, in other words, push the skewness toward positive values, and negative returns decrease the skewness – or push skewness toward negative values. Thus, positive and negative news have different and opposite impacts on skewness. Based on the size of the estimated parameters, the impact of positive news is larger than that of negative news.

---

selection rules do not necessarily pick the same model. Thus, using a different rule leads to a different set of results. While we acknowledge that the global significance level may be lower than the level of the test for each single pair of models, this does not influence our choice of model, since our model selection rule does not account for this contingency.

[11] Please note that in Tables 2, 3, 5, 6, and 8, we abuse the notation to save space. That is, $\kappa_{1+}$ and $\gamma_{1+}$ represent both the coefficients for the ARCH-like terms and the coefficients for positive innovation shocks.

Models M5 to M8 in Tables 2 and 3 report the impact of time variation on the conditional kurtosis. In general, the GARCH-type parameter $\gamma_2$ and the ARCH-type parameter $\gamma_{1,+}$ are statistically significant at the 5 percent level. We observe evidence of an asymmetric impact in good and bad news for conditional kurtosis models. The estimated $\gamma_{1,-}$ parameter is statistically different from zero in the full sample, but this is not true in general for subsample results. In addition, the magnitude of estimated $\gamma_{1,+}$ is generally larger than that of $\gamma_{1,-}$ parameters. Thus, it seems that, in this case, good news is more important than bad news for conditional kurtosis dynamics.

The likelihood test statistics (henceforth LR statistics) show that the models studied differ on how well they fit the data. Overall, model M8, which allows for asymmetric GARCH-type dynamics for both $\tilde{\lambda}_t$ and $\tilde{\eta}_t$, performs the best based on LR tests. That is, based on LR tests carried out, M8 provides a better fit than models M0 to M7 in the full sample and the 1990–99 and 2000–09 subsamples for returns.

Table 4 reports the results from running the Mincer-Zarnowitz predictive regressions introduced in equation (5), where the right-hand-side variables are implied skewness measures from one of the models in equations (14)-(21) and the left-hand-side variable is the realized skewness measure introduced in equation (4). The RRSV measure is based on high-frequency information, while implied skewness measures are based on daily data. Note that model M1 provides the highest $R^2$ in full-sample estimation, but the estimated $\beta_1$ is significantly different from unity. However, the best model should be the one with the highest $R^2$ among all specifications with $\beta_1$ not significantly different from unity. Our results suggest that models M4 and M8 both have $\beta_1$ not statistically different from one, and the highest $R^2$ of 19 percent in full-sample estimation. Both these specifications have time-varying conditional skewness dynamics featuring asymmetry-in-asymmetry, but M8 nests M4 through the conditional kurtosis dynamics. For all models, the hypothesis $\beta_1 = 1$ is not rejected in the second subsample except for specifications M0 and M7, while the joint hypothesis $\beta_0 = 0$ and $\beta_1 = 1$ is not rejected in the third subsample except for specifications M0, M4 and M6. Our preferred model would then be M8 based on full-sample and subsample Mincer-Zarnowitz predictive regression results, added to the fact that M8 is significantly favored over M4 given the LR tests reported in Tables 2 and 3.

Tables 5 and 6 report the estimation results for models M0–M8 when errors are SGED. Similar to what is seen in the results for the GST case, models M0–M8 perform better than the baseline model N across the board. Models M0–M4 demonstrate statistically significant peakedness parameter estimates across models and samples. The statistical evidence in favor of asymmetry-in-asymmetry is quite strong in the full sample. Statistical evidence supporting asymmetry-in-asymmetry is more mixed in the subsamples.

Our estimation results for models M5–M8 imply that, again, evidence in support of asymmetry-in-asymmetry in the conditional skewness is quite strong for the full sample and in the subsamples. In addition, there is significant support for asymmetry in the conditional kurtosis, where good news reduces fat-tailedness. This is not surprising. Arrival of "good news" should reduce market uncertainty, and hence conditional kurtosis. Based on the LR statistics, again model M8 is the preferred model in the full sample and across the subsamples. It appears as if allowing for a rich, asymmetric parameterization adds to model flexibility and hence its performance. It is worth noting that model M8 shares many features with the celebrated Nelson (1991) EGARCH model. We find that in the 1990–99 and 2000–09 subsamples, models M6 and M8 are indistinguishable using LR tests. Based on BIC, model M6 is preferred to M8 in these subsamples. However, as mentioned earlier, we base our judgment mainly on full-sample results, where the LR test clearly picks M8 over M6 and M7.

Table 7 reports the results for Mincer-Zarnowitz predictive regressions where implied relative semi-variance estimates are based on models with SGED errors. In the full sample, models M0–M2, M5 and M8 all have slope parameters that are statistically indistinguishable from one. Joint tests of both slope and intercept parameters being different from one and zero, respectively, cannot be rejected for the above models at the 5 percent confidence level for the full sample, and models M1 and M5 provide the highest $R^2$ of 24 percent. These specifications have ARCH-type time-varying conditional skewness dynamics in common, but M5 nests M1 through the conditional kurtosis dynamics. However, neither of these two models is preferred in any of the subsamples, since the hypothesis $\beta_1 = 1$ is rejected at conventional levels of confidence. Again, as mentioned earlier, we base our judgment mainly on full-sample results, and model M1 would be the preferred specification, since it is more parsimonious and slightly favored over M5 based on LR tests and BIC values in Tables 5 and 6. However, note that the joint hypothesis $\beta_0 = 0$ and $\beta_1 = 1$ is not rejected for model M8, neither in the full sample with an $R^2$ of 14 percent nor in the second and third subsamples with $R^2$s of 7 percent and 16 percent, respectively. In the first subsample, M8 is also the preferred model, with an $R^2$ of 22 percent among all specifications where the hypothesis $\beta_1 = 1$ is not rejected. Then M8 would be the best model, should we extend our model-selection criteria to subsample regression results and the need for fitting the excess-returns data as well.

We report the estimation results for the binormal distribution model in Table 8. Since the standardized binormal distribution is a one-parameter distribution, it does not allow for independent estimation of conditional kurtosis. Thus, we estimate only models M0–M4 for the case of standardized binormal distributed errors. Similar to the previous cases, we find significant supporting evidence for asymmetry-in-asymmetry. Almost all estimated parameters are statistically different from zero. Based on the LR statistics, M3 is the preferred model for the full sample. It allows for symmetric GARCH-type dynamics in the conditional skewness measure. In the 1980–89 subsample, M3 remains the model of choice. In the 1990–99 subsample, M4 is tied with M2 based on the LR test, but M2 is preferred to M4 based on BIC. In the 2000–09 subsample, no model performs better than M0, which allows for constant $\lambda$ and $\eta$.

Table 9 reports model evaluation results through Mincer-Zarnowitz predictive regressions, for models M0–M4 with standardized binormally distributed errors. Across the full sample, models M1, M2 and M4 have estimated $\beta_1$ parameters that are statistically indistinguishable from unity. M1 and M2 have the highest $R^2$ of 25 percent for the full sample. Once we consider subsamples, models M1 and M4 demonstrate similar performance; both models do not reject the hypothesis $\beta_1 = 1$ in all subsamples except 1980–89. However, based on our results reported in Table 8, we know that model M1 is preferred over model M4 in fitting the excess-returns data. Thus, based on Mincer-Zarnowitz predictive regression considerations, we choose M1, which allows for asymmetric ARCH-type dynamics in the conditional skewness. Note, however, that although model M3 has an estimated $\beta_1$ parameter of 0.9262 that is statistically different from unity in the full sample, this value remains quite close to one. Furthermore, $\beta_1$ parameter estimates are not statistically different from one for model M3 in all subsamples. Given these observations, once we accommodate the need for fitting the excess-returns data as well, we may be tempted to choose M3 over M1 based on LR testing and similar predictive results.[12]

Comparing our results for the full sample across the three distributional assumptions, we observe the following: (i) GST-based estimation results imply that Model 8, which allows for asymmetry in both conditional skewness and kurtosis, performs better than alternative models studied in characterizing excess-returns dynamics and in Mincer-Zarnowitz predictive regressions of parametric onto non-parametric measures of relative downside semi-variances. (ii) SGED-based estimation results imply that if we want both reasonable characterization of excess-returns dynamics and strong predictive results for measuring RRSV, then our choice is M8, which allows for asymmetry in both conditional skewness and kurtosis. However, if we are interested only in predictive power for the RRSV, then M1 is the better choice. (iii) Binormal distribution-based estimation results, similar to the SGED case, lead us to choose the parsimonious M1 for predictive purposes only. Once we decide to characterize excess-returns dynamics as well, then we choose the more flexible M3, which allows for symmetric GARCH(1, 1)-type dynamics in the skewness process.

[12] $\chi^2_2$ critical values at 5 percent and 1 percent confidence levels are 5.99 and 9.21, respectively, while the computed LR test statistic is over 11.

We derive our theoretical framework from the literature on realized volatility. Thus, in our testing procedures, we follow that literature closely in that we are concerned with in-sample fitting and not out-of-sample forecast ability. Thus, it appears that, strictly speaking, model M8 with SGED errors provides estimated $\beta_1$s that are statistically close to unity, delivers reasonable $R^2$s and provides very good in-sample characterization of excess-returns dynamics. We conclude that a model that allows for asymmetric GARCH(1, 1)-type dynamics and a SGED is the overall preferred model. In addition, this model provides strong statistical support for asymmetry-in-asymmetry and asymmetry in conditional kurtosis.

In quite a few cases, we could not reject a statistically significant positive difference between intercept parameters $\beta_0$ and 0. While the upward bias observed across models and distributions may seem undesirable at first glance, we keep in mind that the vast literature on the information content of implied volatility faces the same type of problem. Chernov (2007) addresses the bias issue by adding a suitable predictor to the right-hand side of the predictive regression. It is possible that, by addressing the possibility of a missing variable, predictive regressions may improve. However, since M8 coupled with an SGED works well and is largely free of this bias, it is quite possible that this bias is due to distributional assumptions. We will address this interesting issue in future research.

# 6 Conclusions

In this paper, we address an important gap in the recent and growing literature on conditional skewness. In recent years, an increasing number of studies have addressed the importance of modelling higher moments and how their dynamics affect financial risk-management procedures. We propose a methodology to assess the model adequacy of the growing number of proposed parametric measures of the conditional skewness.

We show that, theoretically, conditional skewness matters when a series demonstrates significant relative semi-variance. We then propose a simple and intuitive non-parametric measure of relative semi-variance, which we call realized relative semi-variance, in the spirit of the successful literature on realized variance. We show that, under mild regularity conditions, parametric models of conditional asymmetry approximate the realized relative semi-variance. We also show that our proposed measure and methodology have close ties to the literature on realized variance, on the one hand, and to the recent developments on realized semi-variance on the other.

We then study several parametric models of conditional skewness. Since the performance of the parametric models of skewness crucially depends on the distributional assumptions, we study a range of relevant distributional models, including the skewed generalized Student-$t$ of Hansen (1994) and Jondeau and Rockinger (2003), skewed GED, and binormal distribution in our previous work (Feunou et al. 2013). We find statistically significant evidence in support of asymmetry-in-asymmetry, or the so-called leverage effect in the conditional skewness, and strong support for asymmetric (G)ARCH-type dynamics for both skewness and peakedness processes in the conditional distribution. We conclude that, based on our results, the best model for conditional skewness is one that allows for asymmetric GARCH(1,1) dynamics and admits skewed GED errors, applied to both skewness and peakedness processes in the conditional distribution of returns. This model shares many features with the well-known EGARCH model of Nelson (1991).

# References

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Bangert, U., Goodhew, P. J., Jeynes, C., Wilson, I. H., 1986. Low Energy (2-5 keV) Argon Damage in Silicon. Journal of Physics D: Applied Physics 19, 589–603.

Barndorff-Nielsen, O. E., Kinnebrock, S., Shephard, N., 2010. Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle. Oxford University Press, Ch. Measuring downside risk: realised semivariance, pp. 117–136.

Bekaert, G., Erb, C., Harvey, C. R., Viskanta, T., 1998. Distributional Characteristics of Emerging Market Returns and Asset Allocation. Journal of Portfolio Management, 102–116.

Bollerslev, T., Engle, R. F., Nelson, D. B., 1994. ARCH Models. Vol. 4 of The Handbook of Econometrics. North-Holland, Amsterdam, pp. 2959–3038.

Bowley, A., 1920. Elements of Statistics. Charles Scribners Sons, New York, NY.

Brännäs, K., Nordman, N., 2003. Conditional skewness modelling for stock returns. Applied Economics Letters 10 (11), 725–728.

Brooks, C., Burke, S. P., Heravi, S., Persand, G., 2005. Autoregressive Conditional Kurtosis. Journal of Financial Econometrics 3 (3), 399–421.

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.

Chernov, M., 2007. On the Role of Risk Premia in Volatility Forecasting. Journal of Business & Economic Statistics 25 (4), 381–411.

Christoffersen, P., Heston, S., Jacobs, K., 2006. Option valuation with conditional skewness. Journal of Econometrics 131 (12), 253–284.

Conrad, J., Dittmar, R., Ghysels, E., 2013. Ex-Ante Skewness and Expected Stock Returns. Journal of Finance 68 (1), 85–124.

Davidson, J., 1994. Stochastic Limit Theory, An Introduction for Econometricians, 1st Edition. Advanced Texts in Econometrics. Oxford University Press, Oxford, U.K.

Durham, G. B., Park, Y.-H., 2013. Beyond Stochastic Volatility and Jumps in Returns and Volatility. Journal of Business and Economic Statistics 31 (1), 107–121.

Engle, R. F., Ng, V. K., 1993. Measuring and Testing the Impact of News on Volatility. Journal of Finance 48 (5), 1749–1778.

Feunou, B., Jahan-Parvar, M. R., Tédongap, R., 2013. Modeling Market Downside Volatility. Review of Finance 17 (1), 443–481.

Ghysels, E., Plazzi, A., Valkanov, R., 2011. Conditional Skewness of Stock Market Returns in Developed and Emerging Markets and its Economic Fundamentals. Working Paper, Kenan-Flagler Business School-UNC, and Rady School of Business-UCSD.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a Risk-Return Trade-Off After All. Journal of Financial Economics 76, 509–548.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. Journal of Econometrics 131, 59–95.

Ghysels, E., Sinko, A., Valkanov, R., 2007. MIDAS regressions: further results and new directions. Econometric Reviews 26 (1), 53–90.

Gibbons, J. F., Mylroie, S., 1973. Estimation of Impurity Profiles in Ion-Implanted Amorphous Targets Using Joined Half-Gaussian Distributions. Applied Physics Letters 22, 568–569.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the Relation Between Expected Value and the Volatility of the Nominal Excess Return on Stocks. Journal of Finance 48 (5), 1779–1801.

Grigoletto, M., Lisi, F., 2009. Looking for skewness in financial time series. Econometrics Journal 12 (2), 310–323.

Groeneveld, R., Meeden, G., 1984. Measuring skewness and kurtosis. The Statistician 33, 391–399.

Guidolin, M., Timmermann, A., 2008. International Asset Allocation under Regime Switching, Skew and Kurtosis Preferences. Review of Financial Studies 21, 889–935.

Hansen, B. E., 1994. Autoregressive Conditional Density Estimation. International Economic Review 35 (3), 705–730.

Harvey, C. R., Siddique, A., 1999. Autoregressive Conditional Skewness. Journal of Financial and Quantitative Analysis 34, 465–488.

Harvey, C. R., Siddique, A., 2000. Conditional Skewness in Asset Pricing Tests. Journal of Finance 55, 1263–1295.

Jarque, C. M., Bera, A. K., 1980. Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. Economics Letters 6 (3), 255–259.

Jondeau, E., Rockinger, M., 2003. Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements. Journal of Economic Dynamics and Control 27, 1699–1737.

Kim, T.-H., White, H., 2004. On More Robust Estimation of Skewness and Kurtosis. Finance Research Letters 1 (1), 56–73.

Kimber, A. C., Jeynes, C., 1987. An Application of the Truncated Two-Piece Normal Distribution to the Measurement of Depths of Arsenic Implants in Silicon. Journal of the Royal Statistical Society: Series B 36, 352–357.

Lanne, M., Saikkonen, P., 2007. Modeling conditional skewness in stock returns. European Journal of Finance 13 (8), 691–704.

León, A., Rubio, G., Serna, G., 2005. Autoregressive conditional volatility, skewness and kurtosis. The Quarterly Review of Economics and Finance 45, 599–618.

Mincer, J., Zarnowitz, V., 1969. Jacob Mincer (editor), Economic Forecasts and Expectations. National Bureau of Economic Research, New York, NY, Ch. The Evaluation of Economic Forecasts.

Nelson, D. B., 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica 59, 347–370.

Neuberger, A., 2012. Realized skewness. Review of Financial Studies 25 (11), 3423–3455.

Patton, A. J., 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. Journal of Financial Econometrics 2 (1), 130–168.

Pearson, K., 1895. Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. Philosophical Transactions of the Royal Society of London. A 186, 343–414.

Protter, P., 1992. Stochastic Integration and Differential Equations. Springer-Verlag, Berlin.

Toth, Z., Szentimrey, T., 1990. The Binormal Distribution: A Distribution for Representing Asymmetrical but Normal-Like Weather Elements. Journal of Climate 3 (1), 128–137.

Widder, D., 1946. The Laplace Transform. Princeton University Press, Princeton, NJ.

Wilhelmsson, A., 2009. Value at risk with time varying variance, skewness and kurtosis–the NIG-ACD model. Econometrics Journal 12 (1), 82–104.

# Appendix:

## Proof of Proposition (2.2)

The set-up is similar to that of Andersen et al. (2003) (henceforth, ABDL (2003)). Proposition 1 of ABDL (2003) permits a unique canonical decomposition of the logarithmic asset price process $p = (p(t))_{t\in[0,T]}$,

$$p(t) = p(0) + A(t) + M(t),$$

where $A$ is a finite variation and predictable mean component, and $M$ is a local martingale.

Let $r(t,h) = p(t) - p(t-h)$ denote the continuously compounded return over $[t-h,\ t]$. The cumulative returns process from $t = 0$ onward, $r = (r(t))_{t\in[0,T]}$, is then $r(t) \equiv r(t,t) = p(t) - p(0) = A(t) + M(t)$.

Further assume that the mean process, $\{A(s) - A(t)\}_{s\in[t,t+h]}$, conditional on information at time $t$ is a predetermined function over $[t, t+h]$.

An immediate result of this assumption is the Corollary 1, and hence equation (6) of ABDL (2003), which establishes that

$$Var\left(r(t+h,h)|\mathcal{F}_t\right) = E\left([r,r]_{t+h} - [r,r]_t\,|\mathcal{F}_t\right).$$

In other words, the conditional variance equals the conditional expectation of the quadratic variation of the returns process.

Denote the mode of the distribution of $r(t+h,h)$ conditional on $\mathcal{F}_t$ by $m(t,h)$. Let $i(t+h,h) = 1_{[r(t+h,h)\geq m(t,h)]}$ denote an indicator random process that takes 1 if the returns between $[t,\ t+h]$ are greater than or equal to the conditional mode $m(t,h)$.

Since the mean process, $\{A(s) - A(t)\}_{s\in[t,t+h]}$, conditional on information at time $t$, is a predetermined function over $[t, t+h]$, without loss of generality, we omit $A$ from this point onwards.

We thus have

$$\begin{aligned}
\sigma_u^2(t,h) &\equiv Var\left(r(t+h,h)|\mathcal{F}_t,\ i(t+h,h)=1\right) \\
&= Var\left(M(t+h)-M(t)|\mathcal{F}_t,\ i(t+h,h)=1\right) \\
&= Var\left(M(t+h)|\mathcal{F}_t,\ i(t+h,h)=1\right) \\
&= E\left(M(t+h)^2|\mathcal{F}_t,\ i(t+h,h)=1\right)-\left\{E\left(M(t+h)|\mathcal{F}_t,\ i(t+h,h)=1\right)\right\}^2 \\
&= E\left([M,M]_{t+h}|\mathcal{F}_t,\ i(t+h,h)=1\right)-\left\{E\left(M(t+h)|\mathcal{F}_t,\ i(t+h,h)=1\right)\right\}^2 \\
&= E\left(\frac{[M,M]_{t+h}\, i(t+h,h)}{\pi(t,h)}|\mathcal{F}_t\right)-\left\{E\left(\frac{M(t+h)i(t+h,h)}{\pi(t,h)}|\mathcal{F}_t\right)\right\}^2,
\end{aligned}$$

where $\pi(t,h)\equiv \Pr\left[i(t+h,h)=1|\mathcal{F}_t\right]$.

Using Corollary 3 in Chapter II.6 of Protter (1992), where it is shown that $E\left(M(t+h)^2\right)=E\left([M,M]_{t+h}\right)$, we have

$$\begin{aligned}
\sigma_d^2(t,h) &\equiv Var\left(r(t+h,h)|\mathcal{F}_t,\ i(t+h,h)=1\right) \\
&= E\left(\frac{[M,M]_{t+h}\left(1-i(t+h,h)\right)}{1-\pi(t,h)}|\mathcal{F}_t\right)-\left\{E\left(\frac{M(t+h)\left(1-i(t+h,h)\right)}{1-\pi(t,h)}|\mathcal{F}_t\right)\right\}^2,
\end{aligned}$$

and

$$\begin{aligned}
&\sigma_u^2(t,h)-\sigma_d^2(t,h) \\
=\ & E\left(\frac{[M,M]_{t+h}\, i(t+h,h)}{\pi(t,h)}-\frac{[M,M]_{t+h}\left(1-i(t+h,h)\right)}{1-\pi(t,h)}|\mathcal{F}_t\right) \\
&+\left\{E\left(\frac{M(t+h)\left(1-i(t+h,h)\right)}{1-\pi(t,h)}|\mathcal{F}_t\right)\right\}^2-\left\{E\left(\frac{M(t+h)i(t+h,h)}{\pi(t,h)}|\mathcal{F}_t\right)\right\}^2
\end{aligned}$$

$$\begin{aligned}
&\sigma_u^2(t,h)-\sigma_d^2(t,h) \\
=\ & \frac{1}{\pi(t,h)\left(1-\pi(t,h)\right)}E\left([M,M]_{t+h}\left(i(t+h,h)-\pi(t,h)\right)|\mathcal{F}_t\right) \\
&-\frac{1}{\pi(t,h)\left(1-\pi(t,h)\right)}E\left(M(t+h)\left(i(t+h,h)-\pi(t,h)\right)|\mathcal{F}_t\right)\times \\
&\left(E\left(M(t+h)\left(\frac{\pi(t,h)-\pi(t,h)\,i(t+h,h)+i(t+h,h)-i(t+h,h)\,\pi(t,h)}{\pi(t,h)\left(1-\pi(t,h)\right)}\right)|\mathcal{F}_t\right)\right).
\end{aligned}$$

Since $E\left[i(t+h,h)|\mathcal{F}_t\right]=\pi(t,h)$, we have

$$\begin{aligned}
&\sigma_u^2(t,h)-\sigma_d^2(t,h) \\
=\ & \frac{1}{\pi(t,h)\left(1-\pi(t,h)\right)}E\left(\left([M,M]_{t+h}-[M,M]_t\right)\left(i(t+h,h)-\pi(t,h)\right)|\mathcal{F}_t\right) \\
&-\frac{1}{\pi(t,h)\left(1-\pi(t,h)\right)}E\left(\left(M(t+h)-M(t)\right)\left(i(t+h,h)-\pi(t,h)\right)|\mathcal{F}_t\right)\times \\
&\left(E\left(M(t+h)\left(\frac{\left(1-2\pi(t,h)\right)}{\pi(t,h)\left(1-\pi(t,h)\right)}\left(i(t+h,h)-\pi(t,h)\right)+2\right)|\mathcal{F}_t\right)\right)
\end{aligned}$$

$$\begin{aligned}
&\sigma_u^2(t,h)-\sigma_d^2(t,h)\\
=\;&\frac{1}{\pi(t,h)(1-\pi(t,h))}E\left(\left([M,M]_{t+h}-[M,M]_t\right)(i(t+h,h)-\pi(t,h))|\mathcal{F}_t\right)\\
&-\frac{1}{\pi(t,h)(1-\pi(t,h))}E\left((M(t+h)-M(t))(i(t+h,h)-\pi(t,h))|\mathcal{F}_t\right)\times\\
&\left(\frac{(1-2\pi(t,h))}{\pi(t,h)(1-\pi(t,h))}E\left((M(t+h)-M(t))(i(t+h,h)-\pi(t,h))|\mathcal{F}_t\right)+2M(t)\right)
\end{aligned}$$

$$\begin{aligned}
&\sigma_u^2(t,h)-\sigma_d^2(t,h)\\
=\;&\frac{1}{\pi(t,h)(1-\pi(t,h))}E\left(\left([r,r]_{t+h}-[r,r]_t\right)(i(t+h,h)-\pi(t,h))|\mathcal{F}_t\right)\\
&-\frac{1}{\pi(t,h)(1-\pi(t,h))}E\left(r(t+h,h)(i(t+h,h)-\pi(t,h))|\mathcal{F}_t\right)\times\\
&\left(\frac{(1-2\pi(t,h))}{\pi(t,h)(1-\pi(t,h))}E\left(r(t+h,h)(i(t+h,h)-\pi(t,h))|\mathcal{F}_t\right)+2M(t)\right)
\end{aligned}$$

$$\begin{aligned}
&(\pi(t,h)(1-\pi(t,h)))\left(\sigma_u^2(t,h)-\sigma_d^2(t,h)\right)\\
=\;&E\left(\left([r,r]_{t+h}-[r,r]_t\right)(i(t+h,h)-\pi(t,h))|\mathcal{F}_t\right)\\
&-E\left(r(t+h,h)(i(t+h,h)-\pi(t,h))|\mathcal{F}_t\right)\times\\
&\left(\frac{(1-2\pi(t,h))}{\pi(t,h)(1-\pi(t,h))}E\left(r(t+h,h)(i(t+h,h)-\pi(t,h))|\mathcal{F}_t\right)+2M(t)\right)
\end{aligned}$$

$$\begin{aligned}
&(\pi(t,h)(1-\pi(t,h)))\left(\sigma_u^2(t,h)-\sigma_d^2(t,h)\right)\\
=\;&cov\left([r,r]_{t+h}-[r,r]_t\,,i(t+h,h)|\mathcal{F}_t\right)\\
&-cov\left(r(t+h,h),i(t+h,h)|\mathcal{F}_t\right)\times\\
&\left(\frac{(1-2\pi(t,h))}{\pi(t,h)(1-\pi(t,h))}cov\left(r(t+h,h),i(t+h,h)|\mathcal{F}_t\right)+2M(t)\right).
\end{aligned}$$

Hence, two components drive the conditional skewness. The first component is the conditional covariance between returns and the indicator variable that shows whether the market has an upward or downward movement. The second component is the conditional covariance between the quadratic variation of returns and the same indicator variable. From this point on, we focus on the latter component, since, as we show later in this appendix, it is driven by jumps in the returns. The former component is the skewness induced by the instantaneous correlation between log-returns and volatility:

$$\sigma_u^2(t,h)-\sigma_d^2(t,h)\approx\frac{1}{\pi(t,h)(1-\pi(t,h))}cov\left([r,r]_{t+h}-[r,r]_t\,,i(t+h,h)|\mathcal{F}_t\right).$$

Let $y(t+h,h) \equiv \left(y^{(1)}(t+h,h), y^{(2)}(t+h,h)\right)' \equiv \left([r,r]_{t+h} - [r,r]_t\,,\ i\,(t+h,h)\right)'$.

From Corollary 1 of ABDL (2003), we have

$$cov\left([r,r]_{t+h} - [r,r]_t\,, i\,(t+h,h)\,|\mathcal{F}_t\right) = E\left(\left[y^{(1)}, y^{(2)}\right]_{t+h} - \left[y^{(1)}, y^{(2)}\right]_t |\mathcal{F}_t\right).$$

In other words, the conditional covariance between quadratic return variation and the upside indicator equals the conditional expectation of the quadratic covariation between the quadratic return variation and the upside indicator.

Proposition 2 of ABDL (2003) provides a consistent estimator of $\left[y^{(1)}, y^{(2)}\right]_t$.

Recall that proposition: For an increasing sequence of random partitions of $[0,T]$, $0 = \tau_{m,0} \leq \tau_{m,1} \leq \cdots$, such that $\sup_{j\geq 1}(\tau_{m,j+1} - \tau_{m,j}) \to 0$ and $\sup_{j\geq 1}\tau_{m,j} \to T$ for $m \to \infty$ with probability one, we have

$$\lim_{m\to\infty}\left\{\sum_{j\geq 1}\left[y^{(1)}\left(t\wedge\tau_{m,j}\right) - y^{(1)}\left(t\wedge\tau_{m,j-1}\right)\right]\left[y^{(2)}\left(t\wedge\tau_{m,j}\right) - y^{(2)}\left(t\wedge\tau_{m,j-1}\right)\right]\right\} \to \left[y^{(1)}, y^{(2)}\right]_t,$$

and thus we have

$$\lim_{m\to\infty}\left\{\sum_{j=\underline{\mathrm{j}}_t}^{\bar{j}_{t+h}}\left([r,r]_{\tau_{m,j}} - [r,r]_{\tau_{m,j-1}}\right)\left(i\left(\tau_{m,j}\right) - i\left(\tau_{m,j-1}\right)\right)\right\} \to \left[y^{(1)}, y^{(2)}\right]_{t+h} - \left[y^{(1)}, y^{(2)}\right]_t,$$

where $\underline{\mathrm{j}}_t \equiv \inf_{\tau_{m,j}\geq t}(j)$ , $\bar{j}_{t+h} \equiv \sup_{\tau_{m,j}\leq t+h}(j)$ and $i\,(t) \equiv i\,(t,t) = 1_{[r(t)\geq\delta(t,0)]}$.

Thus, $\sum_{j=\underline{\mathrm{j}}_t}^{\bar{j}_{t+h}}\left([r,r]_{\tau_{m,j}} - [r,r]_{\tau_{m,j-1}}\right)\left(i\left(\tau_{m,j}\right) - i\left(\tau_{m,j-1}\right)\right)$ is a measure of realized skewness.

We now simplify the derived realized skewness expression

$$[r,r]_{\tau_{m,j}} - [r,r]_{\tau_{m,j-1}} \approx \left(p\left(\tau_{m,j}\right) - p\left(\tau_{m,j-1}\right)\right)^2,$$

and we can show that

$$i\left(\tau_{m,j}\right) - i\left(\tau_{m,j-1}\right) = 1_{[p(\tau_{m,j})-p(\tau_{m,j-1})\geq 0]} - 1_{[p(\tau_{m,j})-p(\tau_{m,j-1})\leq 0]},$$

so that the measure of realized skewness becomes

$$\sum_{j=\underline{\mathrm{j}}_t}^{\bar{j}_{t+h}}\left(p\left(\tau_{m,j}\right) - p\left(\tau_{m,j-1}\right)\right)^2\left[1_{[p(\tau_{m,j})-p(\tau_{m,j-1})\geq 0]} - 1_{[p(\tau_{m,j})-p(\tau_{m,j-1})\leq 0]}\right].$$

Thus, using Barndorff-Nielsen et al.'s (2010) (henceforth, BKS (2010)) notation, we have shown that the realized skewness is the difference between the upside realized semi-variance ($RS^{+}$) and the downside realized semi-variance ($RS^{-}$):

$$
\begin{aligned}
RS^{+}(t+h,h) &= \sum_{j=\underline{j}_t}^{\bar{j}_{t+h}} \left(p(\tau_{m,j}) - p(\tau_{m,j-1})\right)^2 1_{p(\tau_{m,j})-p(\tau_{m,j-1})\geq 0} \\
RS^{-}(t+h,h) &= \sum_{j=\underline{j}_t}^{\bar{j}_{t+h}} \left(p(\tau_{m,j}) - p(\tau_{m,j-1})\right)^2 1_{p(\tau_{m,j})-p(\tau_{m,j-1})\leq 0}.
\end{aligned}
$$

To recap, we have

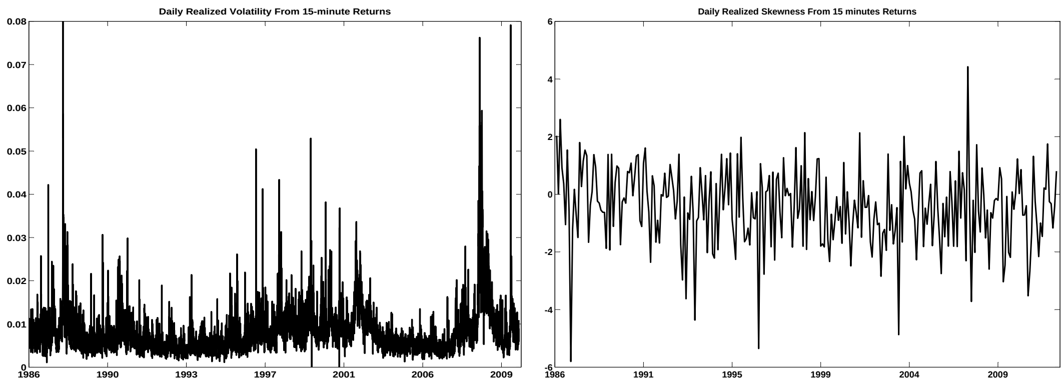
$$
\sigma_u^2(t,h) - \sigma_d^2(t,h) \approx \frac{1}{\pi(t,h)(1-\pi(t,h))} E\left(RS^{+}(t+h,h) - RS^{-}(t+h,h) \,|\mathcal{F}_t\right).
$$

Further using BKS (2010) results, we can provide more insight into the sources of conditional skewness. BKS (2010) show that

$$
RS^{+}(t+h,h) - RS^{-}(t+h,h) \to \sum_{t}^{t+h} (\Delta p_s)^2 \left[1_{\Delta p_s \geq 0} - 1_{\Delta p_s \leq 0}\right],
$$

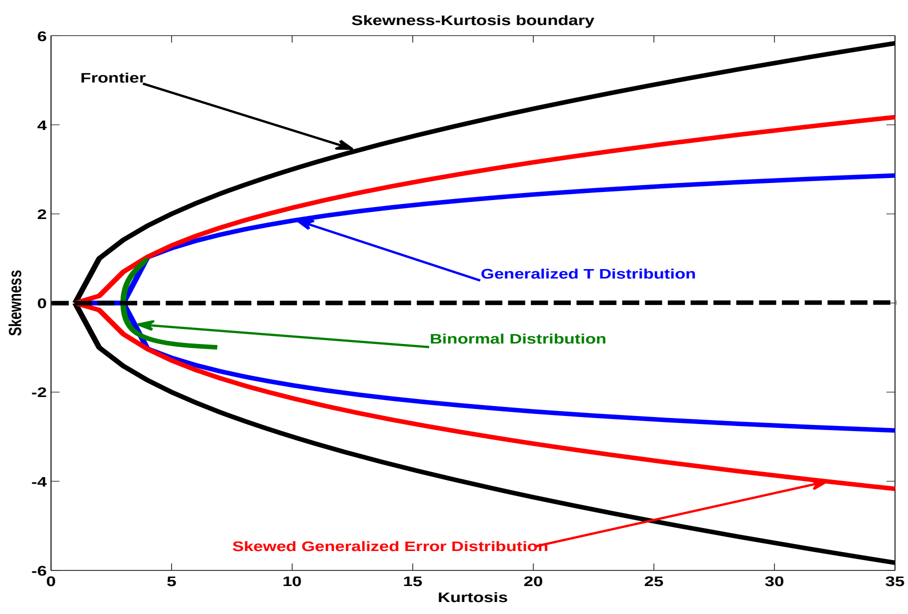
where $\Delta p_s \equiv p(s) - p(s-)$ is the jump at time $s$. Hence positive skewness is driven by the fact that positive jump amplitudes are higher than negative amplitudes, whereas negative skewness is driven by the fact that negative jump amplitudes are higher than positive amplitudes.

Figure 1: Realized volatility and realized relative semi-variance for S&P 500 returns, 1986–2009



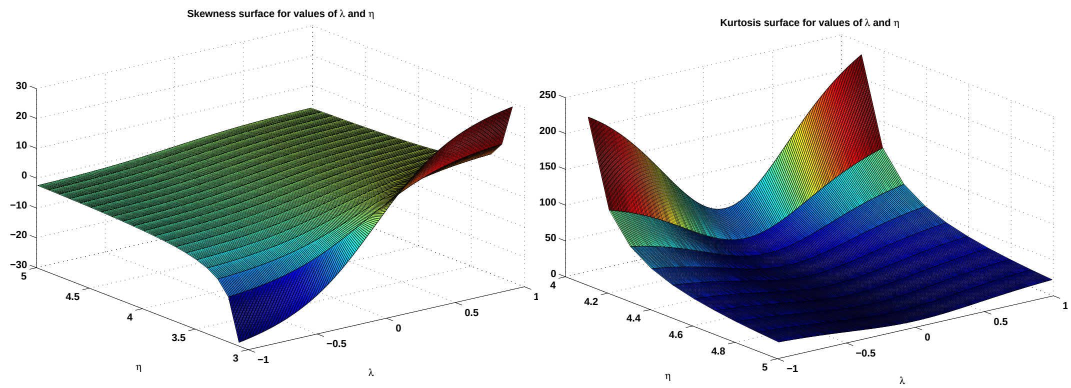

In this figure we plot realized volatility and realized skewness (realized relative semi-variance) for S&P 500 returns, sampled at 15-minute frequency. To provide a more informative illustration, we plot daily realized volatilities and the end of the month realized skewness (every 22nd observation is plotted) values.

Figure 2: Skewness-Kurtosis Boundary for Skewed Generalized Student-$t$, Skewed GED and Binormal Distributions



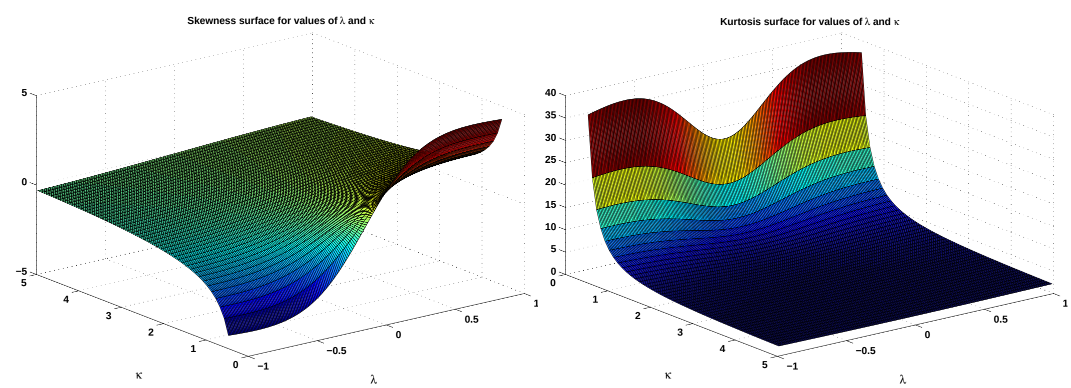

This figure depicts the skewness-kurtosis boundaries for Hansen (1994) skewed generalized Student-$t$ distribution and skewed GED against the theoretical boundary implied by Theorem 12.a, page 134 of Widder (1946). The theoretical boundary is obtained from $\mu_3^2 < \mu_4 - 1$ with $\mu_4 > 0$, where $\mu_3 = E(z_t^3)$ and $\mu_4 = E(z_t^4)$ are the non-centered third and fourth moments of a random variable $\{z_t\}_{t=0}^{\infty}$.

Figure 3: Impact of Skewness and Peakedness Parameters on Skewness and Kurtosis Surfaces for Skewed Generalized Student-$t$ Distribution




This figure shows the contribution of the skewness, $\lambda$, and peakedness parameter, $\eta$, in generating skewness (top panel) and kurtosis (bottom panel) surfaces for the skewed generalized Student-$t$ distribution of Hansen (1994). The range of parameters is chosen such that they do not violate the skewness-kurtosis boundary.

Figure 4: Impact of Skewness and Peakedness Parameters on Skewness and Kurtosis Surfaces for Skewed Generalized Error Distribution




This figure shows the contribution of the skewness, $\lambda$, and peakedness parameter, $\eta$, in generating skewness (top panel) and kurtosis (bottom panel) surfaces for the skewed generalized error distribution (SGED). The range of parameters is chosen such that they do not violate the skewness-kurtosis boundary.

Table 1: Summary Statistics of the Data

| Panel A: Descriptive Statistics, Excess Returns | | | | | |
|---|---|---|---|---|---|
| Return Series | Mean (%) | Std. Dev. (%) | Skewness | Kurtosis | J-B p-Value |
| S&P 500 | 3.48 | 21.94 | -1.24 | 31.87 | 0.01 |
| Panel B: Descriptive Statistics, Relative downside semi-variance | | | | | |
| | RSV (%) | Std. Dev. (%) | Skewness | Kurtosis | J-B p-Value |
| S&P 500 (NP) | 1.84 | 8.71 | 1.21 | 16.36 | 0.01 |

The top panel of this table reports summary statistics of excess returns. Calculation of the returns is based on subtracting the daily 3-month U.S. Treasury bill rate from the log difference of the market total return index. Mean of excess returns and standard deviations are reported as annualized percentages. Excess kurtosis values are reported. The column titled "J-B p-Value" reports $p$-values of Jarque and Bera (1980) tests of normality in percentages. The bottom panel reports the computed statistics of the observed relative semi-variance (RSV) in the data. Reported RSV is based on the mean difference between filtered downside and upside semi-variances. The first column reports the annualized RSV, and the second column is the standard deviation of this quantity. Due to availability of high-frequency data for S&P 500 returns, we report the non-parametric estimate for relative downside semi-variance for the United States denoted as S&P 500 (NP), based on 15-minute returns. The sample period is January 1980 to September 2010. Source: Thomson Reuters Datastream and FRED II data bank at the Federal Reserve Bank of St. Louis.

Table 2: Estimation Results for the Conditional Skewness Dynamics with Generalized Student-$t$ Distribution (1)

| | N | M0 | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1980-2009 | | | | | | | | | | |
| $\kappa_0$ | | -0.1047* | -0.1050* | -0.1147* | -0.0327* | -0.0594* | -0.1109* | -0.1246* | -0.0324* | -0.0655* |
| | | (0.0298) | (0.0296) | (0.0428) | (0.0136) | (0.0282) | (0.0297) | (0.0435) | (0.0128) | (0.0286) |
| $\kappa_{1+}$ | | | 0.1070* | 0.1222* | 0.1080* | 0.1511* | 0.1144* | 0.1211* | 0.1208* | 0.1584* |
| | | | (0.0253) | (0.0544) | (0.0224) | (0.0459) | (0.0260) | (0.0486) | (0.0220) | (0.0443) |
| $\kappa_{1-}$ | | | | 0.0981* | | 0.0814* | | 0.0840* | | 0.0716* |
| | | | | (0.0385) | | (0.0312) | | (0.0426) | | (0.0353) |
| $\kappa_2$ | | | | | 0.6593* | 0.6590* | | | 0.6694* | 0.6643* |
| | | | | | (0.0955) | (0.0915) | | | (0.0823) | (0.0902) |
| $\gamma_0$ | | -1.4635* | -1.4464* | -1.4450* | -1.4660* | -1.4663* | -1.4762* | -1.0870* | -0.7289* | -0.4997* |
| | | (0.1159) | (0.1364) | (0.1368) | (0.1365) | (0.1361) | (0.1567) | (0.1594) | (0.1737) | (0.1947) |
| $\gamma_{1+}$ | | | | | | | -0.5805* | -1.0077* | -0.6848* | -0.9800* |
| | | | | | | | (0.1898) | (0.1191) | (0.1283) | (0.1092) |
| $\gamma_{1-}$ | | | | | | | | 0.2368* | | 0.1909* |
| | | | | | | | | (0.0943) | | (0.0762) |
| $\gamma_2$ | | | | | | | | | 0.5119* | 0.4309* |
| | | | | | | | | | (0.1097) | (0.1158) |
| Log L | 24639.4 | 24837.7 | 24846.0 | 24846.1 | 24852.4 | 24853.1 | 24849.1 | 24867.8 | 24861.6 | 24878.7 |
| LR Stat | | 396.60* | 16.60* | 0.20 | NA | 1.40 | NA | 37.40* | NA | 34.20* |
| BIC | -3.2494 | -3.2732 | -3.2731 | -3.2720 | -3.2728 | -3.2717 | -3.2724 | -3.2725 | -3.2716 | -3.2715 |
| 1980-1989 | | | | | | | | | | |
| $\kappa_0$ | | -0.0157 | -0.0129 | -0.0307 | -0.0022 | -0.0133 | -0.0110 | -0.0678 | -0.0018 | -0.0326 |
| | | (0.0546) | (0.0517) | (0.0732) | (0.0168) | (0.0407) | (0.0512) | (0.0731) | (0.0165) | (0.0493) |
| $\kappa_{1+}$ | | | 0.0853 | 0.1120 | 0.0864* | 0.1056 | 0.0783 | 0.1446 | 0.0760* | 0.1282 |
| | | | (0.0444) | (0.0902) | (0.0341) | (0.0728) | (0.0462) | (0.0947) | (0.0357) | (0.0845) |
| $\kappa_{1-}$ | | | | 0.0670 | | 0.0767 | | 0.0073 | | 0.0481 |
| | | | | (0.0706) | | (0.0469) | | (0.0646) | | (0.0598) |
| $\kappa_2$ | | | | | 0.6741* | 0.6702* | | | 0.6777* | 0.6617* |
| | | | | | (0.1566) | (0.1557) | | | (0.1681) | (0.1757) |
| $\gamma_0$ | | -1.8280* | -1.7839* | -1.7848* | -1.7891* | -1.7887* | -1.7515* | -1.5049* | -0.0407 | -0.6140 |
| | | (0.2053) | (0.2007) | (0.2006) | (0.2012) | (0.2011) | (0.2128) | (0.2418) | (0.0219) | (0.4645) |
| $\gamma_{1+}$ | | | | | | | 0.0591 | -0.6017* | 0.0510 | -0.5835* |
| | | | | | | | (0.0941) | (0.1869) | (0.0294) | (0.1680) |
| $\gamma_{1-}$ | | | | | | | | 0.2116 | | 0.1201 |
| | | | | | | | | (0.1321) | | (0.1017) |
| $\gamma_2$ | | | | | | | | | 0.9779* | 0.5330* |
| | | | | | | | | | (0.0121) | (0.2576) |
| Log L | 8190.1 | 8308.4 | 8310.1 | 8310.2 | 8312.4 | 8312.5 | 8310.3 | 8313.8 | 8314.2 | 8316.4 |
| LR Stat | | 236.60* | 3.40 | 0.20 | NA | 0.20 | NA | 7.00* | NA | 4.40 |
| BIC | -3.2268 | -3.2674 | -3.2650 | -3.2619 | -3.2628 | -3.2597 | -3.2620 | -3.2571 | -3.2573 | -3.2519 |

This table reports the estimated parameters of $\tilde{\lambda}$ and $\tilde{\eta}$ as in equation (10), and where innovations follow a skewed GST process as in Hansen (1994). $N$ represents the baseline normally distributed errors case. Models $M_0$ to $M_8$ represent equations (14) to (21), respectively. To save space, we report estimated values for $\kappa_1$ and $\kappa_{1+}$ on the same row. Estimated standard errors are reported in parentheses, below the estimated parameters. "Log L" represents the computed log likelihood function. We report likelihood ratio test statistics, "LR Stat," with respect to the preceding model. "NA" implies that the models are non-nested, and thus the LR test statistic is not computed. All models are preferred to the benchmark normal model, based on the likelihood ratio test. Thus, we do not report these test statistics. * represents rejection of the null hypothesis in question at the 5% confidence level. "BIC" represents Bayesian information criteria for the estimated models.

Table 3: Estimation Results for the Conditional Skewness Dynamics with Generalized Student-$t$ Distribution (2)

| | N | M0 | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1990-1999 | | | | | | | | | | |
| $\kappa_0$ | | -0.0643 | -0.0582 | -0.1636* | -0.0090 | -0.0788 | -0.0527 | -0.1666* | -0.0022 | -0.1037 |
| | | (0.0515) | (0.0516) | (0.0769) | (0.0156) | (0.0538) | (0.0508) | (0.0770) | (0.0154) | (0.0533) |
| $\kappa_{1+}$ | | | 0.1503* | 0.3335* | 0.1488* | 0.2739* | 0.1572* | 0.3567* | 0.1614* | 0.3524* |
| | | | (0.0457) | (0.1101) | (0.0377) | (0.0982) | (0.0415) | (0.1040) | (0.0339) | (0.0957) |
| $\kappa_{1-}$ | | | | 0.0596 | | 0.0949 | | 0.0686 | | 0.0946 |
| | | | | (0.0683) | | (0.0509) | | (0.0696) | | (0.0528) |
| $\kappa_2$ | | | | | 0.7029* | 0.6455* | | | 0.6945* | 0.5963* |
| | | | | | (0.1141) | (0.1341) | | | (0.0990) | (0.1210) |
| $\gamma_0$ | | -1.5922* | -1.5691* | -1.5289* | -1.6366* | -1.6053* | -1.6505* | -1.1899* | -0.6799* | -0.1532 |
| | | (0.2022) | (0.2221) | (0.2297) | (0.2164) | (0.2221) | (0.2274) | (0.2842) | (0.1876) | (0.2080) |
| $\gamma_{1+}$ | | | | | | | -0.7164* | -1.1082* | -0.7516* | -1.1332* |
| | | | | | | | (0.2075) | (0.2077) | (0.1571) | (0.1588) |
| $\gamma_{1-}$ | | | | | | | | 0.2026 | | 0.1884 |
| | | | | | | | | (0.2201) | | (0.1723) |
| $\gamma_2$ | | | | | | | | | 0.6061* | 0.6387* |
| | | | | | | | | | (0.0957) | (0.0920) |
| Log L | 8628.3 | 8687.3 | 8692.8 | 8694.6 | 8697.6 | 8698.8 | 8696.6 | 8702.7 | 8706.6 | 8714.3 |
| LR Stat | | 118.00* | 11.00* | 3.60 | NA | 2.40 | NA | 12.20* | NA | 15.40* |
| BIC | -3.3989 | -3.4161 | -3.4151 | -3.4127 | -3.4139 | -3.4113 | -3.4135 | -3.4097 | -3.4113 | -3.4081 |
| 2000-2009 | | | | | | | | | | |
| $\kappa_0$ | | -0.2392* | -0.2404* | -0.1073 | -0.1069 | -0.1131 | -0.2593* | -0.0932 | -0.0897 | -0.1127 |
| | | (0.0530) | (0.0530) | (0.0835) | (0.0972) | (0.1023) | (0.0530) | (0.0805) | (0.0877) | (0.1090) |
| $\kappa_{1+}$ | | | 0.0762 | -0.0945 | 0.0818 | -0.0977 | 0.0359 | -0.1622 | 0.0570 | -0.1522 |
| | | | (0.0521) | (0.0983) | (0.0509) | (0.1025) | (0.0491) | (0.0904) | (0.0422) | (0.0935) |
| $\kappa_{1-}$ | | | | 0.2391* | | 0.2392* | | 0.2398* | | 0.2475* |
| | | | | (0.0952) | | (0.0959) | | (0.0820) | | (0.1006) |
| $\kappa_2$ | | | | | 0.5504 | -0.0306 | | | 0.6375 | -0.1292 |
| | | | | | (0.3974) | (0.2909) | | | (0.3412) | (0.2886) |
| $\gamma_0$ | | -0.2152* | -0.2614* | -0.3293* | -0.2537* | -0.3307* | 0.3538* | 0.2914* | -0.0378 | -0.4276 |
| | | (0.0546) | (0.0560) | (0.0741) | (0.0625) | (0.0713) | (0.1140) | (0.1363) | (0.2236) | (0.4257) |
| $\gamma_{1+}$ | | | | | | | -2.0265* | -1.9913* | -1.9479* | -1.7840* |
| | | | | | | | (0.2790) | (0.2322) | (0.2298) | (0.2403) |
| $\gamma_{1-}$ | | | | | | | | 0.7292* | | -4.0525 |
| | | | | | | | | (0.1026) | | (3.5040) |
| $\gamma_2$ | | | | | | | | | 0.4158* | 0.3005 |
| | | | | | | | | | (0.0875) | (0.1588) |
| Log L | 7866.4 | 7890.5 | 7891.7 | 7893.8 | 7892.1 | 7893.9 | 7910.2 | 7916.2 | 7916.5 | 7919.5 |
| LR Stat | | 48.20* | 2.40 | 4.20 | NA | 3.60 | NA | 12.00* | NA | 6.00* |
| BIC | -3.1109 | -3.1143 | -3.1116 | -3.1094 | -3.1087 | -3.1063 | -3.1159 | -3.1121 | -3.1122 | -3.1072 |

This table reports the estimation results for the 1990–99 and 2000–09 subsamples. Refer to the notes to Table 2 for more information.

Table 4: Evaluating the Conditional Relative Downside Variance with Generalized Student-$t$ Distribution

| | M0 | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 |
|---|---|---|---|---|---|---|---|---|---|
| 1980-2009 | | | | | | | | | |
| $\beta_0$ | -3.016E-06 | -1.16E-05* | -1.16E-05* | -1.81E-05* | -1.72E-05* | -1.19E-05* | -1.33E-05* | -1.96E-05* | -1.91E-05* |
| | (2.229E-06) | (1.86E-06) | (1.87E-06) | (2.73E-06) | (1.94E-06) | (1.87E-06) | (1.85E-06) | (1.94E-06) | (1.93E-06) |
| $\beta_1$ | 2.4336* | 1.5751* | 1.5727* | 0.9441* | 1.0250 | 1.5274* | 1.4075* | 0.8826* | 0.9819 |
| | (0.0839) | (0.0338) | (0.0345) | (0.0251) | (0.0277) | (0.0338) | (0.0304) | (0.0244) | (0.0497) |
| $R^2$ | 0.1226 | 0.2655 | 0.2568 | 0.1905 | 0.1858 | 0.2538 | 0.2631 | 0.1788 | 0.1877 |
| $JT$ | 292.89* | 309.64* | 295.44* | 48.692* | 79.044* | 284.57* | 231.66* | 125.66* | 98.71* |
| 1980-1989 | | | | | | | | | |
| $\beta_0$ | -1.37E-06 | -8.69E-07 | -8.90E-08 | -9.59E-06 | -6.36E-06 | -3.08E-07 | -2.74E-06 | -1.09E-05 | -7.90E-06 |
| | (1.17E-05) | (7.96E-06) | (8.07E-06) | (9.40E-06) | (6.64E-06) | (5.70E-06) | (5.53E-06) | (6.72E-06) | (6.50E-06) |
| $\beta_1$ | 2.0431* | 1.7049* | 1.7341* | 0.8763* | 1.0070 | 1.6928* | 1.4798* | 0.8189* | 0.9870* |
| | (0.2120) | (0.0534) | (0.0556) | (0.0414) | (0.0468) | (0.1086) | (0.0450) | (0.0402) | (0.0395) |
| $R^2$ | 0.0868 | 0.5106 | 0.4988 | 0.3145 | 0.3215 | 0.0543 | 0.5254 | 0.2980 | 0.3460 |
| $JT$ | 24.22* | 174.24* | 174.30* | 9.974* | 0.938 | 162.75* | 113.90* | 22.091* | 1.585 |
| 1990-1999 | | | | | | | | | |
| $\beta_0$ | 1.88E-06 | -1.33E-05* | -1.32E-05* | -1.34E-05* | -1.32E-05* | -1.35E-05* | -1.38E-05* | -1.47E-05* | -1.47E-05* |
| | (4.33E-06) | (3.20E-06) | (3.21E-06) | (3.05E-06) | (2.17E-06) | (2.26E-06) | (2.27E-06) | (2.12E-06) | (2.14E-06) |
| $\beta_1$ | 3.2425* | 1.1276 | 1.1382 | 1.0224 | 1.0422 | 1.0847 | 1.0366 | 0.9304* | 1.0414 |
| | (0.3154) | (0.1386) | (0.1387) | (0.1015) | (0.1061) | (0.1355) | (0.1352) | (0.0964) | (0.1005) |
| $R^2$ | 0.0403 | 0.0256 | 0.0260 | 0.0388 | 0.0369 | 0.0248 | 0.0228 | 0.0357 | 0.0318 |
| $JT$ | 50.740* | 18.136* | 17.94* | 19.27* | 37.18* | 36.18* | 37.25* | 48.48* | 47.01* |
| 2000-2009 | | | | | | | | | |
| $\beta_0$ | -1.55E-06 | -4.05E-06 | -8.24E-06 | -7.45E-06 | -8.15E-06* | -3.63E-07 | -9.52E-06* | -2.20E-06 | -7.14E-06* |
| | (5.05E-06) | (4.88E-06) | (4.72E-06) | (4.88E-06) | (3.33E-06) | (3.53E-06) | (3.37E-06) | (3.51E-06) | (3.37E-06) |
| $\beta_1$ | 1.2052* | 1.0808 | 0.9579 | 0.9489 | 0.9637 | 1.0791 | 0.8153* | 1.0192 | 0.9872 |
| | (0.0501) | (0.0426) | (0.0366) | (0.0394) | (0.0367) | (0.0428) | (0.0326) | (0.0410) | (0.0362) |
| $R^2$ | 0.1880 | 0.2046 | 0.2151 | 0.1882 | 0.2165 | 0.2023 | 0.2006 | 0.1980 | 0.2113 |
| $JT$ | 16.91* | 4.296 | 4.379 | 4.012 | 6.693* | 3.422 | 40.18* | 0.611 | 4.613 |

This table reports Mincer-Zarnowitz regression results from $RV^d_{t+1}(h) - RV^u_{t+1}(h) = \beta_0 + \beta_1 (Var_t[r_{t+1}|r_{t+1} < m_t] - Var_t[r_{t+1}|r_{t+1} > m_t])$, where error terms in the parametric model follow a skewed Student-$t$ distribution. $JT$ represents the value of the test statistic built under the null hypothesis that $\beta_0$ and $\beta_1$ are jointly equal to 0 and 1, respectively. * and † denote the rejection of this null hypothesis at the 5 and 10% confidence levels, respectively. The critical values for this test are 5.991 and 4.605, respectively, based on $\chi^2_{df=2}$.

Table 5: Estimation Results for the Conditional Skewness Dynamics with Skewed GED(1)

| | N | M0 | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1980-2009 | | | | | | | | | | |
| $\kappa_0$ | | -0.1347* | -0.1317* | -0.1355* | -0.0450* | -0.0764* | -0.1316* | -0.1370* | -0.0449* | -0.0997* |
| | | (0.0293) | (0.0270) | (0.0346) | (0.0150) | (0.0257) | (0.0268) | (0.0277) | (0.0150) | (0.0273) |
| $\kappa_{1+}$ | | | 0.1218* | 0.1281* | 0.1234* | 0.1717* | 0.1221* | 0.1244* | 0.1236* | 0.1944* |
| | | | (0.0211) | (0.0402) | (0.0186) | (0.0393) | (0.0185) | (0.0293) | (0.0151) | (0.0305) |
| $\kappa_{1-}$ | | | | 0.1188* | | 0.0948* | | 0.1195* | | 0.0751* |
| | | | | (0.0314) | | (0.0323) | | (0.0129) | | (0.0274) |
| $\kappa_2$ | | | | | 0.6347* | 0.6226* | | | 0.6355* | 0.5666* |
| | | | | | (0.0830) | (0.0714) | | | (0.0799) | (0.0936) |
| $\gamma_0$ | | 0.3199* | 0.3215* | 0.3216* | 0.3174* | 0.3168* | 0.3221* | 0.4620* | 0.3184* | 0.3730* |
| | | (0.0070) | (0.0071) | (0.0071) | (0.0071) | (0.0070) | (0.0225) | (0.0229) | (0.0573) | (0.0454) |
| $\gamma_{1+}$ | | | | | | | 0.0034 | -0.2414* | -0.0024 | -0.2345* |
| | | | | | | | (0.0186) | (0.0373) | (0.0216) | (0.0386) |
| $\gamma_{1-}$ | | | | | | | | 0.1280* | | 0.1268* |
| | | | | | | | | (0.0177) | | (0.0216) |
| $\gamma_2$ | | | | | | | | | -0.0053 | 0.2492* |
| | | | | | | | | | (0.1669) | (0.1053) |
| Log L | 24639.4 | 24824.0 | 24838.7 | 24838.7 | 24848.1 | 24848.9 | 24838.7 | 24859.0 | 24848.1 | 24869.7 |
| LR Stat | | 369.20* | 29.40* | 0.00 | NA | 1.60 | NA | 40.60* | NA | 43.20* |
| BIC | -3.2494 | -3.2714 | -3.2722 | -3.2710 | -3.2722 | -3.2711 | -3.2710 | -3.2713 | -3.2698 | -3.2704 |
| 1980-1989 | | | | | | | | | | |
| $\kappa_0$ | | -0.0767 | -0.0525 | -0.0193 | -0.0202 | -0.0030 | -0.0419 | -0.0958 | -0.0199 | -0.0761 |
| | | (0.0513) | (0.0505) | (0.0695) | (0.0176) | (0.0294) | (0.0382) | (0.0724) | (0.0144) | (0.0392) |
| $\kappa_{1+}$ | | | 0.1111* | 0.0717 | 0.1146* | 0.0881* | 0.0856* | 0.1085* | 0.0962* | 0.1565* |
| | | | (0.0395) | (0.0559) | (0.0255) | (0.0241) | (0.0399) | (0.0430) | (0.0237) | (0.0421) |
| $\kappa_{1-}$ | | | | 0.1424* | | 0.1263* | | 0.0279 | | 0.0334 |
| | | | | (0.0518) | | (0.0279) | | (0.0434) | | (0.0506) |
| $\kappa_2$ | | | | | 0.5797* | 0.5741* | | | 0.5613* | 0.5535* |
| | | | | | (0.1352) | (0.1435) | | | (0.1610) | (0.2190) |
| $\gamma_0$ | | 0.3210* | 0.2508* | 0.2507* | 0.2527* | 0.2510* | 0.2743* | 0.3873* | 0.1348 | 0.2560* |
| | | (0.0113) | (0.0090) | (0.0089) | (0.0090) | (0.0090) | (0.0377) | (0.0499) | (0.1313) | (0.0941) |
| $\gamma_{1+}$ | | | | | | | 0.0648* | -0.1384* | 0.0423 | -0.1223* |
| | | | | | | | (0.0290) | (0.0662) | (0.0259) | (0.0645) |
| $\gamma_{1-}$ | | | | | | | | 0.1563* | | 0.1291* |
| | | | | | | | | (0.0379) | | (0.0355) |
| $\gamma_2$ | | | | | | | | | 0.5072 | 0.4137 |
| | | | | | | | | | (0.4725) | (0.2611) |
| Log L | 8190.1 | 8290.8 | 8295.3 | 8295.6 | 8299.0 | 8299.2 | 8298.3 | 8303.6 | 8301.3 | 8305.8 |
| LR Stat | | 201.40* | 9.00* | 0.60 | NA | 0.40 | NA | 10.60* | NA | 9.00* |
| BIC | -3.2268 | -3.2604 | -3.2591 | -3.2561 | -3.2575 | -3.2544 | -3.2572 | -3.2531 | -3.2522 | -3.2478 |

This table reports the estimated parameters of $\tilde{\lambda}$ and $\tilde{\eta}$ as in equation (11), and where innovations follow a skewed GED. $N$ represents the baseline normally distributed errors case. Models $M_0$ to $M_8$ represent equations (14) to (21), respectively. To save space, we report estimated values for $\kappa_1$ and $\kappa_{1+}$ on the same row. Estimated standard errors are reported in parentheses, below the estimated parameters. "Log L" represents the computed log likelihood function. We report likelihood ratio test statistics, "LR Stat," with respect to the preceding model. "NA" implies that the models are non-nested, and thus the LR test statistic is not computed. All models are preferred to the benchmark normal model, based on the likelihood ratio test. Thus, we do not report these test statistics. * represents rejection of the null hypothesis in question at the 5% confidence level. "BIC" represents Bayesian information criteria for the estimated models.

Table 6: Estimation Results for the Conditional Skewness Dynamics with Skewed GED (2)

| | N | M0 | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1990-1999 | | | | | | | | | | |
| $\kappa_0$ | | -0.0767 | -0.0845* | -0.1917* | -0.0121 | -0.0977* | -0.0827 | -0.1934* | -0.0123 | -0.1486* |
| | | (0.0518) | (0.0462) | (0.0571) | (0.0182) | (0.0435) | (0.0469) | (0.0650) | (0.0097) | (0.0347) |
| $\kappa_{1+}$ | | | 0.1631* | 0.3541* | 0.1614* | 0.3099* | 0.1632* | 0.3862* | 0.1607* | 0.4361* |
| | | | (0.0387) | (0.0571) | (0.0374) | (0.0722) | (0.0346) | (0.0673) | (0.0142) | (0.0385) |
| $\kappa_{1-}$ | | | | 0.0757 | | 0.1046* | | 0.1095* | | 0.1175* |
| | | | | (0.0582) | | (0.0381) | | (0.0456) | | (0.0282) |
| $\kappa_2$ | | | | | 0.7051* | 0.5881* | | | 0.7141* | 0.4931* |
| | | | | | (0.1130) | (0.1464) | | | (0.0617) | (0.0656) |
| $\gamma_0$ | | 0.3210* | 0.3154* | 0.3218* | 0.2896* | 0.2987* | 0.3054* | 0.4211* | 0.5355* | 0.2802* |
| | | (0.0123) | (0.0121) | (0.0124) | (0.0116) | (0.0117) | (0.0390) | (0.0512) | (0.0743) | (0.0592) |
| $\gamma_{1+}$ | | | | | | | -0.0412 | -0.2250* | -0.0459* | -0.2925* |
| | | | | | | | (0.0388) | (0.0692) | (0.0212) | (0.0609) |
| $\gamma_{1-}$ | | | | | | | | 0.0791 | | 0.1180* |
| | | | | | | | | (0.0624) | | (0.0428) |
| $\gamma_2$ | | | | | | | | | -0.8665* | 0.5614* |
| | | | | | | | | | (0.0555) | (0.1392) |
| Log L | 8628.3 | 8684.1 | 8692.9 | 8695.4 | 8699.5 | 8701.1 | 8693.4 | 8699.5 | 8702.2 | 8710.0 |
| LR Stat | | 111.60* | 17.60* | 5.00 | NA | 3.20 | NA | 12.20* | NA | 15.60* |
| BIC | -3.3989 | -3.4148 | -3.4151 | -3.4131 | -3.4147 | -3.4122 | -3.4123 | -3.4085 | -3.4096 | -3.4064 |
| 2000-2009 | | | | | | | | | | |
| $\kappa_0$ | | -0.2635* | -0.2622* | -0.1409 | -0.1136 | -0.1457 | -0.2552* | -0.1326* | -0.0878 | -0.1361 |
| | | (0.0492) | (0.0455) | (0.0767) | (0.0951) | (0.0924) | (0.0508) | (0.0661) | (0.0650) | (0.1137) |
| $\kappa_{1+}$ | | | 0.0798 | -0.0695 | 0.0857 | -0.0713 | 0.0536 | -0.1378* | 0.0779 | -0.1522* |
| | | | (0.0451) | (0.0791) | (0.0478) | (0.0617) | (0.0668) | (0.0337) | (0.0615) | (0.0531) |
| $\kappa_{1-}$ | | | | 0.2265* | | 0.2271* | | 0.1903* | | 0.2181* |
| | | | | (0.0997) | | (0.0922) | | (0.0734) | | (0.0667) |
| $\kappa_2$ | | | | | 0.5586 | -0.0239 | | | 0.6498* | -0.1111 |
| | | | | | (0.3577) | (0.2454) | | | (0.2383) | (0.1505) |
| $\gamma_0$ | | 0.4713* | 0.4646* | 0.4589* | 0.4655* | 0.4588* | 0.4678* | 0.6886* | 0.2559* | 0.6867* |
| | | (0.0199) | (0.0197) | (0.0193) | (0.0197) | (0.0193) | (0.0418) | (0.0403) | (0.1258) | (0.0939) |
| $\gamma_{1+}$ | | | | | | | -0.1262* | -0.4590* | -0.1286* | -0.4716* |
| | | | | | | | (0.0515) | (0.0624) | (0.0474) | (0.0640) |
| $\gamma_{1-}$ | | | | | | | | 0.2288* | | 0.2476* |
| | | | | | | | | (0.0297) | | (0.0277) |
| $\gamma_2$ | | | | | | | | | 0.4427 | 0.0302 |
| | | | | | | | | | (0.2648) | (0.1816) |
| Log L | 7866.4 | 7893.4 | 7895.0 | 7896.9 | 7895.6 | 7896.9 | 7898.1 | 7919.5 | 7899.9 | 7919.6 |
| LR Stat | | 54.00* | 3.20 | 3.80 | NA | 2.60 | NA | 42.80* | NA | 39.40* |
| BIC | -3.1109 | -3.1154 | -3.1130 | -3.1106 | -3.1101 | -3.1075 | -3.1111 | -3.1134 | -3.1056 | -3.1072 |

This table reports the estimation results for the 1990–99 and 2000–09 subsamples. Refer to the notes to Table 5 for more information.

Table 7: Evaluating the Conditional Relative Downside Variance with Skewed GED

| | M0 | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 |
|---|---|---|---|---|---|---|---|---|---|
| 1980-2009 | | | | | | | | | |
| $\beta_0$ | -2.99E-06 | -2.36E-06 | -2.44E-06 | -9.72E-06* | -7.80E-06* | -2.41E-06 | 4.17E-06* | -9.80E-06* | 2.02E-06 |
| | (3.25E-06) | (2.79E-06) | (2.80E-06) | (2.80E-06) | (2.03E-06) | (1.97E-06) | (2.09E-06) | (1.98E-06) | (2.26E-06) |
| $\beta_1$ | 1.0033 | 0.9702 | 0.9614 | 0.7091* | 0.6974* | 0.9694 | 0.6680* | 0.7071* | 0.9604 |
| | (0.0380) | (0.0226) | (0.0226) | (0.0186) | (0.0191) | (0.0225) | (0.0167) | (0.0186) | (0.0423) |
| $R^2$ | 0.1037 | 0.2352 | 0.2320 | 0.1948 | 0.1811 | 0.2359 | 0.2100 | 0.1942 | 0.1376 |
| $JT$ | 0.856 | 2.468 | 3.698 | 256.62* | 265.82* | 3.343 | 399.22* | 273.80* | 0.886 |
| 1980-1989 | | | | | | | | | |
| $\beta_0$ | -1.65E-06 | 1.61E-05† | 1.63E-05† | 1.21E-06 | 6.81E-06 | 1.59E-05* | 3.54E-05* | 1.07E-06 | 3.91E-05* |
| | (1.23E-05) | (8.85E-06) | (8.90E-06) | (9.65E-06) | (6.99E-06) | (6.24E-06) | (6.75E-06) | (6.84E-06) | (1.17e-05) |
| $\beta_1$ | 0.84525 | 1.2084* | 1.2036* | 0.7324* | 0.7736* | 1.2065* | 0.9007* | 0.9928 | 1.0314 |
| | (0.1094) | (0.0447) | (0.0450) | (0.0357) | (0.0391) | (0.0444) | (0.0358) | (0.0357) | (0.0311) |
| $R^2$ | 0.0577 | 0.4285 | 0.4226 | 0.3013 | 0.2860 | 0.4306 | 0.3930 | 0.2996 | 0.2199 |
| $JT$ | 2.021 | 25.07* | 23.80* | 56.20* | 34.49* | 28.16* | 32.19* | 0.066 | 12.18* |
| 1990-1999 | | | | | | | | | |
| $\beta_0$ | 2.58E-06 | -7.73E-06* | -7.66E-06* | -8.75E-06* | -7.50E-06* | -7.80E-06* | -4.78E-06 | -8.76E-06* | -1.94E-06 |
| | (4.61E-06) | (3.58E-06) | (3.58E-06) | (3.31E-06) | (3.43E-06) | (5.05E-06) | (4.13E-06) | (3.31E-06) | (4.59E-06) |
| $\beta_1$ | 1.3910† | 0.7908† | 0.7889† | 0.7286* | 0.7109* | 0.7880* | 0.9537 | 0.7280* | 0.9849 |
| | (0.2051) | (0.1210) | (0.1209) | (0.097) | (0.0970) | (0.1210) | (0.0950) | (0.0970) | (0.0690) |
| $R^2$ | 0.0353 | 0.0323 | 0.0327 | 0.0432 | 0.0413 | 0.0325 | 0.0259 | 0.0432 | 0.0244 |
| $JT$ | 3.946 | 7.636* | 7.622* | 14.87* | 13.74* | 7.820* | 1.575 | 14.95* | 0.228 |
| 2000-2009 | | | | | | | | | |
| $\beta_0$ | -1.35E-06 | -1.17E-05* | -1.19E-05* | -1.56E-05* | -1.58E-05* | -1.18E-05* | -6.39E-06 | -1.57E-05* | -8.05E-06 |
| | (5.24E-06) | (4.89E-06) | (4.90E-06) | (4.85E-06) | (4.92E-06) | (4.89E-06) | (5.13E-06) | (4.85E-06) | (5.34E-06) |
| $\beta_1$ | 1.0932 | 0.7952* | 0.7859* | 0.6909* | 0.6314* | 0.7942* | 1.0512 | 0.9156 | 0.9531 |
| | (0.0705) | (0.0500) | (0.0500) | (0.0460) | (0.0440) | (0.0500) | (0.0517) | (0.0465) | (0.0520) |
| $R^2$ | 0.1612 | 0.1654 | 0.1636 | 0.1549 | 0.1414 | 0.1653 | 0.1535 | 0.1548 | 0.1606 |
| $JT$ | 1.814 | 22.22* | 24.14* | 56.33* | 80.56* | 22.49* | 3.505 | 13.89* | 3.085 |

This table reports Mincer-Zarnowitz regression results from $RV^d_{t+1}(h) - RV^u_{t+1}(h) = \beta_0 + \beta_1 (Var_t [r_{t+1}|r_{t+1} < m_t] - Var_t [r_{t+1}|r_{t+1} > m_t])$, where error terms in the parametric model follow a skewed Student-$t$ distribution. * indicates rejection of the null hypothesis that $\beta_i = i, i = 0, 1$ at the 5% confidence level or better. $JT$ represents the value of the test statistic built under the null hypothesis that $\beta_0$ and $\beta_1$ are jointly equal to 0 and 1, respectively. * and † denote the rejection of this null hypothesis at the 5 and 10% confidence levels, respectively. The critical values for this test are 5.991 and 4.605, respectively, based on $\chi^2_{df=2}$.

Table 8: Estimation Results for the Conditional Skewness Dynamics with Binormal Distribution

| | N | M0 | M1 | M2 | M3 | M4 |
|---|---|---|---|---|---|---|
| 1980-2009 | | | | | | |
| $\kappa_0$ | | -0.2435* | -0.1853* | -0.1989* | -0.0955* | -0.1127* |
| | | (0.0329) | (0.0299) | (0.0438) | (0.0237) | (0.0380) |
| $\kappa_{1+}$ | | | 0.2055* | 0.2342* | 0.1819* | 0.2143* |
| | | | (0.0272) | (0.0729) | (0.0252) | (0.0607) |
| $\kappa_{1-}$ | | | | 0.1948* | | 0.1687* |
| | | | | (0.0375) | | (0.0338) |
| $\kappa_2$ | | | | | 0.4187* | 0.4128* |
| | | | | | (0.0960) | (0.0953) |
| Log L | 24639.4 | 24667.0 | 24692.3 | 24692.4 | 24697.8 | 24698.1 |
| LR Stat | | 55.20* | 50.60* | 0.20 | NA | 0.60 |
| BIC | -3.2494 | -3.2530 | -3.2552 | -3.2540 | -3.2547 | -3.2536 |
| 1980-1989 | | | | | | |
| $\kappa_0$ | | -0.2022* | -0.1182* | -0.0595 | -0.0689* | 0.0071 |
| | | (0.0543) | (0.0482) | (0.0766) | (0.0306) | (0.0543) |
| $\kappa_{1+}$ | | | 0.2709* | 0.1556 | 0.2291* | 0.0922 |
| | | | (0.0436) | (0.1211) | (0.0400) | (0.0910) |
| $\kappa_{1-}$ | | | | 0.3230* | | 0.2883* |
| | | | | (0.0758) | | (0.0633) |
| $\kappa_2$ | | | | | 0.3750* | 0.4270* |
| | | | | | (0.1203) | (0.1180) |
| Log L | 8190.1 | 8197.5 | 8213.7 | 8214.2 | 8216.7 | 8218.0 |
| LR Stat | | 14.80* | 32.40* | 1.00 | NA | 2.60 |
| BIC | -3.2268 | -3.2296 | -3.2330 | -3.2301 | -3.2311 | -3.2285 |
| 1990-1999 | | | | | | |
| $\kappa_0$ | | -0.2211* | -0.1529* | -0.3460* | -0.0864 | -0.2925* |
| | | (0.0579) | (0.0524) | (0.0799) | (0.0446) | (0.0880) |
| $\kappa_{1+}$ | | | 0.2079* | 0.6721* | 0.1861* | 0.5903* |
| | | | (0.0440) | (0.1424) | (0.0432) | (0.1516) |
| $\kappa_{1-}$ | | | | 0.0626 | | 0.0683 |
| | | | | (0.0712) | | (0.0658) |
| $\kappa_2$ | | | | | 0.3643 | 0.1669 |
| | | | | | (0.2108) | (0.1573) |
| Log L | 8628.3 | 8635.7 | 8644.7 | 8649.8 | 8646.0 | 8650.4 |
| LR Stat | | 14.80* | 18.00* | 10.20* | NA | 8.80* |
| BIC | -3.3989 | -3.4018 | -3.4023 | -3.4012 | -3.3997 | -3.3983 |
| 2000-2009 | | | | | | |
| $\kappa_0$ | | -0.3209* | -0.3029* | -0.1736 | -0.1140 | -0.1826 |
| | | (0.0618) | (0.0593) | (0.0947) | (0.0855) | (0.1288) |
| $\kappa_{1+}$ | | | 0.0881 | -0.1279 | 0.0919 | -0.1307 |
| | | | (0.0640) | (0.1402) | (0.0556) | (0.1451) |
| $\kappa_{1-}$ | | | | 0.2270* | | 0.2279* |
| | | | | (0.1026) | | (0.1034) |
| $\kappa_2$ | | | | | 0.5806* | -0.0325 |
| | | | | | (0.2780) | (0.3113) |
| Log L | 7866.4 | 7879.6 | 7880.6 | 7882.1 | 7881.3 | 7882.1 |
| LR Stat | | 26.40* | 2.00 | 3.00 | NA | 1.60 |
| BIC | -3.1109 | -3.1162 | -3.1135 | -3.1109 | -3.1106 | -3.1078 |

This table reports the estimated parameters of $\tilde{\lambda}$, where $\tilde{\lambda} = -\ln[\sqrt{\pi/2-1}\lambda + 1]$ and innovations follow a binormal distribution. $N$ represents the baseline normally distributed errors case. Models $M_0$ to $M_4$ represent equations (14) to (17), respectively. To save space, we report estimated values for $\kappa_1$ and $\kappa_{1+}$ on the same row. Estimated standard errors are reported in parentheses, below the estimated parameters. "Log L" represents the computed log likelihood function. We report likelihood ratio test statistics, "LR Stat," with respect to the preceding model. "NA" implies that the models are non-nested, and thus the LR test statistic is not computed. All models are preferred to the benchmark normal model, based on the likelihood ratio test. Thus, we do not report these test statistics. * represents rejection of the null hypothesis in question at the 5% confidence level. "BIC" represents Bayesian information criteria for the estimated models.

Table 9: Evaluating the Conditional Relative Downside Variance for Binormal Distribution

| | M0 | M1 | M2 | M3 | M4 |
|---|---|---|---|---|---|
| 1980-2009 | | | | | |
| $\beta_0$ | -2.30E-06 | -1.06E-05* | -1.09E-05* | -1.36E-05* | -1.35E-05* |
| | (3.13E-06) | (2.66E-06) | (2.67E-06) | (2.72E-06) | (1.93E-06) |
| $\beta_1$ | 1.5177* | 1.1317 | 1.1199 | 0.9262* | 0.9371 |
| | (0.0502) | (0.0849) | (0.0752) | (0.0229) | (0.0435) |
| $R^2$ | 0.1320 | 0.2549 | 0.2474 | 0.2133 | 0.2096 |
| $JT$ | 107.11* | 18.03* | 19.19* | 35.17* | 50.93* |
| 1980-1989 | | | | | |
| $\beta_0$ | 5.70E-06 | -9.84E-06 | -1.05E-05 | -1.06E-05 | -1.30E-05 |
| | (1.11E-05) | (8.08E-06) | (7.86E-06) | (9.04E-06) | (6.53E-06) |
| $\beta_1$ | 2.0913* | 1.2928* | 1.2435* | 0.9860 | 0.7991* |
| | (0.1570) | (0.0423) | (0.0386) | (0.0419) | (0.0362) |
| $R^2$ | 0.1537 | 0.4884 | 0.5148 | 0.3620 | 0.3334 |
| $JT$ | 48.58* | 49.38* | 41.57* | 1.488 | 34.86* |
| 1990-1999 | | | | | |
| $\beta_0$ | 9.24E-07 | -1.44E-05* | -1.85E-05* | -1.34E-05* | -1.79E-05* |
| | (4.24E-06) | (3.10E-06) | (2.93E-06) | (3.07E-06) | (2.09E-06) |
| $\beta_1$ | 2.2048* | 0.8565 | 0.5137* | 0.9437 | 1.1143 |
| | (0.2150) | (0.1029) | (0.0806) | (0.0975) | (0.0856) |
| $R^2$ | 0.0401 | 0.0268 | 0.0158 | 0.0359 | 0.0182 |
| $JT$ | 31.46* | 23.52* | 76.39* | 19.43* | 75.47* |
| 2000-2009 | | | | | |
| $\beta_0$ | -1.26E-06 | -3.28E-06 | -5.06E-06 | -6.55E-06 | -4.97E-06 |
| | (5.05E-06) | (4.90E-06) | (4.76E-06) | 4.88E-06 | (3.36E-06) |
| $\beta_1$ | 1.1713* | 1.0719 | 0.9981 | 0.9395 | 1.0034 |
| | (0.0485) | (0.0424) | (0.0374) | (0.0385) | (0.0374) |
| $R^2$ | 0.1888 | 0.2037 | 0.2217 | 0.1919 | 0.2231 |
| $JT$ | 12.54* | 3.327 | 1.134 | 4.270 | 2.192 |

This table reports Mincer-Zarnowitz regression results from $RV_{t+1}^{d}(h) - RV_{t+1}^{u}(h) = \beta_0 + \beta_1 \left(Var_t\left[r_{t+1}|r_{t+1} < m_t\right] - Var_t\left[r_{t+1}|r_{t+1} > m_t\right]\right)$, where error terms in the parametric model are binormally distributed. * indicates rejection of the null hypothesis that $\beta_i = i, i = 0, 1$ at the 5% confidence level or better. $JT$ represents the value of the test statistic built under the null hypothesis that $\beta_0$ and $\beta_1$ are jointly equal to 0 and 1, respectively. * and † denote the rejection of this null hypothesis at the 5 and 10% confidence levels, respectively. The critical values for this test are 5.991 and 4.605, respectively, based on $\chi^2_{df=2}$.