Testing for Conditional Skewness with *Epsilon-Skew-t* Distributions*

Andrea De Polis†

This draft: December, 2024

Abstract

I develop a parametric test to detect the presence of instability in the third moment of time series data. The test is based on the score function of the flexible epsilon-Skew-t distribution, and belongs to the class of Lagrange Multiplier tests. The test presents appropriate asymptotic properties, as evaluated by means of an extensive Monte Carlo analysis. When applied to the three asset pricing anomalies of Fama and French (1993), the test points at an overwelimg evidence of con- ditional non-Gaussianity at the daily frequency, whereas weaker results are observed at the monthly frequency. These results should be taken as a warning of possible misspecification of asset pricing models based on symmetric likelihoods.

1 Introduction

Motivated by are renewed interest in *Skewed-Student-t* distributions, in this paper we derive the properties of the *Epsilon-Skew-t* distribution of Arellano-Valle et al. (2005) and Gómez et al. (2007).[1] Specifically, we derive a closed form solution for the coefficient of skewness, defined as the rescaled third central moment, and we provide a novel proof of the unbiasedness of the distribution's score vector and its Fisher Information matrix.

There is a long standing tradition in the modeling of higher-order moments in modern financial econometrics and risk management literatures. Models for the conditional distribution of financial returns are commonly estimated assuming *Student-t* innovations, which allow statistical models to capture and exploit the occurrence of extreme observations. Bollerslev (1987) introduced heavy tails in the GARCH setting for conditional heteroskedasticity. More recently, Hansen (1994) and

*I would like to thank Ivan Petrella for extensive discussions about this paper.

†University of Strathclyde and Economic Statistics Centre of Excellence. andrea.de-polis@strath.ac.uk

[1] The *Epsilon* in the name of the distribution derives from the fact that Mudholkar and Hutson (2000) first, and Gómez et al. (2007) then, denote the asymmetry parameter by ε.

Harvey and Siddique (1999) have introduced filters to model the time variation in the skewness of financial returns. Theodossiou (1998), Komunjer (2007) and Zhu and Galbraith (2010) are among the recent contributions that proposed novel distributions to accommodate the modeling of higher-order moments for risk management purposes. There is also a long-standing interest for the modelling of asymmetry in the macroeconomic literature date back to the contributions of Neftci (1984) or Hamilton (1989). Recently, Adrian et al. (2019) have renewed the interest for the reduced-form modelling of the asymmetry in the residual distribution of macroeconomic shocks.

Nevertheless, the presence of time-varying skewness in financial returns has been at best evaluated empirically, by means of model fit statistics. In this paper we introduce a parametric test for the presence of autocorrelated skewness in time series data based on the *Epsilon-Skew-t* distribution. The test belongs to the class of *Lagrange Multiplier* (LM) tests and it is aimed at providing a tool to draw inference on the dynamic properties of the third-order moments of any sample. Interestingly, we show that when modelling heteroskedasticity assuming the location to be fixed, the test for time-varying asymmetry reduces to the classic Box and Pierce (1970) test on the autocorrelation of the conditional score of the asymmetry parameter, estimated under the null hypothesis.

In an application to financial data, we test the presence of time-varying asymmetry in the famous Fama and French (1993) three factors. Our test detects strong evidence in favor of daily time-varying skewness for the anomalies, whereas less striking evidence is observed at the monthly frequency.

The rest of the paper is as follows: Section 2 provides an overview of different methods to introduce skewness in symmetric densities, with a focus on the t distribution. Section 3 illustrates the properties of the *Epsilon-Skew-t* distribution, providing detailed derivations when appropriate. In Section 4 we introduce the parametric test for conditional skewness based on the *Epsilon-Skew-t* density. Section 5 evaluates the properties of the test. In Section 6 we provide an empirical application of the test to financial data. Section 7 concludes.

2 The Skew-t distribution

Skew-t distributions can be constructed in different ways (see, e.g., Jones, 2015). Currently, two methods are the most commonly used: the modulation of the density and the mixture representation. In what follow, we will introduce these two approaches to skew symmetric distributions, but we will focus on the latter in the remainder of the paper.

2.1 Modulation of the density

Azzalini (1985) introduced a simple way to create a family of distributions that i) strictly includes the Normal distribution and ii) is mathematically tractable, by *modulating* the Gaussian distribution. More generally, let $f(y)$ be symmetric about 0, that is $f(y) = f(-y)$, $\forall y \in \mathbb{R}$, $G(\cdot)$ be a symmetric (about 0) distribution function, e.g., $F(-y) = 1 - F(y)$, and $w(\cdot)$ a real valued, odd function such that $w(-y) = -w(y)$; then

$$p(y) = 2f(y)F(w(y)) \tag{1}$$

is an admissible density function, such that $\int_{\mathbb{R}} p(y)dy = 1$ and $p(y) > 0$, $\forall y \in \mathbb{R}$ (see Azzalini, 2013, for further detils). Let t_ν be a symmetric *Student-t* distribution with zero location, unit scale and ν degrees of freedom (Johnson et al., 1995), with distribution function F_ν. By Equation (1), $y \sim Skt_\nu(0, 1, \alpha)$ with density

$$p(y) = 2t_\nu(y)F_{\nu+1}\left(\alpha\sqrt{\frac{\nu+1}{\nu+y^2}}y\right)$$

is a *Skew-t* distribution with ν degrees of freedom, asymmetry parameter α and $w(y) = \left(\alpha\sqrt{\frac{\nu+1}{\nu+y^2}}y\right)$, as in Azzalini and Capitanio (2003). Limiting cases for $p(y)$ encompass the symmetric *Student-t* for $\alpha = 0$, *Skew-Normal* of Azzalini (1985), $Skw{-}\mathcal{N}$, for ν approaching ∞, and the standard Normal when both restrictions apply. The distribution also admits a stochastic representation of the form $Y = \frac{X}{\sqrt{V}}$, with $X \sim Skw{-}\mathcal{N}$ and $V \sim \frac{\chi^2_\nu}{\nu}$.

Di Ciccio and Monti (2011) work out the score function and the Fisher Information matrix for this type of *Skew-t* distributions, highlighting inferential issues related to the singularity of the Information quantity when $\nu \to \infty$ and $\alpha = 0$.

2.2 Mixture representation

Arellano-Valle et al. (2005) generalize the family of skewed distributions built by joining two half-distributions. Consider $a(\varrho)$ and $b(\varrho)$, two positive asymmetry functions of a common asymmetry parameter $\varrho \in \mathbb{R}$, and let U_ϱ be a discrete random variable defined over the support $[-b(\varrho), a(\varrho)]$, with $\mathcal{P}(U_\varrho = a(\varrho)) = \frac{a(\varrho)}{a(\varrho)+b(\varrho)}$ and $\mathcal{P}(U_\varrho = -b(\varrho)) = 1 - \mathcal{P}(U_\varrho = a(\varrho))$. Define $V \sim 2f(y)\{y \geq 0\}$ be a generic *half-distribution* created by truncating the symmetric density $f(y)$ around $y = 0$, independent of U_ϱ.[2] Then, $Y \sim Skew{-}f = U_\varrho V$.

When f is the density function of a *Student-t* with ν degrees of freedom, $y \sim Skt_\nu(\mu, \sigma, \varrho)$ is a

[2]Alternatively, we can define $V \sim |Y|$.

Skew-t random variable with density function of the form:

$$p(y|\mu,\sigma,\varrho,\nu)=\frac{2}{\sigma(a(\varrho)+b(\varrho))}\left[t_\nu\left(\frac{y-\mu}{\sigma a(\varrho)}\right)I\{y\geq\mu\}+t_\nu\left(\frac{y-\mu}{\sigma b(\varrho)}\right)I\{y<\mu\}\right], \tag{2}$$

with location parameter μ, scale σ, and ϱ being the parameter regulating the asymmetry of the distribution with respect to the mode. As for the density in Section 2.1, this distribution collapses to the symmetric, lcation-scale *Student-t* with ν degrees of freedom for $a(\varrho)=b(\varrho)$, to a version of the $Skw-\mathcal{N}$ for $\nu\to\infty$, and to a $\mathcal{N}(\mu,\sigma)$ when both conditions hold.

The generality of this approach allows to generate skewd densities from a vast family of distributions (see, e.g., Arellano-Valle et al., 2005, and referneces therein), and to generate asymmetry in several ways, via the functions $a(\cdot)$ and $b(\cdot)$. For example, by setting $a(\varrho)=\varrho$ and $b(\varrho)=\varrho^{-1}$ we retrieve Fernández and Steel (1998) *Skew-t* distribution. In the following Sections we will focus on the case when $a(\varrho)=1-\varrho$ and $b(\varrho)=1+\varrho$, that corresponds to the *Epsilon-Skew-t* popularized by Gómez et al. (2007).

3 *Epsilon-Skew-t* properties

When $Y\sim Skt_\nu(\mu,\sigma,\varrho)$ and $a(\varrho)=1+\varrho$ and $b(\varrho)=1-\varrho$, we can rewrite Equation (2) as

$$p(y|\mu,\sigma,\varrho,\nu)=\frac{\mathcal{C}}{\sigma}\left[1+\frac{1}{\nu}\left(\frac{y-\mu}{\sigma(1+sgn(y-\mu)\varrho)}\right)^2\right]^{-\frac{1+\nu}{2}}, \tag{3}$$

where $\mathcal{C}=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}$; see Gómez et al. (2007) for some basic properties of Equation (3).

Denoting with F_ν the cumulative distribution function (CDF) of a *Student-t*, the CDF of Y, $H(\cdot)$ reads:

$$H(y|\mu,\sigma,\varrho)=\begin{cases}(1-\varrho)F_\nu\left(\frac{y-\mu}{\sigma(1-\varrho)}\right), & y<\mu\\ (1+\varrho)F_\nu\left(\frac{y-\mu}{\sigma(1+\varrho)}\right)-\varrho, & y\geq\mu,\end{cases} \tag{4}$$

such that $H(\mu)=\frac{1-\varrho}{2}$.[3] Figure 1 illustrate the density function and the distribution function of Skew-t-variates for different levels of the asymmetry parameter. To recover the quantile function, we can invert Equation (4) and notice that the threshold for the quantile is expressed as a function

[3]Notice that this is equivalent to $H(0|\varrho)$ for $\tilde{y}=\frac{y-\mu}{\sigma}$.

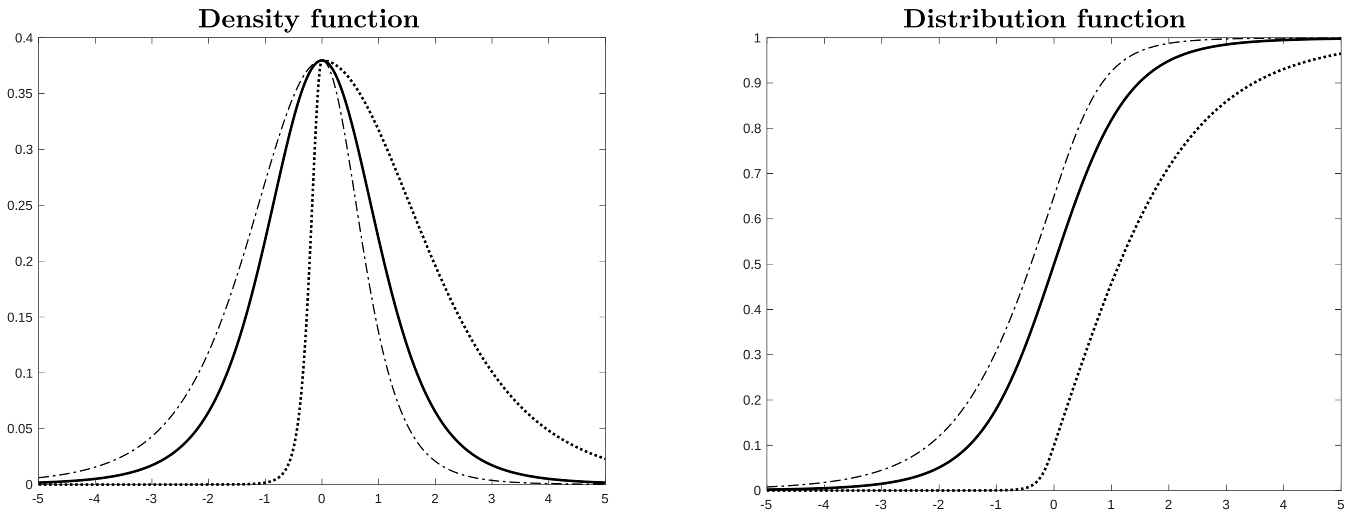

Figure 1: Density and distribution functions

Note: The panels illustrate the density function (left) and the distribution function (right) for three different Skew-t distributions for different values of the asymmetry parameter: $\varrho = 0$, *solid*, $\varrho = -0.3$, *dash-dotted*, $\varrho = 0.8$, *dotted*. In all cases, $\mu = 0$, $\sigma = 1$ and $\nu = 5$.

of the shape parameter:

$$H^{-1}(q|\mu,\sigma,\varrho) = \begin{cases} \mu + \sigma(1-\varrho)F_{\nu}^{-1}\left(\frac{q}{1-\varrho}\right), & \frac{1-\varrho}{2} < q \\ \mu + \sigma(1+\varrho)F_{\nu}^{-1}\left(\frac{q+\varrho}{1+\varrho}\right), & \frac{1-\varrho}{2} \geq q, \end{cases}$$

where F_{ν}^{-1} is the quantile function of a *Student-t*.

3.1 Two-piece representation

In order to recover the moments, the score vector and the Information quantity, it is often convenient to exploit the *two-piece representation* of Equation (2) (Fernández and Steel, 1998). Since any symmetric density on $\mathbb{R}$ can be uniquely determined from a density on $\mathbb{R}^{+}$, the *Skt* distribution can be defined in terms of strictly positive densities (Arellano-Valle et al., 2005), we can re-parametrize the density in Equation (3) as:

$$p(y|\mu,\sigma,\varrho,\nu) = \begin{cases} \frac{\mathcal{C}}{\sigma}\left[1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma_{+}}\right)^{2}\right]^{-\frac{1+\nu}{2}}, & y \geq \mu \\ \frac{\mathcal{C}}{\sigma}\left[1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma_{-}}\right)^{2}\right]^{-\frac{1+\nu}{2}}, & y < \mu \end{cases}$$

where $\sigma_+ = (1+\varrho)\sigma$ and $\sigma_- = (1-\varrho)\sigma$ are the scale parameters of the two *Half-t* densities on each side, and

$$P(y \geq \mu) = \frac{\sigma_+}{\sigma_+ + \sigma_-} = \frac{1+\varrho}{2}, \qquad P(y < \mu) = \frac{\sigma_-}{\sigma_+ + \sigma_-} = \frac{1-\varrho}{2}.$$

3.2 Moments

The two-piece formulation allows to consider separately the two halves of the distribution when taking expectations: for $y = \mu + \sigma\zeta$, where $\zeta \sim Skt_\nu(0, 1, \varrho)$, the moments of y are weighted averages of the moments of $|\zeta|$, where $|\zeta| \sim Ht_\nu$, is an *Half-t* distribution.[4]

Half-t moments. Following Kim (2008), let $\eta \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ and ϕ be the standard Normal probability density function; then for Lemma 2.2 in Kim (2008):

$$\mathbb{E}[\eta^{-\frac{\ell}{2}}\phi\left(\sqrt{\eta}x\right)] = \frac{\Gamma\left(\frac{\nu-\ell}{2}\right)\nu^{\frac{\nu}{2}}}{2^{\frac{(\ell+1)}{2}}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)}(\nu + x^2)^{-\frac{\nu-\ell}{2}}, \quad \ell = 1, 2, 3.$$

Let $Z = |\zeta| \sim Ht_\nu$ or, equivalently, $Z \sim Tt_{(a,b)}(\nu)$, a *truncated-t* distribution with truncation parameters $a = 0$ and $b = \infty$. The moments of the Ht_ν can be obtained as:

$$\mathbb{E}[Z^{k+2}] = \mathbb{E}_\eta[\eta^{-\frac{k+2}{2}}V^{k+2}], \quad k = -1, 0, 1, \ldots,$$

where

$$V^{k+2} = -\frac{(\sqrt{\eta}b)^{k+1}\phi(\sqrt{\eta}b) - (\sqrt{\eta}a)^{k+1}\phi(\sqrt{\eta}a)}{F_\nu(b) - F_\nu(a)} + (k+1)V^k.$$

Proposition 1. *The first four moments of a Half-t distribution with ν degrees of freedom are:*

$$\begin{aligned}
\mathbb{E}[Z] &= \frac{2\nu}{\nu-1}\mathcal{C}(\nu) \\
\mathbb{E}[Z^2] &= \frac{\nu}{\nu-2} \\
\mathbb{E}[Z^3] &= \frac{4\nu^2}{(\nu-3)(\nu-1)}\mathcal{C}(\nu) \\
\mathbb{E}[Z^4] &= 3\frac{\nu^2}{(\nu-4)(\nu-2)}.
\end{aligned}$$

Proof. See Section 8. ■

[4]Notice that the Half-t distribution is a special case of the folded-f distribution (Psarakis and Panaretoes, 1990).

Moments of $\zeta \sim Skt_\nu(0,1,\varrho)$. Define $d_r(\nu) = \mathbb{E}[Z^r]$, $\forall r \in \mathbb{Z}$, then the moments of $\zeta \sim Skt_\nu(0,1,\varrho)$ are computed from

$$\mathbb{E}[\zeta^r] = \frac{1}{2}\left[(1+\varrho)^{r+1} + (-1)^r(1-\varrho)^{r+1}\right] d_r(\nu),\ \forall r \in \mathbb{Z}.$$

The first three moments read:

$$\begin{aligned}
\mu_1 &= 2\varrho\, d_1(\nu) &&= \frac{4\nu}{\nu-1}\mathcal{C}(\nu)\varrho;\\
\mu_2 &= (1+3\varrho^2)d_2(\nu) &&= (1+3\varrho^2)\frac{\nu}{\nu-2};\\
\mu_3 &= 4\varrho(1+\varrho^2)d_3(\nu) &&= \frac{16(1+\varrho^2)\nu^2}{(\nu-3)(\nu-1)}\mathcal{C}(\nu)\varrho;\\
\mu_4 &= (1+10\varrho^2+5\varrho^4)d_4(\nu) &&= \frac{3\nu^2(1+10\varrho^2+5\varrho^4)}{(\nu-4)(\nu-2)}.
\end{aligned}$$

Moments of $y \sim Skt_\nu(\mu,\sigma,\varrho)$. Finally, to recover the moments of $y = \mu + \sigma\zeta$ we can use the following aggregation rule:

$$\mathbb{E}[Y^r] = \sum_{k=0}^{r}\binom{r}{k}\sigma^k\mu^{r-k}\mu_k, \quad r = 1,2,\dots,$$

to obtain:

$$\begin{aligned}
\mathbb{E}[Y] &= \mu + g(\nu)\sigma\varrho,\\
\mathbb{E}[Y^2] &= \mu^2 + 2g(\nu)\mu\sigma\varrho + \sigma^2\frac{\nu(1+3\varrho^2)}{\nu-2},\\
\mathbb{E}[Y^3] &= \mu^3 + 3\mu^2 g(\nu)\sigma\varrho + 3\mu\frac{\nu(1+3\varrho^2)}{\nu-2}\sigma^2 + 16\frac{(1+\varrho^2)g(\nu)\nu}{(\nu-3)}\sigma^3\varrho,\\
\mathbb{E}[Y^4] &= \mu^4 + 4\mu^4 g(\nu)\sigma\varrho + 6\mu^2\frac{\nu(1+3\varrho^2)}{\nu-2}\sigma^2 + 64\frac{(1+\varrho^2)g(\nu)\nu}{(\nu-3)}\sigma^3\varrho + \frac{3\nu^2(1+10\varrho^2+5\varrho^4)}{(\nu-4)(\nu-2)}\sigma^4,
\end{aligned}$$

with $g(\nu) = \frac{4\mathcal{C}(\nu)\nu}{\nu-1}$. Hence, the variance reads:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(Y) &= \sigma^2\left(\frac{\nu(1+3\varrho^2)}{\nu-2} + g(\nu)^2\varrho^2\right)\\
&= \sigma^2\left(\frac{\nu}{\nu-2} + h(\nu)\varrho^2\right),
\end{aligned}$$

with $h(\nu) = \frac{3}{\nu-2} + g(\nu)^2$. The skewness and kurtosis are:

$$\begin{aligned}
\mathbb{S}kew(Y) &= \frac{\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3}{(\mu_2-\mu_1^2)^{\frac{3}{2}}}\\
&= \frac{g(\nu)\varrho\left[\nu(\nu+1) - \varrho^2\left((5-2g(\nu)^2)\nu^2 + (10g(\nu)^2-19)\nu - 12g(\nu)^2\right)\right]}{(\nu-3)(\nu-2)\left(\frac{\nu}{\nu-2} + h(\nu)\varrho^2\right)^{\frac{3}{2}}};
\end{aligned} \tag{5}$$

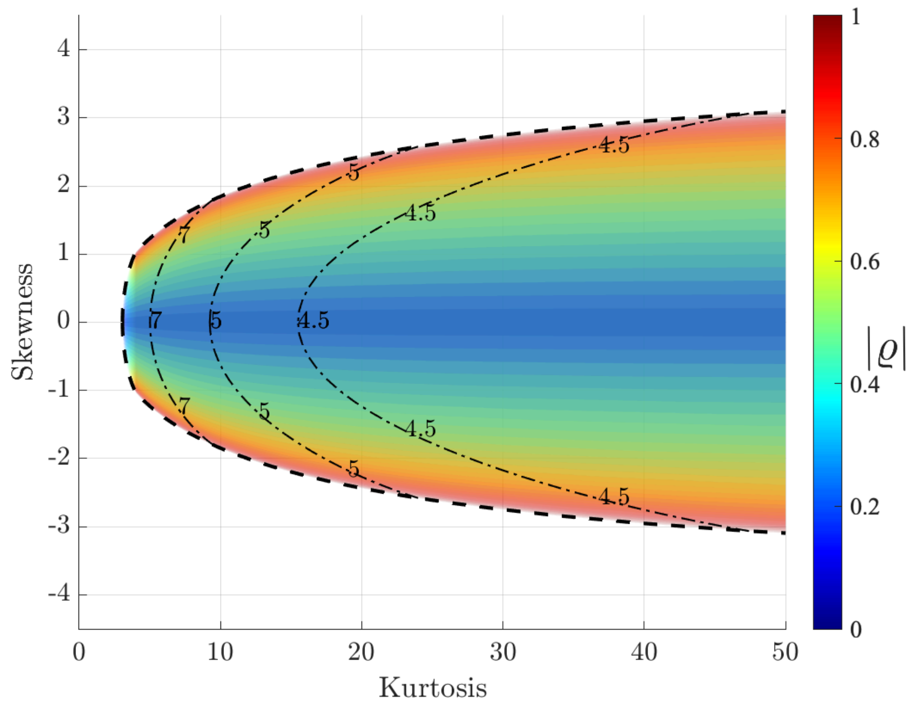

Figure 2: Skewness-Kurtosis bound

Note: The figure reports the skewness-kurtosis bound for the *Epsilon-Skew-t* distribution with $\nu > 4$. Shaded colors relate to the level of asymmetry. Dotted lines highlights the bounds for 4.5, 5 and 7 degrees of freedom.

$$\mathbb{K}urt(Y) = \frac{\mu_4 - 4\mu_1\mu_3 + 6\nu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2}.$$

Figure 2 illustrates the skewness-kurtosis bound for different levels of the asymmetry parameter; dotted lines highlight the bounds for 4.5, 5 and 7 degress of freedom.

3.3 Score vector

Define $\varepsilon = y - \mu$, and denote the positive (negative) "centered" variable as ε_+ (ε_-), and the corresponding positive (negative) standardized version as $\zeta_+ = \frac{\varepsilon_+}{\sigma_+}$ ($\zeta_- = \frac{\varepsilon_-}{\sigma_-}$). Moreover, $h = (1 + sgn(\varepsilon)\varrho)$, with $h_+ = (1+\varrho)$ and $h_- = (1-\varrho)$, thus $h_+\zeta_+ = \frac{\varepsilon_+}{\sigma}$, and $h_-\zeta_- = \frac{\varepsilon_-}{\sigma}$. The log-likelihood function can be expressed as:

$$\ell = \log p(y|\theta) = \log \mathcal{C}(\eta) - \frac{1}{2}\log\sigma^2 - \frac{1+\eta}{2\eta}\log m(\mu,\sigma,\varrho), \qquad (6)$$

where $\eta = \frac{1}{\nu}$, $m(\mu,\sigma,\varrho) = \left(1 + \frac{\eta\zeta^2}{h^2}\right)$ and $\log \mathcal{C}(\eta) = \log\Gamma\left(\frac{\eta+1}{2\eta}\right) - \log\Gamma\left(\frac{1}{2\eta}\right) - \frac{1}{2}\log\left(\frac{1}{\eta}\right) - \frac{1}{2}\log\pi$. Let us also define $w = \frac{(1+\eta)}{(1+sgn(\varepsilon)\varrho)^2+\eta\zeta^2} = \frac{1+\eta}{h^2+\eta\zeta^2}$, a set of weights, common to all the gradients, that downplay the effect of outliers. The following propositions lay out the score vector and its unbiasedness.

Proposition 2. *The elements of score vector with respect to the parameters of location, squared scale*[5] *and shape of the* $Skt_\nu(\mu, \sigma, \varrho)$ *distribution with log-likelihood given by Equation (6) read:*

$$\nabla_\mu = \frac{w\zeta}{\sigma}, \qquad \nabla_{\sigma^2} = \frac{(w\zeta^2 - 1)}{2\sigma^2}, \qquad \nabla_\varrho = \frac{sgn(\varepsilon) w\zeta^2}{(1 + sgn(\varepsilon)\varrho_t)}.$$

Proof. See Section 9. ∎

3.4 Fisher's Information matrix

Proposition 2 highlights that the elements of the gradient vector are functions of the random variable $(w_t\zeta_t^2)$. In order to compute expectations of ∇, it is often convenient to rewrite the gradient as the sum of the semi-gradients computed over the positive and negative semi-support, ∇_+ and ∇_- respectively, such that

$$\nabla = \nabla_+ I\{y \geq \mu\} + \nabla_- I\{y < \mu\}.$$

Therefor, the two components can be expressed as functions of $(w\zeta^2)_\pm = w_\pm \zeta_\pm^2$, with $w_\pm = \frac{1+\eta}{h_\pm^2 + \eta\zeta_\pm^2}$. Following Harvey (2013), it can be shown that $(w\zeta^2)_\pm \sim \mathcal{B}\left(\frac{1}{2}, \frac{1}{2\eta}\right)$, where $\mathcal{B}$ is the Beta distribution (Johnson et al., 1995); some notable results that relate the $(Skew-)t$ and Beta distributions are reported in Section 10.

It is now easy to compute the Information matrix as the expectation of the outer product of the gradients, $\mathcal{I} = \mathbb{E}[\nabla\nabla']$.[6]

Proposition 3. *Taking advantage of the two-piece representation,*

$$\mathbb{E}[\nabla\nabla'] = P(y \geq \mu)\,\mathbb{E}[\nabla_+\nabla_+'] + P(y < \mu)\,\mathbb{E}[\nabla_-\nabla_-'].$$

Therefore,

$$\mathcal{I} = \begin{bmatrix} \frac{(1+\eta)}{\sigma^2(1+3\eta)(1-\varrho^2)} & 0 & \frac{4\mathcal{C}(1+\eta)}{\sigma(1+3\eta)(1-\varrho^2)} \\ 0 & \frac{1}{2\sigma^4(1+3\eta)} & 0 \\ \frac{4\mathcal{C}(1+\eta)}{\sigma(1+3\eta)(1-\varrho^2)} & 0 & \frac{3(1+\eta)}{(1+3\eta)(1-\varrho^2)} \end{bmatrix}.$$

Proof. See Section 11 ∎

[5] We opt for the squared scale parameter for exposure purposes.

[6] Notice that Gómez et al. (2007) derive the same expression using the expectation of the negative Hessian matrix.

4 Testing conditional skewness in heteroskedastic time series

Consider $y_t \sim Skt_\nu(\mu, \sigma_t^2, \varrho_t)$, that is y_t is a realization from a the distribution in Equation (3), where both the squared scale and the asymmetry parameter are allowed to vary over time; without loss of generality we keep the location and the degrees of freedom fixed. We assume that σ_t^2 follows a first order score-driven process (see Harvey, 2013; Creal et al., 2013)

$$\sigma_{t+1}^2 = \omega_\sigma + \phi_\sigma \sigma_t^2 + \kappa_\sigma s_t^{(\sigma)}, \tag{7}$$

where $\omega_\sigma = \frac{\delta_\sigma}{1-\phi_\sigma}$ and $s_t = \mathcal{I}_t^{-1}\nabla_t$, such that $s \sim iid\,(0, \mathcal{I}^{-1})$.[7] We can define

$$\varrho_{t+1} = \omega_\varrho + \kappa_{\varrho,0} s_t + \kappa_{\varrho,1} s_{t-1} + \cdots + \kappa_{\varrho,P} s_{t-P}, \tag{8}$$

for $P < \infty$, and we can verifying that $\varrho_t = \omega_\varrho$, $\forall 0 < t < T$ testing the following hypotheses:[8]

$$\mathcal{H}_0 : \kappa_{\varrho,0} = \kappa_{\varrho,1} = \cdots = \kappa_{\varrho,P};$$
$$\mathcal{H}_1 : \kappa_{\varrho,i} \neq 0,\ i = 1, \ldots, P,$$

by means of the Lagrange multiplier (LM) principle (see, e.g., Engle, 1982).

Let us now define $\theta = [\omega_\varrho\ \psi_\sigma\ \nu]'$, a vector of all the static parameters with $\psi_\sigma = [\omega_\sigma\ \phi_\sigma\ \kappa_\sigma]'$, and $\boldsymbol{\kappa} = [\kappa_{\varrho,0}\ \ldots\ \kappa_{\varrho,P}]'$. Harvey (2013) and Harvey and Thiele (2016) show that when testing for $\mathcal{H}_0$ in a model where another parameter varies over time, the test statistic takes the form[9]

$$\tilde{\zeta}_\varrho(P) = \frac{1}{T}\nabla_{\boldsymbol{\kappa}'}\mathcal{I}_{\boldsymbol{\kappa\kappa}}^{-1}\nabla_{\boldsymbol{\kappa}} + \frac{1}{T}\nabla_{\boldsymbol{\kappa}'}\mathcal{I}_{\boldsymbol{\kappa\kappa}}^{-1}\mathcal{I}_{\boldsymbol{\kappa\theta}}\left(\mathcal{I}_{\boldsymbol{\theta\theta}} - \mathcal{I}_{\boldsymbol{\theta\kappa}}\mathcal{I}_{\boldsymbol{\kappa\kappa}}^{-1}\mathcal{I}_{\boldsymbol{\kappa\theta}}\right)\mathcal{I}_{\boldsymbol{\theta\kappa}}\mathcal{I}_{\boldsymbol{\kappa\kappa}}^{-1}\nabla_{\boldsymbol{\kappa}}. \tag{9}$$

Proposition 4. *In a Skt_ν model with time-varying (squared) scale, the LM test for conditional skewness reduces to*

$$\xi_\varrho(P) = T\sum_{j=0}^{P}\rho^2(j), \tag{10}$$

the Box and Pierce (1970), where $\rho(j)$ is the j^{th} sample autocorrelation of ∇_ϱ.

Proof. See Section 12. ∎

[7] If the innovation distribution is Gaussian (e.g., $\eta \to 0, \varrho = 0$), Equation (7) reduces to the GARCH(1,1) of Bollerslev (1986).

[8] Equation (8) is equivalent to the $MA(P)$ representation of a covariance stationary ($|\phi_\varrho| < 1$) first-order score-driven model for ϱ_t.

[9] See also Calvori et al. (2017).

Table 1: Size and power of the test

	ω_ϱ	*Size*				*Power*							
			10%	5%	1%		10%	5%	1%		10%	5%	1%
		BNp	DGP1			*BNp*	DGP2			*BNp*	DGP3		
T=500	0	0.214	0.116	0.066	0.024	0.178	0.844	0.909	0.962	0.198	0.830	0.890	0.952
	-0.3		0.119	0.070	0.024		0.863	0.917	0.963		0.877	0.933	0.975
	0.6		0.116	0.067	0.023		0.862	0.912	0.961		0.861	0.920	0.975
T=1000	0	0.219	0.107	0.056	0.020	0.208	0.834	0.896	0.953	0.222	0.851	0.907	0.964
	-0.3		0.103	0.058	0.019		0.841	0.903	0.960		0.812	0.889	0.955
	0.6		0.115	0.064	0.020		0.840	0.903	0.959		0.864	0.925	0.975
T=2500	0	0.217	0.105	0.058	0.017	0.220	0.757	0.831	0.920	0.223	0.689	0.783	0.898
	-0.3		0.105	0.054	0.016		0.756	0.834	0.921		0.784	0.869	0.943
	0.6		0.100	0.055	0.016		0.761	0.839	0.925		0.769	0.862	0.951

Note: The table reports the *p-values* relative to the test in Equation (10) for three DGPs, three values of the unconditional asymmetry parameter and three sample sizes. The first DGP assumes $\varrho = \omega_\varrho$, and therefore is used to evaluate the size of the test. DGP2 and DGP5 assumes a single break and continuous time variation respectively, and are therefore used to evaluate the power of the test. For the cases in which $\omega_\varrho = 0$ we also report the *p-value* for Bai and Ng (2005) test for unconditional skewness. All values are obtained from 5000 simulations.

The test is distributed as a χ^2 with $P + 1$ degrees of freedom.

5 Monte Carlo evidence

Here, we evaluate the properties of the size and power of the test via Monte Carlo simulations. Data are generated from a Skt distribution with $\mu = 0$, $\nu = 5$, and scale simulated from an $AR(1)$ process

$$\sigma_t^2 = \omega + \varphi\sigma_{t-1}^2 + \eta_t,\ \eta \sim \mathcal{N}(0, \sigma_\eta^2),$$

with autocorrelation $\varphi = 0.99$ and $\sigma_\eta = 0.2$.

We consider three data generating processes (DGPs) for the asymmetry parameter: i) fix $\varrho = \omega_\varrho$, ii) a single break in ϱ with an jump of 0.2, and iii) a time-varying ϱ, simulated from an $AR(1)$ with an half-line of about 7 periods ($\phi_\varrho = 0.9$) and Gaussian innovations with standard deviation of 0.05. We repeat each GDP for three values of the unconditional asymmetry: $\omega_\varrho = 0, 0.3, -0.5$, and $T = 500, 1000, 2500$ observations. We consider 5000 replication for each combination.

For each replication we estimate the model via Maximum Likelihood techniques; Blasques et al. (2022) establishes the necessary conditions to ensure the stationarity, ergodicity and invertibility for the filter in Equation (7).[10] The model are estimated under the null hypothesis of no serial

[10]Conditional on ϱ, the score in Equation (7) reduces to the one considered in Blasques et al. (2022).

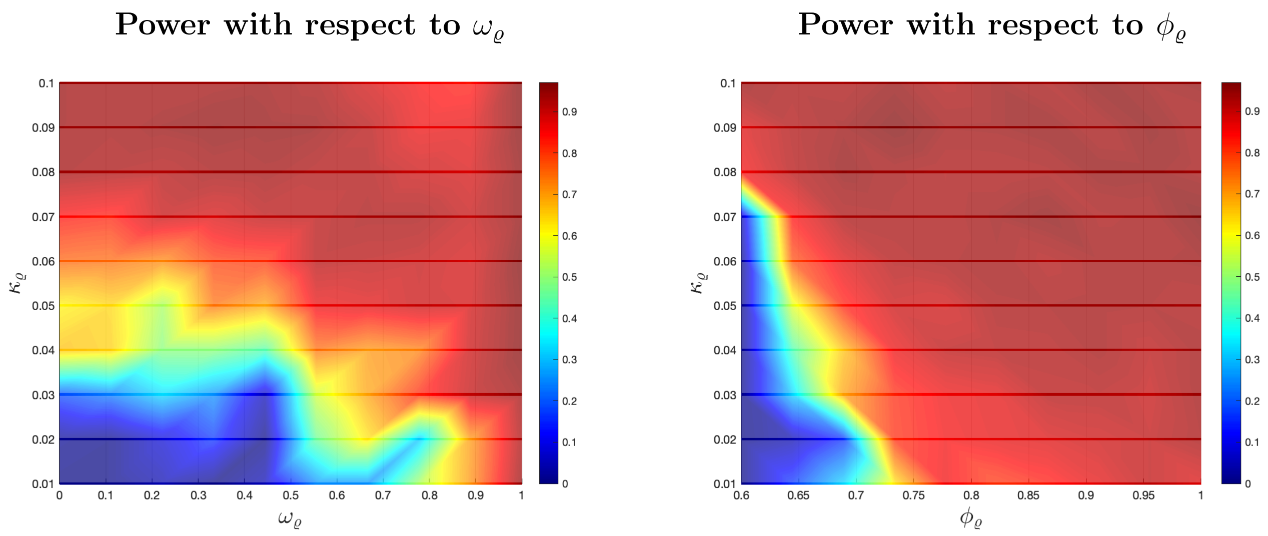

Figure 3: Power of the test

Note: The panels report the power of the test, marked by the color-bar, at the 5% nominal level for combinations of the (ω_ϱ, κ_ϱ) and (ϕ_ϱ, κ_ϱ). Results are obtained from 2000 simulations with T=1000.

correlation in skewness, and Equation (10) is used to test the conditional scaled score for the asymmetry parameter, $s_{\varrho,t}$. We use the automatic selection procedure for the number of lags P introduced by Escanciano and Lobato (2009). This approach uses an information criterion to select the optimal number of lags for the Portmanteau test. Asymptotically, only the first lag is selected, and therefore the test distribution is χ^2_1. When appropriate, we also test for the presence of unconditional skewness in the sample by means of Bai and Ng (2005) test. The results from the exercise are summarised in Table 1. First, we evaluate the size of the test. The first (left) part of the Table reports the empirical probability of rejection for DGP1, which assumes constant asymmetry, for conventional rejection levels. The test appears slightly oversized for $T = 500$; however, nominal levels are obtained as T increases. Different levels of the unconditional asymmetry do not affect the size of the test. When $\omega_\varrho = 0$, Bai and Ng (2005) test for unconditional skewness fails to reject the null hypothesis. Second, we analyze the power of the test by means of DGP 2 and 3. These DGPs represent very different scenarios for ϱ: in DGP 2 the parameter experiences instability only via a single break, whereas in DGP 3 the asymmetry is generated by a first-order autoregressive process, implying continuous variation of the parameter. Overall, the test shows reassuring power values for both DGPs. We further evaluate the power of the test in Figure 3. The panels compare the power of the test compute at the 5% nominal level for different levels of the innovation standard deviation (σ_ϱ) and of the unconditional asymmetry parameter (ω_ϱ, left panel) and of the autocorrelation parameter (ϕ_ϱ, right panel). Already for $\eta_\varrho = 0.04$ the test presents acceptable power levels, especially in cases where the unconditional level is above 0.5 and the autocorrelation in just above 0.7. Above these values, the power fast approaches 1 for any combination of the parameters.

Table 2: Time-varying skewness test for Fama-French 3 factors

	Daily			*Monthly*		
	Rm	SMB	HML	Rm	SMB	HML
	Skewness					
$Skew$	-0.827	-1.032	0.184	-0.743	0.147	-0.093
$E\text{-}Skt\text{-}t$	-0.181	-0.142	0.062	-0.629	0.117	-0.006
$QSkew_5$	-0.064	-0.053	0.027	-0.200	0.030	0.055
$QSkew_1$	-0.048	-0.046	0.016	-0.161	-0.014	0.035
ϱ	-0.063	-0.055	0.024	-0.212	0.049	-0.002
	Test					
BNp	0.036	0.078	0.150	0.006	0.204	0.357
LMp	0.000	0.000	0.000	0.685	0.141	0.000

Note: The top part of the table reports skewness statistics for the considered anomalies. Specifically, we report the sample skewness (Skew), the estimated skewness from the *Epsilon-Skew-t* model ($E-Skt$), Groeneveld and Meeden (1984) quantile skewness at the 5^{th} and 1^{th} quantiles, and the estimated asymmetry parameter, ϱ. In the bottom part we report the *p-values* for Bai and Ng (2005) test for unconditional skewness (BNp), and for the LM test for time-varying skewness (LMp) developed in Section 4.

6 Testing market anomalies

In this Section we propose an application of the test developed above to some of the well-known market anomalies identified in the financial literature (Fama and French, 1992, 1995; Carhart, 1997; Pástor and Stambaugh, 2003, see, e.g.,). Whereas there exists a multitude of such anomalies, the so called *factor zoo* (Cochrane, 2011), in what follow we will focus on the three factors commonly used in the 3-factor regression (3FFF, Fama and French, 1993): the excess return on the market portfolio (Rm), the small-minus-big (SMB) and the high-minus-low (HML). We estimate a score-driven model with time-varying volatility, as in Equation (7), and *Epsilon-Skew-t* innovations,[11] and we test the resulting conditional scaled score for the asymmetry parameter. We run the exercise both at the daily and monthly frequency, for a sample that goes from 1960 to 2022.

First, we analyze the skewness properties of the data. The top panel of Table 2 reports several measures of skewness for the anomalies. Specifically, we report the sample skewness (Skew), the estimated skewness from the *Epsilon-Skew-t* model ($E - Skt$) computes as per Equation (5), two robust measures of skewness, $QSkew_5$ and $QSkew_2$, and the estimated asymmetry parameter, ϱ. It's important to notice that quantile skewness measures (Groeneveld and Meeden, 1984) are bounded in the $[-1\ 1]$ region, similarly to the asymmetry parameter, and therefore are directly comparable. Starting from the daily frequency, the values show a strong presence of negative

[11] Creal et al. (2013) show that when the innovations belong to the class of t densities, the score-driven filter for volatility downplays outliers more aggressively than in the t-GARCH of Bollerslev (1987).

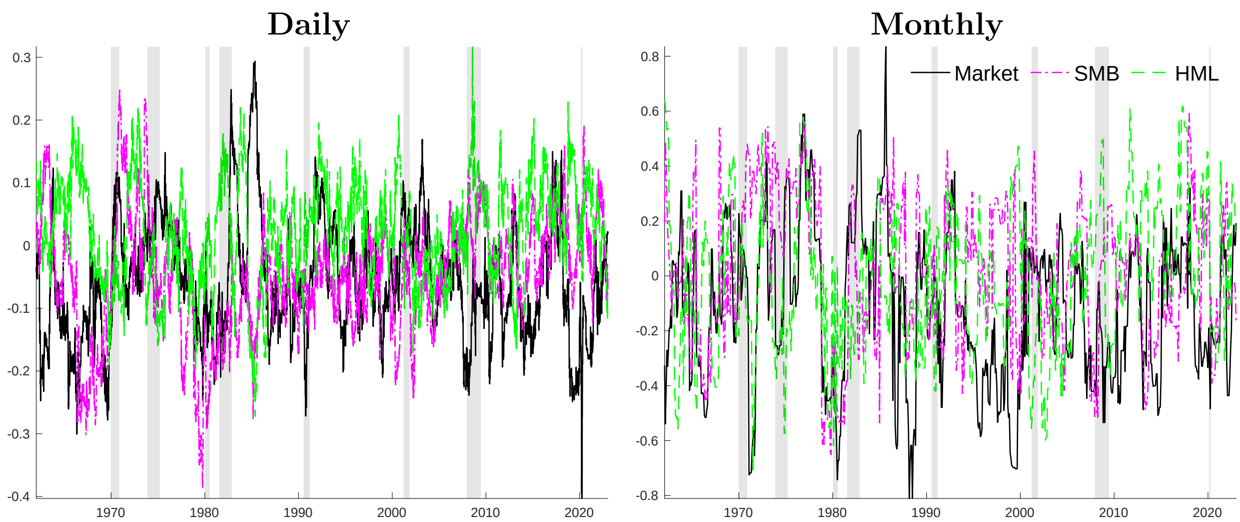

Figure 4: Time-varying skewness

Note: The plots report the time-varying quantile skewness for the three anomalies estimated as 1-year rolling window both for daily and monthly returns.

skewness for the market and SMB, whereas mildly positive asymmetry for HML. Evidence at the monthly frequency appears less striking, with SMB and HML presenting discording signs over different skewness measures. It is however interesting to notice that the skewness measures by the asymmetry parameter is largely consistent with quantile skewness values at the 5^{th} quantile. Before proceeding with the testing, to help visualize these results, in Figure 4 we report rolling measure of quantile skewness computed over windows of 1 year for the daily and monthly returns. The patterns are largely consistent across frequencies.[12]

The bottom rows of Table 2 report the p-values for the Bai and Ng (2005) test for unconditional skewness (BNp) and that for the LM test derived in Section 4 (LMp). At the daily frequency, the presence of unconditional skewness cannot be rejected for the market and SMB, but the LM test shows overwhelming evidence in favor of conditional asymmetry in all three factors. At the monthly frequency, on the other hand, the picture looks rather different. The market seems to feature significant unconditional skewness, but this does not present a strong autocorrelation; SMB is the only one for which skewness is rejected both conditionally and unconditionally, whereas for HML our test strongly rejects the null hypothesis.

7 Conclusions

This paper proposed a review of the *Epsilon-Skew-t* distribution first introduced by Arellano-Valle et al. (2005). Derivations for the probability, distribution and quantile functions are provided,

[12]These plots should be taken as guidance only, as autocorrelation in the series is implicitly inherited by the rolling-windows.

as well as for the moments. We also provide a detailed derivation of the score vector and the associated Information matrix, exploiting the relation between the t distribution and the Beta distribution, which allows to easily evaluate expectations for functions of the (skew-) t variate. Interest for this distribution has recently emerged in the economic and financial literature due to its nice properties and easiness to use in time series applications when skewness is varying over time (see, e.g., Lucas and Zhang, 2016; Delle Monache et al., 2021). Within this context, we also develop a parametric test to evaluate the presence of autocorrelated skewness in time series data. The test belongs to the class of LM tests, and in a context with fixed location, which is rather common in financial applications, the test collapses to a Box-Pierce test on the autocorrelation of the conditional score of the asymmetry parameter. The test presents good asymptotic properties, as evaluated by means of a Monte Carlo simulation exercise.

The test is applied to the three most common asset pricing anomalies. Results point at an overwhelming presence of time-varying asymmetry in these anomalies at the daily frequency, whereas only modest evidence arises at the monthly frequency. Recently, Barroso and Maio (2021) have analysed the risk-return trade-off for some of the well-known market anomalies, reporting mixed evidence. While they show that the 3FFF present some degree of sample skewness, our results should be taken as a warning sign that inconclusive evidence about the presence of a risk-return trade-off could be plagued by the presence of conditional skewness, a point already illustrated by Theodossiou and Savva (2016) in the case of the market portfolio and by Bianchi et al. (2022) for momentum returns.

8 Proof of Proposition 1

Following Kim (2008), the moments of the Ht_ν can be obtained as:

$$\mathbb{E}[Z^{k+2}] = \mathbb{E}_\eta[\eta^{-\frac{k+2}{2}} V^{k+2}], \quad k = -1, 0, 1, \ldots, \tag{8.1}$$

where

$$V^{k+2} = -\frac{(\sqrt{\eta}b)^{k+1}\phi(\sqrt{\eta}b) - (\sqrt{\eta}a)^{k+1}\phi(\sqrt{\eta}a)}{F_\nu(b) - F_\nu(a)} + (k+1)V^k.$$

We can now compute Equation (8.1) for $k = -1, 0, 1, 2$:

- $k = -1$:

$$\mathbb{E}[Z] = 2E_\eta[\eta^{-\frac{1}{2}}\phi(\sqrt{\eta}\cdot 0)] = 2\frac{\Gamma\left(\frac{\nu-1}{2}\right)\nu^{\frac{\nu}{2}}}{2\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)}\nu^{-\frac{\nu-1}{2}} = \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)}\sqrt{\nu};$$

- $k = 0$:

$$\mathbb{E}[Z^2] = \mathbb{E}_\eta[\eta^{-1}] = \frac{\nu}{\nu - 2};$$

- $k = 1$:

$$\begin{aligned}\mathbb{E}[Z^3] = 2\,\mathbb{E}_\eta[\eta^{-\frac{3}{2}}V] = 2\,\mathbb{E}_\eta[\eta^{-\frac{3}{2}}2\phi(\sqrt{\eta}\cdot 0)] &= 4\frac{\Gamma\left(\frac{\nu-3}{2}\right)\nu^{\frac{\nu}{2}}}{4\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)}\nu^{-\frac{\nu-3}{2}} \\ &= \frac{2\Gamma\left(\frac{\nu-1}{2}\right)\nu^{\frac{3}{2}}}{(\nu-3)\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)}.\end{aligned}$$

- $k = 2$

$$\mathbb{E}[Z^4] = 3\,\mathbb{E}_\eta[\eta^{-2}] = 3\frac{\nu^2}{(\nu-4)(\nu-2)}.$$

Hence, $d_1(\nu) = \sqrt{\nu}G(\nu)$, $d_2(\nu) = \frac{\nu}{\nu-2}$, $d_3(\nu) = \frac{2\nu^{\frac{3}{2}}}{\nu-3}G(\nu)$, and $d_4(\nu) = \frac{3\nu^2}{(\nu-4)(\nu-2)}$ with $G(\nu) = \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{1}{2})}$. Given $\frac{\nu-1}{2}\Gamma(\frac{\nu-1}{2}) = \Gamma(\frac{\nu+1}{2})$, then

$$G(\nu) = \frac{2\Gamma(\frac{\nu+1}{2})}{(\nu-1)\Gamma(\frac{\nu}{2})\Gamma(\frac{1}{2})} = \frac{2\nu^{\frac{1}{2}}}{\nu-1}\mathcal{C}(\nu).$$

Therefore, $d_1(\nu) = \frac{2\nu}{\nu-1}\mathcal{C}(\nu)$, $d_2(\nu) = \frac{\nu}{\nu-2}$, $d_3(\nu) = \frac{4\nu^2}{(\nu-3)(\nu-1)}\mathcal{C}(\nu)$, and $d_4(\nu) = \frac{3\nu^2}{(\nu-4)(\nu-2)}$.

9 Proof of Proposition 2

Location The gradient with respect to the location parameter reads:

$$\nabla_\mu = \frac{1}{\sigma}w\zeta.$$

Proof.

$$\begin{aligned}\nabla_\mu = \frac{\partial \ell}{\partial \mu} &= -\left(\frac{1+\eta}{2\eta}\right)\frac{1}{m(\mu,\sigma,\varrho)}\frac{\partial m(\mu,\sigma,\varrho)}{\partial \mu} \\ &= \left(\frac{1+\eta}{2\eta}\right)\frac{h^2}{h^2+\eta\zeta^2}\frac{2\eta\zeta}{h^2\sigma} \\ &= \frac{1}{\sigma}w\zeta.\end{aligned}$$

■

Squared scale The gradient with respect to the squared scale parameter reads:

$$\nabla_{\sigma^2} = \frac{1}{2\sigma^2}(w\zeta^2 - 1).$$

Proof.

$$\begin{aligned}
\nabla_{\sigma^2} = \frac{\partial \ell}{\partial \sigma^2} &= -\frac{1}{2\sigma^2} - \left(\frac{1+\eta}{2\eta}\right)\frac{1}{g(\mu,\sigma^2,\varrho)}\frac{\partial g(\mu,\sigma^2,\varrho)}{\partial \sigma^2} \\
&= -\frac{1}{2\sigma^2} + \left(\frac{1+\eta}{2\eta}\right)\frac{h^2}{h^2+\eta\zeta^2}\frac{\eta\zeta^2}{h^2\sigma^2} \\
&= \frac{1}{2\sigma^2}\left[\frac{(1+\eta)\zeta^2}{h^2+\eta\zeta^2} - 1\right] \\
&= \frac{1}{2\sigma^2}(w\zeta^2 - 1)
\end{aligned}$$

■

Shape The gradient with respect to the shape parameter reads:

$$\nabla_{\varrho,t} = \frac{sgn(\varepsilon_t)}{(1+sgn(\varepsilon_t)\varrho_t)}w_t\zeta_t^2$$

Proof.

$$\begin{aligned}
\nabla_{\varrho,t} = \frac{\partial \ell_t}{\partial \varrho_t} &= -\left(\frac{1+\eta}{2\eta}\right)\frac{1}{g(\mu_t,\sigma_t,\varrho_t)}\frac{\partial g(\mu_t,\sigma_t,\varrho_t)}{\partial \varrho_t} \\
&= -\left(\frac{1+\eta}{2\eta}\right)\frac{h_t^2}{h_t^2+\eta\zeta_t^2}\left(-\frac{2\eta\zeta_t^2 sgn(\varepsilon_t)}{h_t^3}\right) \\
&= \frac{(1+\eta)}{h_t^2+\eta\zeta_t^2}\frac{sgn(\varepsilon_t)}{h_t}\zeta_t^2 \\
&= \frac{sgn(\varepsilon_t)}{(1+sgn(\varepsilon_t)\varrho_t)}w_t\zeta_t^2
\end{aligned}$$

■

10 Useful properties of the Beta and t distributions

Consider the following results:

Corollary 1. *Given a beta distributed variable, $b \sim \mathcal{B}(\alpha,\beta)$, then*

$$\mathbb{E}[b^h(1-b)^k] = \frac{B(\alpha+h,\beta+k)}{B(\alpha,\beta)},$$

where $B(\alpha,\beta)$ is the beta function (Harvey, 2013, pag. 23).

Corollary 2. *Let $T \sim t_{1/\eta}$ be a t-distributed random variable with $\frac{1}{\eta}$ degrees of freedom. We can define*

$$b = \frac{\eta T^2}{1+\eta T^2} \sim \mathcal{B}\left(\frac{1}{2}, \frac{1}{2\eta}\right),$$

and

$$1-b = \frac{1}{1+\eta T^2} \sim \mathcal{B}\left(\frac{1}{2\eta}, \frac{1}{2}\right),$$

(Harvey, 2013, pag. 25).

Corollary 3. *Let $T \sim t_{1/\eta}$ be a t-distributed random variable with $\frac{1}{\eta}$ degrees of freedom. Then,*

$$\frac{\sqrt{\eta}T}{1+\eta T^2} = \frac{\sqrt{\eta}T}{\sqrt{1+\eta T^2}}\frac{1}{\sqrt{1+\eta T^2}} = b^{\frac{1}{2}}(1-b)^{\frac{1}{2}}.$$

Notable results Following the result in Corollary 1, we can compute some notable expectations of the form $\mathbb{E}[b^h(1-b)^k]$, where b is defined as in Corollary 2:

$$\mathbb{E}[b] = \frac{B\left(\frac{3}{2},\frac{1}{2\eta}\right)}{B\left(\frac{1}{2},\frac{1}{2\eta}\right)} = \frac{\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2\eta}\right)}{\Gamma\left(\frac{1+3\eta}{2\eta}\right)}\frac{\Gamma\left(\frac{1+\eta}{2\eta}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2\eta}\right)} = \frac{\eta}{1+\eta}, \tag{10.1}$$

$$\mathbb{E}[b^2] = \frac{B\left(\frac{5}{2},\frac{1}{2\eta}\right)}{B\left(\frac{1}{2},\frac{1}{2\eta}\right)} = \frac{\frac{3}{4}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2\eta}\right)}{\Gamma\left(\frac{1+5\eta}{2\eta}\right)}\frac{\Gamma\left(\frac{1+\eta}{2\eta}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2\eta}\right)} = \frac{3\eta^2}{(1+3\eta)(1+\eta)}, \tag{10.2}$$

$$\mathbb{E}[b^{\frac{1}{2}}(1-b)^{\frac{1}{2}}] = \frac{B\left(1,\frac{1+\eta}{2\eta}\right)}{B\left(\frac{1}{2},\frac{1}{2\eta}\right)} = \frac{\Gamma(1)\Gamma\left(\frac{1+\eta}{2\eta}\right)}{\Gamma\left(\frac{1+\eta}{2\eta}+1\right)}\frac{\Gamma\left(\frac{1+\eta}{2\eta}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2\eta}\right)} = \frac{2\eta}{1+\eta}\frac{\Gamma\left(\frac{1+\eta}{2\eta}\right)}{\sqrt{\pi}\Gamma\left(\frac{1}{2\eta}\right)} = \frac{2\sqrt{\eta}}{1+\eta}\mathcal{C}, \tag{10.3}$$

$$\mathbb{E}[b(1-b)] = \frac{B\left(\frac{3}{2},\frac{1+2\eta}{2\eta}\right)}{B\left(\frac{1}{2},\frac{1}{2\eta}\right)} = \frac{\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\frac{1}{2\eta}\Gamma\left(\frac{1}{2\eta}\right)}{\Gamma\left(\frac{1+5\eta}{2\eta}\right)}\frac{\Gamma\left(\frac{1+\eta}{2\eta}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2\eta}\right)} = \frac{\eta}{(1+3\eta)(1+\eta)}, \tag{10.4}$$

$$\mathbb{E}[b^{\frac{3}{2}}(1-b)^{\frac{1}{2}}] = \frac{B\left(2,\frac{1+\eta}{2\eta}\right)}{B\left(\frac{1}{2},\frac{1}{2\eta}\right)} = \frac{\Gamma\left(\frac{1+\eta}{2\eta}\right)}{\Gamma\left(\frac{1+5\eta}{2\eta}\right)}\frac{\Gamma\left(\frac{1+\eta}{2\eta}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2\eta}\right)} = \frac{4\mathcal{C}\sqrt{\eta}\eta}{(1+3\eta)(1+\eta)}, \tag{10.5}$$

recognizing that $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, and $\Gamma\left(\frac{1+5\eta}{2\eta}\right) = \frac{1+3\eta}{2\eta}\Gamma\left(\frac{1+3\eta}{2\eta}\right) = \frac{(1+3\eta)(1+\eta)}{4\eta^2}\Gamma\left(\frac{1+\eta}{2\eta}\right)$, $\Gamma(2) = \Gamma(1) = 1$, and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. We will extensively use these results in the following derivations.

11 Proof of Proposition 3

$\mathcal{I}_{\mu,\mu}$.

$$\mathcal{I}_{\mu,\mu} = \mathbb{E}[\nabla_\mu \nabla_\mu'] = \frac{1}{\sigma^2}\mathbb{E}\left[w^2\zeta^2\right],$$

taking advantage of of the two-piece representation we get:

$$\begin{aligned}\mathcal{I}_{\mu,\mu} &= P(y \geq \mu)\,\mathbb{E}\left[\nabla_{\mu,+}\nabla_{\mu,+}\right] + P(y < \mu)\,\mathbb{E}\left[\nabla_{\mu,-}\nabla_{\mu,-}\right] \\ &= \frac{1}{\sigma^2}\left(P(y \geq \mu)\,\mathbb{E}[(w^2\zeta^2)_+] + P(y < \mu)\,\mathbb{E}[(w^2\zeta^2)_-]\right) \\ &= \frac{1}{\sigma^2}\,\mathbb{E}[w^2\zeta^2].\end{aligned}$$

Notice that:

$$w^2\zeta^2 = \frac{(1+\eta)^2}{\eta}\frac{\eta\zeta^2}{h^2+\eta\zeta^2}\frac{1}{h^2+\eta\zeta^2} = \frac{(1+\eta)^2\sigma^2}{\eta}\frac{\eta\varepsilon^2}{\sigma^2h^2+\eta\varepsilon^2}\frac{1}{\sigma^2h^2+\eta\varepsilon^2};$$

evaluating over the positive (negative) semi-support:

$$(w^2\zeta^2)_+ = \frac{(1+\eta)^2}{\eta(1+\varrho)^2}\frac{\eta\zeta_+^2}{1+\eta\zeta_+^2}\frac{1}{1+\eta\zeta_+^2}, \quad (w^2\zeta^2)_- = \frac{(1+\eta)^2}{\eta(1-\varrho)^2}\frac{\eta\zeta_-^2}{1+\eta\zeta_-^2}\frac{1}{1+\eta\zeta_-^2}.$$

Using Corollaries 1 and 2 and Equation (10.4) we have:

$$\mathbb{E}[(w^2\zeta^2)_+] = \frac{(1+\eta)}{(1+3\eta)(1+\varrho)^2}, \quad \mathbb{E}[(w^2\zeta^2)_-] = \frac{(1+\eta)}{(1+3\eta)(1-\varrho)^2},$$

and thus,

$$\begin{aligned}\mathcal{I}_{\mu,\mu} &= \frac{(1+\eta)}{\sigma^2(1+3\eta)}\left(\frac{P(y \geq \mu)}{(1-\varrho)^2} + \frac{P(y < \mu)}{(1-\varrho)^2}\right) \\ &= \frac{(1+\eta)}{\sigma^2(1+3\eta)(1+\varrho^2)}.\end{aligned}$$

■

$\mathcal{I}_{\sigma^2,\sigma^2}$.

$$\begin{aligned}\mathcal{I}_{\sigma^2,\sigma^2} = \mathbb{E}[\nabla_{\sigma^2}\nabla'_{\sigma^2}] &= \frac{1}{4\sigma^4}\,\mathbb{E}\left[(w\zeta^2-1)^2\right] \\ &= \frac{1}{4\sigma^4}\,\mathbb{E}\left[1-2w\zeta^2+w^2\zeta^4\right] \\ &= \frac{1}{4\sigma^4}\left(\mathbb{E}\left[w^2\zeta^4\right]-1\right),\end{aligned}$$

as per **??**. Using the two-piece representation:

$$\begin{aligned}\mathcal{I}_{\sigma^2,\sigma^2} &= P(y \geq \mu)\,\mathbb{E}\left[\nabla_{\sigma^2,+}\nabla'_{\sigma^2,+}\right] + P(y < \mu)\,\mathbb{E}\left[\nabla_{\sigma^2,-}\nabla'_{\sigma^2,-}\right] \\ &= \frac{1}{4\sigma^4}\left(P(y \geq \mu)\,\mathbb{E}[(w^2\zeta^4)_+] + P(y < \mu)\,\mathbb{E}[(w^2\zeta^4)_-] - 1\right).\end{aligned}$$

Notice that:

$$w^2\zeta^4 = \frac{(1+\eta)^2}{\eta^2}\left(\frac{\eta\zeta^2}{h^2+\eta\zeta^2}\right)^2 = \frac{(1+\eta)^2}{\eta^2}\left(\frac{\eta\varepsilon^2}{\sigma^2h^2+\eta\varepsilon^2}\right)^2,$$

and evaluating over the positive (negative) semi-support:

$$(w^2\zeta^4)_+ = \frac{(1+\eta)^2}{\eta^2}\left(\frac{\eta\zeta_+^2}{1+\eta\zeta_+^2}\right)^2, \quad (w^2\zeta^4)_- = \frac{(1+\eta)^2}{\eta^2}\left(\frac{\eta\zeta_-^2}{1+\eta\zeta_-^2}\right)^2.$$

Using Corollaries 1 and 2 and Equation (10.2) we have:

$$\mathbb{E}[(w^2\zeta^4)_+] = \mathbb{E}[(w^2\zeta^4)_-] = \frac{(1+\eta)^2}{\eta^2}\,\mathbb{E}\left[b^2\right] = \frac{3(1+\eta)}{(1+3\eta)} \tag{11.1}$$

Thus,

$$\begin{aligned}\mathcal{I}_{\sigma^2,\sigma^2} &= \frac{1}{4\sigma^2}\left(\frac{3(1+\eta)}{(1+3\eta)}-1\right)\\ &= \frac{1}{2\sigma^4(1+3\eta)}\end{aligned}$$

■

$\mathcal{I}_{\varrho,\varrho}$.

$$\mathcal{I}_{\varrho,\varrho} = \mathbb{E}[\nabla_\varrho\nabla_\varrho'] = \frac{1}{h^2}\,\mathbb{E}\left[w^2\zeta^4\right].$$

Using the two-piece representation and Equation (11.1) we can write:

$$\begin{aligned}\mathcal{I}_{\varrho,\varrho} &= P(y\geq\mu)\,\mathbb{E}\left[\nabla_{\varrho,+}\nabla_{\varrho,+}\right] + P(y<\mu)\,\mathbb{E}\left[\nabla_{\varrho,-}\nabla_{\varrho,-}\right]\\ &= \frac{P(y\geq\mu)}{(1+\varrho)^2}\,\mathbb{E}\left[(w^2\zeta^4)_+\right] + \frac{P(y<\mu)}{(1-\varrho)^2}\,\mathbb{E}\left[(w^2\zeta^4)_-\right]\\ &= \frac{3(1+\eta)}{(1+3\eta)(1-\varrho^2)}.\end{aligned}$$

■

$\mathcal{I}_{\mu,\varrho}$.

$$\mathcal{I}_{\mu,\varrho} = \mathbb{E}[\nabla_\mu\nabla_\varrho'] = \frac{sgn(\varepsilon)}{h\sigma}\,\mathbb{E}\left[w^2\zeta^3\right]$$

Using the two-piece representation we have:

$$\begin{aligned}\mathcal{I}_{\varrho,\mu} &= P(y\geq\mu)\,\mathbb{E}\left[\nabla_{\varrho,+}\nabla_{\mu,+}\right] + P(y<\mu)\,\mathbb{E}\left[\nabla_{\varrho,-}\nabla_{\mu,-}\right]\\ &= \frac{P(y\geq\mu)}{(1+\varrho)\sigma}\,\mathbb{E}\left[(w^2\zeta^3)_+\right] - \frac{P(y<\mu)}{(1-\varrho)\sigma}\,\mathbb{E}\left[(w^2\zeta^3)_-\right]\end{aligned}$$

$$= \frac{1}{2\sigma}\left(\mathbb{E}\left[(w^2\zeta^3)_+\right] - \mathbb{E}\left[(w^2\zeta^3)_-\right]\right).$$

Notice that:

$$w^2\zeta^3 = \frac{(1+\eta)^2}{\eta\sqrt{\eta}}\frac{\eta\zeta^2}{h^2+\eta\zeta^2}\frac{\sqrt{\eta}\zeta}{h^2+\eta\zeta^2} = \frac{(1+\eta)^2}{\eta\sqrt{\eta}}\frac{\eta\varepsilon^2}{\sigma^2h^2+\eta\varepsilon^2}\frac{\sqrt{\eta}\sigma\varepsilon}{\sigma^2h^2+\eta\varepsilon^2}$$

and evaluating over the positive (negative) semi-support:

$$(w^2\zeta^3)_+ = \frac{1}{(1+\varrho)}\frac{(1+\eta)^2}{\eta\sqrt{\eta}}\frac{\eta\zeta_+^2}{1+\eta\zeta_+^2}\frac{\sqrt{\eta}\zeta_+}{1+\eta\zeta_+^2}, \quad (w^2\zeta^3)_- = -\frac{1}{(1-\varrho)}\frac{(1+\eta)^2}{\eta\sqrt{\eta}}\frac{\eta\zeta_-^2}{1+\eta\zeta_-^2}\frac{\sqrt{\eta}|\zeta_-|}{1+\eta\zeta_-^2},$$

for $\zeta_- = -|\zeta_-|$. Using Corollaries 1 and 2 and Equation (10.5) we get:

$$\mathbb{E}[(w^2\zeta^3)_+] = \frac{4\mathcal{C}(1+\eta)}{(1+\varrho)(1+3\eta)}, \quad \mathbb{E}[(w^2\zeta^3)_-] = -\frac{4\mathcal{C}(1+\eta)}{(1-\varrho)(1+3\eta)}; \tag{11.2}$$

therefore,

$$\mathcal{I}_{\varrho,\mu} = \frac{4\mathcal{C}(1+\eta)}{\sigma(1+3\eta)(1-\varrho^2)}$$

■

$\mathcal{I}_{\mu,\sigma^2}$.

$$\mathcal{I}_{\mu,\sigma^2} = \mathbb{E}\left[\nabla_\mu\nabla_{\sigma^2}\right] = \frac{1}{2\sigma^3}\{\mathbb{E}[w^2\zeta^3] - \mathbb{E}[w\zeta]\}$$
$$= \frac{1}{2\sigma^3}\mathbb{E}[w^2\zeta^3],$$

as per equation **??**. Using the two-piece representation and Equation (11.2) we obtain:

$$\mathcal{I}_{\mu,\sigma^2} = P(y\geq\mu)\,\mathbb{E}[\nabla_{\mu,+}\nabla_{\sigma^2,+}] + P(y<\mu)\,\mathbb{E}[\nabla_{\mu,+}\nabla_{\sigma^2,-}]$$
$$= \frac{1}{2\sigma^3}\left(P(y\geq\mu)\,\mathbb{E}\left[(w^2\zeta^3)_+\right] + P(y<\mu)\,\mathbb{E}\left[(w^2\zeta^3)_-\right]\right) = 0.$$

■

$\mathcal{I}_{\varrho,\sigma^2}$.

$$\mathcal{I}_{\varrho,\sigma^2} = \mathbb{E}\left[\nabla_\varrho\nabla_{\sigma^2}\right] = \frac{sgn(\varepsilon)}{2h\sigma^2}\left(\mathbb{E}[w^2\zeta^4] - \mathbb{E}[w\zeta^2]\right).$$

Using the two-piece representation we get:

$$\mathcal{I}_{\varrho,\sigma^2} = P(y\geq\mu)\,\mathbb{E}\left[\nabla_{\varrho,+}\nabla_{\sigma^2,+}\right] + P(y<\mu)\,\mathbb{E}\left[\nabla_{\varrho,-}\nabla_{\sigma^2,-}\right]$$

$$= \frac{1}{4\sigma^2}\left(\mathbb{E}\left[(w^2\zeta^4)_+\right] - \mathbb{E}\left[(w^2\zeta^4)_-\right]\right) - \frac{1}{4\sigma^2}\left(\mathbb{E}\left[(w\zeta^2)_+\right] - \mathbb{E}\left[(w\zeta^2)_-\right]\right) = 0,$$

after using **??** and Equation (11.1). ■

12 Proof of Proposition 4

Starting from

$$\tilde{\zeta}_\varrho(P) = \frac{1}{T}\nabla_{\boldsymbol{\kappa}'}\mathcal{I}^{-1}_{\boldsymbol{\kappa}\boldsymbol{\kappa}}\nabla_{\boldsymbol{\kappa}} + \frac{1}{T}\nabla_{\boldsymbol{\kappa}'}\mathcal{I}^{-1}_{\boldsymbol{\kappa}\boldsymbol{\kappa}}\mathcal{I}_{\boldsymbol{\kappa}\boldsymbol{\theta}}\left(\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\kappa}}\mathcal{I}^{-1}_{\boldsymbol{\kappa}\boldsymbol{\kappa}}\mathcal{I}_{\boldsymbol{\kappa}\boldsymbol{\theta}}\right)\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\kappa}}\mathcal{I}^{-1}_{\boldsymbol{\kappa}\boldsymbol{\kappa}}\nabla_{\boldsymbol{\kappa}}, \tag{12.1}$$

we can first show that $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\kappa}} = \mathbf{0}$, and then that $\frac{1}{T}\nabla_{\boldsymbol{\kappa}'}\mathcal{I}^{-1}_{\boldsymbol{\kappa}\boldsymbol{\kappa}}\nabla_{\boldsymbol{\kappa}} = Q_\nabla(P)$, the *portmanteau* test statistics of Box and Pierce (1970).

First, let us derive the elements of $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{k}} = \left[\mathcal{I}_{\boldsymbol{k}\nu}\ \mathcal{I}_{\boldsymbol{k}\psi_\sigma}\ \mathcal{I}_{\boldsymbol{k}\omega_\varrho}\right]$:

$$\begin{aligned}
\mathcal{I}_{\boldsymbol{k}\nu} &= \mathbb{E}\left[\left(\frac{\partial p_t}{\partial \varrho_{t+1}}\frac{\partial \varrho_{t+1}}{\partial \boldsymbol{k}}\right)\left(\frac{\partial p_t}{\partial \sigma^2_{t+1}}\frac{\partial \sigma^2_{t+1}}{\partial \nu} + \frac{\partial p_t}{\partial \nu}\right)\right];\\
&= \mathcal{I}_{\sigma\varrho}\,\mathbb{E}\left[\frac{\partial \varrho_{t+1}}{\partial \boldsymbol{k}}\frac{\partial \sigma^2_{t+1}}{\partial \nu}\right]\\
\mathcal{I}_{\boldsymbol{k}\psi_\sigma} &= \mathbb{E}\left[\left(\frac{\partial p_t}{\partial \varrho_{t+1}}\frac{\partial \varrho_{t+1}}{\partial \boldsymbol{k}}\right)\left(\frac{\partial p_t}{\partial \sigma^2_{t+1}}\frac{\partial \sigma^2_{t+1}}{\partial \boldsymbol{\psi}_\sigma}\right)\right]\\
&= \mathcal{I}_{\sigma\varrho}\left[\frac{\partial \varrho_{t+1}}{\partial \boldsymbol{k}}\frac{\partial \sigma^2_{t+1}}{\partial \boldsymbol{\psi}_\sigma}\right];\\
\mathcal{I}_{\boldsymbol{k}\omega_\varrho} &= \mathbb{E}\left[\left(\frac{\partial p_t}{\partial \varrho_{t+1}}\frac{\partial \varrho_{t+1}}{\partial \boldsymbol{k}}\right)\left(\frac{\partial p_t}{\partial \varrho_{t+1}}\frac{\partial \varrho_{t+1}}{\partial \omega_\varrho} + \frac{\partial p_t}{\partial \sigma^2_{t+1}}\frac{\partial \sigma^2_{t+1}}{\partial \omega_\varrho}\right)\right]\\
&= \mathcal{I}_{\sigma\varrho}\,\mathbb{E}\left[\frac{\partial \varrho_{t+1}}{\partial \boldsymbol{k}}\frac{\partial \sigma^2_{t+1}}{\partial \omega_\varrho}\right] + \mathcal{I}_{\varrho\varrho}\,\mathbb{E}[\boldsymbol{s}_{\varrho,t}]\\
&= \mathcal{I}_{\sigma\varrho}\,\mathbb{E}\left[\frac{\partial \varrho_{t+1}}{\partial \boldsymbol{k}}\frac{\partial \sigma^2_{t+1}}{\partial \omega_\varrho}\right].
\end{aligned}$$

Therefore, for Proposition 3, $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{k}} = \mathbb{E}\left[\frac{\partial \varrho_{t+1}}{\partial \boldsymbol{k}}\frac{\partial \sigma^2_{t+1}}{\partial \boldsymbol{\theta}}\right]'\mathcal{I}_{\sigma\varrho} = \mathbf{0}$.

Thus, Equation (12.1) reduces to $\tilde{\zeta}_\varrho(P) = \frac{1}{T}\nabla_{\boldsymbol{\kappa}'}\mathcal{I}^{-1}_{\boldsymbol{\kappa}\boldsymbol{\kappa}}\nabla_{\boldsymbol{\kappa}}$. Following Harvey (2013), for ϱ_t covariance stationary and applying the law of iterated expectations,

$$\frac{\partial \varrho_t}{\partial k_j} = \sum_{i=1}^{P}\kappa_{i-1}\frac{\partial s_{t-i}}{\partial \kappa_j} + s_{t-j-1},$$

such that $\frac{\partial \varrho_t}{\partial \boldsymbol{k}'}|_{\mathcal{H}_0} = \boldsymbol{s}_{t-1} = [s_{t-1}, \ldots, s_{t-P}]'$, and so under the null, $\mathcal{I}_{\boldsymbol{k}\boldsymbol{k}} = \varsigma^4_\varrho I_P$. Therefore, for

$\frac{\partial p}{\partial \kappa_j} = \sum_t \frac{\partial \log p_t}{\partial \varrho_t} \frac{\partial \varrho_t}{\partial \kappa_j} = \sum_t s_t s_{t-1-j},\ j = 0, \ldots, P,$

$$\xi_{\varrho}(P) = \frac{1}{T} \sum_t \sum_{j=0}^{P} \frac{(s_t s_{t-1-j})^2}{\varsigma_s^4} = T \sum_{j=0}^{P} \rho^2(j), \qquad (12.2)$$

which is the *portmanteau* statistics of Box and Pierce (1970), $Q(P)$, with a χ_1^2 limiting distribution, and $\rho(j)$ is the j^{th} sample autocorrelation of the score.

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019): "Vulnerable Growth," *American Economic Review*, 109, 1263–89.

Arellano-Valle, R. B., H. W. Gómez, and F. A. Quintana (2005): "Statistical inference for a general class of asymmetric distributions," *Journal of Statistical Planning and Inference*, 128, 427–443.

Azzalini, A. (1985): "A class of distributions which includes the normal ones," *Scandinavian journal of statistics*, 171–178.

——— (2013): *The skew-normal and related families*, vol. 3, Cambridge University Press.

Azzalini, A. and A. Capitanio (2003): "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 367–389.

Bai, J. and S. Ng (2005): "Tests for skewness, kurtosis, and normality for time series data," *Journal of Business & Economic Statistics*, 23, 49–60.

Barroso, P. and P. F. Maio (2021): "The risk-return tradeoff among equity factors," *Available at SSRN 2909085.*

Bianchi, D., A. De Polis, and I. Petrella (2022): "Taming Momentum Crashes," *Available at SSRN 4182040.*

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022): "Maximum likelihood estimation for score-driven models," *Journal of Econometrics*, 227, 325–346.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

——— (1987): "A conditionally heteroskedastic time series model for speculative prices and rates of return," *Review of Economics and Statistics*, 69, 542–547.

Box, G. E. and D. A. Pierce (1970): "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American statistical Association*, 65, 1509–1526.

Calvori, F., D. Creal, S. J. Koopman, and A. Lucas (2017): "Testing for parameter instability across different modeling frameworks," *Journal of Financial Econometrics*, 15, 223–246.

Carhart, M. M. (1997): "On persistence in mutual fund performance," *The Journal of finance*, 52, 57–82.

Cochrane, J. H. (2011): "Presidential address: Discount rates," *The Journal of finance*, 66, 1047–1108.

Creal, D., S. J. Koopman, and A. Lucas (2013): "Generalized autoregressive score models with applications," *Journal of Applied Econometrics*, 28, 777–795.

Delle Monache, D., A. De Polis, and I. Petrella (2021): "Modeling and forecasting macroeconomic downside risk," *Bank of Italy Temi di Discussione (Working Paper) No*, 1324.

Di Ciccio, T. J. and A. C. Monti (2011): "Inferential aspects of the skew t-distribution," *Quaderni di statistica*, 13, 1–21.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: Journal of the Econometric Society*, 987–1007.

Escanciano, J. C. and I. N. Lobato (2009): "An automatic portmanteau test for serial correlation," *Journal of Econometrics*, 151, 140–149.

Fama, E. F. and K. R. French (1992): "The cross-section of expected stock returns," *the Journal of Finance*, 47, 427–465.

——— (1993): "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, 33, 3–56.

——— (1995): "Size and book-to-market factors in earnings and returns," *The journal of finance*, 50, 131–155.

Fernández, C. and M. F. Steel (1998): "On Bayesian modeling of fat tails and skewness," *Journal of the American Statistical Association*, 93, 359–371.

Gómez, H. W., F. J. Torres, and H. Bolfarine (2007): "Large-sample inference for the epsilon-skew-t distribution," *Communications in Statistics—Theory and Methods*, 36, 73–81.

Groeneveld, R. A. and G. Meeden (1984): "Measuring skewness and kurtosis," *Journal of the Royal Statistical Society: Series D (The Statistician)*, 33, 391–399.

Hamilton, J. D. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica: Journal of the Econometric Society*, 357–384.

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

Harvey, A. and S. Thiele (2016): "Testing against changing correlation," *Journal of Empirical Finance*, 38, 575–589.

Harvey, A. C. (2013): *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, vol. 52, Cambridge University Press.

Harvey, C. R. and A. Siddique (1999): "Autoregressive conditional skewness," *Journal of Financial and Quantitative Analysis*, 34, 465–487.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995): *Continuous univariate distributions, volume 2*, vol. 289, John wiley & sons.

Jones, M. (2015): "On families of distributions with shape parameters," *International Statistical Review*, 83, 175–192.

Kim, H.-J. (2008): "Moments of truncated Student-t distribution," *Journal of the Korean Statistical Society*, 37, 81–87.

Komunjer, I. (2007): "Asymmetric power distribution: Theory and applications to risk measurement," *Journal of applied econometrics*, 22, 891–921.

Lucas, A. and X. Zhang (2016): "Score-driven exponentially weighted moving averages and Value-at-Risk forecasting," *International Journal of Forecasting*, 32, 293–302.

Mudholkar, G. S. and A. D. Hutson (2000): "The epsilon–skew–normal distribution for analyzing near-normal data," *Journal of Statistical Planning and Inference*, 83, 291–309.

Neftci, S. N. (1984): "Are economic time series asymmetric over the business cycle?" *Journal of Political Economy*, 92, 307–328.

Pástor, L. and R. F. Stambaugh (2003): "Liquidity risk and expected stock returns," *Journal of Political economy*, 111, 642–685.

Psarakis, S. and J. Panaretoes (1990): "The folded t distribution," *Communications in Statistics-Theory and Methods*, 19, 2717–2734.

Theodossiou, P. (1998): "Financial data and the skewed generalized t distribution," *Management Science*, 44, 1650–1661.

Theodossiou, P. and C. S. Savva (2016): "Skewness and the relation between risk and return," *Management Science*, 62, 1598–1609.

Zhu, D. and J. W. Galbraith (2010): "A generalized asymmetric Student-t distribution with application to financial econometrics," *Journal of Econometrics*, 157, 297–305.