Negative skewness of asset returns with positive time-varying risk premia

Christian M. Hafner* and Dimitra Kyriakopoulou†

Université catholique de Louvain

February 28, 2018

Preliminary - Please do not distribute

Abstract

The marginal distribution of financial time series such as returns is often negatively skewed. We investigate the relation between positive time-varying risk premia and the unconditional skewness of returns. We show that if the error distribution is symmetric, the negative unconditional asymmetry of returns should be the outcome of a negative correlation between their first two conditional moments. Following one of the implications of the intertemporal capital asset pricing model (ICAPM) of Merton (1973), there is a positive and linear relationship between risk and expected returns. Under an EGARCH-in-Mean specification, we propose to use an asymmetric error distribution in order to match the unconditional asymmetry of asset returns.

Keywords: Exponential GARCH, in-mean, risk premium, ICAPM, unconditional skewness, asymmetric distribution, portfolio selection.

*Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA), Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium, e-mail: christian.hafner@uclouvain.be.

†Center for Operations Research and Econometrics (CORE), Université catholique de Louvain, Voie du Roman Pays 34, Bte L1.03.01, B-1348 Louvain-la-Neuve, Belgium, e-mail: dimitra.kyriakopoulou@uclouvain.be.

1 Introduction

Portfolio selection and risk management are important problems that investors and portfolio managers face. Modelling the volatility of returns and studying the risk and return trade-off have gained increased interest in financial econometrics. The optimal portfolio construction assumes that investors are risk averse, meaning that given two portfolios that offer the same expected return, investors will prefer the less risky one, see Markowitz (1952). This is well known as a mean-variance analysis, as the expected return is maximized for a given level of risk, defined as variance. The distributional characteristics of returns, for example the unconditional skewness, are able to create new challenges in the classical portfolio theory of Markowitz. Apart from the mean and variance predictability, portfolio choice can be also made with skewness information. Such a perspective can create for investors the notion of skewness averse, as their financial decisions can be affected by properties of the return distribution.

It is often argued that the marginal distribution of financial time series such as returns is negatively skewed. Asymmetries or nonlinearities in the conditional mean are important towards meeting the objection to generate skewed marginal distributions. In that case, one solution seems to be the use of an asymmetric distribution, which can make the third-moment structure of the observations to be particularly flexible. Another suggestion comes from the economic theory, see for instance Hong and Stein (2003) and Campbell and Hentschel (1992), explaining skewness in the marginal distribution of returns. The dynamic behavior of economic agents plays an important role for explaining the negative skewness. This practically means that the model of use should also accommodate conditional asymmetry, apart from an asymmetric error density, to match the unconditional skewness of returns. This paper addresses the unconditional skewness of returns from the perspective of time-varying conditional first and second moments, together with the use of a more flexible conditional distribution for the returns.

We explore distribution classes, like the one by Azzalini (1985), and the Asymmetric Exponential Power Distribution (AEPD) proposed by Zhu and Zinde-Walsh (2009), which are able to accommodate skewness while nesting also the normal distribution that is typically used in estimation. The AEPD family extends the generalized error distribution (GED) to cases where the data exhibits asymmetry, and also accommodates fat tails. Other classes of distributions with the desired properties of accommodating heavy tails and skewness are the Skewed Exponential Power Distribution (SEPD) classes proposed by Komunjer (2007). The AEPD family can be seen as a fully-asymmetric GED that is capable to capture also the asymmetry in the tails. This is quite important since some applications in finance and risk management for portfolios such as S&P500 and Nasdaq have shown that ex post innovations from estimated GARCH models, even with a leverage effect, are not normally distributed, see e.g. Bradley and Taqqu (2003) and Christoffersen (2003), for Nasdaq and S&P500, respectively.

It is interested to compare between different distributions for the errors and also between different GARCH-M-type specifications, symmetric and asymmetric ones. In the former case, theoretical properties are studied under different distributional assumptions, while in the latter case, the impact of an asymmetric GARCH model on the conditional moment structure of returns is explored, which will shed light on the asymmetries in their marginal distribution.

2 The EGARCH-in-Mean model

Conditional heteroskedastic models, the so-called GARCH models, have been extensively and successfully used to model financial asset returns. We consider a parametrization of the conditional mean and variance that follows the Exponential GARCH (EGARCH) model of Nelson (1991). It also allows for the relation between risk and return in the conditional mean to capture time-varying properties of risk premium, leading to the EGARCH-in-Mean (EGARCH-M) model of the following form

$$Y_t = \lambda_0 + \lambda_1 h_t + \varepsilon_t, \qquad \varepsilon_t = \sqrt{h_t} Z_t \tag{1}$$
$$\log h_t = \omega + \gamma Z_{t-1} + \delta \left| Z_{t-1} \right| + \beta \log h_{t-1}, \quad t \in \mathbb{Z}$$

where Y_t represents a return process and ε_t is a zero mean white noise error process with conditional variance h_t. Also Z_t are the i.i.d. innovations with conditional mean and variance zero and unit, respectively. For simplicity, we focus on first-order model which is often found adequate in modelling volatility in returns.

One of the implications of the intertemporal capital asset pricing model (ICAPM) of Merton (1973) is a positive and linear relation between the conditional expectation of the returns and their conditional variance, which is interpreted as a linear-in-variance risk premium. Hafner and Kyriakopoulou (2017) showed that exponential-type of GARCH models are more natural to deal with linear-in-variance risk premia as they can avoid restrictions of the classical GARCH models and provided the asymptotic theory of the quasi-maximum likelihood estimator. Also exponential models within the GARCH class allow for richer dynamics incorporating the so-called leverage effect and they do not impose any positive restrictions on the parameters that could entail statistical and estimation difficulties.

The conditional mean and variance of Y_t given the information set available at time $t-1$, are given by

$$\mathbb{E}_{t-1}\left(Y_t\right) = \lambda_0 + \lambda_1 h_t,$$

and

$$\mathrm{Var}_{t-1}\left(Y_t\right) = \mathbb{E}_{t-1}\left(Y_t - \mathbb{E}_{t-1}\left(Y_t\right)\right)^2 = h_t,$$

so that shocks have a nonzero effect on both the conditional mean and conditional variance. As a consequence, the degree of asymmetry for the conditional mean is controlled by the asymmetry of the EGARCH process. We assume that the conditional distribution of the return process is modelled as an asymmetric standardized distribution, e.g. a skew-normal or exponential power distribution[1], whose details are given below, so that $Z_t \sim i.i.d.\,(0,1)$, i.e. the distribution is normalized with $\mathbb{E}_{t-1}\left(Z_t\right) = 0$ and $\mathrm{Var}_{t-1}\left(Z_t\right) = 1$.

3 The unconditional third moment structure

Tha main purpose of the paper is to investigate the implications of the EGARCH-M model on the third moment structure of the marginal distribution. We examine under which conditions the marginal distribution of returns can be skewed, how much, and of what sign. He *et al.* (2008) found that using an asymmetric or nonlinear specification

[1] The generalized error distribution (GED) is also called the exponential (or generalized) power distribution or the generalized Laplace distribution.

for the conditional mean is of greater importance to produce unconditional skewness for the returns, than the properties of the conditional variance itself. They also assumed a symmetric density for the errors, focusing only on the conditional mean and its implications on the skewness. In our paper, we allow the density of the errors to be asymmetric, examining under which parameter conditions this density can potentially explain skewness in returns. We are mainly interested in negatively skewed marginal distributions, as there is strong empirical evidence, using data from financial markets, that returns have negative skewness. In Table 1 we summarize some descriptive statistics of some of the largest stock indices. These are daily return series for the period 2/1/1986 - 30/12/2016, except for CAC40, which is for the period 9/7/1987 - 30/12/2016. We notice that for all cases the unconditional skewness is negative and for some it is even more severe. So, this paper can be seen as generalization of He et al. (2008).

	Obs.	Mean	Variance	Skewness	Kurtosis
S&P500	7815	0.037	1.321	−0.826	23.981
FTSE100	7847	0.027	1.246	−0.297	11.615
DAX	7830	0.037	2.073	−0.130	8.487
CAC40	7464	0.026	1.958	−0.018	8.248
Dow Jones	7815	0.039	1.252	−1.076	32.158

Table 1: Distributional properties of daily stock market returns. Returns have been multiplied by 100.

We study the third moment structure assuming that the average of the conditional mean of returns is nonzero, i.e. $\mathbb{E}\left(\mathbb{E}_{t-1}\left(Y_t\right)\right) \neq 0$. He *et al.* (2008) simplified their analysis by assuming that the unconditional mean, $\mathbb{E}\left(Y_t\right)$, is 0. Also, they parameterized the conditional standard deviation instead of the conditional variance, while we are interested in the latter case as this is in accordance with the implications of Merton (1973) and also in relation with the estimation theory for the EGARCH-M by Hafner and Kyriakopoulou (2017). Following Penaranda and Wu (2017), the excess returns can be decomposed as

$$Y_t = \mathbb{E}_{t-1}\left(Y_t\right) + \varepsilon_t \tag{2}$$

and we can also decompose the deviation of the excess return with respect to its unconditional mean as

$$Y_t - \mathbb{E}(Y_t) = d_{t-1} + \varepsilon_t, \tag{3}$$

where

$$d_{t-1} = \mathbb{E}_{t-1}(Y_t) - \mathbb{E}(Y_t).$$

The moments of the term d_{t-1}, that is about the conditional mean of returns, will help to determine the moment structure of the returns. We have

$$\begin{aligned}
\mathbb{E}d_{t-1} &= \mathbb{E}\left[\mathbb{E}_{t-1}(Y_t) - \mathbb{E}(Y_t)\right] = 0, \\
\mathbb{E}\left(d_{t-1}^2\right) &= \mathbb{E}\left[\mathbb{E}_{t-1}(Y_t) - \mathbb{E}(Y_t)\right]^2 = \mathrm{Var}\left(\mathbb{E}_{t-1}(Y_t)\right), \\
\mathbb{E}\left(d_{t-1}^3\right) &= \mathbb{E}\left[\mathbb{E}_{t-1}(Y_t) - \mathbb{E}(Y_t)\right]^3 = \mathrm{Sk}\left(\mathbb{E}_{t-1}(Y_t)\right).
\end{aligned}$$

From (2) and (3), the unconditional mean and variance of returns are given by

$$\begin{aligned}
\mathbb{E}(Y_t) &= \mathbb{E}\left(\mathbb{E}_{t-1}(Y_t)\right), \\
\mathrm{Var}(Y_t) &= \mathbb{E}\left[Y_t - \mathbb{E}(Y_t)\right]^2 = \mathbb{E}\left(\varepsilon_t^2\right) + \mathbb{E}\left(d_{t-1}^2\right) \\
&= \mathbb{E}\left(\mathrm{Var}_{t-1}(Y_t)\right) + \mathrm{Var}\left(\mathbb{E}_{t-1}(Y_t)\right).
\end{aligned}$$

Lemma 1 *The unconditional first and second-order moments of Y_t for the EGARCH$(1,1)$-M model are given by*

$$\begin{aligned}
\mathbb{E}(Y_t) &= \lambda_0 + \lambda_1 \mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i} + \delta\left|Z_{t-i}\right|\right)\right]\right\}, \\
\mathrm{Var}(Y_t) &= \mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i} + \delta\left|Z_{t-i}\right|\right)\right]\right\} + \lambda_1^2\left[\mathbb{E}h_t^2 - \left(\mathbb{E}h_t\right)^2\right].
\end{aligned}$$

Proof. See Section 6. ■

In terms of skewness, the unconditional coefficient of asymmetry (standardized skewness) of an excess return is

$$\mathrm{Sk}(Y_t) = \frac{\mathbb{E}\left[Y_t - \mathbb{E}(Y_t)\right]^3}{\left[\mathrm{Var}(Y_t)\right]^{3/2}}, \tag{4}$$

where the third centered moment in the numerator can be decomposed as

$$\begin{aligned}\mathbb{E}\left[Y_t - \mathbb{E}\left(Y_t\right)\right]^3 &= \mathbb{E}\left(\varepsilon_t^3\right) + \mathbb{E}\left(d_{t-1}^3\right) + 3\mathbb{E}\left(\varepsilon_t^2 d_{t-1}\right) \\ &= \mathbb{E}\left[\mathbb{E}_{t-1}\left(Y_t - \mathbb{E}_{t-1}\left(Y_t\right)\right)^3\right] \\ &+ \mathbb{E}\left[\mathbb{E}_{t-1}\left(Y_t\right) - \mathbb{E}\left(Y_t\right)\right]^3 \\ &+ 3\mathrm{Cov}\left(\mathrm{Var}_{t-1}\left(Y_t\right), \mathbb{E}_{t-1}\left(Y_t\right)\right),\end{aligned} \tag{5}$$

where the first component is the average of the conditional third moment of Y_t, the second component is the third moment (asymmetry) of $\mathbb{E}_{t-1}\left(Y_t\right)$, i.e. $\mathbb{E}\left(d_{t-1}^3\right) = \mathrm{Sk}\left(\mathbb{E}_{t-1}\left(Y_t\right)\right)$, and the third component is three times the covariance between the conditional first and second moments of Y_t. So, the first component depends on the conditional distribution of Z_t, the second component is the level of skewness in the conditional mean that comes from the asymmetry of the conditional variance, h_t, and the third component is the time-varying risk premium as the co-movement between return volatility and expected returns.

Penaranda and Wu (2017) showed that even if Y_t is conditionally symmetric, and hence the first component is zero, the other two components may yield asymmetry in the unconditional distribution of Y_t. When the conditional mean is time-invariant and Y_t is conditionally symmetric, the three above compoments are zero and thus $\mathbb{E}\left[Y_t - \mathbb{E}\left(Y_t\right)\right]^3 = 0$, which implies that $\mathrm{Sk}\left(Y_t\right) = 0$ in (4). Therefore, only assuming that the conditional second moment is time-varying does not imply unconditional skewness. However, assuming also that the conditional mean is time-varying results in the skewed marginal distribution for the observations, see He *et al.* (2008) who studied the term $\mathrm{Sk}\left(\mathbb{E}_{t-1}\left(Y_t\right)\right)$, i.e. the second component in (5), considering processes with a nonconstant conditional mean, but normal errors and therefore the first component in (5) was zero in their paper.

Lemma 2 *The unconditional third-order moment of Y_t for the EGARCH$(1,1)$-M model* *is given by*

$$\begin{aligned}\operatorname{Sk}(Y_t) &= \frac{\mathbb{E}\left[\mathbb{E}_{t-1}\left(Z_t^3\right)\mathbb{E}_{t-1}\left(\sqrt{h_t}\right)^3\right]}{\left[\operatorname{Var}(Y_t)\right]^{3/2}} \\ &+ \frac{\lambda_1^3\mathbb{E}\left[h_t - \mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i}+\delta\left|Z_{t-i}\right|\right)\right]\right\}\right]^3}{\left[\operatorname{Var}(Y_t)\right]^{3/2}} \\ &+ \frac{3\lambda_1\left[\mathbb{E}h_t^2 - \left(\mathbb{E}h_t\right)^2\right]}{\left[\operatorname{Var}(Y_t)\right]^{3/2}}\end{aligned} \tag{6}$$

Proof. See Section 6 ■

We expect the last component in (6), the covariance, to be nonnegative because we have assumed a positive correlation between the conditional mean and variance, which represents a positive risk premium that is implied by (1). Periods of high expected returns are associated with periods of high volatility. Hence, the unconditional negative skewness of the returns will depend on the conditional third moment of the innovations Z_t and the asymmetry of the conditional mean, $\mathbb{E}\left(d_{t-1}^3\right)$. Since the true conditional distribution of the returns is skewed, the use of a symmetric distribution is inappropriate. The correct specification of the conditional distribution is important for the quasi-maximum likelihood estimation. Engle and Gonzalez-Rivera (1991) have showed that the inefficiency of the QMLE may be substantial when the true distribution is asymmetric, while instead using the normal distribution.

We have seen that a skewed marginal distribution can be a result of some type of asymmetric or nonlinear behavior in the process for the conditional mean. This makes the question of testing for asymmetries and nonlinearities in the conditional mean important.

4 Simulations

The unconditional skewness of returns, $\operatorname{Sk}(Y_t)$, from an EGARCH-M process with normal errors and parameters $\omega = -0.08$, $\gamma = -0.06$, $\delta = 0.15$, $\beta = 0.98$ (solid line) and $\beta = 0.9$ (dashed line) is plotted in Figure 1, as a function of λ_1 and assuming also $\lambda_0 = 0.02$. When $\lambda_1 = 0$, the skewness is 0 and the curves intersect the horizontal axis. When the risk premium as a function of the conditional variance enters the conditional mean equation, which implies that the conditional mean now is also time-varying, we can see from the graph that the degree of skewness increases so that the marginal distribution of Y_t becomes skewed. It is also apparent that the range of possible skewness increases with the rate of persistence of the EGARCH process. Notice that when the risk premium is positive, the asymmetry is also positive, while the opposite holds as well. In order to explain the negative asymmetry in the marginal distribution of the returns, when also assuming a positive relation between risk and expected return, this exercise shows that the conditional distrubution of the errors should be asymmetric.

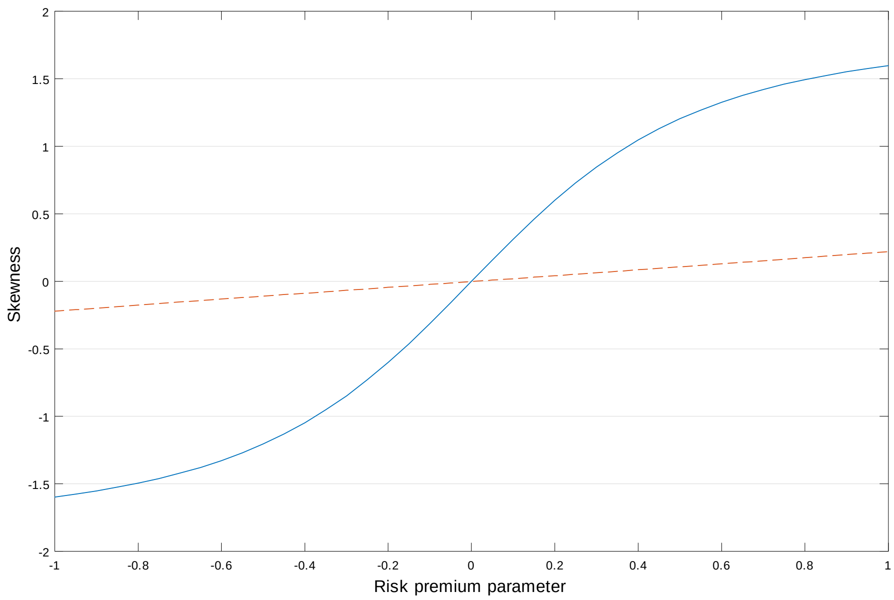

Figure 1: Unconditional skewness of returns as a function of the risk premium parameter for EGARCH-M parameters $\omega = -0.08, \gamma = -0.06, \delta = 0.15$, $\beta = 0.98$ (solid blue line) and $\beta = 0.90$ (dashed red line), with normal errors.

When the errors follow a skew-normal distribution, we plot in next figures the unconditional skewness of Y_t with the same parameter values as before, assuming that the shape parameter (α) of the distribution is either -1 (solid blue lines) or -5 (dashed red lines), indicating a negative skewness for the errors. The case $\alpha = -1$ is representative for some stock markets, and the case of $\alpha = -5$ is already indicative of the convergence to the half-normal distribution on $\mathbb{R}^-$ which occurs for $\alpha \to -\infty$. Figures 2 and 3 compare between lower (-1) and higher (-5) skewness, again for two cases of persistence, i.e. when there is high persistence $(\beta = 0.98)$ in Figure 2 and lower persistence $(\beta = 0.90)$ in Figure 3. Notice also that the first component in (5) is now nonzero. As we can see, the unconditional skewness when dropping normality is not anymore symmetric around zero as in the previous case with normal errors. When $\lambda_1 = 0$, the unconditional skewness of Y_t, that is given by the first term in (6), is equal to -0.1658 (blue line) and -0.9877 (red line). As it is expected, when the shape parameter increases, the skewness is even higher. What is most important here is the fact that the unconditional skewness is still negative with some positive values of the risk premium parameter. For instance, when the skewness parameter of the error distrbution is -5, we can still generate negative unconditional skewness of the returns when the risk premium takes values in the interval $(0, 0.18)$.

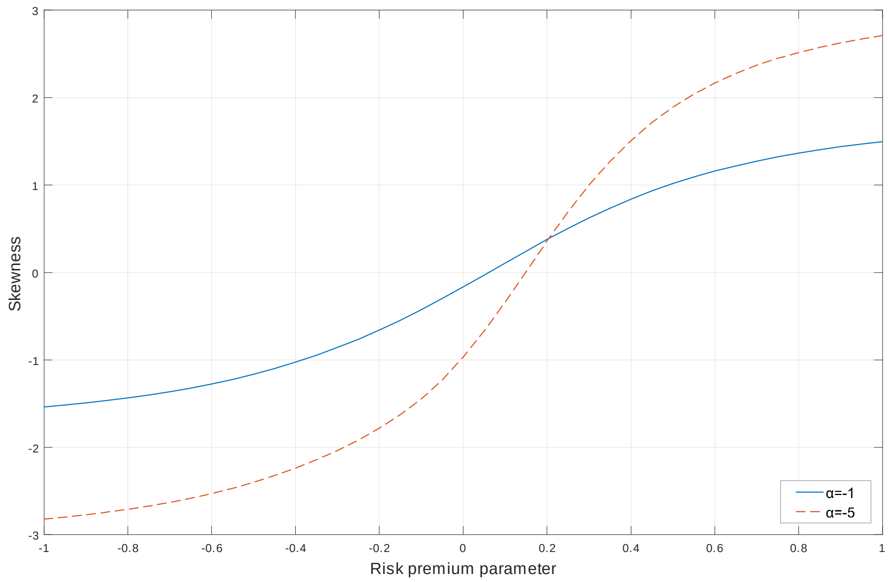

Figure 2: Unconditional skewness of returns as a function of the risk premium parameter for EGARCH-M parameters $\omega = -0.08, \gamma = -0.06, \delta = 0.15, \beta = 0.98$ with skew-normal (SN) errors and shape parameters $\alpha = -1$ (solid blue line), and $\alpha = -5$ (dashed red line).

The case of lower persistence is quite different than the previous one, as it is plotted in Figure 3. Here we see that we can get negative unconditional skewness with higher values of risk premium, as the shape parameter of the SN distribution increases.

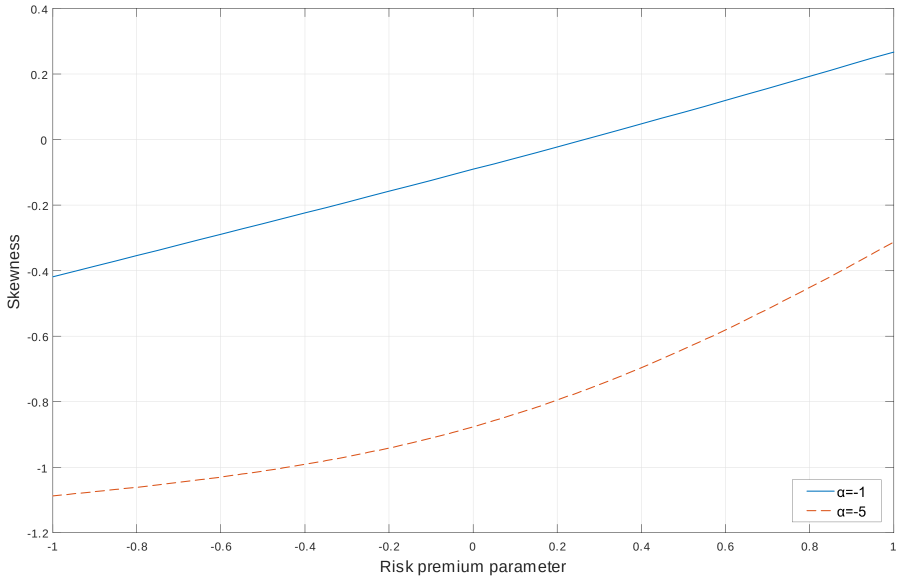

Figure 3: Unconditional skewness of returns as a function of the risk premium parameter for EGARCH-M parameters $\omega = -0.08, \gamma = -0.06, \delta = 0.15, \beta = 0.90$ with skew-normal (SN) errors and shape parameters $\alpha = -1$ (solid blue line), and $\alpha = -5$ (dashed red line).

4.1 Shock impact curve (SIC)

He *et al.* (2008) generalized the News Impact Curve (NIC) of Engle and Ng (1993) by considering the conditional mean structure. The NIC for the EGARCH process is plotted in Figure 4, showing how past shocks affect the current volatility.

The new tool is called Shock Impact Curve (SIC) and describes the impact of a shock on the conditional mean squared error (CMSE) of the returns. The conditional mean of Y_t is given by

$$\mathbb{E}_{t-1}(Y_t) = \lambda_0 + \lambda_1 h_t,$$

so it is a function of the conditional variance and therefore of past shocks. We begin with

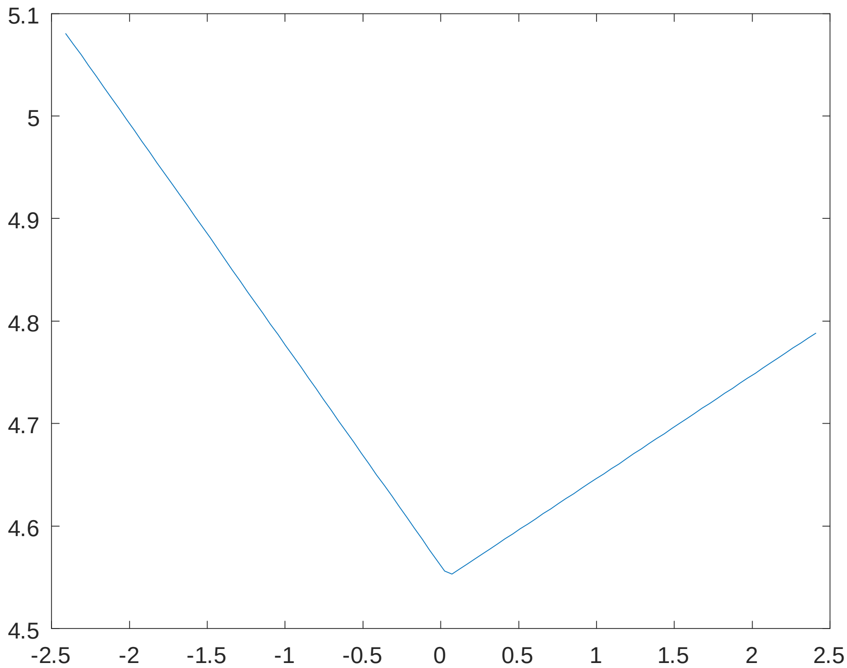

Figure 4: News Impact Curve for the EGARCH(1,1) process with parameters $\omega = -0.08, \gamma = -0.06, \delta = 0.15, \beta = 0.98$.

the conditional variance of the returns, where

$$\text{Var}_{t-1}(Y_t) = \mathbb{E}_{t-1}(Y_t)^2 - (\mathbb{E}_{t-1}(Y_t))^2 .$$

Following He *et al.* (2008), the SIC is defined as

$$\mathbb{E}_{t-1}^{\text{SIC}}(Y_t)^2 = (\mathbb{E}_{t-1}(Y_t))^2 + \text{Var}_{t-1}(Y_t) ,$$

where $\mathbb{E}_{t-1}(Y_t)$ and $\text{Var}_{t-1}(Y_t) = h_t$ are replaced with their unconditional counterparts, for instance $\mathbb{E}_{t-1}(Y_t)$ is replaced by $\mathbb{E}(Y_t) = \lambda_0 + \lambda_1 \mathbb{E} h_t$ and $\text{Var}_{t-1}(Y_t)$ by

$$\begin{aligned} \text{Var}(Y_t) &= \mathbb{E}(\text{Var}_{t-1}(Y_t)) + \text{Var}(\mathbb{E}_{t-1}(Y_t)) \\ &= \mathbb{E} h_t + \lambda_1^2 \left[\mathbb{E} h_t^2 - (\mathbb{E} h_t)^2\right] , \end{aligned}$$

see Lemma 1. Hence,

$$\begin{aligned}\mathbb{E}_{t-1}^{\text{SIC}}\left(Y_t\right)^2 &= \left(\lambda_0+\lambda_1\mathbb{E}h_t\right)^2+\mathbb{E}h_t+\lambda_1^2\left[\mathbb{E}h_t^2-\left(\mathbb{E}h_t\right)^2\right]\\ &= \lambda_0^2+2\lambda_0\lambda_1\mathbb{E}h_t+\lambda_1^2\left(\mathbb{E}h_t\right)^2+\mathbb{E}h_t+\lambda_1^2\left[\mathbb{E}h_t^2-\left(\mathbb{E}h_t\right)^2\right]\\ &= \lambda_0^2+\left(1+2\lambda_0\lambda_1\right)\mathbb{E}h_t+\lambda_1^2\mathbb{E}h_t^2.\end{aligned}$$

All in all, this tool helps us to see how the past news affect not only the current volatility but also the magnitude of today's returns. This can be illustrated by the means of squared returns and their relationship with the conditional mean and conditional variance. If $\mathbb{E}_{t-1}\left(Y_t\right)=0$, SIC coincides with NIC. On the other hand, if the conditional mean is nonzero, the impact of news and that of a shock on the next return can have different shapes.

5 Skewed distributions

5.1 The skew-normal (SN) distribution (Azzalini, 1985)

Azzalini's class of skew-normal distributions has been proved useful for modeling the skewness observed in many financial time series. The skew-normal (SN) distribution has the following properties: it enjoys the 'strict inclusion' property of the normal density, it is mathematically tractable and has a wide range of the indices of skewness and kurtosis. This class of continuous probability distributions generalises the normal distribution to allow for non-zero skewness. A random variable Z is skew-normal with asymmetry parameter α, denoted by $SN\left(\alpha\right)$, if it has density function

$$f\left(z;\alpha\right)=2\phi\left(z\right)\Phi\left(\alpha z\right),\quad -\infty<z<\infty,\quad \alpha\in\mathbb{R} \tag{7}$$

where ϕ and Φ are the standard normal density and distribution function, respectively. Thus, the density of the skew-normal can be interpreted as a normal density times a weight factor given by $2\Phi\left(\alpha z\right)$. If $\alpha=0$, it is the $N\left(0,1\right)$ density. As $\alpha\rightarrow\infty$, $f\left(z;\alpha\right)$ tends to the half-normal density. For positive values of α we obtain a distribution skewed to the right (the weight will be larger for positive z), and for negative α a distribution skewed to the left (the weight will be larger for negative z). The distribution function of (7) is

$$\Phi(z;\alpha) = 2\int_{-\infty}^{z}\int_{-\infty}^{\alpha t}\phi(t)\,\phi(u)\,dudt.$$

We can write the distribution function based on the Owen's function as follows

$$\Phi(z;\alpha) = \Phi(z) - 2T(z,\alpha), \tag{8}$$

where the function $T(z,\alpha)$ studied by Owen (1956) gives the integral of the standard normal bivariate density over region bounded by the lines $x = z$, $y = 0$, $y = \alpha t$ in the (x, y) plane. This function has the following properties:

$$-T(z,\alpha) = T(z,-\alpha),$$
$$T(-z,\alpha) = T(z,\alpha),$$
$$2T(z,1) = \Phi(z)\,\Phi(-z).$$

Taking into account the previous properties, (8) holds also for negative values of z and α, i.e. it is the general expression of the distribution function of (7).

By Lemma 2 of Azzalini (1985), we have that for a variable

$$Z = \xi + wX,$$

where X is a continuous i.i.d. random variable with density function (7) and mean and variance given by $b\mu$ and $1-(b\mu)^2$, respectively, assuming that $\xi = 0$, and $w = 1$, as it is often the case, the moment generating function of Z is

$$M(t) = 2\exp\left(t^2/2\right)\Phi(\mu t),$$

where

$$\mu = \frac{\alpha}{\sqrt{(1+\alpha^2)}}, \quad \mu \in (-1,1).$$

We can then obtain the first three (conditional) moments as

$$\mathbb{E}(Z) = b\mu,$$
$$\mathrm{Var}(Z) = 1-(b\mu)^2,$$
$$\mathbb{E}\left(Z^3\right) = b\left(3\mu - \mu^3\right),$$

where $b = \sqrt{\frac{2}{\pi}}$. So, Z has a nonzero mean for $\alpha \neq 0$.

The third and fourth standardized cumulants, i.e. skewness and kurtosis, respectively are denoted by γ_1, and γ_2, and given by

$$\begin{aligned}\gamma_1(Z) &= 1/2\,(4-\pi)\,\text{sign}(\alpha)\left[\frac{(\mathbb{E}(Z))^2}{\text{Var}(Z)}\right]^{3/2} \\ &= \frac{(4-\pi)\,\text{sign}(\alpha)}{2}\frac{\left(\mu\sqrt{\frac{2}{\pi}}\right)^3}{\left(1-\frac{2\mu^2}{\pi}\right)^{3/2}} \\ &= \frac{\sqrt{2}\,(4-\pi)\,\mu^3\text{sign}(\alpha)}{(\pi-2\mu^2)^{3/2}} \in (-0.995, 0.995)\,,\end{aligned}$$

and

$$\begin{aligned}\gamma_2(Z) &= 2\,(\pi-3)\left[\frac{(\mathbb{E}(Z))^2}{\text{Var}(Z)}\right]^{2} \\ &= 2\,(\pi-3)\,\frac{\left(\mu\sqrt{\frac{2}{\pi}}\right)^4}{\left(1-\frac{2\mu^2}{\pi}\right)^{2}} \\ &= \frac{8\,(\pi-3)\,\mu^4}{(\pi-2\mu^2)^{2}} \in [0, 0.869)\,, \quad \mu \to \pm 1\end{aligned}$$

Our element of interest is the innovations term Z_t for which we assume that is distributed as i.i.d. skew-normal with zero mean, unit variance, and unconditional skewness s, i.e. $Z_t \sim i.i.d.SN\,(0,1,s)$ and is the standardized and centered version of the skew-normal random variable Z, that is

$$Z_t = \frac{Z - b\mu}{\sqrt{1-(b\mu)^2}},$$

where s is given by $\gamma_1(Z)$. For more discussion on the moments of the SN distribution, see Henze (1986), Martínez *et al.* (2008), and Haas (2012).

Figure 5 shows how skewness, γ_1 and kurtosis, γ_2 relate to each other and to α, the shape parameter. Since we are interested in cases where there is negative skewness, we plot for negative values of α, however for positive ones the curve is just mirrored on the opposite side of the vertical axis.

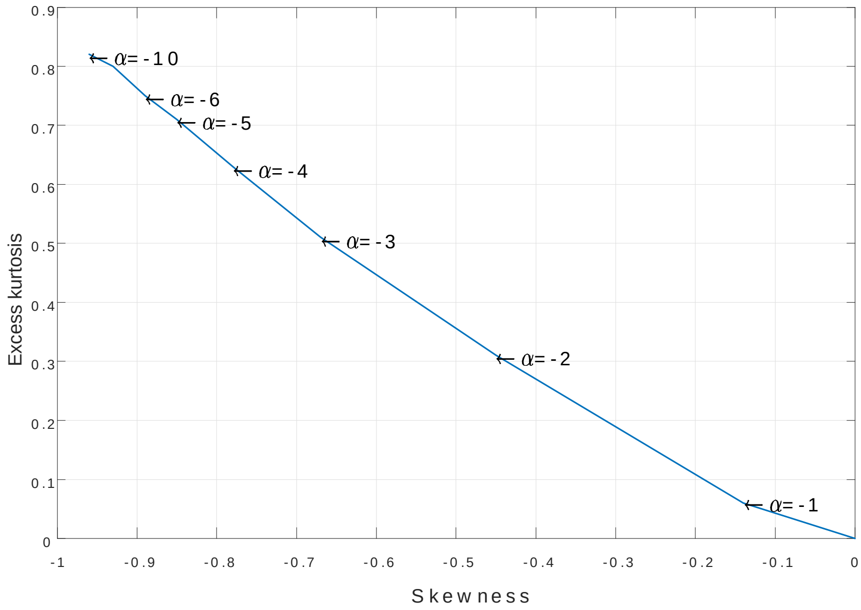

Figure 5: The SN distribution: grid of skewness and kurtosis as the shape parameter α ranges from -10 to 0, with arrows corresponding to choices of α.

5.2 The asymmetric exponential power distribution (AEPD)

The density function of the Exponential Power Distribution (EPD), also known as the Generalized Error Distribution (GED) with three parameters, v, μ, and σ, is usually defined as

$$f_{EP}(x) = \frac{1}{\sigma} \frac{1}{2v^{1/v}\Gamma(1+1/v)} \exp\left(-\frac{1}{v}\left|\frac{x-\mu}{\sigma}\right|^v\right),$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are the location and scale parameters respectively, and v is the shape parameter. Also $\Gamma(\cdot)$ stands for the gamma function. When v gets smaller and smaller, the EPD becomes more and more heavy-tailed and leptokurtic. With $v = 2$, $v = 1$, and $v \to +\infty$, the EPD reduces to the normal, Laplace and uniform distributions, respectively.

The Asymmetric EPD (AEPD) density proposed by Zhu and Zinde-Walsh (2009) with parameter vector $\beta = (\alpha, v_1, v_2, \mu, \sigma)^\intercal$ combines the flexible tail decay property of GED, measured by v, with the asymmetry and has the following form

$$f_{AEP}(x) = \begin{cases} \left(\frac{\alpha}{\alpha^*}\right) \frac{1}{\sigma} \frac{1}{2v_1^{1/v_1}\Gamma(1+1/v_1)} \exp\left(-\frac{1}{v_1}\left|\frac{x-\mu}{2\alpha^*\sigma}\right|^{v_1}\right), & \text{if } x \leq \mu; \\ \left(\frac{1-\alpha}{1-\alpha^*}\right) \frac{1}{\sigma} \frac{1}{2v_2^{1/v_2}\Gamma(1+1/v_2)} \exp\left(-\frac{1}{v_2}\left|\frac{x-\mu}{2(1-\alpha^*)\sigma}\right|^{v_2}\right), & \text{if } x > \mu, \end{cases} \quad (9)$$

where μ and σ are still the location and scale parameters as before, $\alpha \in (0,1)$ is the skewness parameter with $\alpha = 1/2$ for the symmetric case, $v_1 > 0$ and $v_2 > 0$ are the left and right tail parameters and α^* is defined as

$$\alpha^* = \left[\frac{\alpha}{2v_1^{1/v_1}\Gamma(1+1/v_1)}\right] / \left[\frac{\alpha}{2v_1^{1/v_1}\Gamma(1+1/v_1)} + \frac{1-\alpha}{2v_2^{1/v_2}\Gamma(1+1/v_2)}\right].$$

This parameter provides scale adjustments to the left and right parts of the density so as to ensure continuity of the density under changes of shape parameters (α, v_1, v_2). Also, if $v_1 = v_2 = v$ it implies $\alpha^* = \alpha$. Obviously, the AEPD collapses to GED with $\alpha = 1/2$ and $v_1 = v_2$.

A convenient reparametrization of (9) is obtained by rescaling, such that

$$f_{AEP}(x) = \begin{cases} \frac{1}{\sigma} \exp\left(-\frac{1}{v_1}\left|\frac{x-\mu}{2\alpha\sigma K_{v_1}}\right|^{v_1}\right), & \text{if } x \leq \mu; \\ \frac{1}{\sigma} \exp\left(-\frac{1}{v_2}\left|\frac{x-\mu}{2(1-\alpha)\sigma K_{v_2}}\right|^{v_2}\right), & \text{if } x > \mu, \end{cases}$$

where $K_{v_{(\cdot)}} = \frac{1}{2v_{(\cdot)}^{1/v_{(\cdot)}}\Gamma\left(1+1/v_{(\cdot)}\right)}$ for $v_{(1)}, v_{(2)}$. This density is used to derive a closed form expression for the information matrix of the maximum likelihood estimator.

Suppose that X is a random variable with the AEPD density but also standardized, i.e. $\mu = 0$ and $\sigma = 1$, which is often the case in financial applications. Zhu and Zinde-Walsh (2009) showed that the AEPD class has desired properties: interpretable parameters to represent location, scale and shape, closed-form expressions for the moments as well as for value at risk (VaR) and expected shortfall (ES).

6 Proofs

Proof of Lemma 1. We have that

$$\begin{aligned}
\mathbb{E}(Y_t) &= \mathbb{E}(\mathbb{E}_{t-1}(Y_t)) \\
&= \mathbb{E}(\lambda_0 + \lambda_1 h_t) \\
&= \lambda_0 + \lambda_1 \mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i} + \delta\left|Z_{t-i}\right|\right)\right]\right\},
\end{aligned}$$

and as $\varepsilon_t^2 = Z_t^2 h_t$

$$\begin{aligned}
\mathrm{Var}(Y_t) &= \mathbb{E}\left[Y_t - \mathbb{E}(Y_t)\right]^2 = \mathbb{E}\left(\varepsilon_t^2\right) + \mathbb{E}\left(d_{t-1}^2\right) \\
&= \mathbb{E}\left(\mathrm{Var}_{t-1}(Y_t)\right) + \mathrm{Var}\left(\mathbb{E}_{t-1}(Y_t)\right) \\
&= \mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i} + \delta\left|Z_{t-i}\right|\right)\right]\right\} + \mathrm{Var}(\lambda_0 + \lambda_1 h_t) \\
&= \mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i} + \delta\left|Z_{t-i}\right|\right)\right]\right\} + \mathbb{E}\left[(\lambda_0 + \lambda_1 h_t)^2\right] - \left[\mathbb{E}(\lambda_0 + \lambda_1 h_t)\right]^2 \\
&= \mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i} + \delta\left|Z_{t-i}\right|\right)\right]\right\} + \mathbb{E}\left[\lambda_0^2 + 2\lambda_0\lambda_1 h_t + \lambda_1^2 h_t^2\right] \\
&- \left[\lambda_0^2 + 2\lambda_0\lambda_1 \mathbb{E}h_t + \lambda_1^2\left(\mathbb{E}h_t\right)^2\right] \\
&= \mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i} + \delta\left|Z_{t-i}\right|\right)\right]\right\} + \lambda_1^2\left[\mathbb{E}h_t^2 - \left(\mathbb{E}h_t\right)^2\right],
\end{aligned}$$

making use of the multiplicative form of the EGARCH process, as we are interested in the moment structure of $\{h_t\}$ instead of that of $\{\log h_t\}$. ■

Lemma 3 *Moments of h_t.*

$$\begin{aligned}
\mathbb{E}h_t &= \mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i} + \delta\left|Z_{t-i}\right|\right)\right]\right\}, \\
\mathrm{Var}h_t &= \mathbb{E}h_t^2 - \left(\mathbb{E}h_t\right)^2, \\
\mathrm{Sk}h_t &= \frac{\mathbb{E}\left[h_t - \mathbb{E}h_t\right]^3}{\left[\mathrm{Var}h_t\right]^{3/2}},
\end{aligned}$$

where for $m \geq 2$

$$\mathbb{E}h_t^m = \mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)^m \prod_{i=1}^{\infty}\exp\left[\beta^{m(i-1)}\left(\gamma Z_{t-i} + \delta\left|Z_{t-i}\right|\right)^m\right]\right\}.$$

To evaluate the above expectations, we must make a distributional assumption about Z_t. Next Proposition deals with Azzalini (1985)'s class of a skew-normal distribution.

Proposition 1 *Let us assume that* $Z_t \sim i.i.d.SN(0,1,\alpha)$*, with* $\mathbb{E}(Z_t) = 0$, $\mathrm{Var}(Z_t) = 1$*, and* $\alpha < 0$*. Then*

$$\mathbb{E}\{\exp[\beta(\gamma Z_t + \delta|Z_t|)]\} = \Phi(\alpha\beta(\delta+\gamma))\exp\left(\beta^2(\delta+\gamma)^2/2\right) + \Phi(\alpha\beta(\delta-\gamma))\exp\left(\beta^2(\delta-\gamma)^2/2\right)$$

Proof. The density of a skew-normal random variable z if given by

$$f(z;\alpha) = 2\phi(z)\Phi(\alpha z),$$

where ϕ and Φ are the standard normal density and distribution function, respectively, i.e.

$$\phi(z) = \exp\left(-z^2/2\right)/\sqrt{2\pi}, \quad \Phi(\alpha z) = \int_{-\infty}^{\alpha z}\phi(t)\,dt,$$

and α is the shape parameter. If $\alpha = 0$, then $z \sim N(0,1)$. As α increases, the skewness of the distribution also increases.

Then

$$\mathbb{E}\{\exp[\beta(\gamma Z_t + \delta|Z_t|)]\} = \Phi(\alpha\beta(\delta+\gamma))\exp\left(\beta^2(\delta+\gamma)^2/2\right) + \Phi(\alpha\beta(\delta-\gamma))\exp\left(\beta^2(\delta-\gamma)^2/2\right).$$

■

Proof of Lemma 2. In (5) we have that the conditional third centered moment of Y_t is

$$\mathbb{E}_{t-1}[Y_t - \mathbb{E}_{t-1}(Y_t)]^3 = \mathbb{E}_{t-1}\left(\sqrt{h_t}Z_t\right)^3 = \mathbb{E}_{t-1}\left(h_t^{3/2}\right)\mathbb{E}_{t-1}\left(Z_t^3\right),$$

since Z_t is independent of h_t. Hence, the first component in the unconditional skewness which is the average of the conditional third moment of Y_t is

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}_{t-1}(Y_t - \mathbb{E}_{t-1}(Y_t))^3\right] &= \mathbb{E}\left[\mathbb{E}_{t-1}\left(h_t^{3/2}\right)\right]\mathbb{E}\left[\mathbb{E}_{t-1}\left(Z_t^3\right)\right] \\ &= \mathbb{E}\left(h_t^{3/2}\right)\mathbb{E}\left(Z_t^3\right),\end{aligned}$$

where $\mathbb{E}\left(h_t^{3/2}\right) = [\mathrm{Var}_{t-1}(\varepsilon_t)]^{3/2}$, where in general $[\mathrm{Var}_{t-1}(\varepsilon_t)]^{3/2} = [\mathrm{Var}_{t-1}(Y_t)]^{3/2}$, but in our setting $[\mathrm{Var}_{t-1}(\varepsilon_t)]^{3/2} \neq [\mathrm{Var}_{t-1}(Y_t)]^{3/2}$ due to the risk premium. The third moment (asymmetry) of $\mathbb{E}_{t-1}(Y_t)$ is

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}_{t-1}(Y_t)-\mathbb{E}(Y_t)\right]^3 &= \mathbb{E}\left[\lambda_0+\lambda_1 h_t-\mathbb{E}(Y_t)\right]^3 \\ &= \mathbb{E}\left[\lambda_1 h_t-\lambda_1\mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i}+\delta\left|Z_{t-i}\right|\right)\right]\right\}\right]^3 \\ &= \lambda_1^3\mathbb{E}\left[h_t-\mathbb{E}\left\{\exp\left(\frac{\omega}{1-\beta}\right)\prod_{i=1}^{\infty}\exp\left[\beta^{i-1}\left(\gamma Z_{t-i}+\delta\left|Z_{t-i}\right|\right)\right]\right\}\right]^3,\end{aligned}$$

and the covariance between the conditional first and second moments of Y_t is

$$\text{Cov}\left(\text{Var}_{t-1}(Y_t),\mathbb{E}_{t-1}(Y_t)\right)=\lambda_1\left[\mathbb{E}h_t^2-\left(\mathbb{E}h_t\right)^2\right].$$

Notice also that from the law of iterated expectations, $\mathbb{E}(Y_t)=\mathbb{E}\left(\mathbb{E}_{t-1}(Y_t)\right)$. ■

References

[1] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12: 171-178.

[2] Bradley, B. O., Taqqu, M. S. (2003). Financial risk and heavy tails. In S. T. Rachev (Ed.), *Handbook of Heavy-Tailed Distributions in Finance* 35-103, Amsterdam: Elsevier.

[3] Campbell, J. Y., Hentschel, L. (1992). No news is good news. *Journal of Financial Economics* 31: 281-318.

[4] Christoffersen, P. F. (2003). *Elements of Financial Risk Management.* Academic Press.

[5] Engle, R. F., Gonzalez-Rivera, G. (1991). Semiparametric ARCH models. *Journal of Business and Economic Statistics* 9: 345-359.

[6] Engle, R. F., Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance* 48: 1749-1778.

[7] Haas, M. (2012). A note on the moments of the skew-normal distribution. *Economics Bulletin* 32: 3306-3312.

[8] Hafner, C. M., Kyriakopoulou, D. (2017). Exponential-type GARCH models with linear-in-variance risk premium. *Working Paper.*

[9] He, C., Silvennoinen, A., Teräsvirta, T. (2008). Parameterizing unconditional skewness in models for financial time series. *Journal of Financial Econometrics* 6: 208-230.

[10] Henze, N. (1986). A probabilistic presentation of the skew-normal distribution. *Scandinavian Journal of Statistics* 13: 271-275.

[11] Hong, H., Stein, J. C. (2003). Differences of opinion, short-sales constraints, and market crashes. *Review of Financial Studies* 16: 487-525.

[12] Komunjer, I. (2007). Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics* 22: 891-921.

[13] Markowitz, H. (1952). Portfolio selection. *Journal of Finance* 7: 77-91.

[14] Martínez, E. H., Varela, H., Gómez, H. W., Bolfarine, H. (2008). A note on the likelihood and moments of the skew-normal distribution. *SORT* 32: 57-66.

[15] Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica* 41: 867-886.

[16] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59: 347-370.

[17] Owen, D. B. (1956). Tables for computing bivariate normal probabilities. *Annals of Mathematical Statistics* 27: 1075–1090.

[18] Penaranda, F, Wu, L. (2017). Predictability and performance. *Working Paper.*

[19] Zhu, D., Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics* 148: 86-99.