

# RECONCILING NEGATIVE RETURN SKEWNESS WITH POSITIVE TIME- VARYING RISK PREMIA

Dimitra Kyriakopoulou, Christian M. Hafner

REPRINT | 2022 / 31

## **ISBA**

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : [lidam-library@uclouvain.be](mailto:lidam-library@uclouvain.be)

<https://uclouvain.be/en/research-institutes/lidam/isba/publication.html>

# Reconciling negative return skewness with positive time-varying risk premia

Dimitra Kyriakopoulou\* and Christian M. Hafner†

April 19, 2022

## Abstract

One of the implications of the intertemporal capital asset pricing model (ICAPM) is a positive and linear relationship between the conditional mean and conditional variance of returns to the market portfolio. Empirically, however, it is often observed that there is a negative skewness in equity returns. This paper shows that a negative skewness is only compatible with a positive risk premium if the innovation distribution is asymmetric with a negative skewness. We extend recent work using the EGARCH-in-Mean specification to allow for asymmetric innovations, and give results for the unconditional skewness of returns. We apply the model to the prediction of Value-at-Risk of the largest stock market indices, and demonstrate its good performance.

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

---

\*Corresponding author. National and Kapodistrian University of Athens (NKUA), Department of Economics, 1 Sofokleous Str., 105 59 Athens, Greece, e-mail: dimkyriak@econ.uoa.gr.

†Louvain Institute of Data Analysis and Modeling in economics and statistics (LIDAM), and ISBA, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium, e-mail: christian.hafner@uclouvain.be.

# 1 Introduction

Portfolio selection and risk management are important problems that investors and portfolio managers face. 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 and optimization can be also made with higher order moment information, like skewness. Such a perspective can create for investors the notion of skewness corrected risk estimate, as their financial decisions can be affected by properties of the return distribution.

Modelling and forecasting the volatility of financial time series such as returns have gained increased interest in financial econometrics. Another substantial topic in this literature is the trade-off between risk and expected return, which has been studied by Engle et al. (1987) in the so-called Generalized Autoregressive Conditional Heteroskedastic (GARCH) framework proposing the use of the GARCH-in-Mean models. It is often argued that the marginal distribution of returns is negatively skewed. Asymmetries or nonlinearities in the conditional mean are important towards meeting the objective to generate skewed marginal distributions. In that case, one solution comes from the economic theory, see for instance Campbell and Hentschel (1992) and Hong and Stein (2003), explaining skewness in the marginal distribution of returns. The dynamic behavior of economic agents plays an important role for explaining the negative skewness. Another suggestion seems to be the use of an asymmetric distribution, which can make the third-moment structure of the observations to be particularly flexible. This practically means that the model of use can accommodate an asymmetric error density, apart from conditional asymmetry, 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, which nest the normal case that is typically used in estimation, like the Generalized Error Distribution (GED), see Subbotin (1923) and Nelson

(1991), and the Asymmetric Power Distribution (APD) proposed by Komunjer (2007). They both regulate for their tail thickness that is of great importance for a range of application areas in finance where the distribution of the observed data is known to have fatter tails than the normal ones. The APD family extends GED to cases where the data exhibits asymmetry. It is interesting to compare different distributions for the errors and also cases with and without risk premium. In the distribution case, theoretical properties are studied under different distributional assumptions, while in the model case, the impact of an asymmetric GARCH model with positive time-varying risk premia on the conditional moment structure of returns is explored, which will shed light on the asymmetries in their marginal distribution.

The paper also aims to examine whether the risk premium in the context of GARCH modelling helps to accurately estimate risk measures used in portfolio optimization. Asymptotic expansions, such as the Cornish-Fisher approximation, are useful to derive a more precise risk estimate, as the approximation of the quantiles of the unknown distribution is based on the first four moments and hence, this will provide a skewness corrected risk estimate for the optimized portfolio.

The paper is organized as follows. In Section 2, we present the model with asymmetric innovations and discuss the estimation of the model parameters. We study the third moment structure of the marginal distribution of returns with a numerical approximation in Section 3. Section 4 discusses applications to Value-at-Risk and Section 5 delivers an empirical application to stock market indices. Section 6 concludes. Proofs and discussion of the error distributions are collected in the Appendix.

## **2 The EGARCH-in-Mean model with asymmetric innovations**

Conditionally heteroskedastic models have been extensively 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 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

$$\begin{aligned} Y_t &= \lambda_0 + \lambda_1 h_t + \varepsilon_t, & \varepsilon_t &= \sqrt{h_t} Z_t \\ \log h_t &= \omega + \gamma Z_{t-1} + \delta |Z_{t-1}| + \beta \log h_{t-1}, & t &\in \mathbb{Z} \end{aligned} \quad (1)$$

where  $Y_t$  represents the return process and  $\varepsilon_t$  is a zero mean error process with conditional variance  $h_t$ . Also,  $Z_t$  are i.i.d. innovations with zero mean and unit variance. For simplicity, we focus on the first order which is often found adequate in modelling volatility in returns. For its estimation theory and asymptotic properties of the Quasi Maximum Likelihood estimator we refer to Hafner and Kyriakopoulou (2019).

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 returns and their conditional variance, which is interpreted as a linear-in-variance risk premium. Hafner and Kyriakopoulou (2019) showed that exponential-type GARCH models are more natural to deal with linear-in-variance risk premia as they can avoid restrictions of the classical GARCH models when proving asymptotics. Moreover, 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 strict stationarity and ergodicity of the model in (1) follow from the conditions

$$|\beta| < 1 \quad \text{and} \quad \mathbb{E} [\log^+ (\omega + \gamma Z_t + \delta |Z_t|)] < \infty, \quad (2)$$

see Hafner and Kyriakopoulou (2019) and Straumann and Mikosch (2006). In Appendix A, we prove that  $\{h_t\}$  and  $\{\varepsilon_t\}$  have finite unconditional moments of arbitrary order under the assumption that the distribution of  $Z_t$  is both asymmetric and fat tailed. Nelson (1991) proved the finiteness of all unconditional moments if the innovations distribution is a member of the symmetric Generalized Error Distribution (GED) family.

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

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

and

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

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 either belongs to the class of GED, which is symmetric but allows for fatter tails than the normal distribution, or the Asymmetric Power Distribution (APD), which has an additional parameter for the asymmetry, see Appendix C for the details, standardized so that  $Z_t \sim i.i.d.D(0, 1)$ , i.e.  $\mathbb{E}_{t-1}(Z_t) = 0$ ,  $\text{Var}_{t-1}(Z_t) = 1$ .

The model parameters are estimated by maximum likelihood (ML), assuming a distribution for the innovations  $Z_t$ . First we consider the case where  $Z_t$  follows the GED distribution, with mean zero and unit variance, and with shape parameter  $\nu > 0$ . Let  $\theta^{\text{GED}}$  denote the parameter vector, i.e.  $\theta^{\text{GED}} = (\lambda_0, \lambda_1, \omega, \gamma, \delta, \beta, \nu)'$  of the GED-EGARCH(1, 1)-M process in (1). We denote the ML estimator (MLE) by  $\hat{\theta}_T^{\text{GED}}$ , where  $T$  is the sample size. The MLE  $\hat{\theta}_T^{\text{GED}}$  is obtained by maximizing the GED log-likelihood, i.e.

$$L_T(\theta^{\text{GED}}) = \log\left(\frac{\nu}{\zeta}\right) - \left(1 + \frac{1}{\nu}\right) \log(2) - \log \Gamma\left(\frac{1}{\nu}\right) - \frac{1}{2} T^{-1} \sum_{t=1}^T \left[ \log h_t + \left| \frac{Z_t}{\zeta} \right|^\nu \right],$$

where  $\zeta \equiv \sqrt{2^{-2/\nu} \Gamma(1/\nu) / \Gamma(3/\nu)}$  and  $\Gamma(\cdot)$  is the gamma function.

Next, we assume alternatively that  $Z_t$  follows the standardized APD with parameters  $\alpha$ ,  $0 < \alpha < 1$ , for the asymmetry and  $\nu$ ,  $\nu > 0$ , for the shape. Let  $\theta^{\text{APD}}$  denote the parameter vector, i.e.  $\theta^{\text{APD}} = (\lambda_0, \lambda_1, \omega, \gamma, \delta, \beta, \alpha, \nu)'$  of the APD-EGARCH(1, 1)-M process in (1). The MLE  $\hat{\theta}_T^{\text{APD}}$  is obtained by maximizing the APD log-likelihood function

$$L_T(\theta^{\text{APD}}) = T^{-1} \sum_{t=1}^T [-\log h_t + \log f_Z(Z_t | \alpha, \nu)],$$

where  $f_Z(\cdot | \alpha, \nu)$  is the standard density function of the APD used in financial applications, as in Komunjer (2007), see Appendix C.2. These log-likelihood functions are used in the empirical section to estimate the model parameters.

### 3 Conditional and unconditional skewness

One of the main tasks is to investigate the implications of the EGARCH-M process on the third moment structure of the marginal distribution of returns. We examine under which conditions the marginal distribution can be skewed, how much, and of what sign. He et al. (2008) found that using an asymmetric or nonlinear specification 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, the model defined in (1) represents an asymmetric model for the conditional mean. We also 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 in the related literature that stock returns have negative skewness, see also Table 1 in Section 5. We have noticed that for all cases the unconditional skewness is negative and for some it is quite severe. Therefore, this paper can be seen as an extension of He et al. (2008).

We study the third moment structure assuming that the average of the conditional mean of returns is nonzero, i.e.  $\mathbb{E}(\mathbb{E}_{t-1}(Y_t)) \neq 0$ . He et al. (2008) simplified their analysis by assuming that the unconditional mean,  $\mathbb{E}(Y_t)$ , is 0. The moment structure of first order EGARCH models has been studied by He et al. (2002). Following Penaranda and Wu (2017), the excess returns can be decomposed as

$$Y_t = \mathbb{E}_{t-1}(Y_t) + \varepsilon_t \quad (3)$$

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

$$Y_t - \mathbb{E}(Y_t) = d_{t-1} + \varepsilon_t, \quad (4)$$

where

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

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

$$\begin{aligned} \mathbb{E}d_{t-1} &= \mathbb{E}[\mathbb{E}_{t-1}(Y_t) - \mathbb{E}(Y_t)] = 0, \\ \mathbb{E}(d_{t-1}^2) &= \mathbb{E}[\mathbb{E}_{t-1}(Y_t) - \mathbb{E}(Y_t)]^2 = \text{Var}(\mathbb{E}_{t-1}(Y_t)), \\ \mathbb{E}(d_{t-1}^3) &= \mathbb{E}[\mathbb{E}_{t-1}(Y_t) - \mathbb{E}(Y_t)]^3 = \text{Asy}(\mathbb{E}_{t-1}(Y_t)), \end{aligned}$$

where  $\text{Asy}(\cdot)$  stands for the third centered moment of the conditional mean. From (3) and

(4), the unconditional mean and variance of returns are given by

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

**Lemma 1** *Given the stationarity conditions in (2), the unconditional first and second-order moments of  $Y_t$  for the EGARCH(1,1)-M model are given by*

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

**Proof.** See Appendix B. ■

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

$$\text{Sk}(Y_t) = \frac{\mathbb{E}[Y_t - \mathbb{E}(Y_t)]^3}{[\text{Var}(Y_t)]^{3/2}}, \quad (5)$$

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

$$\begin{aligned}\mathbb{E}[Y_t - \mathbb{E}(Y_t)]^3 &= \mathbb{E}(\varepsilon_t^3) + \mathbb{E}(d_{t-1}^3) + 3\mathbb{E}(\varepsilon_t^2 d_{t-1}) \\ &= \mathbb{E}[\mathbb{E}_{t-1}(Y_t - \mathbb{E}_{t-1}(Y_t))^3] \\ &\quad + \mathbb{E}[\mathbb{E}_{t-1}(Y_t) - \mathbb{E}(Y_t)]^3 \\ &\quad + 3\text{Cov}(\text{Var}_{t-1}(Y_t), \mathbb{E}_{t-1}(Y_t)).\end{aligned} \quad (6)$$

The first component is the average of the conditional third moment of  $Y_t$ , the second component is the third moment (asymmetry) of  $\mathbb{E}_{t-1}(Y_t)$ , i.e.  $\mathbb{E}(d_{t-1}^3) = \text{Asy}(\mathbb{E}_{t-1}(Y_t))$ , 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 components above are zero and thus  $\mathbb{E}[Y_t - \mathbb{E}(Y_t)]^3 = 0$ , which implies that  $\text{Sk}(Y_t) = 0$  in (5). 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  $\text{Asy}(\mathbb{E}_{t-1}(Y_t))$ , i.e. the second component in (6), considering processes with a nonconstant conditional mean, but normal errors and therefore the first component in (6) was zero in their paper.

**Lemma 2** *Given the stationarity conditions in (2), the unconditional skewness of  $Y_t$  for the EGARCH(1,1)-M model is given by*

$$\begin{aligned} \text{Sk}(Y_t) = & \frac{\mathbb{E}(Z_t^3) \mathbb{E}(\sqrt{h_t})^3}{[\text{Var}(Y_t)]^{3/2}} \\ & + \frac{\lambda_1^3 \mathbb{E} \left[ h_t - \mathbb{E} \left\{ \exp \left( \frac{\omega}{1-\beta} \right) \prod_{i=1}^{\infty} \exp [\beta^{i-1} (\gamma Z_{t-i} + \delta |Z_{t-i}|)] \right\} \right]^3}{[\text{Var}(Y_t)]^{3/2}} \\ & + \frac{3\lambda_1 \text{Var}(h_t)}{[\text{Var}(Y_t)]^{3/2}} \end{aligned} \quad (7)$$

**Proof.** See Appendix B. ■

We expect the last component in (7) 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 our model (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,  $\mathbb{E}(d_{t-1}^3)$ . Since the distribution of 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 shown 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.

### 3.1 Numerical approximation of unconditional skewness

The unconditional skewness of returns in (5),  $\text{Sk}(Y_t)$ , from an EGARCH-M process with parameters  $\omega = -0.08$ ,  $\gamma = -0.06$ ,  $\delta = 0.15$ ,  $\beta = 0.98$  and  $\lambda_0 = 0.02$  is plotted in Figure 1, as a function of  $\lambda_1$  and assuming that the innovations follow the standardized APD. The sample size of returns is 10,000 and the number of simulations is 5,000. We consider three cases, i.e. the normality case with  $\alpha = 0.5$  and  $\nu = 2$  (dashed black line), together with other two special cases where there is negative asymmetry, i.e.  $\alpha = 0.9$ , but with either shorter tails ( $\nu = 2$ ) with the solid bold blue line or fatter tails ( $\nu = 1.4$ ) with the solid yellow line. For the normality case ( $\alpha = 0.5$  and  $\nu = 2$ ), when  $\lambda_1 = 0$ , the skewness is 0 and the dotted black curve intersects 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. Other experiments have shown 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 simulation exercise shows that the conditional distribution of the errors should be asymmetric.

When there is asymmetry in the distribution of the innovations, we have seen that the first component in (6) is nonzero. It is obvious that the unconditional skewness of returns is not anymore symmetric around zero as in the previous case with normality. When  $\lambda_1 = 0$ , the unconditional skewness of  $Y_t$ , that is given by the first term in (7), is equal to  $-1.1064$  (blue line) and  $-1.7182$  (yellow 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 distribution is 0.9 and the tails get fatter, we can generate negative unconditional skewness of the returns when the risk premium takes values in the interval  $(0, 0.18)$ . We have also plotted the case of lower persistence (not displayed) to compare between higher and lower GARCH parameter values. We have seen that we can get negative unconditional skewness with

higher values of risk premium, as the shape parameter of the APD distribution increases.

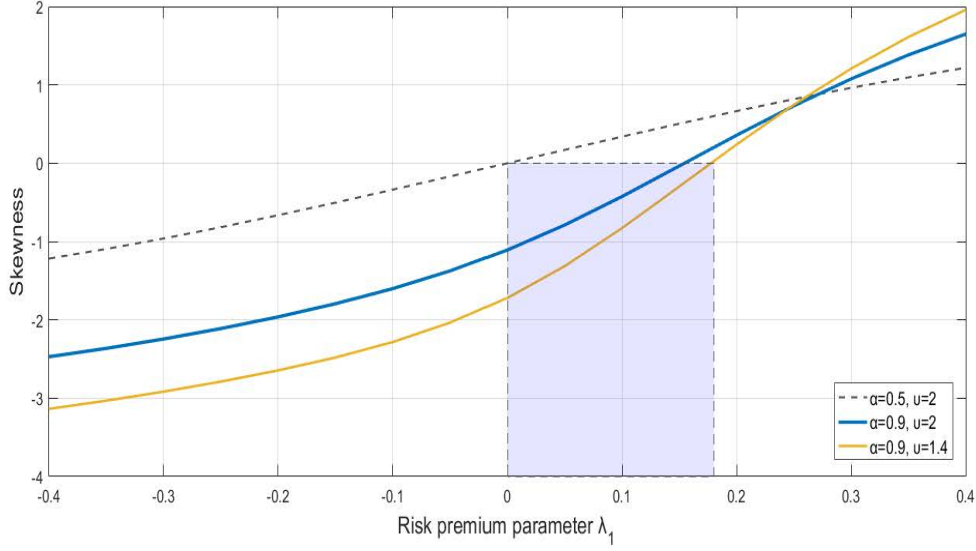


Figure 1: *Unconditional skewness of returns as a function of the risk premium parameter for EGARCH-M parameters  $\omega = -0.08$ ,  $\gamma = -0.06$ ,  $\delta = 0.15$ ,  $\beta = 0.98$ , and  $\lambda_0 = 0.02$ , with APD errors and parameters  $\alpha = 0.5$ ,  $\nu = 2$  (dashed black line),  $\alpha = 0.9$ ,  $\nu = 2$  (solid bold blue line), and  $\alpha = 0.9$ ,  $\nu = 1.4$  (solid yellow line).*

## 4 Applications to Value-at-Risk

A prominent measure of financial risk in investment portfolios is the Value-at-Risk (VaR), known also as downside risk, which is widely used in financial risk management applications. It is a measure of the potential loss of a portfolio that is encountered over a fixed time period and with a fixed small probability. Typically, risk is controlled and managed over short periods and this requires a conditional VaR forecast estimate that employs information up to time  $t$ .

A  $\vartheta\%$   $n$ -period conditional VaR forecast estimate is the maximum potential loss  $l$  of an investment portfolio that can not be exceeded, with a  $(1 - \vartheta)\%$  confidence level, if the portfolio is left unmanaged during the  $n$ -period, using the conditioning information available at time  $t$ , denoted by  $\mathcal{F}_t$ . It can therefore be derived by the computation of the  $\vartheta$ -quantile of the  $n$ -period portfolio return distribution at time  $t + n$  conditional on  $\mathcal{F}_t$ , where  $\vartheta$  is typically chosen as 1% or 5%. The portfolio VaR, expressed as a loss measure,

is then defined to be the negative of this  $\vartheta$ -quantile.

To predict the downside risk, for example in period  $t + 1$  conditional on  $\mathcal{F}_t$ , we need to make a 1-step-ahead forecast of the conditional distribution of returns, which is time-varying due to the time-variation of the conditional mean and variance. Therefore, forecasting the conditional distribution amounts to estimating the parameters of the model using data available at time  $t$ , and then forecasting the conditional mean and variance of  $Y_{t+1}$ , denoted by  $\hat{m}_{t+1|t}$  and  $\hat{h}_{t+1|t}$  respectively. Based on these forecasts at time  $t$ , the 1-step-ahead forecast of the VaR is given by

$$\widehat{\text{VaR}}_{t+1|t}(\vartheta) = - \left[ \hat{m}_{t+1|t} + \sqrt{\hat{h}_{t+1|t}} \hat{q}_Z(\vartheta) \right],$$

where  $\hat{q}_Z(\vartheta)$  is the  $\vartheta$ -quantile of the innovations distribution, after having estimated the parameter vector of the distribution of  $Z_t$ . We propose to use the fully parametric EGARCH-in-Mean specification given by (1), with normal and alternative distributions for the innovations, namely the GED and asymmetric power distributions. Hence, if we use normality, the parameter vector of the distribution of  $Z_t$  is empty, if we use GED it contains the shape parameter  $\nu$ , and if we assume the APD it will contain the skewness  $\alpha$  and the shape  $\nu$  parameters.

Additional to the parameterizations of the innovation distribution, we use asymptotic expansions for the VaR estimation based on the first four moments, assuming that they exist. An analytic approximation to the conditional VaR for EGARCH-M returns is the Cornish-Fisher expansion. This method allows quantiles to be estimated from standard normal quantiles at high significance levels, given only the first four conditional moments of the portfolio return, and also be adjusted due to the discrepancies arising from normal quantiles.

**Cornish-Fisher expansion (Cornish and Fisher, 1937)** *The fourth order Cornish-Fisher approximation to the  $\vartheta$ -quantile of an empirical distribution with mean zero and variance one is*

$$x_\vartheta \approx z_\vartheta + \frac{\hat{\tau}}{6} (z_\vartheta^2 - 1) + \frac{\hat{\kappa}}{24} z_\vartheta (z_\vartheta^2 - 3) - \frac{\hat{\tau}^2}{36} z_\vartheta (2z_\vartheta^2 - 5),$$

where  $z_\vartheta = \Phi^{-1}(\vartheta)$ , i.e. the  $\vartheta$ -quantile of a standard normal distribution, and  $\hat{\tau}$  and  $\hat{\kappa}$  denote the skewness and excess kurtosis of the empirical distribution. Hence, it is a Taylor series expansion around the  $\vartheta$ -quantile of a standard normal distribution.

In our empirical Value-at-Risk application, we will compare the performance of the Cornish-Fisher expansion with those of the parametric distributions.

## 5 Empirical application

In this section we address the empirical application by first estimating the EGARCH-M parameters under the GED and APD distributional assumptions. Results for VaR prediction are also presented performing an out-of-sample analysis. We evaluate the quality of the VaR forecasts by applying the dynamic quantile test of Engle and Manganelli (2004) and the failure ratio tests of coverage probability of Christoffersen (1998) and Kupiec (1995).

### 5.1 Data and descriptive statistics

We use data of daily return series of some of the largest stock indices: S&P 500, CAC 40, DAX, FTSE 100 and Dow Jones. Our sample covers the period from January 2, 1986 to December 30, 2016, except for CAC40, which is for the period from July 9, 1987 to December 30, 2016. In Table 1 we summarize some descriptive statistics of the time series, the unconditional moments of the returns and also normality and dependence tests. All series exhibit negative skewness between -0.02 and -1.08. Excess kurtosis is present in each return series ranging from 5.25 (CAC 40) to 29.16 (Dow Jones). Consequently, the Jarque-Bera test rejects the hypothesis of normality for all series. The significance of Ljung-Box statistics indicates significant autocorrelation for all series. This may be explained by a time-varying risk premium, which is one component of the EGARCH-M model.

### 5.2 Estimation results

We now turn our attention to estimate the parameters of the EGARCH-M model by maximum likelihood<sup>1</sup>, assuming that the conditional distribution of the return process is modelled as either GED or APD. We compare these distributions with the normal one in Table 2, where the ML estimates of the parameters are displayed, as well as their standard errors

---

<sup>1</sup>The ML estimation is implemented in Matlab R2016a with the routine "fmincon" and initial values  $\theta_0^{\text{GED}} = (0.002, 0.07, -0.08, -0.06, 0.15, 0.98, 2)$ , and  $\theta_0^{\text{APD}} = (0.002, 0.07, -0.08, -0.06, 0.15, 0.98, 0.5, 2)$ .

|           | Obs. | Mean  | Variance | Skewness | Kurtosis | Jarque-Bera | Ljung-Box |
|-----------|------|-------|----------|----------|----------|-------------|-----------|
| S&P 500   | 7815 | 0.037 | 1.321    | −0.826   | 23.981   | 144230      | 100.681   |
| FTSE 100  | 7847 | 0.027 | 1.246    | −0.297   | 11.615   | 24380       | 78.101    |
| DAX       | 7830 | 0.037 | 2.073    | −0.130   | 8.487    | 9846        | 51.194    |
| CAC 40    | 7464 | 0.026 | 1.958    | −0.018   | 8.248    | 8566        | 60.376    |
| Dow Jones | 7815 | 0.039 | 1.252    | −1.076   | 32.158   | 278360      | 89.134    |

Table 1: *Distributional and other statistical properties of daily stock market returns. Returns have been multiplied by 100. Ljung-Box statistics are calculated for 20 lags. Critical values for 95% and 99% confidence levels are 31.4 and 37.6 based on a chi-squared distribution with 20 degrees of freedom.*

in parentheses. Estimated values  $\hat{\alpha}$  of the APD asymmetry parameter range from 0.535 (CAC 40) to 0.546 (FTSE 100) for the APD, significantly different from 0.5 (symmetry), justifying that there is negative asymmetry. To compare the goodness of fit of the different models, we employ the maximum log-likelihood value (LL) criterion. In all cases, this measure ranks the APD-EGARCH-M model as the best, followed by GED-EGARCH-M. The significant return autocorrelation disappears when looking at standardized residuals of the estimated EGARCH-M models, with Ljung-Box statistics that become insignificant. This confirms that the serial correlation of index returns can, at least partially, be explained by time-varying risk premia.

### 5.3 Results for VaR prediction

To perform an out-of-sample analysis, and check also the predictive performance of the VaR, we divide the data sample  $T$  in two periods. In that way, we create an estimation window and a testing window. We utilize an initial sample of  $N$  observations, chosen as 10% of the full sample  $T$ , for precise initial estimation of the EGARCH-M parameters and the remaining ones are used as an out-of-sample testing window for the 1-day VaR estimates and their performance. In the related literature there are different procedures, either using a fixed estimation window, i.e. the VaR forecasts are computed with the estimated parameters kept fixed during the out-of-sample period, and so the model is estimated only once, or using a rolling estimation window of  $T - N$  observations in which the EGARCH-M

|           | $\omega$ | $\gamma$ | $\delta$ | $\beta$ | $\lambda_0$ | $\lambda_1$   | $\alpha$     | $\nu$   | LL             |
|-----------|----------|----------|----------|---------|-------------|---------------|--------------|---------|----------------|
| S&P 500   |          |          |          |         |             |               |              |         |                |
| Normal    | −0.106   | −0.118   | 0.134    | 0.974   | 0.000       | <b>0.036</b>  | −            | −       | −10435         |
|           | (0.008)  | (0.008)  | (0.010)  | (0.003) | (0.011)     | (0.014)       | −            | −       |                |
| GED       | −0.106   | −0.117   | 0.131    | 0.977   | 0.013       | <b>0.043</b>  | −            | 1.340   | −10245         |
|           | (0.012)  | (0.009)  | (0.011)  | (0.003) | (0.012)     | (0.013)       | −            | (0.028) |                |
| APD       | −0.105   | −0.120   | 0.133    | 0.975   | −0.006      | <b>0.040</b>  | <b>0.544</b> | 1.363   | − <b>10228</b> |
|           | (0.008)  | (0.009)  | (0.014)  | (0.003) | (0.008)     | (0.011)       | (0.007)      | (0.023) |                |
| CAC 40    |          |          |          |         |             |               |              |         |                |
| Normal    | −0.091   | −0.100   | 0.130    | 0.972   | −0.040      | <b>0.045</b>  | −            | −       | −11944         |
|           | (0.004)  | (0.006)  | (0.007)  | (0.003) | (0.004)     | (0.006)       | −            | −       |                |
| GED       | −0.093   | −0.103   | 0.130    | 0.974   | 0.022       | <b>0.040</b>  | −            | 1.549   | −11873         |
|           | (0.008)  | (0.008)  | (0.011)  | (0.004) | (0.020)     | (0.013)       | −            | (0.032) |                |
| APD       | −0.091   | −0.103   | 0.129    | 0.973   | −0.035      | <b>0.038*</b> | <b>0.535</b> | 1.570   | − <b>11862</b> |
|           | (0.008)  | (0.008)  | (0.010)  | (0.004) | (0.025)     | (0.015)       | (0.009)      | (0.044) |                |
| DAX       |          |          |          |         |             |               |              |         |                |
| Normal    | −0.112   | −0.100   | 0.161    | 0.971   | 0.014       | 0.018         | −            | −       | −12623         |
|           | (0.006)  | (0.004)  | (0.006)  | (0.002) | (0.009)     | (0.016)       | −            | −       |                |
| GED       | −0.113   | −0.093   | 0.156    | 0.978   | 0.042       | <b>0.015</b>  | −            | 1.453   | −12494         |
|           | (0.009)  | (0.008)  | (0.012)  | (0.003) | (0.010)     | (0.006)       | −            | (0.017) |                |
| APD       | −0.112   | −0.094   | 0.158    | 0.975   | 0.012       | 0.020         | <b>0.545</b> | 1.466   | − <b>12476</b> |
|           | (0.000)  | (0.025)  | (0.010)  | (0.004) | (0.014)     | (0.026)       | (0.011)      | (0.025) |                |
| FTSE 100  |          |          |          |         |             |               |              |         |                |
| Normal    | −0.122   | −0.094   | 0.153    | 0.976   | −0.010      | <b>0.036</b>  | −            | −       | −10615         |
|           | (0.008)  | (0.005)  | (0.013)  | (0.002) | (0.016)     | (0.012)       | −            | −       |                |
| GED       | −0.120   | −0.093   | 0.148    | 0.978   | −0.001      | <b>0.033*</b> | −            | 1.583   | −10532         |
|           | (0.067)  | (0.019)  | (0.015)  | (0.007) | (0.015)     | (0.024)       | −            | (0.054) |                |
| APD       | −0.117   | −0.092   | 0.146    | 0.978   | −0.011      | 0.035         | <b>0.546</b> | 1.597   | − <b>10530</b> |
|           | (0.008)  | (0.009)  | (0.013)  | (0.003) | (0.012)     | (0.036)       | (0.016)      | (0.029) |                |
| Dow Jones |          |          |          |         |             |               |              |         |                |
| Normal    | −0.113   | −0.113   | 0.142    | 0.972   | 0.005       | <b>0.036</b>  | −            | −       | −10321         |
|           | (0.005)  | (0.003)  | (0.006)  | (0.001) | (0.472)     | (0.014)       | −            | −       |                |
| GED       | −0.111   | −0.109   | 0.137    | 0.976   | 0.020       | <b>0.038</b>  | −            | 1.339   | −10123         |
|           | (0.010)  | (0.012)  | (0.013)  | (0.005) | (0.016)     | (0.010)       | −            | (0.018) |                |
| APD       | −0.108   | −0.110   | 0.135    | 0.975   | 0.001       | 0.036         | <b>0.538</b> | 1.357   | − <b>10106</b> |
|           | (0.007)  | (0.007)  | (0.068)  | (0.036) | (0.494)     | (0.563)       | (0.128)      | (0.454) |                |

Table 2: *Estimation results for the EGARCH(1,1)-M under normal and asymmetric innovations. Standard errors are given in parentheses. Risk premium and skewness estimates that are significant at all levels are displayed in bold, while 1 asterisk is only at 5% level. LL is the log-likelihood value, and with bold is the higher one among the distributions.*

parameters are reestimated  $N$  times moving the estimation window 1-step forward each time, and so provide 1-step-ahead VaR forecasts. We recursively evaluate  $\widehat{\text{VaR}}_{t+1|t}(\vartheta)$ ,  $N \leq t \leq T-1$ , for 1-step-ahead forecasts and for the probabilities  $\vartheta = 0.01, 0.025, 0.05, 0.1$ .

To evaluate the quality of the forecasts, there are standard statistical backtesting procedures that compare *ex ante* VaR forecasts with *ex post* realized returns. If the model is correctly specified, we expect  $\vartheta \times 100\%$  of the observed  $Y_{t+1}$ -values, i.e.  $Y_{N+1}, \dots, Y_T$ , to be smaller than the negative estimated  $\widehat{\text{VaR}}_{t+1|t}(\vartheta)$  that is implied by the model. This is given by the observed frequency defined as

$$f_{1\text{-step-ahead}}(\vartheta) = \frac{1}{T-N} \sum_{t=N}^{T-1} I\left(Y_{t+1} \leq -\widehat{\text{VaR}}_{t+1|t}(\vartheta)\right),$$

where  $I(\cdot)$  is the indicator function, so that the expected frequency is  $\mathbb{E}[f_{1\text{-step-ahead}}(\vartheta)] = \vartheta$ . If it is lower (higher) than  $\vartheta$ , then the model tends to overestimate (underestimate) the risk.

The unconditional coverage (UC) test of Christoffersen (1998) and Kupiec (1995) tests the actual fraction of VaR violations against the theoretical fraction  $\vartheta$ . It is a test of the hypothesis  $H_0 : \mathbb{E}[f_{1\text{-step-ahead}}(\vartheta)] = \vartheta$  against the alternative  $H_1 : \mathbb{E}[f_{1\text{-step-ahead}}(\vartheta)] > \vartheta$ . The dynamic quantile (DQ) test of Engle and Manganelli (2004) is regression-based and tests whether the VaR violations are predictable by past information. If the VaR forecast works correctly, the violation  $I\left(Y_{t+1} \leq -\widehat{\text{VaR}}_{t+1|t}(\vartheta)\right)$  should be unrelated to the forecasted VaR values. The test is based on the process  $\text{Hit}_t$  defined as

$$\text{Hit}_t(\vartheta) \equiv I\left(Y_{t+1} \leq -\widehat{\text{VaR}}_{t+1|t}(\vartheta)\right) - \vartheta,$$

which assumes the value  $\text{Hit}_t(\vartheta) = 1 - \vartheta$  if there is a violation, and  $\text{Hit}_t(\vartheta) = -\vartheta$  if there is not. Also, it holds that  $\mathbb{E}_{t-1}[\text{Hit}_t(\vartheta)] = 0$ . The idea is to regress current violations on past violations in order to test whether the exceedances are independent or not. That is, we estimate the linear regression model

$$\text{Hit}_t(\vartheta) = b_0 + \sum_{k=1}^K b_k \text{Hit}_{t-k}(\vartheta) + e_t,$$

where  $e_t$  is the error term, and then we test the joint hypothesis

$$H_0 : b_0 = b_1 = \dots = b_K = 0,$$

where the DQ test statistic is asymptotically distributed as  $\chi_{K+1}^2$ , see Engle and Manganelli (2004).

In Tables 3-4 in Appendix D we report the values of the empirical coverage (% of violations), together with the failure ratio tests of Christoffersen (1998) and Kupiec (1995), as well as the dynamic quantile test results for each of the three forecasting models, N-EGARCH-M, GED-EGARCH-M, and APD-EGARCH-M, and for the Cornish-Fisher expansion.<sup>2</sup> Not surprisingly, the Gaussian model fails both the Christoffersen and DQ tests for a VaR level that is sufficiently far in the tail (1%), both for the Dow Jones and for the S&P 500. The GED and Cornish-Fisher also fail at this level for the S&P 500, while the model using APD clearly shows the best performance. For the Dow Jones, which is less asymmetric, both GED and APD seem appropriate, while Cornish-Fisher does not pass the failure ratio tests at (1%). Our conclusion is that the EGARCH-M model with parametric fat-tailed innovation distributions is able to generate VaR forecasts that pass classical diagnostic tests. Additional asymmetry may be necessary to account for skewness, as in the case of the S&P 500 returns.

## 6 Conclusions

In this paper, we investigate the relation between positive time-varying risk premia and the unconditional skewness of returns. 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. The marginal distribution of financial time series is often negatively skewed. Under a fully parametric EGARCH-in-Mean specification, we propose to use an asymmetric error distribution in order to match the unconditional asymmetry of asset returns. This is a flexible and sufficient framework to allow for unconditional and conditional skewness, which are both important for risk management applications, using a skewness-corrected risk measure of Value-at-Risk. Applications to VaR of the largest stock market indices complement the analysis. The normal case fails the *DQ* test and mainly for extreme tails (1%), while the VaR forecast works correctly under the APD case that is chosen especially for 1%-percentile.

---

<sup>2</sup>To economize on space we report the results for the Dow Jones and the S&P 500, the results of the other indices are similar and available upon request.

## Acknowledgements

We are grateful to the participants of the seminar series of the department of Economics, Athens University of Economics and Business, the department of Economics, University of Crete, the 12th International Conference on Computational and Financial Econometrics (Pisa), the 1st Quantitative Finance and Financial Econometrics (QFFE) 2018 conference (Marseille), and the IAS Focused Program on Time Series Analysis of Higher Moments and Distributions of Financial Data (Hong Kong). All remaining errors are solely our responsibility. The first author would like to acknowledge financial support from the MOVE-IN Louvain Post-doctoral Fellowship, co-funded by the Marie Curie Actions of the European Commission.

## Appendix A Moments of $h_t$

**Existence of moments of  $h_t$**  Let  $Z$  be a standardized APD random variable with probability density function be given by

$$f(z) = \begin{cases} \frac{\nu}{\kappa} \frac{\Gamma(2/\nu)}{\Gamma(1/\nu)^2} \exp\left(-\left(\frac{\Gamma(2/\nu)}{\alpha\Gamma(1/\nu)}\right)^\nu \left|\frac{z}{\kappa} + 1 - 2\alpha\right|^\nu\right), & \text{if } z \leq -(1 - 2\alpha)\kappa; \\ \frac{\nu}{\kappa} \frac{\Gamma(2/\nu)}{\Gamma(1/\nu)^2} \exp\left(-\left(\frac{\Gamma(2/\nu)}{(1-\alpha)\Gamma(1/\nu)}\right)^\nu \left|\frac{z}{\kappa} + 1 - 2\alpha\right|^\nu\right), & \text{if } z > -(1 - 2\alpha)\kappa, \end{cases}$$

where  $\alpha \in (0, 1)$ ,  $\nu > 0$ , and  $\kappa$  a positive constant defined as

$$\kappa = \Gamma(2/\nu) [\Gamma(3/\nu) \Gamma(1/\nu) [1 - 3\alpha + 3\alpha^2] - \Gamma(2/\nu)^2 [1 - 2\alpha]^2]^{-1/2}.$$

For the existence of moments for  $h_t$ , where

$$h_t = \exp\left\{\frac{\omega}{1 - \beta}\right\} \prod_{i=0}^{\infty} \exp\left\{\beta^i (\gamma Z_{t-1-i} + \delta |Z_{t-1-i}|)\right\},$$

we need mainly to calculate

$$\mathbb{E}[\exp(\beta(\gamma Z + \delta |Z|))].$$

We have that

$$\begin{aligned} \mathbb{E}[\exp(\beta(\gamma Z + \delta |Z|))] &= \mathbb{E}[\exp(\beta(\gamma + \delta)Z)] + \mathbb{E}[\exp(\beta(\delta - \gamma)Z)] \\ &= \frac{\nu}{\kappa} \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})^2} \int_{-(1-2\alpha)\kappa}^{\infty} e^{[\beta(\gamma+\delta)z]} \exp\left[-\left(\frac{\Gamma(\frac{2}{\nu})}{(1-\alpha)\Gamma(\frac{1}{\nu})}\right)^\nu \left|\frac{z}{\kappa} + 1 - 2\alpha\right|^\nu\right] dz \\ &\quad + \frac{\nu}{\kappa} \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})^2} \int_{-(1-2\alpha)\kappa}^{\infty} e^{[\beta(\delta-\gamma)z]} \exp\left[-\left(\frac{\Gamma(2/\nu)}{\alpha\Gamma(1/\nu)}\right)^\nu \left|\frac{z}{\kappa} + 1 - 2\alpha\right|^\nu\right] dz. \end{aligned}$$

In the sequel, set

$$z + (1 - 2\alpha)\kappa = w \Rightarrow dz = dw.$$

Hence,

$$\begin{aligned}\mathbb{E}[\exp(\beta(\gamma Z + \delta|Z|))] &= \frac{\nu}{\kappa} \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})^2} \int_0^\infty e^{[\beta(\gamma+\delta)(w-(1-2\alpha)\kappa)]} \exp\left[-\left(\frac{\Gamma(\frac{2}{\nu})}{(1-\alpha)\Gamma(\frac{1}{\nu})}\right)^\nu \left(\frac{w}{\kappa}\right)^\nu\right] dw \\ &\quad + \frac{\nu}{\kappa} \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})^2} \int_0^\infty e^{[\beta(\delta-\gamma)(w-(1-2\alpha)\kappa)]} \exp\left[-\left(\frac{\Gamma(\frac{2}{\nu})}{\alpha\Gamma(\frac{1}{\nu})}\right)^\nu \left(\frac{w}{\kappa}\right)^\nu\right] dw.\end{aligned}$$

Also, set

$$\begin{aligned}\left[\frac{\Gamma(2/v)}{(1-\alpha)\Gamma(1/v)} \left(\frac{w}{\kappa}\right)^\nu\right]^\nu &= t \Rightarrow \\ \left(\frac{w}{\kappa}\right)^\nu &= t \left[\frac{(1-\alpha)\Gamma(1/v)}{\Gamma(2/v)}\right]^\nu \Rightarrow \\ w &= t^{1/\nu} \frac{\kappa(1-\alpha)\Gamma(1/v)}{\Gamma(2/v)} \Rightarrow \\ dw &= \frac{\kappa(1-\alpha)\Gamma(1/v)}{\Gamma(2/v)} \frac{1}{\nu} t^{1/\nu-1} dt.\end{aligned}$$

By substituting and collecting terms, we have

$$\begin{aligned}\mathbb{E}[\exp(\beta(\gamma Z + \delta|Z|))] &= \frac{\nu}{\kappa} \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})^2} \frac{1}{\nu} \frac{\kappa(1-\alpha)\Gamma(\frac{1}{\nu})}{\Gamma(\frac{2}{\nu})} \int_0^\infty \exp\left[\beta(\gamma+\delta) \left(\frac{t^{1/\nu} \frac{\kappa(1-\alpha)\Gamma(\frac{1}{\nu})}{\Gamma(\frac{2}{\nu})}}{-(1-2\alpha)\kappa}\right)\right] t^{\frac{1}{\nu}-1} e^{-t} dt \\ &\quad + \frac{\nu}{\kappa} \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})^2} \frac{1}{\nu} \frac{\kappa\alpha\Gamma(\frac{1}{\nu})}{\Gamma(\frac{2}{\nu})} \int_0^\infty \exp\left[\beta(\delta-\gamma) \left(\frac{t^{1/\nu} \frac{\kappa\alpha\Gamma(\frac{1}{\nu})}{\Gamma(\frac{2}{\nu})}}{-(1-2\alpha)\kappa}\right)\right] t^{\frac{1}{\nu}-1} e^{-t} dt \\ &= \frac{1-\alpha}{\Gamma(\frac{1}{\nu})} \int_0^\infty \exp\left[\beta(\gamma+\delta) \left(t^{1/\nu} \frac{\kappa(1-\alpha)\Gamma(\frac{1}{\nu})}{\Gamma(\frac{2}{\nu})} - (1-2\alpha)\kappa\right)\right] t^{\frac{1}{\nu}-1} e^{-t} dt \\ &\quad + \frac{\alpha}{\Gamma(\frac{1}{\nu})} \int_0^\infty \exp\left[\beta(\delta-\gamma) \left(t^{1/\nu} \frac{\kappa\alpha\Gamma(\frac{1}{\nu})}{\Gamma(\frac{2}{\nu})} - (1-2\alpha)\kappa\right)\right] t^{1/\nu-1} e^{-t} dt \\ &= \frac{1-\alpha}{\Gamma(\frac{1}{\nu})} e^{-\beta(\gamma+\delta)(1-2\alpha)\kappa} \int_0^\infty t^{(1/\nu-1)} e^{-t} e^{\beta(\gamma+\delta) \frac{\kappa(1-\alpha)\Gamma(1/\nu)}{\Gamma(2/\nu)} t^{1/\nu}} dt \\ &\quad + \frac{\alpha}{\Gamma(\frac{1}{\nu})} e^{-\beta(\delta-\gamma)(1-2\alpha)\kappa} \int_0^\infty t^{(1/\nu-1)} e^{-t} e^{\beta(\delta-\gamma) \frac{\kappa\alpha\Gamma(1/\nu)}{\Gamma(2/\nu)} t^{1/\nu}} dt.\end{aligned}$$

Applying a Taylor series expansion of the exponential terms, i.e.

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

we get:

$$\begin{aligned}
\mathbb{E} [\exp (\beta (\gamma Z + \delta |Z|))] &= \frac{1-\alpha}{\Gamma(1/\nu)} e^{-\beta(\gamma+\delta)(1-2\alpha)\kappa} \int_0^\infty \sum_{j=0}^\infty \left[ \beta(\gamma+\delta) \frac{\kappa(1-\alpha)\Gamma(1/\nu)}{\Gamma(2/\nu)} \right]^j \frac{e^{-t} t^{((1+j)/\nu-1)}}{\Gamma(j+1)} dt \\
&+ \frac{\alpha}{\Gamma(1/\nu)} e^{-\beta(\delta-\gamma)(1-2\alpha)\kappa} \int_0^\infty \sum_{j=0}^\infty \left[ \beta(\delta-\gamma) \frac{\kappa\alpha\Gamma(1/\nu)}{\Gamma(2/\nu)} \right]^j \frac{e^{-t} t^{((1+j)/\nu-1)}}{\Gamma(j+1)} dt \\
&= (1-\alpha) e^{-\beta(\gamma+\delta)(1-2\alpha)\kappa} \sum_{j=0}^\infty \left[ \beta(\gamma+\delta) \frac{\kappa(1-\alpha)\Gamma(1/\nu)}{\Gamma(2/\nu)} \right]^j \frac{\Gamma[(j+1)/\nu]}{\Gamma(1/\nu)\Gamma(j+1)} \\
&+ \alpha e^{-\beta(\delta-\gamma)(1-2\alpha)\kappa} \sum_{j=0}^\infty \left[ \beta(\delta-\gamma) \frac{\kappa\alpha\Gamma(1/\nu)}{\Gamma(2/\nu)} \right]^j \frac{\Gamma[(j+1)/\nu]}{\Gamma(1/\nu)\Gamma(j+1)}, \tag{8}
\end{aligned}$$

by also using  $\int_0^\infty t^{(\nu-1)} e^{-mt} dt = \frac{1}{m^\nu} \Gamma(\nu)$  as before. Hence, the finiteness of the previous expression follows by proving the convergence of  $\sum_{j=0}^\infty \left[ \beta(\gamma+\delta) \frac{\kappa(1-\alpha)\Gamma(1/\nu)}{\Gamma(2/\nu)} \right]^j \frac{\Gamma[(j+1)/\nu]}{\Gamma(1/\nu)\Gamma(j+1)}$  and  $\sum_{j=0}^\infty \left[ \beta(\delta-\gamma) \frac{\kappa\alpha\Gamma(1/\nu)}{\Gamma(2/\nu)} \right]^j \frac{\Gamma[(j+1)/\nu]}{\Gamma(1/\nu)\Gamma(j+1)}$  along the lines of Nelson (1991) and proof of Theorem A1.2, p. 366. The series converges when  $\nu > 1$ , as for GED. The proof can be generalized for any arbitrary order of moments, as in Nelson (1991).

## Appendix B Proofs

**Proof of Lemma 1.** We have that

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

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

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

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  $\{\log h_t\}$ .

Moments of  $h_t$  :

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

where for  $m \geq 2$

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

see also He et al. (2002), Theorem 2. As it is mentioned there, analytical expressions may be difficult to derive. To evaluate the above expectation, we must make a distributional assumption about the innovations. When  $Z_t$  is a standardized APD process, the above result combined with (8) gives the unconditional first- and second-order moments of  $Y_t$  for the EGARCH(1,1)-M. ■

**Proof of Lemma 2.** In (6) we have that the conditional third centered moment of  $Y_t$  is

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

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

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

where  $\mathbb{E} (h_t^{3/2}) = [\text{Var}_{t-1} (\varepsilon_t)]^{3/2}$ , where  $[\text{Var}_{t-1} (\varepsilon_t)]^{3/2} = [\text{Var}_{t-1} (Y_t)]^{3/2}$ . The third moment (asymmetry) of  $\mathbb{E}_{t-1} (Y_t)$  is

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

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

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

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

## Appendix C Innovation distributions

The following discussion is on the innovation distributions used in the paper, i.e. the Generalized Error Distribution and the Asymmetric Power Distribution which is a generalization of the first.

### Appendix C.1 The Generalized Error Distribution (GED)

The Generalized Error Distribution (GED) has been proposed by Nelson (1991) when dealing with the EGARCH process. The GED includes the normal as a special case but allows for fat tails while maintaining finiteness of all moments of the random variable. Its standardized version for a random variable  $z$  is given by

$$f(z) = \frac{\nu \exp \{-1/2 |z/\zeta|^\nu\}}{\zeta 2^{(1+1/\nu)} \Gamma(1/\nu)}, \quad z \in \mathbb{R}, \quad \nu > 0,$$

where  $\zeta \equiv \sqrt{2^{-2/\nu} \Gamma(1/\nu) / \Gamma(3/\nu)}$  and  $\Gamma$  is the gamma function. The next Figure 2 shows the GED density under different shape parameter values.

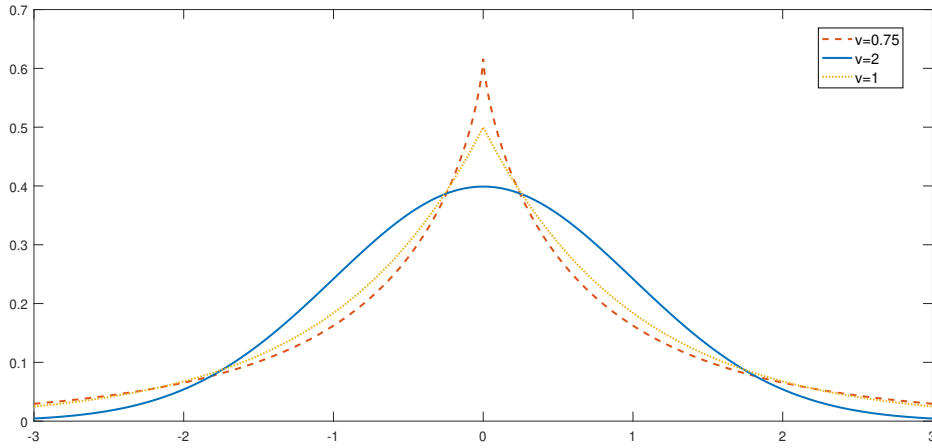


Figure 2: *The standard GED density for different values of the tail parameter, i.e.  $\nu = 2$  for short tails and  $\nu = 0.1$  and  $0.75$  for fat tails.*

However, the GED does not allow for any asymmetry in the data, which can potentially affect the precision of tail risk measures, like the Value-at-Risk or the Expected Shortfall. This drawback is particularly severe in the context of financial time series which are known to have non-zero skewness.

## Appendix C.2 The asymmetric power distribution (APD)

A generalization of the GED to include cases where the data exhibits asymmetry is a flexible family of densities that was proposed by Komunjer (2007) and is called the asymmetric power distribution (APD). This distribution combines the flexible tail decay property of the GED family, measured by a shape parameter denoted by  $\nu$ , with the asymmetry, quantified by a skewness parameter  $\alpha \in (0, 1)$ . Negative skewness corresponds to values of  $0.5 < \alpha < 1$ , and positive one to values of  $0 < \alpha < 0.5$ . The GED family corresponds to the special symmetric case of the APD with  $\alpha = 0.5$ . It also includes distributions that change gradually as the shape parameter decreases, from short-tailed distributions, when  $\nu \in [2, +\infty)$ , to fat-tailed ones, when  $\nu \in (0, 2)$ . Figure 3 plots the standard APD density for fixed values of the tail parameter  $\nu$  and the skewness parameter  $\alpha$ .

**Definition of APD, see Komunjer (2007)** Consider a function  $f : \mathbb{R} \rightarrow \mathbb{R}^+$ ,  $x \mapsto$

$f(x)$  such that

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

where  $\Gamma(\cdot)$  the gamma function,  $0 < \alpha < 1$ ,  $\nu > 0$  and  $\alpha^* = \frac{2\alpha^\nu(1-\alpha)^\nu}{\alpha^\nu + (1-\alpha)^\nu}$ . The function  $f(\cdot)$  is a probability density function and any random variable with density  $f(\cdot)$  is called standard APD.

The moments of a standard APD random variable  $U$  (see Lemma 4 in Komunjer, 2007) can be calculated by

$$\mathbb{E}(U^r) = \frac{\Gamma((1+r)/\nu)}{\Gamma(1/\nu)} \frac{(1-\alpha)^{1+r} + (-1)^r \alpha^{1+r}}{(\alpha^*)^{r/\nu}}.$$

Since in many financial applications the variable of interest, here  $Z$ , is often standardized, i.e. it has mean zero and variance one, that is  $\mu = 0$  and  $\sigma = 1$  in eq. (9), the pdf is then given by

$$f_{APD_Z}(x) = \begin{cases} \frac{\nu}{\kappa} \frac{\Gamma(2/\nu)}{\Gamma(1/\nu)^2} \exp\left(-\left(\frac{\Gamma(2/\nu)}{\alpha\Gamma(1/\nu)}\right)^\nu \left|\frac{x}{\kappa} + 1 - 2\alpha\right|^\nu\right), & \text{if } x \leq -(1-2\alpha)\kappa; \\ \frac{\nu}{\kappa} \frac{\Gamma(2/\nu)}{\Gamma(1/\nu)^2} \exp\left(-\left(\frac{\Gamma(2/\nu)}{(1-\alpha)\Gamma(1/\nu)}\right)^\nu \left|\frac{x}{\kappa} + 1 - 2\alpha\right|^\nu\right), & \text{if } x > -(1-2\alpha)\kappa, \end{cases},$$

where  $a$  and  $\nu$  as above, and  $\kappa$  a positive constant defined as

$$\kappa = \Gamma(2/\nu) [\Gamma(3/\nu) \Gamma(1/\nu) [1 - 3\alpha + 3\alpha^2] - \Gamma(2/\nu)^2 [1 - 2\alpha]^2]^{-1/2}.$$

Based on Table I of Komunjer (2007) about the moments of a standard APD random variable, and considering the case of  $\nu = 2$  (Gaussian), Figure 4 shows the skewness and excess kurtosis as functions of  $\alpha$ , the skewness parameter.

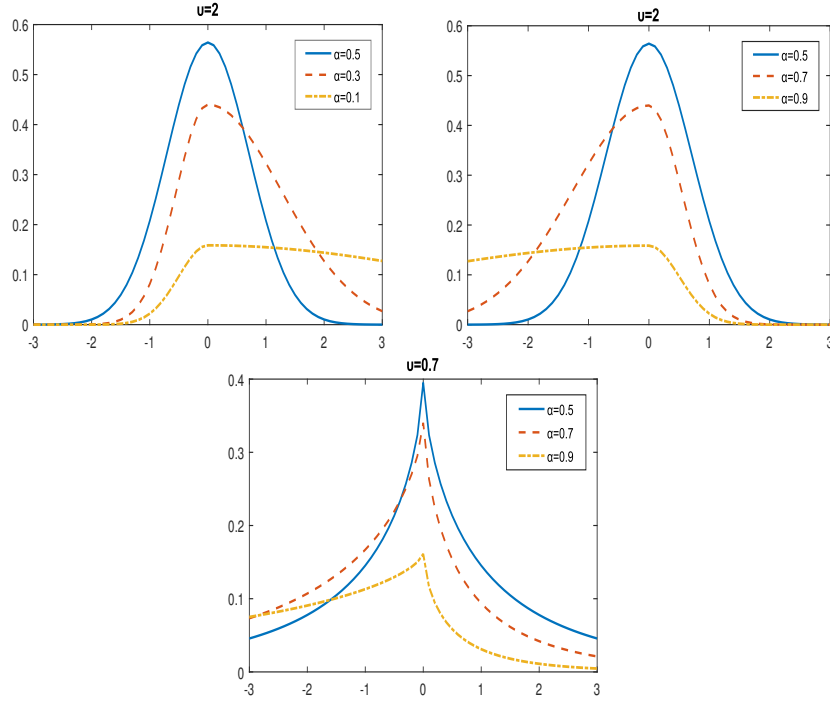


Figure 3: *The standard APD density for different values of the skewness parameter, i.e.  $\alpha = 0.1, 0.3$  for positive skewness,  $\alpha = 0.5$ , for zero skewness, and  $\alpha = 0.7, 0.9$  for negative skewness, and the tail parameter, i.e.  $\nu = 2$  for short tails and  $\nu = 0.7$  for fat tails.*

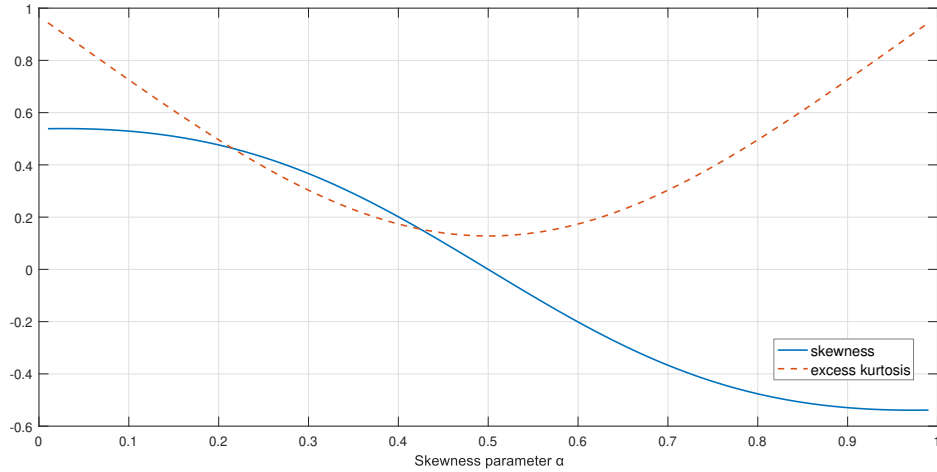


Figure 4: *The APD distribution: skewness (solid blue line) and excess kurtosis (dashed red line) as functions of the skewness parameter  $\alpha$ .*

## Appendix D VaR Tables

| Predictive performance for VaR     |                        |         |                |                |            |            |            |  |
|------------------------------------|------------------------|---------|----------------|----------------|------------|------------|------------|--|
| S&P 500 [Out-of-sample size = 800] |                        |         |                |                |            |            |            |  |
|                                    | $100 \times \vartheta$ | % Viol. | $LR_{cc}$ Chr. | $LR_{PF}$ Kup. | $DQ$ 1 lag | $DQ$ 3 lag | $DQ$ 8 lag |  |
| Normal                             | 1                      | 1.5     | 8.345          | 1.751          | 30.333     | 34.198     | 41.463     |  |
|                                    |                        |         | (0.015)        | (0.186)        | (0.000)    | (0.000)    | (0.000)    |  |
|                                    | 5                      | 4.9     | 0.124          | 0.107          | 0.122      | 7.971      | 17.210     |  |
|                                    |                        |         | (0.94)         | (0.744)        | (0.941)    | (0.093)    | (0.045)    |  |
| GED                                | 1                      | 1.4     | 8.324          | 1.017          | 32.962     | 37.103     | 37.584     |  |
|                                    |                        |         | (0.016)        | (0.313)        | (0.000)    | (0.000)    | (0.000)    |  |
|                                    | 5                      | 4.9     | 0.03           | 0.026          | 0.029      | 7.516      | 16.102     |  |
|                                    |                        |         | (0.986)        | (0.871)        | (0.986)    | (0.111)    | (0.065)    |  |
| APD                                | 1                      | 1       | 3.401          | 0.003          | 10.762     | 11.249     | 11.543     |  |
|                                    |                        |         | (0.183)        | (0.957)        | (0.005)    | (0.024)    | (0.240)    |  |
|                                    | 5                      | 3.8     | 3.033          | 2.489          | 2.803      | 9.846      | 14.316     |  |
|                                    |                        |         | (0.219)        | (0.115)        | (0.246)    | (0.043)    | (0.112)    |  |
| Cornish-Fisher                     | 1                      | 0       | NaN            | NaN            | 8.071      | 8.050      | 8.000      |  |
|                                    |                        |         |                |                | (0.018)    | (0.089)    | (0.534)    |  |
|                                    | 5                      | 5.5     | 0.289          | 0.231          | 0.296      | 4.828      | 12.44      |  |
|                                    |                        |         | (0.865)        | (0.631)        | (0.862)    | (0.305)    | (0.189)    |  |

Table 3: *Predictive performance for the VaR of daily S&P 500 returns for two target probabilities, 1% and 5%. This table shows the empirical coverage (% violations) for the three models, i.e. Normal-EGARCH-M, GED-EGARCH-M, and APD-EGARCH-M, as well as EGARCH-M with Cornish-Fisher approximation of the innovation distribution. The models are estimated with a moving window of 7015 daily observations. The reported tests are the failure ratio tests of Christoffersen (1998) and Kupiec (1995), and the dynamic quantile (DQ) test of Engle and Manganelli (2004) for different number of lags. Significance levels (p-values) are displayed in parentheses. The results are based on 800 forecasts.*

| Predictive performance for VaR       |                        |         |                |                |            |            |            |
|--------------------------------------|------------------------|---------|----------------|----------------|------------|------------|------------|
| Dow Jones [Out-of-sample size = 800] |                        |         |                |                |            |            |            |
|                                      | $100 \times \vartheta$ | % Viol. | $LR_{cc}$ Chr. | $LR_{PF}$ Kup. | $DQ$ 1 lag | $DQ$ 3 lag | $DQ$ 8 lag |
| Normal                               | 1                      | 1.5     | 8.345          | 1.751          | 30.333     | 55.040     | 56.250     |
|                                      |                        |         | (0.015)        | (0.186)        | (0.000)    | (0.000)    | (0.000)    |
|                                      | 5                      | 4.4     | 0.815          | 0.686          | 0.782      | 9.960      | 17.536     |
|                                      |                        |         | (0.665)        | (0.408)        | (0.676)    | (0.041)    | (0.041)    |
| GED                                  | 1                      | 1       | 0.162          | 0.000          | 0.082      | 10.903     | 11.566     |
|                                      |                        |         | (0.922)        | (1.000)        | (0.960)    | (0.028)    | (0.239)    |
|                                      | 5                      | 4.7     | 1.12           | 0.243          | 1.217      | 9.728      | 18.824     |
|                                      |                        |         | (0.571)        | (0.622)        | (0.544)    | (0.045)    | (0.027)    |
| APD                                  | 1                      | 0.4     | 4.157          | 4.146          | 3.152      | 3.144      | 3.123      |
|                                      |                        |         | (0.125)        | (0.042)        | (0.207)    | (0.534)    | (0.96)     |
|                                      | 5                      | 4.1     | 1.636          | 1.368          | 1.543      | 4.621      | 19.525     |
|                                      |                        |         | (0.441)        | (0.242)        | (0.462)    | (0.328)    | (0.021)    |
| Cornish-Fisher                       | 1                      | 0.1     | 9.888          | 9.903          | 6.177      | 6.157      | 6.109      |
|                                      |                        |         | (0.007)        | (0.002)        | (0.046)    | (0.188)    | (0.729)    |
|                                      | 5                      | 4.5     | 0.816          | 0.686          | 0.783      | 9.960      | 17.536     |
|                                      |                        |         | (0.665)        | (0.408)        | (0.676)    | (0.041)    | (0.041)    |

Table 4: *Predictive performance for the VaR of daily Dow Jones returns for two target probabilities, 1% and 5%. Further notes as in Table 3.*

## References

- [1] Campbell, J. Y., Hentschel, L. (1992). No news is good news. *Journal of Financial Economics* 31: 281-318.
- [2] Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review* 39: 841-862.
- [3] Cornish, E., Fisher, R., (1937). Moments and cumulants in the specification of distributions. *Revue de l'Institut International de Statistiques* 5: 307-320.
- [4] Engle, R. F., Gonzalez-Rivera, G. (1991). Semiparametric ARCH models. *Journal of Business & Economic Statistics* 9: 345-359.
- [5] Engle, R. F., Lilien, D. M., Robins, R. P. (1987). Estimating time-varying risk premia in the term structure: the ARCH-M model. *Econometrica* 55: 391-407.
- [6] Engle, R. F., Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 22: 367-381.
- [7] Hafner, C. M., Kyriakopoulou, D. (2019). Exponential-type GARCH models with linear-in-variance risk premium. *Journal of Business & Economic Statistics*, <https://doi.org/10.1080/07350015.2019.1691564>.
- [8] He, C., Silvennoinen, A., Teräsvirta, T. (2008). Parameterizing unconditional skewness in models for financial time series. *Journal of Financial Econometrics* 6: 208-230.
- [9] He, C., Teräsvirta, T., Malmsten, H. (2002). Moment structure of a family of first-order exponential GARCH models. *Econometric Theory* 18: 868-885.
- [10] Hong, H., Stein, J. C. (2003). Differences of opinion, short-sales constraints, and market crashes. *Review of Financial Studies* 16: 487-525.
- [11] Komunjer, I. (2007). Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics* 22: 891-921.
- [12] Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3: 73-84.

- [13] Markowitz, H. (1952). Portfolio selection. *Journal of Finance* 7: 77-91.
- [14] Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica* 41: 867-886.
- [15] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59: 347-370.
- [16] Penaranda, F, Wu, L. (2017). Predictability and performance. *Working Paper*.
- [17] Straumann, D., Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroskedastic time series: a stochastic recurrence equations approach. *Annals of Statistics* 34: 2449-2495.
- [18] Subbotin, M. T. (1923). On the law of frequency of errors. *Mathematicheskii Sbornik* 31: 296-301.