Forecasting the Density of Asset Returns[*]

by

Trino-Manuel Ñíguez
London School of Economics and Political Science

and

Javier Perote
Universidad Rey Juan Carlos, Madrid

Contents:
Abstract
1. Introduction
2. The Positive Edgeworth-Sargan density
3. Evaluating density forecasts
4. Empirical application
5. Concluding remarks
Appendix
References

Abstract

In this paper we introduce a transformation of the Edgeworth-Sargan series expansion of the Gaussian distribution, that we call Positive Edgeworth-Sargan (PES). The main advantage of this new density is that it is well defined for all values in the parameter space, as well as it integrates up to one. We include an illustrative empirical application to compare its performance with other distributions, including the Gaussian and the Student's t, to forecast the full density of daily exchange-rate returns by using graphical procedures. Our results show that the proposed function outperforms the other two models for density forecasting, then providing more reliable value-at-risk forecasts.

Corresponding author:
Javier Perote
Universidad Rey Juan Carlos
Dept. de Economia Aplicada II y Fundamentos del Analisis Economico
Fac. De Ciencias Juridicas y Sociales
Campus de Vicalvaro
Pº de los Artilleros s/n.
28032 Madrid
tel: +34 91 488 7684
email: Perote@fcjs.urjc.es

1 Introduction

In the last years, it seems to have achieved a consensus in the financial econometrics literature on the greater convenience to evaluate the ability of models for forecasting down-side risk in terms of the full forecasted density rather than in terms of point or interval forecasts. Specifically because the latter depend on the election of a proxy for the unobservable "true" value of the target variable, which may be either very costly to obtain –as in the case of the realised volatility, or very noisy –as in the case of the squared or absolute residual; see Andersen and Bollerslev (1998) and Hansen and Lunde (2003), among others. Thus, the evaluation of models according to their accuracy for density forecasting –estimation of the full probability distribution of the random variable, based on the methodology in Rosenblatt (1952) can provide more reliable results.

The problem of fitting the probability distribution of financial asset returns has been widely addressed in the literature by using different families of probability densities, including: a) parametric estimation: e.g. the Student's t (Bollerslev, 1987), the Normal Gaussian and mixtures of Normals (Hamilton, 1991), the Pareto Stable (Mittnik and Rachev, 1993), the Generalised Beta (McDonald and Xu, 1995), the skewed Student's t (Harvey and Siddique, 1999; Lambert and Laurent, 2001), the GED (León and Mora, 1999), and the Pearson type IV distribution (Premaratne and Bera, 2001), among others; b) non-parametric estimation (Engle and Gonzalez-Rivera, 1991) and, c) semi-non-parametric estimation: e.g. the Edgeworth-Sargan (hereafter ES) (Sargan, 1976; Gallant and Tauchen, 1989; Mauleón, 1997; and Mauleón and Perote, 2000). A common characteristic of all these approaches is that they can account for the well-known thick tails feature of financial data.

However, it is a well-known fact that a better fit does not guarantee more accurate predictions (see e.g. Nelson, 1976). Furthermore, from a practical perspective, the out-of-sample performance of a model is of greater concern in the financial markets, e.g. for policy makers or risk managers of financial institutions, who are due to decide on the model (among hundreds of alternatives) used to build the risk predictions of their investment portfolios, on which to update their decisions on capital allocation.

On the other hand, there also are many papers in the literature that deal with forecasting market risk, in which, usually, some or various distributions, able to capture leptokurtosis, are assumed for the returns and evaluated according to their performance; see e.g. Mittnik and Paolella (2000), Giot and Laurent (2003, 2004), and Ñíguez (2003, 2004). It is worth noting that,

the ES density has been much less used, despite its proven flexibility for fitting the thick tails of financial return distributions; some examples include Perote and Del Brío (2003), and Baixaulí and Alvarez (2003) for value-at-risk (VaR hereafter) forecasting; and Mauleón (1998), and Bao et al. (2004) for density forecasting.

The main purpose of this study is to compare the out-of-sample density forecasting performance of a transformation of the ES distribution that we call Positive Edgeworth-Sargan (hereafter PES), with other densities, most notably the Student's t. This is because both functions can capture leptokurtosis; possibly the main feature of financial returns distributions. This new formulation is really a density because cannot take negative values, overcoming the well-known shortcoming of the ES distribution, and integrates up to one.

After a preliminary comparison of the models according to their performance for volatility forecasting by using loss functions, we evaluate density forecasts considering graphical procedures, including: the probability integral transformation in Diebold et al. (1998) and the p-value and p-value discrepancy plots in Davidson and MacKinnon (1998); see also Crnkovic and Drachman (1997); Diebold et al. (1999), Granger (1999a, 1999b), Granger and Pesaran (1999a, 1999b), Berkowitz (2001), Raaij and Raunig (2002) and Bao et al. (2004), among others, for alternative procedures on evaluating density forecasts; and Tay and Wallis (2000) for a complete survey on density forecasting.

The remainder of the paper proceeds as follows. Section 2 introduces our proposed PES probability distribution. Section 3 lays out the setting for the evaluation of the density forecasts. Finally, we present an illustrative empirical application, including VaR calculations, to exchange-rate (FX hereafter) returns in Section 4, followed by our conclusions.

2   The Positive Edgeworth-Sargan density

In this section we introduce the PES probability distribution. It is a reparameterised transformation of the ES density to guarantee positivity for all values of its parameters.[1] The family of ES densities is generally defined in terms of the derivatives of the standard normal, $g(x)$, (Eq. (1)); or more usually in terms, directly, of weighted sums of s order Hermite polynomials,

---

[1]Note that the so-called ES density is only a real density for a subset of the parameter space.

$H_s(x)$ (Eq. (2)),[2]

$$\frac{d^s g(x)}{dx^s} = (-1)^s H_s(x) g(x), \tag{1}$$

$$H_s(x) = x^s - \frac{s!}{2(s-2)!} x^{s-2} + \frac{s!}{2^2 2!(s-4)!} x^{s-4} - ..., \quad s = 1, 2, ... \tag{2}$$

These Hermite polynomials satisfy interesting theoretical properties that allow to define density functions in terms of their weighted sums. In particular, the $H_s(x)$ are orthogonal with respect to the scalar product weighted with the normal density (Eq. (3)). Moreover, when considering the squared polynomials the corresponding integral is $s!$ (Eq. (4)). On the other hand, the derivatives of the Gaussian distribution vanish as the variable tends to infinitive (Eqs. (5)), whilst the derivatives of $H_s(x)$ can be expressed in terms of the lower order polynomials (Eq. (6)),[3]

$$\int H_s(x) H_j(x) g(x) dx = 0 \ \forall s \geq 0, \quad \forall j \geq 0 \text{ and } s \neq j; \tag{3}$$

$$\int H_s(x)^2 g(x) dx = s!, \quad \forall s \geq 0; \tag{4}$$

$$H_s(x) g(x) \underset{x \to \pm\infty}{\to} 0, \quad \forall s \geq 0; \tag{5}$$

$$\frac{dH_s(x)}{dx^s} = s H_{s-1}(x), \quad \forall s \geq 1. \tag{6}$$

Given these properties, it is shown, fairly easily, that the standard ES distribution (defined in Eq. (7)) integrates up to one, and its moments are directly related to the density parameters, δ_s; see Mauleón and Perote (2000) for further details on this density.

$$h(x) = \left[1 + \sum_{s=1}^{q} \delta_s H_s(x) \right] g(x). \tag{7}$$

Nevertheless, as we have mentioned above this distribution is not strictly a density since positive definiteness is not guaranteed.

---

[2] The Hermite expansions are based on Gram-Charlier (1905) or Edgeworth (1907) series.

[3] See Kendall and Stuart (1977) for further details on these theoretical properties.

On the other hand, ensuring positivity is of crucial importance from an applied perspective, since evaluating out-of-sample forecasting performance requires estimation stability through long samples which may have extreme values or structural changes, as e.g. asset returns data. Consequently, there are different papers in the literature that try to give a solution to this still unsolved problem; e.g. Mauleón and Perote (2000) emphasised on the optimization procedure, Jondeau and Rockinger (2001) focused on parameter constraints, and León et al. (2004) on reformulations based on the methodology of Gallant and Tauchen (1989).[4] In this paper we adopt the latter strategy to propose the standard PES density,[5] which is defined as follows,

$$f(x, \boldsymbol{\theta}) = \frac{1}{w} \left[1 + \sum_{s=1}^{q} d_s^2 H_s(x)^2 \right] g(x), \qquad (8)$$

where $\boldsymbol{\theta}$ is a vector of parameters, $\boldsymbol{\theta} = (d_1, ..., d_q)$, and w the constant that makes the density to integrate up to one and is given by,

$$w = \int \left[1 + \sum_{s=1}^{q} d_s^2 H_s(x)^2 \right] g(x)dx = 1 + \sum_{s=1}^{q} d_s^2 s!. \qquad (9)$$

Note that this density is well defined for all values of the $\boldsymbol{\theta}$ parametric space and integrates up to one. Even more, its probability distribution function can be also computed according to the Eq. (10) as a direct application of the properties (3) to (6), see Proof 1 in the Appendix.

$$\int_{-\infty}^{a} \frac{1}{w} \left[1 + \sum_{s=1}^{q} d_s^2 H_s(x)^2 \right] g(x)dx$$
$$= \int_{-\infty}^{a} g(x)dx + \frac{g(a)}{w} \sum_{s=1}^{q} \sum_{i=0}^{s-1} d_s^2 \frac{s!}{(s-i)!} H_{s-1}(a) H_{s-i-1}(a). \qquad (10)$$

In Section 4, we carry out an empirical application to show that this representation is more useful than other functions widely used in financial econometrics (such as e.g. the Student's t), to forecast the full density of high-frequency asset returns .

---

[4]See also Barton and Dennis (1952) for the conditons under which Edgeworth expansions are positive.

[5]Note that the standard PES as defined in Eq. (8), as the general ES density and unlike the standard normal, has not unit variance.

3 Evaluating density forecasts

This section sets up the framework for the evaluation of the models according to their ability for density forecasting. Firstly, we use the graphical procedures in Diebold et at. (1998) which are based on the statistical properties of the probability integral transformations established in Rosenblatt (1952). Secondly, the results are complemented by using the graphical techniques proposed in Davidson and MacKinnon (DMc hereafter) (1998), which are more powerful to discriminate among good performer models.

Let $\widehat{u}_t$ denote the residuals obtained by filtering the returns of a financial asset, r_t, which are assumed to follow an autoregressive moving average (ARMA) process whose disturbances, u_t, are distributed according to either a standardised PES or a standardised Student's t (Bollerslev, 1987) distribution, with conditional variance following a generalised autoregressive conditional heteroskedastic (GARCH) process. Note that by the standardised PES we denote the density of $u_t = x_t\sqrt{h_t}$ where x_t is a variable whose density is given in Eq. (8). Moreover, we truncate the density in the eight order Hermite polynomial and constrain to zero all odd parameters.[6] Therefore, the whole conditional density of u_t is modelled as shown either in Eqs. (11), (13), (14) and (15) for the PES case; or in Eqs. (12), (14) and (16) for the Student's t case,[7]

$$u_t/\Omega_{t-1} \sim PES(0, \sigma_t^2); \tag{11}$$

$$u_t/\Omega_{t-1} \sim t_\nu(0, h_t); \tag{12}$$

$$\sigma_t^2 = kh_t = k\alpha_0 + k\alpha_1 u_{t-1}^2 + \alpha_2 kh_{t-1} = \alpha_0^* + \alpha_1^* u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2; \tag{13}$$

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 h_{t-1}; \tag{14}$$

---

[6]Hereafter, for the sake of simplicity, when we mention the PES distribution we refer to this particular PES density which is defined in Eq. (15).

[7]We also consider the case of the normal distribution, $u_t/\Omega_{t-1} \sim N(0, h_t)$, which is nested on the PES when $d_s = 0 \ \forall s$, as baseline model.

$$f_t^1(u_t, \boldsymbol{\theta}) = \frac{\left[1 + \sum_s d_s^2 H_s \left(\frac{u_t}{\sqrt{h_t}}\right)^2\right]}{\left[1 + \sum_s d_s^2 s!\right] \sqrt{2\pi h_t}} e^{-\frac{1}{2}\frac{u_t^2}{h_t}}, \quad s = 2, 4, 6, 8; \qquad (15)$$

$$f_t^2(u_t, \boldsymbol{\vartheta}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi(\nu-2)h_t}} \left[1 + \frac{u_t^2}{(v-2)h_t}\right]^{-\frac{n+1}{2}}, \qquad (16)$$

where Ω_{t-1} denotes the information set available at time $t - 1$, $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ are finite-dimension unknown parameter vectors, and k is the variance of a variable distributed as the standard PES (see Eq. (8)) constrained to $d_s = 0$ for all $s \neq 2, 4, 6, 8$, which is given by,

$$k = \frac{1 + 10d_2^2 + 216d_4^2 + 9360d_6^2 + 685440d_8^2}{1 + 2d_2^2 + 24d_4^2 + 720d_6^2 + 40320d_8^2} \qquad (17)$$

(see Proof 2 in the Appendix).[8]

It is worth mentioning that we based on the results in Perote (1999), and Mauleón and Perote (2000) to define the PES density, since they show that for many asset return series, the odd parameters in the ES function were not statistically significant, providing the resulting density a quite parsimonious description of their full thick-tailed distributions. On the other hand, it is also important to mention that a better performance could be achieved by considering ES densities with conditional higher moments (see e.g. Harvey and Siddique, 1999; Brooks et al., 2002; and León et al., 2004) although we found that in that case guaranteeing positive definiteness turned out to be elusive, remaining this issue as very interesting for future research.

The predictive accuracy of these two models, as well as the $N(0, h_t)$, are compared according to the graphical methods proposed by Diebold et al. (1998) and DMc (1998). These methododologies are based on the cumulative distribution functions (c.d.f.), $\left\{\widehat{F}_t^i(\widehat{u}_t)\right\}_{t=T+1}^{T+N}$ $i = 1, 2$, evaluated at any given realization of the variable $\widehat{u}_t$, of a sequence of density forecasts $\left\{\widehat{f}_t^i(u_t)\right\}_{t=T+1}^{T+N}$, $i = 1, 2$ –see Eqs. (15) and (16). Thus, the method focuses on calculating the probability integral transform, p_t, of the realization of the process taken with respect to the density forecast; see Eq. (18),

---

[8]Note that the stationarity conditions of the conditional variance process in Eq. (13) are different from the ones of the GARCH(1,1) in Eq. (14).

$$p_t^i = \widehat{F}_t^i(\widehat{u}_t) = \int_{-\infty}^{\widehat{u}_t} \widehat{f}_t^i(u_t) du_t. \tag{18}$$

If $f_t^i(u_t)$ is correctly specified then the probability integral transform series associated with the density $f_t^i(u_t)$, denoted as $\{p_t^i\}_{t=\mathrm{T}+1}^{\mathrm{T}+\mathrm{N}}$, are distributed as i.i.d. $U(0,1)$, so its histogram should reflect that fact. Even more, the null hypothesis of i.i.d. $U(0,1)$ can be tested by computing a confidence interval for every histogram bin (taking into account its binomial structure). In addition, the potentially linear and non-linear forms of time dependence of p_t^i may also be analysed by plotting the empirical correlograms of the series $(p_t^i - \overline{p}^i)^j$, $j = 1, 2, 3, 4$, as suggested by Diebold et al. (1998). Note that detected dependence in the powers of $(p_t^i - \overline{p})$ may reflect conditional moments misspecification. Moreover the reinterpretation of this graphical method in terms of the p-value plots and p-value discrepancy plots proposed by DMc (1998),[9] is useful since p_t^i is the p-value corresponding to the quantile $\widehat{u}_t$ of the forecasted density. Therefore, to determine to what extent one specification perform better than another across every quantile the empirical cumulative distribution of p_t^i may also be plotted and under the correct specification the difference between the c.d.f. of p_t^i and the 45^0 line should tend to zero asymptotically. In particular, the empirical distribution function of p_t^i can be easily computed as

$$\widehat{P}_{p_t^i}^i(y_i) = \frac{1}{\mathrm{N}} \sum_{t=1}^{\mathrm{N}} I(p_t^i \leq y_m), \tag{19}$$

where $I(p_t^i \leq y_m)$ is an indicator function that takes the value 1 if its argument is true and 0 otherwise, and y_m is an arbitrary grid of m points.[10] Alternatively, the p-value discrepancy plot (i.e. plotting $\widehat{P}_{p_t^i}^i(y_m) - y_m$ against y_m) is more revealing when it is necessary to discriminate between alternative specifications that perform similarly in terms of the p-value plot (see e.g. Fiorentini et al., 2003). Note that, under correct density specification, the variable $\widehat{P}_{p_t^i}^i(y_m) - y_m$ must converge to zero.

---

[9]Note that DMc (1998) used the method to compare the size and power of hypothesis tests, while following Fiorentini et al. (2003) we use it to discriminate among alternative models according to their performance for forecasting the full density of the asset returns.

[10]We use the following M points grid, $y_m = 0.001, 0.002, ..., 0.01, 0.015, ..., 0.99, 0.991, ..., 0.999$ ($M = 215$), since it highlights the goodness-of-fit in the distribution tails.

4   Empirical application

4.1   Data and inference

The data set consists of continuously compounded daily returns in percentage of an equally-weighted portfolio based on five major currencies, including: the Deutsche mark (DEM), the Japanese yen (JPY), the British pound (GBP), the Swiss franc (SWF) and the Swedish krona (SDK) from January 4, 1983 to December 17, 2001, for a total of $4,882$ observations, depicted in Figure 4.1. Descriptive statistics for the portfolio returns, r_t, are presented in Figure 4.2, which show that their unconditional distribution is not Normal (the null of the Jarque-Bera test is rejected at the 0 per cent level), and clearly leptokurtic (the sample kurtosis is much greater than 3), what justify the assumption of fat-tailed distributions.
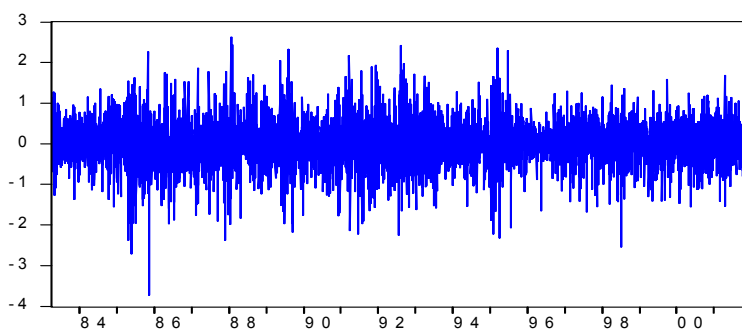
Figure 1: Equally weighted portfolio daily returns series, r_t. In-sample 4/01/1983 - 8/18/1986 (observations 882). Out-of-sample 8/11/1986 - 12/17/2001 (observations 4,000).
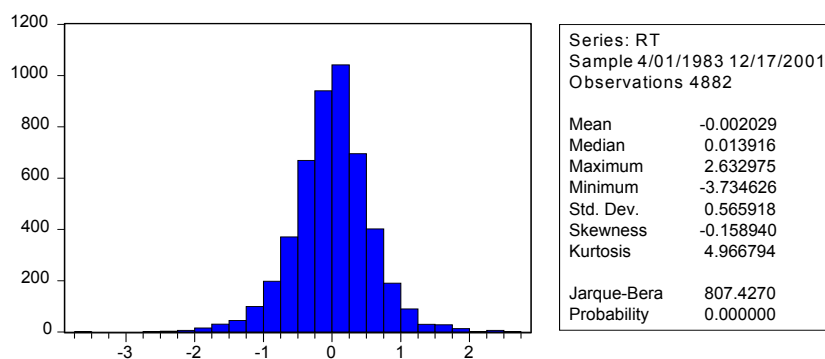
Figure 2: Statistical information of daily portfolio returns.

8

The empirical analysis is outlined as follows: The estimation procedure is performed in two stages: (i) Firstly, the r_t series is filtered by using an AR(1) process, estimated by OLS and selected according to the Information Criterion of Akaike (AIC), (Akaike, 1973). The resulting series of residuals, $\widehat{u}_t$, is split into the in-sample period, which covers the first T= 882 observations, and the out-of-sample period, which includes the last N= 4,000 observations; (ii) the in-sample $\widehat{u}_t$ series is then used as the input for the estimation of the models parameters.[11] Then, stage (ii) is repeated N times, taking a rolling window of size T.

As for the estimation technique, all parameters are estimated by the quasi-maximum lilkelihood (QML) procedure, which provides consistent and asymptotically normal estimates for the three models,[12] and robust standard errors according to the methodology of White (1982) and Weiss (1986). It is worth noting that the sign of the residuals does not affect the parameter estimation process neither through the conditional variance nor the logarithm of the likelihood function.

The in-sample performance of the models is measured by using the mean of the AICs obtained through the N estimations.

Table 1 shows the models estimation results. We observe the usual small structure in the conditional mean of r_t; the slope is statistically significant although the coefficient for the unconditional mean, i.e. ϕ_0, is not statistically different from 0. It is worth pointing out that the three models provide similar significant estimates of the conditional variance equation. The sum of the estimates of either α_1 (or α_1^* in the PES model) and α_2 is near 0.98 in all cases indicating covariance stationarity and strong persistence in the conditional variance.[13]

It is interesting to note that although the three models provide a very similar estimation of the dynamics of the conditional variance, it is observed a slight difference between the estimates of the variance equation from the Gaussian model and the ones obtained when the thick-tailed distributions are assumed. Thus, there are changes in the magnitude of $\widehat{\alpha}_1$, $\widehat{\alpha}_1^*$, and $\widehat{\alpha}_2$ with respect to the Gaussian model, although their sums (i.e. $\widehat{\alpha}_1 + \widehat{\alpha}_2$ for the

---

[11]The AIC is calculated as $2*(n-\ln L)/\text{T}$, where n denotes the number of model parameters, and $\ln L$ the value of the likelihhod function in the parameter estimates.

[12]Note that the first and second moments are well defined for all models.

[13]The estimates of k, $\widehat{\alpha}_0^*$ and $\widehat{\alpha}_1^*$ are: 1.25033, 0.00493 and 0.05616, respectively.

Gaussian and the Student's t, and $\widehat{\alpha}_1^* + \widehat{\alpha}_2$ for the PES) remain unchanged.

In relation to the fit of the distribution tails, we can see that the estimated degrees of freedom, $\widehat{\nu}$, is around 10, even after correcting for volatility clustering. It is also found that the estimation of the Hermite polynomial weights in the PES distribution (with only the coefficient d_6 not statistically significant) corroborates the existence of leptokurtosis in the returns conditional distribution.

It is worth mentioning that although d_6 is not significant for the first in-sample window (4/01/1983 - 8/10/1986), we detect that all four d coefficients are significant for many in-sample windows through the out-sample period, what points out that those sample present more extreme values, showing the model itself flexible enough to capture the higher kurtosis of those distributions. This fact would justify the inclusion of this parameter or even others weighting higher order Hermite polynomials. It also proves the great flexibility of the PES density to fit the tails in the density, in what constitutes the main difference from the Student's t distribution, all even for the small size of the in-sample window used in this application. Moreover, the optimization of the likelihood function based on the PES density function do not present convergence difficulties even for initial values far away for the optimum, overcoming the usual convergence problems presented by other ES expansions of the normal density proposed in the literature.

In relation to the overall in-sample goodness-of-fit (N estimations), we observe that according to the mean of the AICs, the Student's t and the PES models provide a very similar fit, being the one of the former slightly better and both clearly outperforming the one provided by the Gaussian model. Note that the fact of that either d_6 or d_8, or both, are not significant for some in-sample windows, may lead to misleading conclusions regarding the in-sample goodness-of-fit according to the AIC. Observe that, from a practical perspective, if one/some estimates are found to be not significant they may be ruled out, and re-estimate the model without it/them, what may change the magnitude of the AIC in favour of the PES model. This flexibility is one the advantages of the ES-type distributions in relation to the Student's t.

Next, we measure the forecasting ability of the models for density and VaR forecasting. The aim is to shed light on the appropriateness of the three distributions for practical risk management applications.

Table 1: Estimation results.

Mean Equation: $r_t = \phi_0 + \phi_1 r_{t-1} + u_t; \ u_t = h_t^{1/2} x_t; \quad u_t/\Omega_{t-1} \sim N(0, h_t);$
$u_t/\Omega_{t-1} \sim t_\nu(0, h_t); \ u_t/\Omega_{t-1} \sim PES(0, \sigma_t^2);$

Variance Equations: $h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 h_{t-1};$
$\sigma_t^2 = k h_t = \alpha_0^* + \alpha_1^* u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2.$

	Parameters	Gaussian	Student's t	PES
Mean equation	ϕ_0	-0.0019 (-0.23)	-0.0019 (-0.23)	-0.0019 (-0.23)
	ϕ_1	0.0394 (2.75)	0.0394 (2.75)	0.0394 (2.75)
Variance equation	α_0	0.0052 (1.96)	0.0047 (1.76)	0.0039 (1.78)
	α_1	0.0642 (4.28)	0.0574 (3.81)	0.0449 (3.83)
	α_2	0.9227 (50.51)	0.9302 (52.06)	0.9309 (53.62)
Weights	d_2	-	-	0.1499 (3.34)
	d_4	-	-	0.0161 (1.67)
	d_6	-	-	0.0000 (0.00)
	d_8	-	-	-0.0002 (-1.80)
DoF	ν	-	10.173 (3.24)	-
AIC		1.6557	1.6154	1.6184

The Table presents the models estimation results. The reported coefficients shown in the first two rows of the table are OLS estimates of the AR(1) process for the percentage daily returns from January 4, 1983 to December 17, 2001, of an equally-weighted portfolio based on five currencies, including: the Deutsche mark (DEM), the Japanese yen (JPY), the British pound (GBP), the Swiss franc (SWF) and the Swedish krona (SDK). The coefficients shown in the rest rows of the table are QML estimates of the GARCH(1,1) model under the Gaussian, the Student's t or the PES distribution for the AR(1) in-sample residuals, 4/01/1983 - 8/10/1986. d_s, $s = 2, 4, 6, 8$, denotes the weight parameter of the order s Hermite polynomial in the PES distribution. DoF denotes degrees of freedom, and AIC is the mean of the AICs of the N estimations through the whole out-of-sample period. See Eq. (17) for the function of k on the d_s. QML t-statistics are in parenthesis below the parameter estimates.

4.2 Forecasting analysis

The out-of-sample forecasting performance is analysed preliminarily for the conditional volatility variable, by using the usual symmetric loss functions, i.e. the mean squared prediction error (MSPE) and the mean absolute prediction error (MAPE), with respect to the squared residuals, $\widehat{u}_{t+i}^2$, $i = 1,...,$N. To conclude this preliminary part, the significativity of the difference between the loss functions is checked by performing the test of Diebold and Mariano [DM] (1995). This test assumes no differences between the loss-functions of two alternative models under the null hypothesis. The test statistic is $DM_s = \overline{y}/\sqrt{2\pi\widehat{\varphi}_y\left(\omega=0\right)/\text{N}}$, where $\overline{y}$ is the sample mean of the differences in the forecasting errors, and $\widehat{\varphi}_y\left(\omega=0\right)$ is the spectral density function of the forecasting error differences evaluated at the zero frequency (long run variance). This statistic is asymptotically distributed as a standard normal under the null. We compute $\widehat{\varphi}_y\left(\omega=0\right)$ using the heteroskedasticity and autocorrelation consistent estimator of Newey and West (1987).[14]

Table 2 shows the values of the loss functions and their corresponding DM's test t statistics. The "best" to "worst" performance models ranking in relation to the prediction error functions is: Gaussian, PES and Student's t. However, the null of the DM's test is accepted in all cases except for the case of the MAPE from the Student's t and PES models. As a conclusion, we find that the three models provide, overall, the same performance, being very difficult to discriminate among them, in accordance to their similarity in the estimation conditional variance dynamics.

We note that this result might seem counterintuitive since one would, a priory, expect significant differences between the forecasting capacity of the PES and the Student's t models in relation to the Gaussian. A possible explanation for that, is that the daily squared residuals constitute very noisy proxies for the "true" volatility, so its use may give misleading results; see e.g. Andersen and Bollerslev (1998), and Andersen et al. (2001). This fact has led to the development of alternative techniques for providing more reliable evaluations of volatility models, including the aforementioned graphical procedures.

Thus, it seems that the higher flexibility of the Student's t and PES distributions would help only to forecast measures related to the distribution tails of the variable. As mentioned in the introduction, one of the reasons for using these densities is because they are capable of capturing the

---

[14]Note that as the forecast horizon is one period ahead, the Newey and West estimator of $\widehat{\varphi}_y(\omega=0)$ is given by the sample variance.

Table 2: Out-of-sample volatility forecasting performance.

	Gaussian	Student's t	PES
MSPE	0.3373 (-0.049)†	0.3384 (0.003)*	0.3383 (-0.046)**
MAPE	0.3305 (-0.265)†	0.3353 (0.202)*	0.3311 (-0.030)**

Out-of-sample volatility forecasting performance. Sample 8/11/1988 - 12/17/2001. Predictions 4,000. The table reports the mean squared prediction error (MSPE) and mean absolute prediction error (MAPE), for one-period ahead forecasts of the conditional variance.

† Denotes t-statistics of the DM's test for the significativity of the difference between the loss functions from the Gaussian and Student's t models.

* Denotes t-statistics of the DM's test for the significativity of the difference between the loss functions from the Student's t and PES models.

** Denotes t-statistics of the DM's test for the significativity of the difference between the loss functions from the Gaussian and PES models.

leptokurtosis shown by the data, what is crucial when computing the quantiles of the one-period-ahead predictive distributions of portfolio returns required in VaR calculations. To determine to what extent the PES is more useful than the Student's t (and the Normal) we have calculated the probability integral transform, p_t, of u_t with respect to density forecasts produced under the assumptions of that u_t is conditionally distributed either as a PES, Student's t or Normal, whose conditional variances are modelled to follow a GARCH(1,1) process. Thus, if the model is correctly specified, p_t should be uniformly distributed, $p_t \sim U(0,1)$.

Figures 3a, 4a and 5a provide estimates of the density of p_t under the different distributional assumptions. The figures show that there is a significant difference between the performance of the PES and Student's t models and the Gaussian, the latter providing clearly the least accurate density forecasts (as revealed by the butterfly shape of its histogram), and the PES distribution showing itself as the most appropriate for forecasting the full density of the portfolio. Observe that for the 5 per cent confidence levels shown in the tables, the PES model provides the histogram of p_t that better adjust to the one of a uniformly distributed variable, since it remains within the corresponding confidence interval bounds. Moreover, this analysis is complemented with the study of the correlograms of the first four powers of the series $p_t - \overline{p}$ (displayed in Figures 3b, 4b and 5b) which reveal that the conditional mean and variance processes adequately capture the moments conditional dynamics.[15]

---

[15]Observe that in our case it would not be necessary to model conditional skewness and kurtosis dynamics; see Harvey and Siddique (1999) and León et al. (2004) for details.
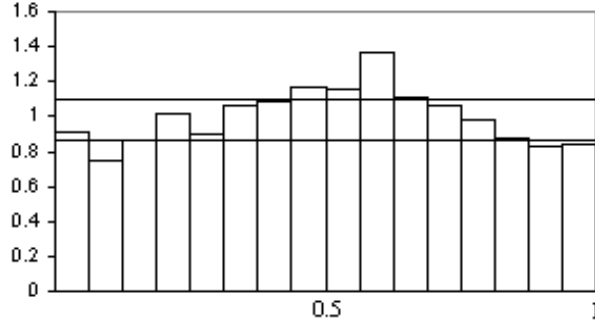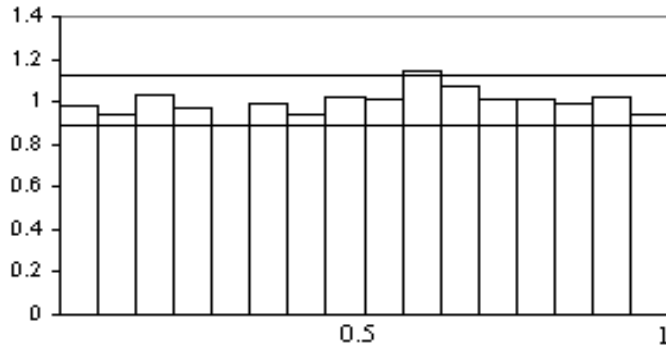
Figure 3a: Estimates of the density of the probability integral transform of $\widehat{u}_t$ with respect to density forecasts obtained with the Gaussian model.
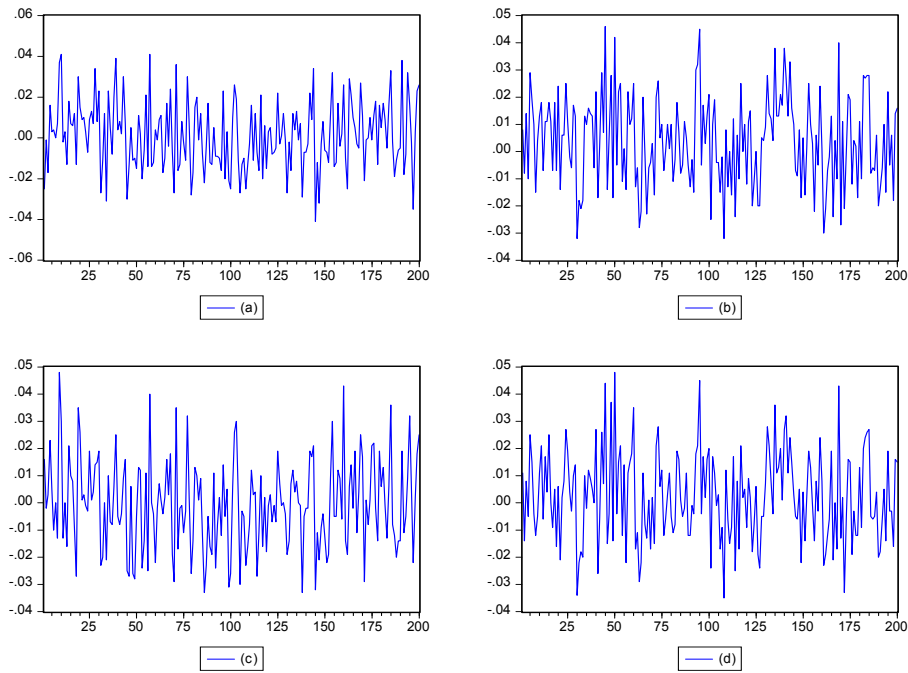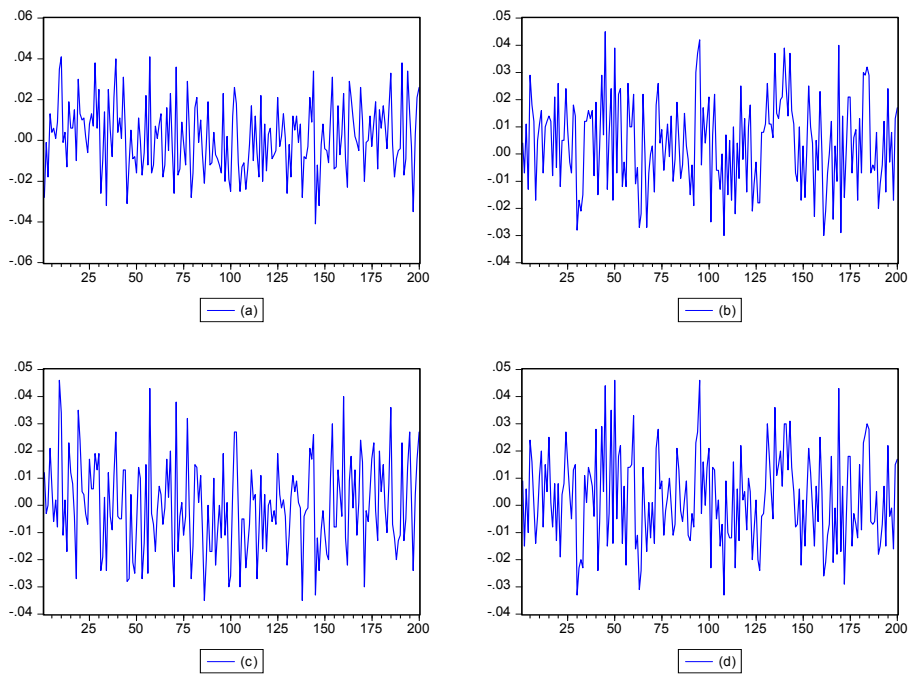

Figure 3b: Panels (a) to (d) show autocorrelations of series $(p_t - \overline{p})$, $(p_t - \overline{p})^2$, $(p_t - \overline{p})^3$, $(p_t - \overline{p})^4$ where p_t is the probability integral transform of $\widehat{u}_t$ with respect to density forecasts obtained with the Gaussian model.

14

Figure 4a: Estimates of the density of the probability integral transform of $\widehat{u}_t$ with respect to density forecasts obtained with the Student's t model.

Figure 4b: Panels (a) to (d) show autocorrelations of series $(p_t - \overline{p})$, $(p_t - \overline{p})^2$, $(p_t - \overline{p})^3$, $(p_t - \overline{p})^4$ where p_t is the probability integral transform of $\widehat{u}_t$ with respect to density forecasts obtained with the Student's t model.
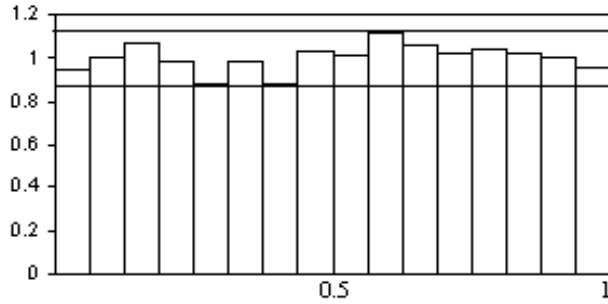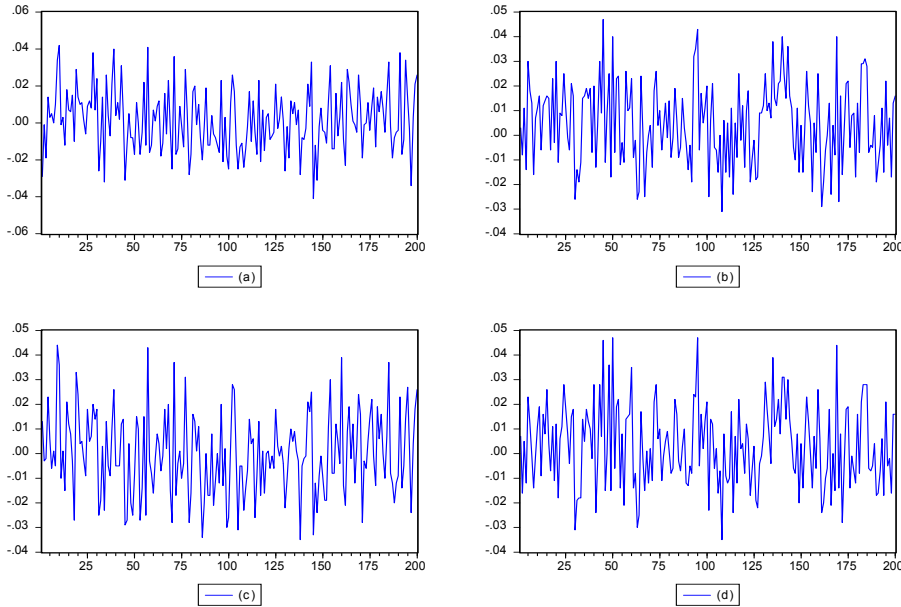
15

Figure 5a: Estimates of the density of the probability integral transform of $\widehat{u}_t$ with respect to density forecasts obtained with the PES model.


Figure 5b: Panels (a) to (d) show autocorrelations of series $(p_t - \overline{p})$, $(p_t - \overline{p})^2$, $(p_t - \overline{p})^3$, $(p_t - \overline{p})^4$ where p_t is the probability integral transform of $\widehat{u}_t$ with respect to density forecasts obtained with the PES model.

Furthermore, we use DMc's (1998) p-value discrepancy plots to try to discriminate better between the best performer models and complement the former findings. Figures 6 and 7 show that the Gaussian distribution performs much worst than the other two, but do not provide enough information to allow us to chose between the Student's t and the PES. So, according to this criteria the PES and the Student's t provide very similar accurate forecasts of the full conditional distribution.

16
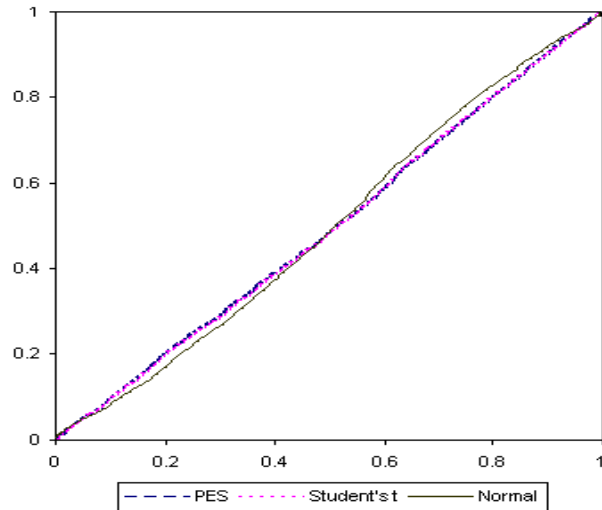
Figure 6: P Value Plot of the empirical c.d.f. of the probability integral transform of $\widehat{u}_t$ obtained under the Gaussian, Student's t and PES models.
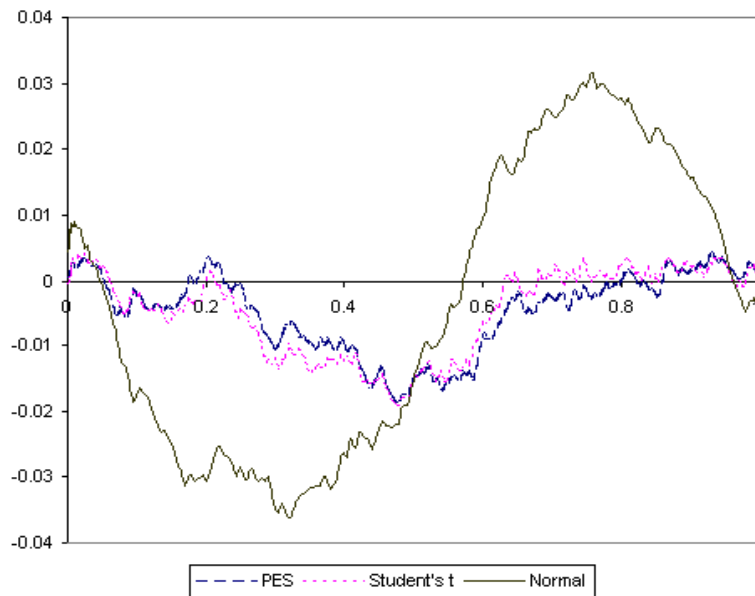
Figure 7: P Value Discrepancy Plot of the empirical c.d.f. of the probability integral transform of $\widehat{u}_t$ obtained under the PES, Student's t and Gaussian models.

As a conclusion the graphical analysis show that it is difficult to decide on the PES and the Student's t as the most accurate distribution for forecasting

17

the conditional density of the portfolio returns. They both seem to forecast very well the full density including the shape of the tails, although given the results in Figures 3 to 5 the PES would be slightly more appropriate.

Finally, we compute one-day-ahead VaR forecasts, denoted as $\widehat{VaR}_{m,\mathrm{T}+1}^{\alpha}$, under the different specifications, m, through the whole out-sample period, for various confidence levels $\alpha = 0.1, 0.05, 0.025, 0.01$. The aim of this application is to comment further on the specific models differences in forecasting the FX returns distribution tails, what may be of great interest for financial risk practitioners. VaR forecasts are calculated as the $100 \cdot \alpha$ order quantile of the one-day-ahead forecasted conditional distribution of the returns when model m is assumed, so it is given by,

$$\widehat{VaR}_{m,\mathrm{T}+1}^{.01} = \widehat{\mu}_{\mathrm{T}+1} - \widehat{h}_{m,\mathrm{T}+1}^{1/2} \cdot \widehat{\delta}_{m,\mathrm{T}+1}, \tag{20}$$

where $\widehat{\mu}_{\mathrm{T}+1}$ and $\widehat{h}_{m,\mathrm{T}+1}^{1/2}$ are the forecasts of the conditional mean and conditional standard deviation, respectively, of the distribution under model m,[16] and $\widehat{\delta}_{m,\mathrm{T}+1}$ is the quantile of the forecasted standard distribution under model m. The calculation of $\widehat{\delta}_{m,\mathrm{T}+1}$ under either the Gaussian or the Student's t conditional distributions is quite straightforward; see e.g. Tsay (2002). On the other hand when the PES is assumed, $\widehat{\delta}_{m,\mathrm{T}+1}$ is given by the upper limit of the integral in Eq. (10) for a given probability level α and for $s = 2, 4, 6, 8$.

Table 3: Value-at-Risk forecasting.

	Gaussian		Student's t		PES	
	l_1	l_2	l_1	l_2	l_1	l_2
$\overline{VaR}_s^{.1}$	-0.8006	-0.6245	-0.7781	-0.6051	-0.7540	-0.5830
$\overline{VaR}_s^{.05}$	-1.0224	-0.8035	-1.0429	-0.8172	-1.0243	-0.7904
$\overline{VaR}_s^{.025}$	-1.2147	-0.9587	-1.3079	-1.0285	-1.2961	-1.0128
$\overline{VaR}_s^{.01}$	-1.4383	-1.1392	-1.6747	-1.3193	-1.6538	-1.3125

The table reports the means of the one-day-ahead VaR forecasts for the confidence level $\alpha = 0.1, 0.05, 0.025, 0.01$, denoted as $\overline{VaR}^{\alpha}$, obtained with the different models. The out-of-sample period is split into two equal subperiods, l_i, $i = 1, 2$, and the means of the VaR forecasts are calculated for each l_i.

---

[16]It is worth mentioning that $\widehat{\mu}_{t+1}$ is obtained recursively through a T size rolling window using the selected AR(1) process for r_t, so it is given by: $\mathbf{E}_t(r_{t+1}) = \widehat{\mu}_{t+1} = \widehat{\phi}_0 + \widehat{\phi}_1 r_t$, where $\mathbf{E}_t(\cdot)$ denotes the mathematical expectation conditional on all information avalaible at time t.

18
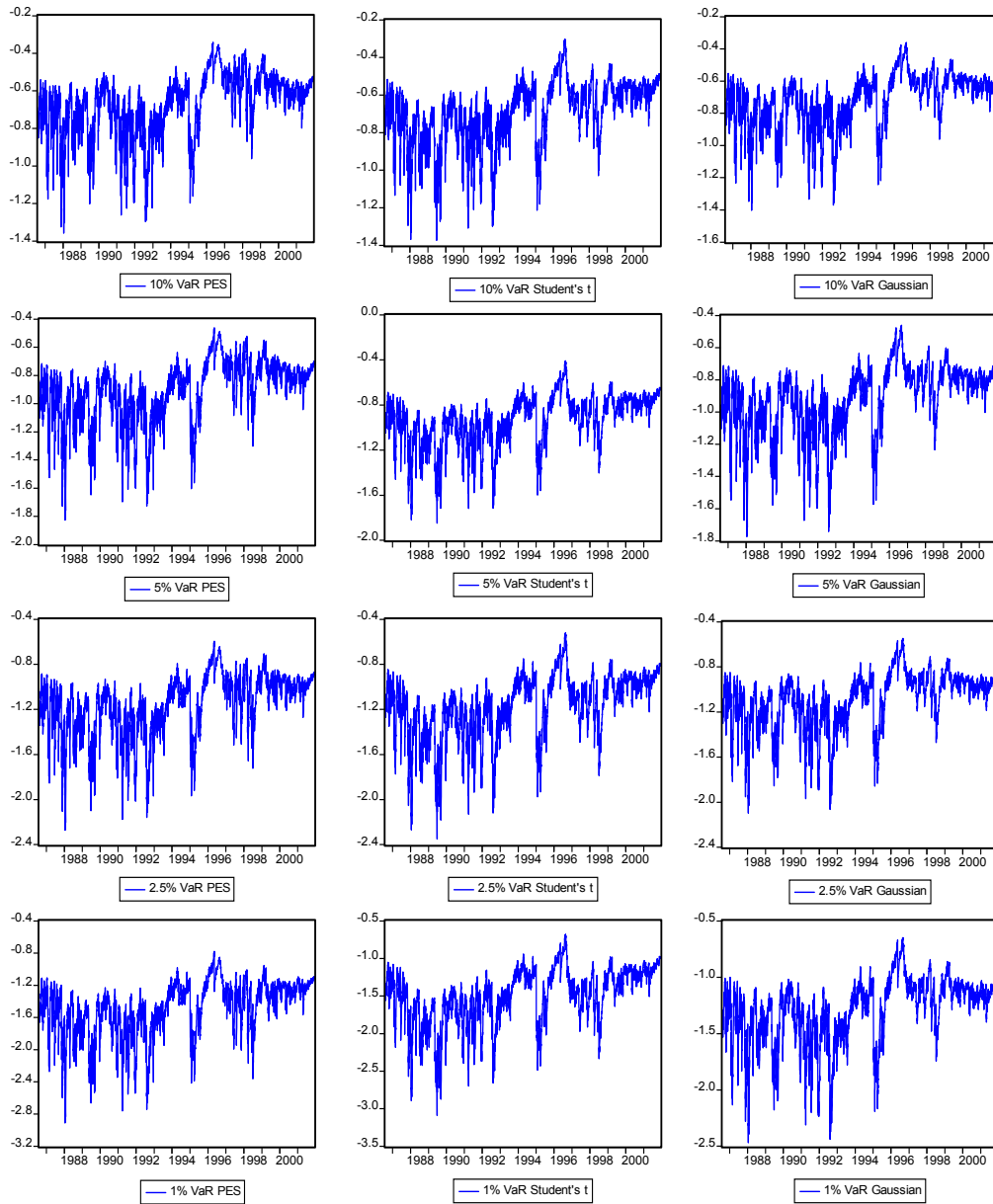
Figure 8: One-day-ahead VaR forecasts for 10, 5, 2.5 and 1 per cent confidence levels obtained with the different models through the whole out-of-sample period 8/11/1986 - 12/17/2001 (forecasts 4,000).

Table 3 above reports the means of the one-day-ahead VaR forecasts obtained with the different models. The out-of-sample period is split into

two equal subperiods and the means of the VaR forecasts are calculated for each of then and for the different confidence levels. It is found that the portfolio risk (VaR in absolute value) forecast obtained under the Gaussian distribution is greater than the one obtained under the other two distributions only for $\alpha = 0.1$. On the other hand, under the Student's t distribution the risk is over-forecasted in relation to the PES distribution for all confidence levels. These results can also be observed in the plots of the VaR forecasts, through the whole out-of-sample period presented in Figures 8.

5   Concluding remarks

The Edgeworth-Sargan (ES) density has been proved to fit accurately the density of returns for most high-frequency financial variables, specifically for its flexibility to describe the tails shape of leptokurtic distributions. Nevertheless, it presents an important drawback because positivity is not guaranteed for all values in the parametric space. When the ES is used for fitting densities, it is a minor shortcoming, but in the density forecasting context, it requires a further general solution.

In this paper we have tackled this problem by proposing a transformation of the ES distribution that we call Positive Edgeworth-Sargan (PES). More specifically, we study its statistical properties and provide an expression to calculate its probability distribution function. In addition we examine its practical applicability by discussing an empirical example for exchange-rates returns.

Our results show that the PES distribution perform quite well in fitting and forecasting the full conditional distribution of the portfolio returns in relation to other widely used distributions in empirical finance, such as the Student's t and the Normal. The analysis was carried out by using the graphical methods proposed in Diebold et al. (1998) and Davidson and MacKinnon (1998). Furthermore, as one of the reasons for using the PES distribution may be to calculate the quantiles of the forecasted distributions of portfolio returns required in VaR analysis, and to determine to what extent the PES is more useful than the Student's t and the Normal, we have computed 4,000 one-day-ahead VaR forecasts for the FX portfolio returns by using the three distributions. The PES is found to provide less conservative measures of downside risk in relation to the Student's t distribution for all confidence levels, and in relation to the Gaussian for 5, 2.5 and 1 per cent levels.

Appendix

This appendix includes the proofs of some interesting properties concerning the PES density proposed in this paper.

Proof 1:

The proof provides an explicit form for the standard PES probability distribution function in terms of the c.d.f. of the standard normal, $g(x)$, as follows

$$\int_{-\infty}^{a} \frac{1}{w} \left[1 + \sum_{s=1}^{q} d_s^2 H_s(x)^2 \right] g(x) dx$$

$$= \frac{1}{w} \int_{-\infty}^{a} g(x) dx + \frac{1}{w} \sum_{s=1}^{q} d_s^2 \int_{-\infty}^{a} H_s(x) H_s(x) g(x) dx$$

$$= \frac{1}{w} \int_{-\infty}^{a} g(x) dx + \frac{1}{w} \sum_{s=1}^{q} d_s^2 \left(\sum_{i=0}^{s-1} \frac{s!}{(s-i)!} H_{s-1}(x) H_{s-i-1}(x) g(x) \bigg|_{-\infty}^{a} \right)$$

$$+ \frac{1}{w} \sum_{s=1}^{q} d_s^2 \int_{-\infty}^{a} g(x) dx$$

$$= \int_{-\infty}^{a} g(x) dx + \frac{g(a)}{w} \sum_{s=1}^{q} d_s^2 \sum_{i=0}^{s-1} \frac{s!}{(s-i)!} H_{s-1}(a) H_{s-i-1}(a),$$

where the integral $\int H_s(x) H_s(x) g(x) dx$ is solved by parts as indicated below,

$$\int H_s(x) H_s(x) g(x) dx$$

$$= -H_s(x) H_{s-1}(x) g(x) + \int H_{s-1}(x) g(x) \frac{d H_s(x)}{dx} dx$$

$$= -H_s(x) H_{s-1}(x) g(x) + s \int H_{s-1}(x) H_{s-1}(x) g(x) dx,$$

since,

$$\int H_s(x) g(x) dx$$

$$= \int (-1)^s \frac{d^s(x)}{dx^s} dx = (-1)^s \frac{d^{s-1}(x)}{dx^{s-1}}$$

$$= (-1)^s (-1)^{s-1} H_{s-1}(x) g(x) = -H_{s-1}(x) g(x).$$

Therefore, by repeating the same argument recursively we obtain,

$$\int H_s(x)H_s(x)g(x)dx = \sum_{i=0}^{s-1} \frac{s!}{(s-i)!}H_{s-1}(x)H_{s-i-1}(x)g(x) + \int g(x)\,dx.$$

Proof 2:

The variance of the standard PES constrained to $d_s = 0$ for all $s \neq 2, 4, 6, 8$, denoted by k, is given by,

$$
\begin{aligned}
k &= \mathbf{E}\left[x^2\right] = \frac{1}{w}\int x^2 [1 + d_2^2\left(x^2-1\right)^2 \\
&\quad + d_4^2\left(x^4-6x^2+3\right)^2 + d_6^2\left(x^6-15x^4+45x^2-45\right)^2 \\
&\quad + d_8^2\left(x^8-28x^6+220x^4-420x^2+105\right)^2]g(x) \\
&= \frac{1}{w}[\mu_2 + d_2^2\left(\mu_6-2\mu_4+\mu_2\right) + d_4^2\left(\mu_{10}-12\mu_8+42\mu_6-36\mu_4+9\mu_2\right) \\
&\quad + d_6^2(\mu_{14}-30\mu_{12}+315\mu_{10}-1380\mu_8+2475\mu_6-1350\mu_4+225\mu_2) \\
&\quad + d_8^2(\mu_{18}-56\mu_{16}+1204\mu_{14}-12600\mu_{12}+67830\mu_{10}-182280\mu_8 \\
&\quad + 220500\mu_6-88200\mu_4+11025\mu_2)] \\
&= \frac{1+10d_2^2+216d_4^2+9360d_6^2+685440d_8^2}{1+2d_2^2+24d_4^2+720d_6^2+40320d_8^2},
\end{aligned}
$$

where $\mathbf{E}(\cdot)$ is the unconditional mathematical expectation.[17]

---

[17]Note that if $x \sim N(0,1)$ then $\mu_s = \mathbf{E}\left[x^s\right] = \frac{s!}{(s/2)!}\frac{1}{2^{s/2}}$ for all s even.

References

[1] Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csáki (eds.), *Second International Symposium on Information Theory*. Budapest: Akadémiai Kiadó.

[2] Andersen TG, Bollerslev T. 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39 (4), 885-905.

[3] Andersen TG, Bollerslev T, Diebold FX, Labys P. 2003. Modeling and forecasting realized volatility. *Econometrica* 71 (2), 579-625.

[4] Bao Y, Tae-Hwy L, Saltoglu B. 2004. A test for density forecast comparison with applications to risk management. Working Paper. *UC Riverside.*

[5] Barton DE, Dennis KER. 1952. The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal. *Biometrika* 39, 425-427.

[6] Baixaulí S, Alvarez S. 2003. Accurate measures of value at risk fitting fat tails. Discussion Papers in Economics. *University of York.*

[7] Berkowitz J. 2001. Testing density forecasts with applications to risk management. *Journal of Business and Economics Statistics* 19, 465-474.

[8] Bollerslev T. 1987. A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542-547.

[9] Brooks C, Burke SP, Persand G. 2002. Autoregressive conditional kurtosis. ISMA Centre Discussion Paper in Finance 2002-05. *University of Reading.*

[10] Charlier CV. 1905. Uber die darstellung willkurlicher funktionen. *Arvik fur Mathematik Astronomi och fysik* 9, 1-13.

[11] Davidson R, MacKinnon JG. 1998. Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School of Economic & Social Studies* 66 (1), 1-26.

[12] Diebold FX, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economics Statistics* 13, 253-264.

[13] Diebold FX, Gunther TA, Tay SA. 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39 (4), 863-883.

[14] Diebold FX, Hahn J, Tay SA. 1999. Multivariate density forecasts evaluation and calibration in financial risk management: high-frequency returns of foreign exchange. *Review of Economics and Statisrics* 81, 661-673.

[15] Edgeworth FY. 1907. On the representation of statistical frequency by series. *Journal of the Royal Statistical Society*, series A, 80.

[16] Engle RF, Bollerslev T. 1986. Modeling the persistence of conditional variances. *Econometric Reviews* 5, 1-50.

[17] Engle RF, González-Rivera G. 1991. Semiparametric ARCH models. *Journal of Business and Economic Statistics* 9, 345-359.

[18] Fiorentini G, Sentana E, Calzolari G. 2003. Maximum likelihood estimation and inference in multivariate conditional heteroskedastic dynamic regression models with Student t innovations. *Journal of Business and Economic Statistics* 21 (4), 532-546.

[19] Gallant AR, Tauchen G. 1989. Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica* 57 (5) 5, 1091-1120.

[20] Giot P, Laurent S. 2003. Value-at-risk for long and short trading positions. *Journal of Applied Econometrics* 18 (6), 641-666.

[21] Giot P, Laurent S. 2004. Modeling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, forthcoming.

[22] Granger CWJ. 1999a. Outline of forecast theory using generalized cost functions. *Spanish Economic Review* 1, 161-173.

[23] Granger CWJ. 1999b. *Empirical Modeling in Economics*. Cambridge: Cambridge University Press.

[24] Granger CWJ, Pesaran MH. 2000a. A decision theoretic approach to forecasting evaluation. In W. S. Chan, W. K. Li and Howell Tong (eds.), *Statistics and Finance: An Interface*. London: Imperial College Press.

[25] Granger CWJ, Pesaran MH. 2000b. Economic and statistical measures of forecast accuracy. *Journal of Forecasting* 19, 537-560.

[26] Hansen PR, Lunde A. 2001. Choosing the best volatility models: the model confidence set approach. *Oxford Bulletin of Economics and Statistics* 65, 839-861.

[27] Hamilton JD. 1991. A quasi-bayesian approach to estimating parameters for mixtures of normal distributions. *Journal of Business and Economics Statistics* 9, 27-39.

[28] Harvey CR, Siddique A. 1999. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* 34, 465-487.

[29] Lambert P, Laurent S. 2001. Modeling financial time series using GARCH-type models and skewed Student density. Discussion Paper 01-25. *Université Catholique de Louvain and Université de Liege.*

[30] León A, Mora J. 1999. Modeling conditional heteroscedasticity: application to the IBEX 35 stock-return index. *Spanish Economic Review* 1, 215-238..

[31] León A, Rubio G, Serna G. 2004. Autoregressive conditional volatility, skewness and kurtosis. WP-AD 2004-13. *IVIE.*

[32] Mauleón I. 1997. Instability and long memory in conditional variances. *Journal de la Societe de Statistique de Paris* 134, 67-88.

[33] Mauleón I. 1998. A forecasting stability test under Edgeworth-Sargan distributed errors. *ESEM*, Berlin, (mimeo).

[34] Mauleón I, Perote J. 2000. Testing densities with financial data: An empirical comparison of the Edgeworth-Sargan density to the Student's *t*. *European Journal of Finance* 6, 225-239.

[35] McDonald J, Xu I. 1995. A generalization of the beta distribution with applications. *Journal of Econometrics* 66, 133-152.

[36] Mittnik S, Paolella MS. 2000. Conditional density and Value-at-Risk prediction of Asian currency exchange rates. *Journal of Forecasting* 19, 313-333..

[37] Mittnik S, Rachev SV. 1993. Modeling asset returns with alternative Stable distributions. *Econometric Reviews* 12, 261-330.

[38] Nelson CR. 1976. *Applied time series analysis*. San Francisco: Holden-Day.

[39] Ñíguez TM. 2003. Volatility and VaR forecasting for the IBEX-35 stock-return index using FIGARCH-type processes and different evaluation criteria. WP-AD 2003-33. *IVIE*.

[40] Ñíguez TM. 2004. Forecasting kurtosis. *University of Alicante and LSE*, (mimeo).

[41] Perote J. 1999. Estimation and testing of financial densities: the case of the Edgeworth-Sargan density. Unpublised PhD dissertation. *University of Salamanca.*.

[42] Perote J, del Brío E. 2003. Measuring value-at-risk under the conditional Edgeworth-Sargan distribution. *Finance Letters* 1 (3), 23-40.

[43] Premaratne G, Bera AK. 2001. Modeling asymmetry and excess kurtosis in stock return data. Working Paper. Department of Economics. *University of Illinois*.

[44] Rosenblatt M. 1952. Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23, 470-472.

[45] Raaij G, Rauning B. 2002. Evaluating density forecast with an application to stock market returns. WP 08/02. *Economic Research Centre of the Deutsche Bundesbank*.

[46] Sargan JD. 1976. Econometric estimators and the Edgeworth approximation. *Econometrica* 44, 421-448.

[47] Tay AS, Wallis KF. 2000. Density forecasting: a survey. *Journal of Forecasting* 19, 235-254.

[48] Tsay RS. 2002. *Analysis of financial time series*. New York: Wiley.

[49] Weiss AA. 1986. Asymptotic theory of ARCH models: estimation and testing. *Econometric Theory* 2, 107-131.

[50] White H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1-26.