



Munich Personal RePEc Archive

## **Financial Density Selection**

Marin, J. Miguel and Sucarrat, Genaro

Universidad Carlos III de Madrid, BI Norwegian Business School

31 August 2012

Online at <https://mpra.ub.uni-muenchen.de/66839/>

MPRA Paper No. 66839, posted 23 Sep 2015 13:39 UTC

## Financial Density Selection \*

J. Miguel Marín<sup>†</sup> and Genaro Sucarrat<sup>‡</sup>

31 August 2011

### Abstract

We propose and study simple but flexible methods for density selection of skewed versions of the two most popular density classes in finance, the exponential power distribution and the  $t$  distribution. For the first type of method, which simply consists of selecting a density by means of an information criterion, the Schwarz criterion stands out since it performs well across density categories, and in particular when the Data Generating Process is normal. For the second type of method, General-to-Specific density selection, the simulations suggest that it can improve the recovery rate in predictable ways by changing the significance level. This is useful because it enables us to increase (reduce) the recovery rate of non-normal densities by increasing (reducing) the significance level, if one wishes to do so. The third type of method is a generalisation of the second type, such that it can be applied across an arbitrary number of density classes, nested or non-nested. Finally, the methods are illustrated in an empirical application.

*JEL Classification:* C16, C52, C58

*Keywords:* Financial returns, density selection, skewed exponential power distribution, skewed  $t$  distribution, general-to-specific density selection

---

\*We are thankful to Jonas Andersson, Mark Steel, and participants at the FIBE 2011 (Bergen) and CMS 2011 (Neuchatel) conferences for useful comments, suggestions and questions. We are also thankful for comments and suggestions on a precursor to this paper at the MAF 2010 conference (Ravello). Funding from The Bank of Spain Excellence Program is gratefully acknowledged.

<sup>†</sup>Department of Statistics, Universidad Carlos III de Madrid (Spain). Email: [jmmarin@econ.uc3m.es](mailto:jmmarin@econ.uc3m.es).

<sup>‡</sup>Corresponding author. Department of Economics, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway. Email: [genaro.sucarrat@bi.no](mailto:genaro.sucarrat@bi.no), Phone: +47+46 41 07 79, Fax: +47+23 26 47 88. Webpage: <http://www.sucarrat.net/>.

# 1 Introduction

Financial returns are often characterised by heavier tails than the normal—possibly skewed—even after demeaning and standardisation. One could consider modelling everything simultaneously, say, by means of an Autoregressive Moving Average (ARMA) specification in the mean, an Autoregressive Conditional Heteroscedasticity (ARCH) type specification in the volatility, and a conditional standardised density that admits both skewed and heavy-tailed errors. In practice, however, reliable estimation of all this jointly may not be feasible due to the complexity of the problem, and the complexity becomes even more severe if one wishes to include additional regressors in either the mean or volatility specifications, or in both. This motivates the modelling of the density of the conditional, standardised error in a separate, subsequent step. Another reason for modelling the density of the conditional, standardised error separately, is that many practitioners prefer using simple ARCH models like the RiskMetrics and Equally Weighted Moving Average (EqWMA) specifications in predicting volatility due to simplicity and/or comparability reasons. In both of these models the specification of a density of the conditional, standardised error would take place separately. A third and very common situation in which the specification of a density of the conditional, standardised error becomes a separate subsequent step, is when estimation is undertaken by means of Quasi Maximum Likelihood (QML) methods.

In this paper we propose and evaluate simple density selection methods of the conditional, standardised error. Marín and Sucarrat (2011) have recently studied the properties of two density selection methods in this setting for the standardised Skewed Exponential Power (SEP) distribution.<sup>1</sup> The first method simply consists of choosing the density that minimises an appropriately chosen information criterion. The second method can be viewed as an analogue to multi-path General-to-Specific (GETS) model selection in a regression context, see Campos et al. (2005) for a comprehensive overview of GETS model selection in regression analysis. In summary, multi-path GETS in regression analysis combines repeated backwards elimination (with continuous diagnostic checking and parsimonious encompassing tests of each terminal specification) with the use of an information criterion as a tie-breaker in the case of multiple terminal specifications. The attractiveness of this modelling strategy is that the recovery rate can be altered in controlled ways via the significance level.

We extend the work by Marín and Sucarrat (2011) in several ways, of which the most important are three. First, we extend their methods to the standardised Skewed  $t$  (ST) distribution. The normal (N) and skewed normal (SN) can both be obtained as special cases in both the SEP and ST classes, but the SEP is unable to produce as heavy tails as the ST. Also, the  $t$  distribution is arguably the most

---

<sup>1</sup>In financial econometrics, primarily because of Harvey (1981) and Nelson (1991), the Exponential Power distribution is also commonly known as the Generalised Error Distribution (GED).

popular heavy-tailed distribution among practitioners. So density selection methods that consider both SEP and ST densities are desirable. Second, we propose and evaluate a way in which the second method can be applied across an arbitrary number of density classes, nested or non-nested. For expository brevity we restrict ourselves to the SEP and ST classes in the illustration. Finally, we explore the practical usefulness of the methods in an empirical application.

The rest of this paper contains four sections. The next section, section 2, outlines the statistical framework. Section 3 studies the finite sample properties of the methods, whereas section 4 contains the empirical application. Finally, section 5 concludes.

## 2 Statistical framework

The generic ARCH model is given by

$$r_t = \mu_t + \epsilon_t, \quad (1)$$

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim IID(0, 1), \quad (2)$$

$$\sigma_t^2 = Var(r_t | \mathcal{I}_{t-1}), \quad (3)$$

where  $\mathcal{I}_t$  is the information at time  $t$ ,  $\epsilon_t$  is the error of the mean specification  $\mu_t$ ,  $\sigma_t^2$  is the conditional variance and  $\{z_t\}$  is an Independently and Identically Distributed (IID) sequence with mean zero and unit variance. Typically,  $\mathcal{I}_t = \{I_t, I_{t-1}, \dots\}$  with  $I_t = \{r_t, \sigma_t, z_t\}$ , as for example when the mean specification  $\mu_t$  is an Autoregressive Moving Average (ARMA) model. The most common specification of  $\sigma_t^2$  is Bollerslev's (1986) Generalised ARCH (GARCH) model with both the ARCH and GARCH orders equal to 1, that is,  $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ .

Arguably the two most common conditional densities in financial practice are the Exponential Power (EP) and Student's  $t$  (T) distributions. The former nests the normal as a special case, whereas in the latter the normal is obtained as the degrees of freedom parameter goes to infinity. The main advantage of the  $t$  distribution is that it is capable of producing heavier tails than the EP distribution, whereas the main advantages of the former are that it is more flexible (it can produce thicker *and* thinner tails than the normal), and that it exhibits better properties in many exponential ARCH models, see Nelson (1991), Harvey and Chakravarty (2010), and Sucarrat and Escribano (2010). The EP distribution of order  $p$  is usually parametrised as

$$EP(z, p, \mu, \sigma) = \frac{1}{2p^{1/p} \Gamma(1 + 1/p) \sigma} \exp\left(-\frac{|z - \mu|^p}{p\sigma^p}\right) \quad (4)$$

with  $\mu \in (-\infty, \infty)$ ,  $\sigma > 0$  and  $p \in (0, \infty)$ .  $\mu$  is a location parameter,  $\sigma$  is a scale parameter and  $p$  is a shape parameter. The normal distribution is obtained when  $p =$

2, whereas fatter (thinner) tails are produced when  $p < 2$  ( $p > 2$ ). In particular, the double exponential distribution, also known as the Laplace distribution, is obtained when  $p = 1$ , whereas  $p \rightarrow \infty$  yields a uniform distribution. The standardised EP density of Nelson (1991) such that  $E(z) = 0$  and  $Var(z) = 1$  is obtained by setting

$$\mu = 0, \quad \sigma = \frac{\Gamma(1/p)^{1/2}}{p^{1/p}\Gamma(3/p)^{1/2}}, \quad (5)$$

whereas the standardised  $t$  distribution with  $\nu > 2$  degrees of freedom such that  $E(z) = 0$  and  $Var(z) = 1$  can be expressed as

$$t(z, \nu) = \frac{1}{\sqrt{\nu - 2} B(\frac{1}{2}, \frac{\nu}{2}) \left(1 + \frac{z^2}{\nu - 2}\right)^{\frac{\nu + 1}{2}}}, \quad (6)$$

where  $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$  is the Beta function.

Arguably the two most common approaches to the skewing of the EP and  $t$  distributions are that of Azzalini (1986) on the one hand, and that of Fernández and Steel (1998) on the other. Examples of the former approach are DiCiccio and Monti (2004) for the EP, and Azzalini and Capitanio (2003) for the Student's  $t$ . Examples of the second approach are Fernández et al. (1995), Theodossiou (2000), Komunjer (2007), and Zhu and Zinde-Walsh (2009) for the EP, and Fernández and Steel (1998), Bauwens and Laurent (2005), and Zhu and Galbraith (2010) for the  $t$  distribution (Aas and Haff (2006) contains a brief review of different skewing approaches to the  $t$  distribution in finance). The main advantage of the Azzalini (1986) method is that it enables some elegant and attractive manipulation properties. Unfortunately, however, it is not clear that ML estimation provides consistent parameter estimates, see the discussions in Zhu and Zinde-Walsh (2009, p. 90), and in Azzalini and Capitanio (2003, pp. 384-385). By contrast, consistency (and asymptotic normality) of ML estimation for the Fernández and Steel (1998) method is proved for the SEP by Zhu and Zinde-Walsh (2009) when the shape parameter  $p$  is greater than 1, and for the  $t$  distribution in Zhu and Galbraith (2010). Moreover, the Fernández and Steel (1998) method is conceptually simpler and readily applicable to unimodal and symmetric densities in general. For these reasons we study our density selection methods using the skewing method of Fernández and Steel (1998).<sup>2</sup>

According to the Fernández and Steel (1998) method, if  $f(z)$  is a probability density function that is unimodal and symmetric about 0, then

$$f(z|\gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} \left[ f\left(\frac{z}{\gamma}\right) I_{[0, \infty)}(z) + f(z\gamma) I_{(-\infty, 0)}(z) \right]$$

---

<sup>2</sup>An interesting research question is to compare the Fernández and Steel (1998) method with the Azzalini (1986) method. Unfortunately, however, a rigorous comparison is beyond the scope of this paper.

is a skewed probability density function, where  $I_{(\cdot)}(z)$  is an indicator function, and where  $\gamma \in (0, \infty)$ . Symmetry is attained when  $\gamma = 1$ , whereas  $\gamma < 1$  and  $\gamma > 1$  produce left and right skewness, respectively. That is, heavier tails to the left and right, respectively. From the formula for the  $r$ th. (positive) integer moment (Fernández and Steele, 1998 p. 360) it follows (assuming  $f(z)$  is a standardised density such that  $E(z) = 0$  and  $Var(z) = 1$ ) that

$$\begin{aligned} M_r &= 2 \int_0^\infty z^r f(z) dz && (r\text{th. absolute moment, note: } M_2 = 1), \\ \mu_\gamma &= M_1(\gamma - 1/\gamma) && (\text{mean}), \\ \sigma_\gamma^2 &= (1 - M_1^2)(\gamma^2 + 1/\gamma^2) + 2M_1^2 - 1 && (\text{variance}). \end{aligned}$$

Next, the change of variable  $z^* = (z - \mu_\gamma)/\sigma_\gamma$  yields the standardised density

$$f^*(z^*) = \frac{2\sigma_\gamma}{\gamma + \frac{1}{\gamma}} f(z_{\mu_\gamma\sigma_\gamma}),$$

where

$$z_{\mu_\gamma\sigma_\gamma} = \frac{\sigma_\gamma z + \mu_\gamma}{\gamma \text{sign}(\sigma_\gamma z + \mu_\gamma)}.$$

Henceforth, for notational convenience, we will not make a distinction between  $z$  and  $z^*$ . The variable  $z$  will always satisfy  $E(z) = 0$  and  $Var(z) = 1$ .

Figure 1 provides a graphical comparison of the different types of densities. Both the EP and T densities have more mass at the mode and in the tails than the normal density, whereas the EP density has more mass at the mode than the T density. This comes at the cost of the mass in the tails, in which the T density is heavier than the EP. With respect to the effect of skewing, it is clear that skewing the density to the left, that is, generating heavier tails to the left intended to capture the tendency of financial returns to be larger in absolute value when negative, shifts the mode to the right and away from zero.

Studying the properties of a density selection algorithm necessitates a numerically robust estimation algorithm. The main properties of our ML estimation code,<sup>3</sup> which is available on request, are contained in table 1. It should be noted that we restrict the parameter space numerically, so that only the values in the regions  $\gamma \in [0.6, 5]$ ,  $p \in [1, 3]$  and  $\nu \in [4.0001, 50]$  are considered. This improves the estimation accuracy substantially in small samples, and speeds up estimation. The region  $\gamma \in [0.6, 5]$  covers the range of skewness values that (we believe) are likely to be encountered in practice, both for SEP and ST densities. Further, we believe that it is very unlikely to encounter a value greater than the upper bound of 3 on the shape parameter  $p$  of the SEP distribution, and the lower bound of 1 ensures that ML estimation provides consistent estimates of the SEP. Similarly, for the ST, the lower bound of  $\nu = 4.0001$  ensures that the fourth moment exists, whereas the

---

<sup>3</sup>Some of our code is adapted code from the R package `fGarch`, see Würtz and Chalabi (2009).

upper bound of  $\nu = 50$  produces a density that for most financial purposes is approximately equal to the normal. The initial values of the algorithm are  $\gamma = 1$  and  $p = 2$  for SEP densities, and  $\gamma = 1$  and  $\nu = 10$  for ST densities.

### 3 Financial density selection

In the investigation of our density selection algorithms we will make use of six different Data Generating Processes (DGPs): (1)  $z \sim N(0, 1)$ , (2)  $z \sim SN(\gamma = 0.7)$ , (3)  $z \sim EP(p = 1.1)$ , (4)  $z \sim T(\nu = 5)$ , (5)  $z \sim SEP(\gamma = 0.7, p = 1.1)$  and (6)  $z \sim ST(\gamma = 0.7, \nu = 5)$ . For expository brevity we will sometimes refer to these DGPs as N, SN, EP, T, SEP and ST, respectively. The values  $p = 1.1$ ,  $\nu = 5$  and  $\gamma = 0.7$  are at the border of what one is likely to encounter in practice.

#### 3.1 Density selection by means of information criteria

Choosing the density that minimises an appropriate information criterion results in consistent density selection. However, the success rate may not be very high in finite samples. Here, our objective is to shed light on this by comparing the performance of three commonly used information criteria: The Schwarz (1978) criterion (SC),<sup>4</sup> the Akaike (1974) criterion (AIC) and the Hannan and Quinn (1979) criterion (HQ). There are numerous other criteria that we could consider as well, but as a start we will only focus on the three most common ones. The three information criteria we compute as

$$\begin{aligned} \text{SC} &= -2\text{Log}L/T^* + k(\log T^*)/T^* \\ \text{AIC} &= -2\text{Log}L/T^* + 2k/T^* \\ \text{HQ} &= -2\text{Log}L/T^* + 2k \log[\log(T^*)]/T^* \end{aligned}$$

where  $\text{Log}L$  is the empirical log-likelihood,  $T^*$  is the number of observations, and where  $k = 0$  for N,  $k = 1$  for SN, EP and T, and where  $k = 2$  for SEP and ST.

Table 2 contains the probabilities of recovering the right density under different DGPs, in three different experiments. In the first experiment (A), four candidate densities  $\{N, SN, EP, SEP\}$  nested in the SEP are considered under four different DGPs: N, SN, EP and SEP. In the second experiment (B), four candidate densities  $\{N, SN, T, ST\}$  nested in the ST are considered under four different DGPs: N, SN, T and ST. Finally, in the third experiment (C), all the six densities  $\{N, SN, EP, T, SEP, ST\}$  are considered as candidates under six different DGPs: N, SN, EP, T, SEP and ST. The results of the table suggest that the SC criterion has the best overall performance, since it performs well across density types and experiments, and since it the preferred criterion in most cases when the sample size is 300 or larger. Consistent model selection, that is,  $p(DGP) \rightarrow 1$  as  $T^* \rightarrow \infty$ , is

---

<sup>4</sup>The SC is also known as the Bayesian Information Criterion or BIC.

attained faster for SC than for the other criteria in almost every instance. Of course, this is to some extent due to the large differences between the DGPs (smaller differences would presumably result in lower recovery rates). When the sample size is smaller than 300, then the experiments A and B suggest that the HQ criterion should be preferred. The AIC is sometimes slightly better than HQ in small samples (100 observations), but the probabilities increase slower than for HQ. Overall, then, since one typically has access to relatively large samples of at least several hundred observations (at least) in finance, the simulations suggest that the SC is the preferred information criterion. This is particularly the case if one has a preference for the normal density (as is often the case in QML estimation, inference and prediction, and in the high-frequency finance literature), since the SC recovers the normal with a very high probability already for a moderate number of observations.

### 3.2 GETS density selection of the SEP distribution

The General-to-Specific (GETS) density selection algorithm that we propose starts with the unrestricted estimate of a skewed, heavy-tailed density. If the unrestricted starting density is of the SEP type, then two different simplification paths are subsequently considered. The first path consists of first testing the restriction  $p = 2$  and then, if  $p = 2$  is rejected in the first test,  $\gamma = 1$ . In other words, simplification along a path stops when a null is rejected. Similarly, the second path consists of first testing the restriction  $\gamma = 1$  and then, if  $\gamma = 1$  is rejected in the first test,  $p = 2$ . In concise notation, the first path considered is thus  $SEP \rightarrow SN \rightarrow N$  whereas the second is  $SEP \rightarrow EP \rightarrow N$ . Inference is by means of likelihood ratio (LR) tests, and sometimes simplification can result in two different terminal models, say, SN and EP, or SN and N, and so on. In such cases the model with the lowest value on the chosen information criterion is selected. As the sample size  $T^*$  goes to infinity, this density selection algorithm has some very useful and known properties, namely that the probabilities of recovering the DGP depends on the significance level  $\alpha$  in the following ways:

$$\lim_{T^* \rightarrow \infty} p(DGP|N) = (1 - \alpha)^2 \quad (7)$$

$$\lim_{T^* \rightarrow \infty} p(DGP|SN) = (1 - \alpha) \quad (8)$$

$$\lim_{T^* \rightarrow \infty} p(DGP|EP) = (1 - \alpha) \quad (9)$$

$$\lim_{T^* \rightarrow \infty} p(DGP|SEP) = 1 \quad (10)$$

That is, when the DGP is equal to N, then the probability of recovering the DGP tends to  $(1 - \alpha)^2$  as the sample size goes to infinity (the squared expression is due to the asymptotic independence of the two test-statistics). For example, for the significance levels 10% and 5%, respectively, the probability  $p(DGP|N)$  tends to 0.81 and 0.9025, respectively. Similarly, if the DGP is SN, then the probability of



recovering SN tends to  $(1 - \alpha)$ , and so on. The usefulness of these properties is that one can use the significance level  $\alpha$  to “push” the algorithm either towards or away from normality, if one wishes to do so. For example, the simulations in the previous subsection showed that the SC criterion performs very well in both small and large samples when  $DGP = N$ . However, when the DGP is not normal, then the SC criterion does not always recover the DGP more often than the other criteria. Hence one may increase the recovering probabilities when the DGP is not normal (or alternatively when the cost of falsely characterising the density as non-normal is not considered very large) in a controlled and predictable way by simply increasing the significance level.

Table 3 contains the finite sample probabilities of rejecting the various null hypotheses involved in GETS density selection, under different DGPs. The table suggests that the empirical size is close to the nominal size when the DGP is nested in the null hypothesis. When the DGP is neither nested in the null nor contained in the alternative, as for example in the test  $H_0 : N, H_1 : EP$  when  $DGP = SN$ , then the empirical size (0.10) is misleading since it neither approaches the nominal size nor 1. However, it should be stressed that this is *not* a problem for the GETS algorithm due to its multiple path nature. To see this consider the two paths considered by the algorithm,  $SEP \rightarrow EP \rightarrow N$  and  $SEP \rightarrow SN \rightarrow N$ . Suppose now that  $DGP = SN$ . In path 1 the skewness will be detected with asymptotic probability 1 already in the first test, whereas in the second path the terminal model will be SN with asymptotic probability equal to  $(1 - \alpha)$ . Asymptotically, then, we will have two terminal models, SEP and SN, in  $100 \cdot (1 - \alpha)\%$  of the cases. Since the GETS algorithm uses an information criterion to select among several terminal models, this means the asymptotic probability of recovering the DGP when it is equal to SN is  $(1 - \alpha)$ , which is exactly as predicted.

Figure 2 suggests indeed that the asymptotic values are attained as the sample gets larger. The figure contains the probabilities of recovering the DGP with an SC criterion, and the probabilities of recovering the DGP using GETS density selection combined with an SC criterion. The first thing to note is that the asymptotic probabilities are (approximately) attained relatively fast: At 100 observations at the earliest, in the case when  $DGP = N$ , and at about 300 to 500 observations at the latest. Of course, this convergence will be slower when the DGPs differ less. The second thing to note is that there are notable gains to be made in small samples. For example, when  $DGP = SEP$  then there is a gain of about 12 percentage points when the sample size is 100 observations. In finance, where one would expect departure from normality, this can be a very useful gain. Also, when the DGP is not normal, the gain is likely to be larger in small samples if either the AIC or HQ is used instead of the SC.

### 3.3 GETS density selection of the ST distribution

GETS density selection of the ST class starts with an unrestricted estimate of an ST density. Next, two different simplification paths are considered,  $ST \rightarrow SN \rightarrow N$  and  $ST \rightarrow T \rightarrow N$ . Again, inference is by means of likelihood ratio (LR) tests, simplification along a path stops when a null hypothesis is rejected, and in cases with two terminal models then an information criterion is used as tie-breaker. As the sample size  $T^*$  goes to infinity, one would like this density selection algorithm to exhibit the same sort of useful and known properties as in the SEP case, namely that the asymptotic probabilities of recovering the DGP depends are given by:

$$\lim_{T^* \rightarrow \infty} p(DGP|N) = (1 - \alpha)^2 \quad (11)$$

$$\lim_{T^* \rightarrow \infty} p(DGP|SN) = (1 - \alpha) \quad (12)$$

$$\lim_{T^* \rightarrow \infty} p(DGP|T) = (1 - \alpha) \quad (13)$$

$$\lim_{T^* \rightarrow \infty} p(DGP|ST) = 1 \quad (14)$$

Table 4 contains the finite sample probabilities of rejecting the various null hypotheses involved in GETS density selection, under different DGPs. In contrast to the SEP case, the table does not suggest that the empirical size is always close to the nominal size when it should. Additional simulations (not reported but available on request) suggest this happens when the degrees of freedom parameter  $\nu$  of the DGP is on the boundary of the parameter space, that is, when the DGP is either N or SN, because then  $\nu = \infty$ . Nevertheless, the empirical sizes are reasonably close to the nominal size, so the probabilities given by (11)-(14) can be viewed as indicative. And, more importantly, GETS density selection will still exhibit the property that the significance level can be used to either push density selection towards or away from the normal density.

Figure 3 provides a confirmation of this. The figure contains the probabilities of recovering the DGP with an SC criterion, and the probabilities of recovering the DGP using GETS density selection combined with an SC criterion for 5% and 10% significance levels. Only in two cases, when  $DGP = T$  and  $DGP = ST$ , does GETS density selection yield large sample recovery probabilities that correspond to those suggested by (11)-(14). Nevertheless, in the other two instances the large sample probabilities are not far away from their theoretical counterparts, so the theoretical values can be used as rough indications. Finally, the figures suggest there can be even greater small sample gains for the ST case compared with the SEP case. When  $DGP = ST$  and the number of observations is 100, for example, then GETS density selection with a 10% significance level can increase the recovery rate with about 20 percentage points. This is almost the double compared with the SEP case.

### 3.4 GETS density selection over the SEP and ST distributions

The GETS density selection algorithm we have outlined for the SEP and ST classes can straightforwardly be generalised to select across an arbitrary number of classes, nested or non-nested. Here, for expository brevity, we will restrict ourselves to only two classes, namely the SEP and ST. The algorithm then proceeds in three steps. First, undertake GETS density selection in the first class. This will produce a terminal model from the first class, winner number 1. Next, undertake GETS density selection in the second class. This will yield a second terminal model, winner number 2 (which need not differ from winner number 1). Finally, the third step consists of choosing the density with the smallest value on the chosen information criterion among the two winners. This algorithm exhibits known asymptotic properties: The probability of recovering the DGP is simply the highest across density classes (this is explained in more detail below). Similarly, if one considers GETS density selection across three different classes (nested or non-nested) instead of two, then the algorithm proceeds in four steps and the probability of recovering the DGP is the maximum across the three classes. And so on.

We now study in more detail the case where we select across two classes, SEP and ST. Let  $p_{SEP}(DGP|N)$  denote the asymptotic probability of finding N in the SEP class when DGP=N with GETS density selection using a significance level of  $\alpha_{SEP}$ , let  $p_{SEP}(DGP|SN)$  denote the asymptotic probability of finding SN when DGP=SN using the same significance level, and so on. Similarly, let  $p_{ST}(DGP|N)$  denote the asymptotic probability of finding N in the ST class when the DGP=N with GETS density selection using a significance level of  $\alpha_{ST}$ , let  $p_{ST}(DGP|SN)$  denote the asymptotic probability of finding SN when DGP=SN using the same significance level, and so on. The significance levels  $\alpha_{SEP}$  and  $\alpha_{ST}$  need not be equal, but we assume so for simplicity. The asymptotic probabilities of finding the DGP are then:

$$p(DGP|N) = \max \{p_{SEP}(DGP|N), p_{ST}(DGP|N)\} \quad (15)$$

$$p(DGP|SN) = \max \{p_{SEP}(DGP|SN), p_{ST}(DGP|SN)\} \quad (16)$$

$$p(DGP|EP) = \max \{p_{SEP}(DGP|EP), p_{ST}(DGP|EP)\} \quad (17)$$

$$p(DGP|T) = \max \{p_{SEP}(DGP|T), p_{ST}(DGP|T)\} \quad (18)$$

$$p(DGP|SEP) = \max \{p_{SEP}(DGP|SEP), p_{ST}(DGP|SEP)\} \quad (19)$$

$$p(DGP|ST) = \max \{p_{SEP}(DGP|ST), p_{ST}(DGP|ST)\} \quad (20)$$

For example, suppose DGP=N and that a significance level of 5% is used in both the SEP and ST classes. The simulations in figure 2 from subsection 3.2 suggest  $p_{SEP}(DGP|N)$  is equal to  $(1 - \alpha_{SEP})^2 = 0.9025$ , whereas the simulations in figure 3 suggest  $p_{ST}(DGP|N)$  is approximately equal to 0.93. Accordingly,  $p(DGP|N)$  is approximately equal to 0.93, the maximum of 0.9025 and 0.93. Similarly, suppose

DGP=EP and that a significance level of 5% is used. The simulations in figure 2 from subsection 3.2 suggested  $p_{SEP}(DGP|EP) = (1 - \alpha_{SEP}) = 0.95$ , whereas  $p_{ST}(DGP|N) = 0$  since EP is not contained within the ST class. Accordingly,  $p(DGP|EP)$  is equal to 0.95. And so on.

Figure 4 summarises the finite sample properties of the algorithm. The results suggest indeed that the asymptotic properties are reproduced, although more observations are needed in order to reach the asymptotic limits. For example, for  $T^* = 1500$  the top-left graph shows that  $p(DGP|N)$  is about 0.94 and 0.88 for the 5% and 10% levels, respectively, which is exactly what the simulations from the two previous subsections predict. Similarly, when DGP=EP the left graph in the middle shows that  $p(DGP|EP)$  is about 0.94 and 0.89 for the 5% and 10% levels, respectively, which is almost exactly as predicted. A result found in the previous subsections that is not reproduced to the same extent, however, is the gain in small samples. To recall, in the two previous subsections the simulations suggested that GETS density selection increases the probability of recovering non-normal DGPs in small samples by increasing the significance level. This is still the case when the DGP is either SN, SEP or ST. But it is not the case when the DGP is either EP or T. Presumably this is because the algorithm finds it difficult to distinguish between the type of fat-tailed density when there is no skewness.

## 4 Empirical application

In order to illustrate the financial density selection methods studied in the previous section, we will apply them to three different types of financial returns series: The daily return of a European stock market index (the FTSE100, abbreviated FT100), the daily return of the USD per Euro (USD/EUR) exchange rate, and the daily return of the North-Western European Brent Blend spot oilprice. All the data span the period 1 January 2001 to 21 September 2010, and are freely available on the internet.<sup>5</sup> Graphs of the returns in percent, which we denote  $r_t^{ft100}$ ,  $r_t^{usdeur}$  and  $r_t^{oilp}$ , are contained in figure 5. Note that the returns are computed as their log-difference multiplied by 100. That is,  $r_t = (\ln S_t - \ln S_{t-1}) \cdot 100$ , where  $S_t$  is the nominal value at  $t$  of the asset in question.

From the graphs in figure 5, and from the descriptive statistics in panel A of table 5, it is clear that the three returns series exhibit the usual features that typically characterise financial returns: ARCH, marked excess kurtosis compared with the normal, and skewness. Another common characteristics, a negative (but weak) first-order autocorrelation, typically interpreted as a return-reversal effect, appears to be present in the FT100 return series, but not in the two others.

Before applying the density selection methods from the previous section, we fit

---

<sup>5</sup><http://yahoo.finance.com/> (the FTSE 100 series), the European Central bank's webpage <http://www.ecb.int/> (USD/EUR) and the US Energy Information Administration's statistics webpage <http://www.eia.doe.gov/emeu/international/> (oilprice).

an ARMA-GARCH model to each return series by means of Quasi Maximum Likelihood (QML).<sup>6</sup> Skewness or “density asymmetry” is potentially related to leverage (“volatility-asymmetry”). So we opt for the GARCH model of Glosten et al. (1993), which is a variant of a first-order threshold specification (Zakoïan (1994)).<sup>7</sup> Specifically, all the models we fit are contained in:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \epsilon_t, \quad (21)$$

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim IID(0, 1), \quad (22)$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \lambda \epsilon_{t-1}^2 I_{(\epsilon_{t-1} < 0)}. \quad (23)$$

The exact choice of the mean specification results from a comparison of the values on the SC information criterion (using a normal log-likelihood made up of the standardised residuals) from six models: An ARMA(0,0), an AR(1), an MA(1), an ARMA(1,1), an MA(5) and an ARMA(1,5). A constant is included in all six models, and the winning specification for each return series together with the estimation and diagnostic results are contained in panel B of table 5. As is commonly the case, the ARMA-GARCH specification takes away the autoregressive correlation in returns, and most of the ARCH in the residuals as measured by the  $AR_1$  and  $ARCH_1$  diagnostic tests. Also, the excess kurtosis as measured by the sample kurtosis  $K$  is reduced. The leverage term  $\epsilon_{t-1}^2 I_{(\epsilon_{t-1} < 0)}$  is statistically significant at all conventional significance levels for the FT100 and oilprice models, but not for the exchange rate model.

Panel C of table 5 contains the results of applying the density selection methods on the standardised residuals  $\{\hat{z}_t\}$ . The methods suggest that skewness is present in the conditional densities of the FT100 and oilprice return series, but not in that of exchange rate returns. A possible reason for this is the inherent symmetry in currency speculation. The estimated skewness coefficients lie in the range 0.84-1.00 (if we include the results of the exchange rate, that is, 1.00 = no skewness), which implies negative skewness when less than 1. As for fat-tailedness, apart from the IC selection in the FT100 case, the results suggest that all the conditional densities are fat-tailed compared with the normal. The estimated shape parameters are 1.37, 1.57 and 1.83 in the SEP class, and 7.11, 11.9 and 27.3 in the ST class.

---

<sup>6</sup>The estimation of the models is undertaken in EViews 6. The standard errors are those of Bollerslev and Wooldridge (1992), which are consistent under both normality and non-normality of the conditional errors.

<sup>7</sup>Another popular ARCH model that contains leverage in the volatility specification is Nelson’s (1991) EGARCH model. Unfortunately, however, Nelson’s EGARCH model is practically useless when the errors are  $t$  distributed, since a necessary condition for EGARCH stability in this case is that the ARCH parameter is negative, see Nelson (1991, p. 365). This is a very restrictive assumption empirically and the primary reason why Nelson used an EP density instead of a  $T$ . Moreover, although Straumann and Mikosch (2006, p. 2452) prove consistency of QML for the first order EGARCH in principle, it is not clear that the estimator always works in practice (even when the errors are not  $t$  distributed) due to the possibility of non-invertibility, see Sorokin (2010). For these reasons we opt for a threshold ARCH model instead.

These are moderate magnitudes for many purposes, which suggests that there is not too much heavy-tailedness left once ARCH and negative skewness have been appropriately accounted for. The daily returns of small companies would possibly exhibit substantially heavier tails. Finally, it is of interest to note that IC selection and SEP-ST-GETS selection gives two different winners in the FT100 case: SN in IC selection and ST in SEP-ST-GETS selection. The estimate of  $\nu$  equal to 27 for the ST suggests that the difference in winner can be interpreted as IC selection not being capable of detecting the extra fat-tailedness.

## 5 Conclusions

We have proposed and studied simple but flexible methods for financial density selection of the standardised Skewed Exponential Power (SEP) class of distributions, and of the standardised Skewed Student's  $t$  (ST) class of distributions. The Monte Carlo simulations of the first method, which consists of selecting the density that minimises a chosen information criterion, suggest that the Schwarz (1978) criterion (SC) provides the best overall performance. The SC does well across density categories and sample sizes, it usually provides the fastest attainment of consistent model selection, and the probability of recovering the Data Generating Process (DGP) is very high even in small samples when the DGP is normal. The latter property is very often desirable in finance, because of the prominent role played by the normal density in estimation, inference and prediction. Contrarily, if the cost of selecting a normal density when the DGP in fact is non-normal, then the Monte Carlo simulations suggest that the Hannan and Quinn (1979) (HQ) criterion is the most suitable criterion in small samples.

The second type of density selection method that we propose and study is General-to-Specific (GETS) density selection, which combines sequential (classical) hypothesis testing with the use of an information criterion as a tie-breaker in the case of multiple terminals. This is useful because it enables us to control the recovery rate in predictable ways by increasing or reducing the significance level. In the SEP case, the simulations showed that the asymptotic, theoretical recovery probabilities were attained already at reasonably small samples. So the asymptotic values are likely to be reasonably indicative as to how the significance level can be used to control the recovery rate under different DGPs. In the ST case, the finite sample sizes of the hypotheses tests involved in the GETS search are not always as close to their nominal counterpart in finite samples. So, in the ST case, the asymptotic properties must be viewed as rough guides. Nevertheless, the main property of GETS density selection, namely that the significance level can be used to control the probability of recovering the DGP, is still present although the magnitude is not as predictable as in the SEP case.

The third type of density selection method that we propose is a generalisation of the second type. The GETS algorithm we have outlined for the SEP and ST classes,

respectively, can straightforwardly be generalised to select across an arbitrary number of classes, nested or non-nested, and the probability of recovering the DGP is simply the maximum across density classes. The simulations suggest indeed that the asymptotic properties are reproduced, although more observations are needed in order to reach the asymptotic limits. A result found in the previous subsections that is not reproduced to the same extent, however, is the gain in small samples achieved by GETS density selection. Presumably this is because it is difficult to distinguish between the type of fat-tailedness (EP *vs.* T) when there is no skewness.

## References

- Aas, K. and D. H. Haff (2006). The Generalized Hyperbolic Skew Student's  $t$ -Distribution. *Journal of Financial Econometrics* 4, 275–309.
- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19, 716–723.
- Azzalini, A. (1986). Further results on a class of distributions that includes the normal ones. *Statistica* 46, 199–208.
- Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew  $t$ -distribution. *Journal of the Royal Statistical Society B* 65, 367–389.
- Bauwens, L. and S. Laurent (2005). A new class of multivariate skew densities, with applications to garch models. *Journal of Business and Economic Statistics* 23, 346–354.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T. and J. Wooldridge (1992). Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances. *Econometric Reviews* 11, 143–172.
- Campos, J., D. F. Hendry, and N. R. Ericsson (Eds.) (2005). *General-to-Specific Modeling. Volumes 1 and 2*. Cheltenham: Edward Elgar Publishing.
- DiCiccio, T. J. and A. C. Monti (2004). Inferential Aspects of the Skew Exponential Power Distribution. *Journal of the American Statistical Association* 99, 439–450.
- Fernández, C., J. Osiewalski, and M. Steel (1995). Modeling and Inference with  $v$ -Spherical Distributions. *Journal of the American Statistical Association* 90, 1331–1340.

- Fernández, C. and M. Steel (1998). On Bayesian Modelling of Fat Tails and Skewness. *Journal of the American Statistical Association* 93, 359–371.
- Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance* 48, 1779–1801.
- Hannan, E. and B. Quinn (1979). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society. Series B* 41, 190–195.
- Harvey, A. C. (1981). *The Econometric Analysis of Time Series*. London: Philip Allan.
- Harvey, A. C. and T. Chakravarty (2010). Beta-t-EGARCH. Unpublished working paper.
- Komunjer, I. (2007). Asymmetric power distribution: Theory and application to risk measurement. *Journal of Applied Econometrics* 22, 891–921.
- Ljung, G. and G. Box (1979). On a Measure of Lack of Fit in Time Series Models. *Biometrika* 66, 265–270.
- Marín, J. M. and G. Sucarrat (2011). Modelling the Skewed Exponential Power Distribution in Finance. In C. Perna and M. Sibill (Eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*. Springer, in press.
- Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59, 347–370.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics* 6, 461–464.
- Sorokin, A. (2010). Non-invertibility in Some Heteroscedastic Models. Unpublished working paper.
- Straumann, D. and T. Mikosch (2006). Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach. *The Annals of Statistics* 34, 2449–2495.
- Sucarrat, G. and Á. Escibano (2010). The Power Log-GARCH Model. <http://www.sucarrat.net/>.
- Theodossiou, P. (2000). Skewed generalized error distribution of financial assets and option pricing. SSRN working paper.
- Würtz, D. and Y. Chalabi (2009). Package fGarch. [www.rmetrics.org](http://www.rmetrics.org). Downloadable via <http://cran.r-project.org/web/packages/fGarch/index.html>.



- Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.
- Zhu, D. and J. W. Galbraith (2010). A Generalized Asymmetric Student t Distribution with Application to Financial Econometrics. *Journal of Econometrics*, 297–305.
- Zhu, D. and V. Zinde-Walsh (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics* 148, 86–99.

Table 1: Numerical performance of ML estimation

Density	$T^*$	$\gamma$	$p$	$M(\hat{\gamma})$	$V(\hat{\gamma})$	$M(\hat{p})$	$V(\hat{p})$	$M(its)$
SEP	100	1.0	2.0	1.027	0.060	2.097	0.243	7.93
			1.1	1.007	0.022	1.152	0.044	17.31
	200		2.0	1.007	0.014	2.060	0.120	7.72
			1.1	1.003	0.004	1.118	0.017	17.34
	500		2.0	1.003	0.005	2.024	0.046	8.24
			1.1	1.000	0.002	1.104	0.007	16.94
	1000		2.0	1.001	0.002	2.006	0.020	8.61
			1.1	1.001	0.001	1.101	0.004	16.46
	100	0.7	2.0	0.709	0.012	2.084	0.229	8.26
			1.1	0.700	0.005	1.144	0.032	16.49
	200		2.0	0.701	0.006	2.056	0.117	8.75
			1.1	0.699	0.002	1.119	0.015	17.14
	500		2.0	0.698	0.002	2.027	0.044	9.55
			1.1	0.700	0.001	1.102	0.006	17.37
	1000		2.0	0.698	0.001	2.011	0.020	9.93
			1.1	0.700	0.000	1.100	0.003	16.90
Density	$T^*$	$\gamma$	$\nu$	$M(\hat{\gamma})$	$V(\hat{\gamma})$	$M(\hat{\nu})$	$V(\hat{\nu})$	$M(its)$
ST	100	1.0	20.0	1.011	0.029	30.338	377.24	13.57
			5.0	1.012	0.022	8.438	108.58	11.57
	200		20.0	1.009	0.013	29.178	328.44	12.89
			5.0	1.003	0.009	6.221	22.48	10.80
	500		20.0	1.001	0.005	27.083	243.33	12.03
			5.0	1.001	0.003	5.226	1.16	10.40
	1000		20.0	1.001	0.002	25.195	173.18	11.64
			5.0	1.000	0.002	5.117	0.42	10.64
	100	0.7	20.0	0.717	0.013	28.942	364.13	13.85
			5.0	0.705	0.009	7.249	60.51	11.58
	200		20.0	0.703	0.005	29.477	317.38	13.96
			5.0	0.698	0.004	5.814	13.21	11.08
	500		20.0	0.700	0.002	26.448	227.53	13.52
			5.0	0.698	0.002	5.173	0.85	9.61
	1000		20.0	0.698	0.001	25.326	170.63	12.76
			5.0	0.700	0.001	5.060	0.31	9.51

Simulations (2000 replications) in R with ML estimation implemented via the `nlminb()` function.  $T^*$ , sample size.  $M(\cdot)$ , sample mean.  $V(\cdot)$ , sample variance.  $its$ , number of iterations.

Table 2: Probabilities of recovering the right density for different information criteria, under different DGPs

		<i>DGP = N</i>			<i>DGP = SN</i>			<i>DGP = EP</i>			<i>DGP = SEP</i>		
<i>T*</i>		<i>SC</i>	<i>AIC</i>	<i>HQ</i>	<i>SC</i>	<i>AIC</i>	<i>HQ</i>	<i>SC</i>	<i>AIC</i>	<i>HQ</i>	<i>SC</i>	<i>AIC</i>	<i>HQ</i>
Experiment A	100	0.93	0.68	0.85	0.55	0.69	0.64	0.74	0.76	0.81	0.77	0.93	0.87
	200	0.95	0.71	0.87	0.81	0.81	0.86	0.95	0.84	0.91	0.97	1.00	0.99
	300	0.97	0.69	0.89	0.93	0.83	0.92	0.98	0.84	0.93	1.00	1.00	1.00
	500	0.97	0.72	0.89	0.98	0.84	0.94	0.99	0.84	0.94	1.00	1.00	1.00
	1000	0.98	0.71	0.91	0.99	0.83	0.95	0.99	0.84	0.95	1.00	1.00	1.00
		<i>DGP = N</i>			<i>DGP = SN</i>			<i>DGP = T</i>			<i>DGP = ST</i>		
<i>T*</i>		<i>SC</i>	<i>AIC</i>	<i>HQ</i>	<i>SC</i>	<i>AIC</i>	<i>HQ</i>	<i>SC</i>	<i>AIC</i>	<i>HQ</i>	<i>SC</i>	<i>AIC</i>	<i>HQ</i>
Experiment B	100	0.95	0.77	0.88	0.54	0.76	0.69	0.58	0.67	0.65	0.48	0.74	0.64
	200	0.96	0.78	0.90	0.82	0.91	0.89	0.82	0.78	0.85	0.84	0.96	0.91
	300	0.97	0.78	0.91	0.95	0.93	0.96	0.92	0.83	0.92	0.96	0.99	0.98
	500	0.98	0.78	0.93	0.99	0.94	0.98	0.98	0.84	0.93	1.00	1.00	1.00
	1000	0.99	0.81	0.94	0.99	0.94	0.98	0.99	0.83	0.95	1.00	1.00	1.00
		<i>DGP = N</i>		<i>DGP = SN</i>		<i>DGP = EP</i>		<i>DGP = T</i>		<i>DGP = SEP</i>		<i>DGP = ST</i>	
<i>T*</i>		<i>SC</i>	<i>HQ</i>	<i>SC</i>	<i>HQ</i>	<i>SC</i>	<i>HQ</i>	<i>SC</i>	<i>HQ</i>	<i>SC</i>	<i>HQ</i>	<i>SC</i>	<i>HQ</i>
Experiment C	100	0.93	0.84	0.57	0.63	0.60	0.61	0.42	0.47	0.64	0.69	0.32	0.38
	200	0.95	0.86	0.79	0.84	0.78	0.76	0.62	0.62	0.81	0.84	0.59	0.68
	300	0.96	0.87	0.93	0.91	0.84	0.82	0.74	0.71	0.88	0.88	0.74	0.73
	500	0.98	0.89	0.98	0.94	0.92	0.88	0.84	0.79	0.92	0.93	0.84	0.85
	1000	0.98	0.91	0.99	0.94	0.97	0.93	0.93	0.89	0.98	0.98	0.94	0.93

Simulations (2000 replications) in R. In experiment A the candidate set is  $\{N, SN, EP, SEP\}$ , in experiment B the candidate set is  $\{N, SN, T, ST\}$ , and in experiment C the candidate set is  $\{N, SN, EP, T, SEP, ST\}$ . The 3rd. number in the second column (0.69) in experiment A means that the probability of selecting N among the four candidates N, SN, EP and SEP is 0.69 for  $T^* = 300$  when  $DGP = N$  using AIC. The 2nd. number in the sixth column (0.89) in experiment B means that the probability of selecting SN among the four candidates N, SN, T and ST is 0.89 for  $T^* = 200$  when  $DGP = SN$  using HQ. The 1st. number in the ninth column (0.64) in experiment C means that the probability of selecting SEP among the six candidates N, SN, EP, T, SEP and ST is 0.64 for  $T^* = 100$  when  $DGP = SEP$  using SC. And so on.

Table 3: Rejection probabilities of  $H_0$  under different SEP density DGPs for the tests involved in GETS density selection, using a nominal significance level of 5%

DGP	$T$	$H_0 : N$	$H_0 : N$	$H_0 : SN$	$H_0 : EP$
		$H_1 : SN$	$H_1 : EP$	$H_1 : SEP$	$H_1 : SEP$
$N$	100	0.06	0.05	0.06	0.07
	200	0.05	0.04	0.05	0.06
	300	0.05	0.05	0.05	0.06
	500	0.06	0.05	0.05	0.06
	1000	0.06	0.06	0.06	0.06
	10000	0.05	0.05	0.05	0.05
$SN$	100	0.64	0.06	0.05	0.63
	200	0.89	0.07	0.06	0.89
	300	0.98	0.08	0.05	0.97
	500	1.00	0.07	0.05	1.00
	1000	1.00	0.07	0.05	1.00
	10000	1.00	0.10	0.04	1.00
$EP$	100	0.10	0.83	0.83	0.06
	200	0.09	0.98	0.98	0.05
	300	0.10	1.00	1.00	0.05
	500	0.11	1.00	1.00	0.05
	1000	0.12	1.00	1.00	0.06
	10000	0.11	1.00	1.00	0.04
$SEP$	100	0.86	0.68	0.85	0.93
	200	0.98	0.92	0.99	1.00
	300	1.00	0.98	1.00	1.00
	500	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00
	10000	1.00	1.00	1.00	1.00

Simulations (2000 replications) in R. The top-left number (0.06) should be interpreted as the probability of accepting  $H_1 : SN$  when  $H_0 : N$ ,  $LR \sim \chi^2(1)$ , for DGP = N. The fourth number in row seven (0.63) should be interpreted as the probability of accepting  $H_1 : SEP$  when  $H_0 : EP$ ,  $LR \sim \chi^2(1)$ , for DGP = SN. The fifth number from the bottom in column three (0.99), should be interpreted as the probability of accepting  $H_1 : SEP$  when  $H_0 : SN$ ,  $LR \sim \chi^2(1)$ , for DGP = SEP. And so on.

Table 4: Rejection probabilities of  $H_0$  under different ST density DGPs for the tests involved in GETS density selection, using a nominal significance level of 5%

DGP	$T$	$H_0 : N$ $H_1 : SN$	$H_0 : N$ $H_1 : T$	$H_0 : SN$ $H_1 : ST$	$H_0 : T$ $H_1 : ST$
$N$	100	0.06	0.02	0.02	0.06
	200	0.06	0.02	0.02	0.05
	300	0.06	0.02	0.02	0.05
	500	0.05	0.02	0.02	0.05
	1000	0.05	0.03	0.03	0.05
	10000	0.05	0.02	0.02	0.04
	$SN$	100	0.64	0.05	0.02
200		0.90	0.07	0.02	0.89
300		0.98	0.08	0.02	0.97
500		1.00	0.12	0.02	1.00
1000		1.00	0.18	0.02	1.00
10000		1.00	0.84	0.02	1.00
$T$		100	0.19	0.68	0.66
	200	0.20	0.89	0.88	0.06
	300	0.22	0.97	0.97	0.05
	500	0.23	1.00	1.00	0.06
	1000	0.25	1.00	1.00	0.05
	10000	0.32	1.00	1.00	0.04
	$ST$	100	0.76	0.68	0.68
200		0.95	0.91	0.93	0.97
300		0.98	0.97	0.99	0.99
500		1.00	1.00	1.00	1.00
1000		1.00	1.00	1.00	1.00
10000		1.00	1.00	1.00	1.00

Simulations (2000 replications) in R. The first number in the second row (0.06) should be interpreted as the probability of accepting  $H_1 : SN$  when  $H_0 : N$ ,  $LR \sim \chi^2(1)$ , for DGP = N. The second number in row eight (0.07) should be interpreted as the probability of accepting  $H_1 : T$  when  $H_0 : N$ ,  $LR \sim \chi^2(1)$ , for DGP = SN. The third number in the fifth (from the bottom) row (0.93), should be interpreted as the probability of accepting  $H_1 : ST$  when  $H_0 : SN$ ,  $LR \sim \chi^2(1)$ , for DGP = ST. And so on.

Table 5: Descriptive statistics, model estimation results and diagnostics, and density selection results of empirical application

Panel A: Descriptive statistics											
	$M$	$SD$	$Sk(r_t)$	$K(r_t)$	$AR_1$ [ $pval$ ]	$ARCH_1$ [ $pval$ ]	$T^*$				
$r_t^{ft100}$	-0.004	1.346	-0.099	9.264	-0.07 [0.00]	0.24 [0.00]	2454				
$r_t^{usdeur}$	0.013	0.652	-0.057	5.936	0.00 [0.85]	0.26 [0.00]	2486				
$r_t^{oilp}$	0.054	2.542	-0.420	7.834	-0.02 [0.43]	0.14 [0.00]	2299				
Panel B: Model estimates and diagnostics											
	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$Sk(\hat{z}_t)$	$K(\hat{z}_t)$	$AR_1$ [ $pval$ ]	$ARCH_1$ [ $pval$ ]	
$r_t^{ft100}$	0.006 (0.016)	-0.070 (0.020)	0.014 (0.003)	-0.008 (0.014)	0.916 (0.010)	0.159 (0.024)	-0.34	3.39	0.01 [0.64]	-0.04 [0.03]	
$r_t^{usdeur}$	0.024 (0.011)		0.002 (0.001)	0.035 (0.012)	0.960 (0.008)	0.0002 (0.013)	-0.01	3.73	0.00 [0.84]	0.00 [0.98]	
$r_t^{oilp}$	0.074 (0.048)		0.104 (0.041)	0.007 (0.013)	0.950 (0.016)	0.049 (0.015)	-0.53	6.58	-0.02 [0.34]	-0.01 [0.69]	
Panel C: Results of density selection											
	IC	SEP-GETS	ST-GETS	SEP-ST-GETS	KS [ $pval$ ]						
$\hat{z}_t^{ft100}$	$SN(0.84)$	$SEP(1.83, 0.84)$	$ST(27.3, 0.84)$	$ST(27.3, 0.84)$	0.014 [0.74]						
$\hat{z}_t^{usdeur}$	$EP(1.57)$	$EP(1.57)$	$T(11.9)$	$EP(1.57)$	0.015 [0.64]						
$\hat{z}_t^{oilp}$	$ST(7.27, 0.90)$	$SEP(1.37, 0.90)$	$ST(7.11, 0.90)$	$ST(7.11, 0.90)$	0.021 [0.26]						

Panels A and B:  $M$ , sample mean;  $SD$ , sample standard deviation;  $T^*$ , number of non-missing observations;  $Sk$ , sample skewness;  $K$ , sample kurtosis;  $AR_1$  and  $ARCH_1$ , first order sample autocorrelation of  $r_t$  and  $r_t^2$ , respectively, with  $p$ -value (in square brackets) from a Ljung and Box (1979) test for first order autocorrelation. Panel C: IC, result of density selection over the candidate set  $\{N, SN, EP, T, SEP, ST\}$  using the Schwarz information criterion (SC); SEP-GETS, result of GETS density selection (using SC and a 5% significance level) in the SEP class; ST-GETS, result of GETS density selection (using SC and a 5% significance level) in the ST class; SEP-ST-GETS, result of GETS density selection (using SC and a 5% significance level) across both the SEP and ST classes. KS [ $pval$ ], the Kolmogorov-Smirnov test statistic and associated  $p$ -value of the standardised residuals against the fitted density obtained by SEP-ST-GETS density selection.

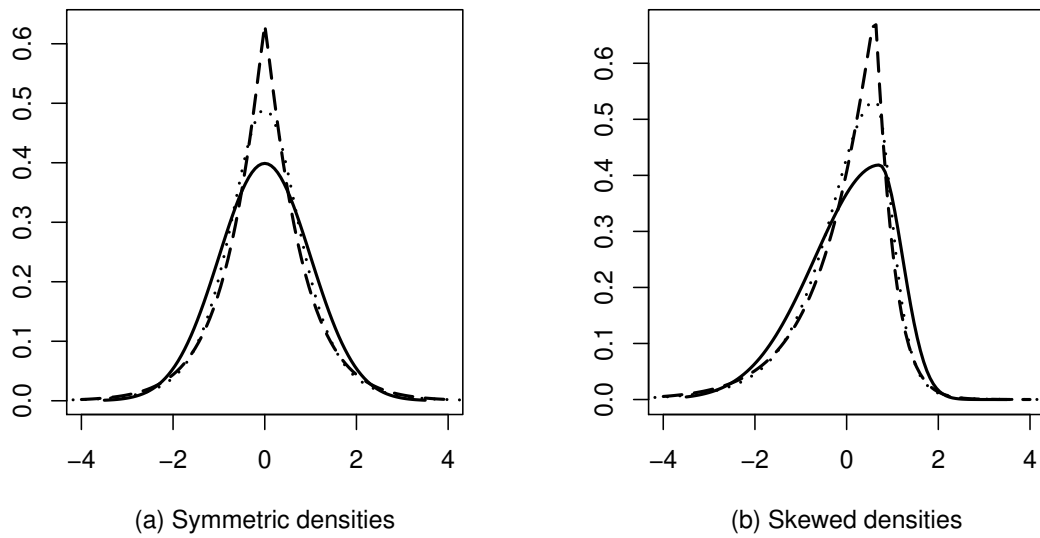


Figure 1: (a) Symmetric densities ( $\gamma = 1$ ): Standardised Normal (solid line), standardised EP with  $p = 1.1$  (dashed line) and standardised T with  $\nu = 5$  (dotted line). (b) Left skewed densities ( $\gamma = 0.6$ ): Standardised Skewed Normal (solid line), standardised Skewed EP with  $p = 1.1$  (dashed line), and standardised Skewed T with  $\nu = 5$  (dotted line)

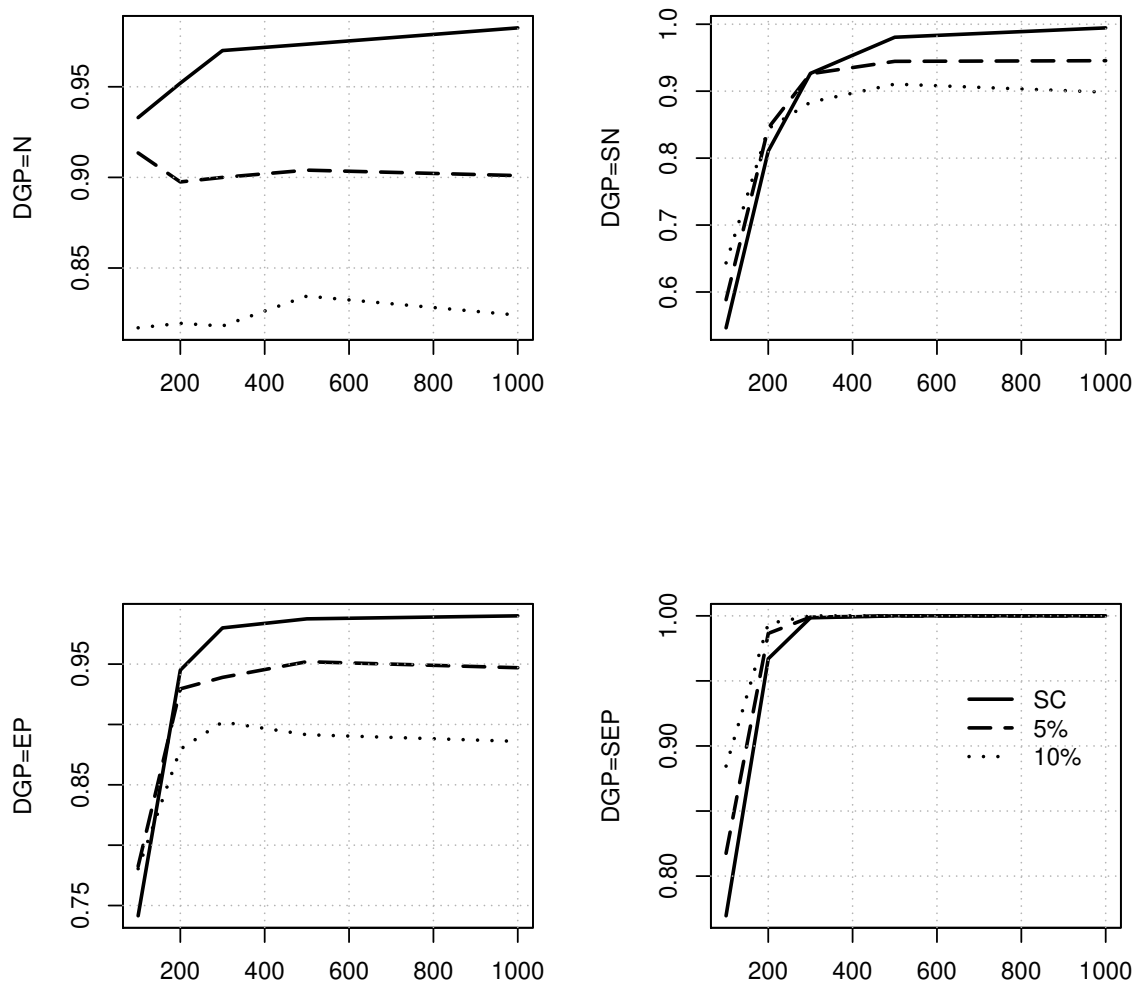


Figure 2: Probabilities of recovering the correct Skewed Exponential Power distribution DGP by means of an SC information criterion (solid line), and by means of GETS density selection combined with an SC criterion using 5% (dashed line) and 10% (dotted line) significance levels, respectively, for different sample sizes (horizontal axis). All simulations (2000 replications) in R



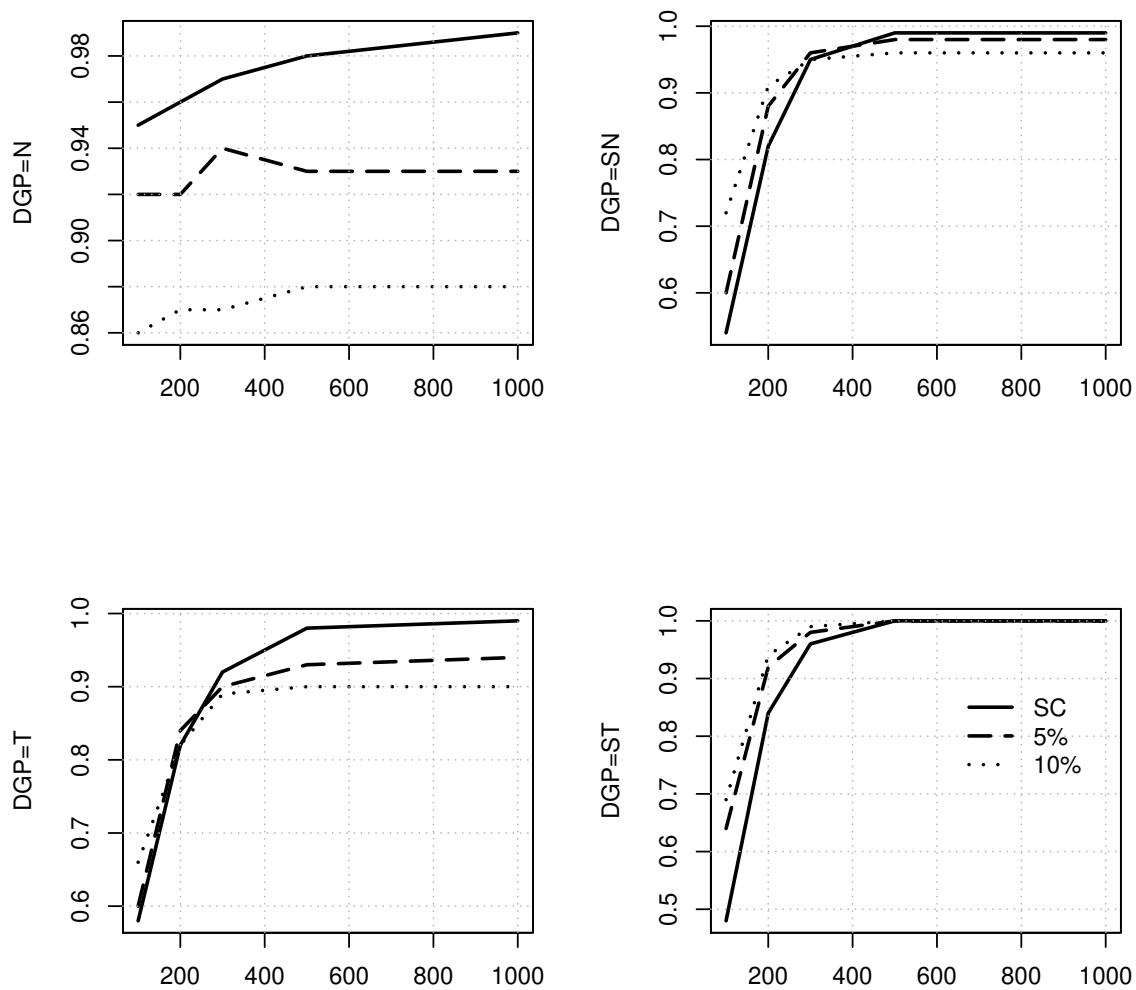


Figure 3: Probabilities of recovering the correct Skewed Student's  $t$  distribution DGP by means of an SC information criterion (solid line), and by means of GETS density selection combined with an SC criterion using 5% (dashed line) and 10% (dotted line) significance levels, respectively, for different sample sizes (horizontal axis). All simulations (2000 replications) in R.

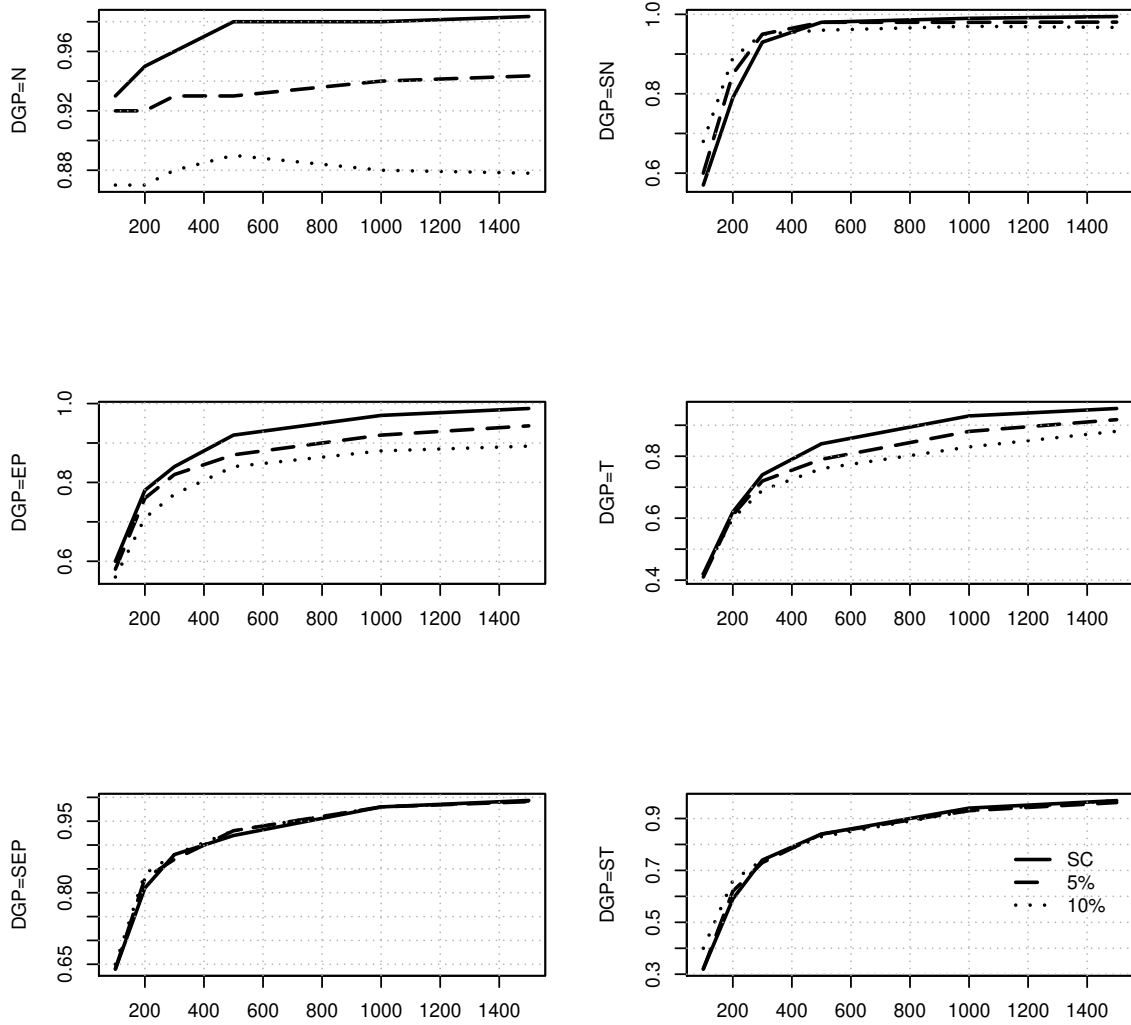


Figure 4: Probabilities of recovering the DGP from a SEP or ST density class by means of an SC information criterion (solid line), and by means of GETS density selection combined with an SC criterion using 5% (dashed line) and 10% (dotted line) significance levels, respectively, for different sample sizes (horizontal axis). All simulations (2000 replications) in R

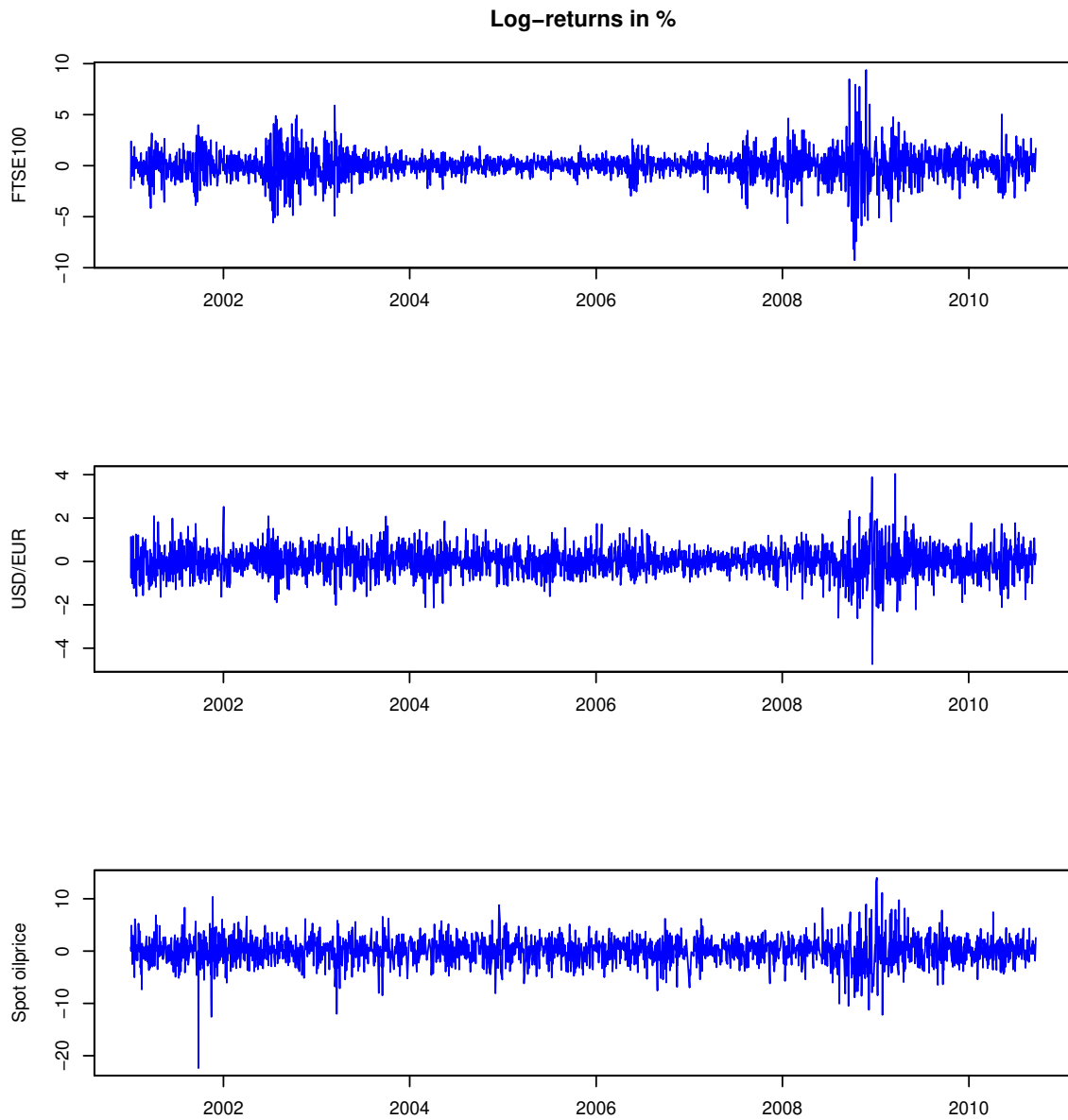


Figure 5: Daily log-returns in % of selected financial time series 1 January 2001 to 21 September 2010: The FTSE100 stock market index (2455 observations), the USD/EUR exchange rate (2487 observations) and the Spot Brent Blend oilprice (2300 observations)

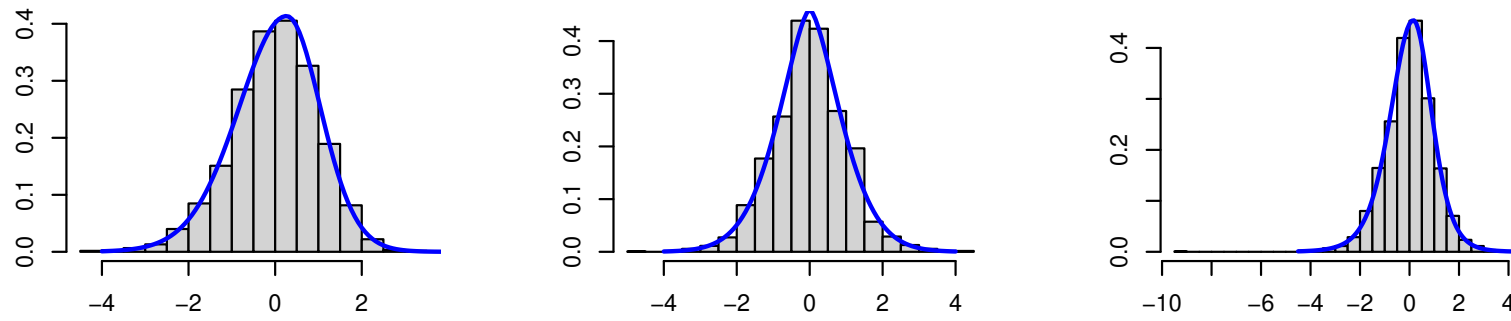


Figure 6: Histograms and fitted distributions (obtained from SEP-ST-GETS density selection) of the standardised residuals  $\{\hat{z}_t^{(\cdot)}\}$  of the estimated models in Panel B of Table 5