

Modelling the Skewed Exponential Power Distribution in Finance

Miguel Marín and Genaro Sucarrat

28 September 2010

Abstract We study the properties of two methods for financial density selection of the Skewed Exponential Power (SEP) distribution. The simulations suggest the two methods can be of great use in financial practice, since the recovery probabilities are sufficiently high in finite samples. For the first method, which simply consists of selecting a density by means of an information criterion, the Schwarz criterion stands out as performing well across density categories, and in particular when the Data Generating Process (DGP) is normal. In smaller samples the simulations suggest that our second method, General-to-Specific (GETS) density selection, can improve the recovery rate in predictable ways by changing the significance level. This is useful because it enables us to increase the recovery rate of a chosen density category, if one wishes to do so.

1 Introduction

Financial returns are often characterised by autoregressive conditional heteroscedasticity (ARCH), and by heavier tails than the normal—possibly skewed—even after standardising the returns. One may consider modelling everything simultaneously, say, by means of an ARCH type model that admits both skewed and heavy-tailed errors. However, in practice this is not always desirable. For example, many practitioners prefer using simple ARCH models like the RiskMetrics and Equally Weighted Moving Average (EqWMA) specifications in predicting volatility, models that do not require the specification of a density on the standardised error. Modelling the density of the standardised error thus becomes a separate step. Also, if the mod-

Juan Miguel Marín
Departamento de Estadística, Universidad Carlos III de Madrid, e-mail: jmmarin@est-econ.uc3m.es

Genaro Sucarrat
Department of Economics, BI Norwegian School of Management e-mail: genaro.sucarrat@bi.no

elling problem involves many explanatory variables in addition to the ARCH and density structures, as when modelling the relative change in daily electricity prices for example, or if the density is simply too complex for reliable estimation in practice, then simultaneous estimation and inference can be numerically inefficient or impossible in practice. This motivates modelling the density of the standardised errors in a separate step.

Here we propose and evaluate simple methods that model a standardised Skewed Exponential Power (SEP) distribution.¹ The SEP is attractive since the normal can be obtained as a special case through parameter restrictions, and since the SEP can produce densities that are both more and less heavy-tailed than the normal. The latter property is a real—although uncommon—possibility, in particular for low frequency financial returns and for models with explanatory variables. Also, the moments of the EP distribution exist under weaker assumptions than many other heavy-tailed distributions, say, the Student's t .

We study the finite sample properties of two density specification search algorithms through simulation. The first density search algorithm we study consists simply of choosing, among four densities, the density that minimises an appropriately chosen information criterion. The four densities are all nested within the standardised SEP: The standard normal (N) density, the standardised skew-normal (SN) density, the standardised symmetric exponential power (EP) density and the standardised skewed exponential power (SEP) density. The second density search algorithm we study can be viewed as a density selection analogue to multi-path General-to-Specific (GETS) model selection, see Campos et al. (2005) for a comprehensive overview of GETS model selection in regression analysis. Multi-path GETS combines repeated backwards stepwise regression (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.

The rest of this chapter contains two sections. The next section outlines the statistical framework, and the final section contains our simulations.

2 Statistical framework

The generic ARCH model is given by

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

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

$$\sigma_t^2 = Var(r_t | \mathcal{F}_t), \quad (3)$$

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

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

The Exponential Power (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 $p \in (0, \infty)$. μ is a location parameter, σ 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) is obtained by setting

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

which means $E(z) = 0$ and $Var(z) = 1$.

Following Zhu and Zinde-Walsh (2009) we may distinguish between two main approaches to the skewing of an EP distribution. The method of Azzalini (1986) on the one hand, and the method of Fernández et al. (1995), Fernández and Steel (1998), Theodossiou (2000) and Komunjer (2007) on the other. 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 Zhu and Zinde-Walsh (2009, p. 90). By contrast, consistency of ML estimation for the second method, which we will refer to as the Fernández and Steel (1998) method, is proved by Zhu and Zinde-Walsh (2009) when the shape parameter p is greater than 1. Moreover, the Fernández and Steel (1998) method is conceptually simpler and readily applicable to other densities. For these reasons we skew the standardised EP distribution by means of the Fernández and Steel (1998) method.

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

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

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 the standardised version of (4) such that (5) is satisfied) that:

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

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

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

where

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

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

Studying the properties of a density selection algorithm necessitates a numerically robust estimation algorithm, and the main properties of our ML code,² 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 $\gamma \in [0.6, 5]$ and $p \in [1, 3]$ are considered. These values cover the range of values that (we believe) are likely to be encountered in empirical practice, and restricting the search space improves the estimation accuracy substantially in small samples. The initial values of the algorithm are always $\gamma = 1$ and $p = 2$, which correspond to the symmetric standard normal density.

3 Financial density selection

We study the finite sample performance of two density selection algorithms under four different Data Generating Processes (DGPs): (1) $z \sim N(0, 1)$, (2) $z \sim SN(\gamma = 0.7)$, (3) $z \sim EP(p = 1.1)$ and (4) $z \sim SEP(\gamma = 0.7, p = 1.1)$. For expository brevity we will refer to these four DGPs as N, SN, EP and SEP, respectively. The values $\gamma = 0.7$ and $p = 1.1$ are at the border of what one is likely to encounter in practice.

² Our code is a modified version of code from the `fGarch` package, see Würtz and Chalabi (2009).

Table 1 Numerical performance of ML estimation

T	γ	p	$M(\hat{\gamma})$	$V(\hat{\gamma})$	$M(\hat{p})$	$V(\hat{p})$	$M(\textit{iters})$	$\% \textit{no-conv.}$
100	1.0	2.0	1.024	0.051	2.113	0.261	8.00	0
		1.1	1.008	0.014	1.146	0.039	17.29	25
200		2.0	1.005	0.014	2.059	0.125	7.70	0
		1.1	1.001	0.005	1.123	0.017	17.16	24
500		2.0	1.002	0.005	2.022	0.044	8.21	0
		1.1	1.002	0.204	1.104	0.007	16.91	20
1000		2.0	1.001	0.002	2.014	0.021	8.62	0
		1.1	1.000	0.001	1.102	0.004	16.43	19
100	0.7	2.0	0.711	0.013	2.089	0.220	8.26	0
		1.1	0.703	0.006	1.135	0.028	16.34	26
200		2.0	0.702	0.006	2.053	0.118	8.79	0
		1.1	0.699	0.003	1.122	0.015	16.98	26
500		2.0	0.699	0.002	2.024	0.043	9.53	0
		1.1	0.699	0.001	1.104	0.006	17.36	23
1000		2.0	0.698	0.001	2.009	0.020	9.89	0
		1.1	0.699	0.000	1.101	0.003	16.92	20

Simulations (2000 replications) in R with ML estimation implemented via the `nlm` function. T is the sample size, $M(\cdot)$ and $V(\cdot)$ denote the mean and sample variance, respectively, \textit{iters} is short for iterations and $\% \textit{no-conv.}$ is the percent of time that the algorithm did not converge.

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 different information criteria: The Schwarz (1978) criterion (SC),³ the Akaike (1974) criterion (AIC) and the Hannan and Quinn (1979) criterion (HQ). The three criteria we compute as

$$\begin{aligned} \text{SC:} & \quad -2\log l/T + k(\log T)/T \\ \text{AIC:} & \quad -2\log l/T + 2k/T \\ \text{HQ:} & \quad -2\log l/T + 2k\log[\log(T)]/T \end{aligned}$$

where $\log l$ is the empirical log-likelihood, and where $k = 0, k = 1, k = 1$ and $k = 2$ for N, SN, EP and SEP, respectively.

Table 2 contains the probabilities of recovering the right density for four different DGPs. The SC criterion has the best overall performance, since it performs well in all four cases, and since it performs well in both small and large samples. Also, consistent model selection is attained relatively fast in all four cases. Indeed, the simulations suggest that when the sample size is greater than 300, then SC is the preferred information criterion. Of course, this is to some extent because of the large differences between the four densities (smaller differences would presumably

³ The SC is also known as the Bayesian Information Criterion (BIC).

result in lower recover rates). When the sample size is smaller than 300, however, then the simulations suggest that the HQ criterion should be preferred. The AIC is sometimes slightly better than HQ in small samples, but the probabilities increase slower than for HQ, in particular when the DGP is normal.

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

T	$DGP = N:$			$DGP = SN:$			$DGP = EP:$			$DGP = SEP:$		
	SC	AIC	HQ	SC	AIC	HQ	SC	AIC	HQ	SC	AIC	HQ
100	0.93	0.69	0.83	0.53	0.68	0.64	0.74	0.75	0.80	0.77	0.91	0.87
200	0.96	0.72	0.87	0.81	0.81	0.86	0.94	0.82	0.91	0.96	1.00	0.99
300	0.96	0.69	0.88	0.93	0.83	0.93	0.97	0.83	0.93	1.00	1.00	1.00
500	0.97	0.71	0.88	0.98	0.84	0.94	0.98	0.84	0.95	1.00	1.00	1.00
1000	0.98	0.71	0.89	0.99	0.83	0.95	0.99	0.84	0.94	1.00	1.00	1.00

Simulations (2000 replications) in R

3.2 GETS density selection

The GETS density selection algorithm that we propose starts with the unrestricted estimate of a SEP density. Next, two different simplification paths are considered. The first path consists of first testing the restriction $p = 2$ and then $\gamma = 1$, and the second path consists of first testing the restriction $\gamma = 1$ and then $p = 2$. Inference is by means of likelihood ratio (LR) tests, and simplification along a path stops when a null hypothesis is rejected. Table 3 contains the finite sample probabilities of rejecting the various null hypotheses that makes up GETS density selection under different DGPs. The table suggests that the empirical size is close to the nominal size in finite samples, when the alternative does not coincide with the DGP. Sometimes, simplification can result in two different terminal models, say, SN and EP , or SN and N , or EP and N . 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 α :

$$p(DGP|N) \rightarrow (1 - \alpha)^2$$

$$p(DGP|SN) \rightarrow (1 - \alpha)$$

$$p(DGP|EP) \rightarrow (1 - \alpha)$$

$$p(DGP|SEP) \rightarrow 1$$

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. For example, for the nominal sizes 10% and 5% the probability $p(DGP|N)$ tends to 0.81 and 0.9025, respectively.

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 α to “push” the algorithm either towards or away from normality, if one wishes to do so. For example, the simulations above showed that the SC criterion performs very well in both small and large sample sizes when the DGP is normal. 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 large) in a controlled and predictable way by simply increasing the significance level.

Figure 1 provides a snapshot of how GETS density selection actually works in practice. 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 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 the DGP is equal to SEP then there is a gain of about 13 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. Indeed, the gain might even be larger when the departure from normality is not as large as in the simulations.

Acknowledgements We are thankful to participants at the MAF 2010 conference (Ravello) for useful comments, suggestions and questions. Funding from The Bank of Spain Excellence Program is gratefully acknowledged.

References

- 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.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.
- Campos, J., D. F. Hendry, and N. R. Ericsson (Eds.) (2005). *General-to-Specific Modeling. Volumes 1 and 2*. Cheltenham: Edward Elgar Publishing.
- 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.

Table 3 Rejection probabilities of H_0 under different DGPs, for a nominal level of 5%

DGP	T	$H_0 : N,$			$H_0 : SN,$	$H_0 : EP,$
		$H_1 : SN$	$H_1 : EP$	$H_1 : SEP$	$H_1 : SEP$	$H_1 : SEP$
N	100	0.05	0.05	0.06	0.06	0.07
	200	0.04	0.05	0.05	0.06	0.05
	500	0.05	0.05	0.05	0.05	0.05
	1000	0.05	0.05	0.06	0.05	0.05
SN	100	0.64	0.06	0.54	0.05	0.63
	200	0.90	0.07	0.84	0.05	0.89
	500	1.00	0.08	1.00	0.05	1.00
	1000	1.00	0.07	1.00	0.04	1.00
EP	100	0.10	0.82	0.76	0.82	0.06
	200	0.09	0.98	0.96	0.98	0.05
	500	0.11	1.00	1.00	1.00	0.05
	1000	0.12	1.00	1.00	1.00	0.05
SEP	100	0.85	0.68	0.96	0.86	0.93
	200	0.99	0.92	1.00	0.99	1.00
	500	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00

Simulations (2000 replications) in R. The top-left number (0.05) should be interpreted as the probability of accepting $H_1 : SN$ when $H_0 : N$, $LR \sim \chi^2(1)$, for a DGP equal to N. The third number in row six (0.84) should be interpreted as the probability of accepting $H_1 : SEP$ when $H_0 : N$, $LR \sim \chi^2(2)$, for a DGP equal to SN. The bottom-right number (1.00), should be interpreted as the probability of accepting $H_1 : SEP$ when $H_0 : EP$, $LR \sim \chi^2(1)$, for a DGP equal to SEP. And so on.

- 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.
- Komunjer, I. (2007). Asymmetric power distribution: Theory and application to risk measurement. *Journal of Applied Economics* 22, 891–921.
- 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.
- 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. Downloadable via <http://cran.r-project.org/web/packages/fGarch/index.html>.
- Zhu, D. and V. Zinde-Walsh (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics* 148, 86–99.

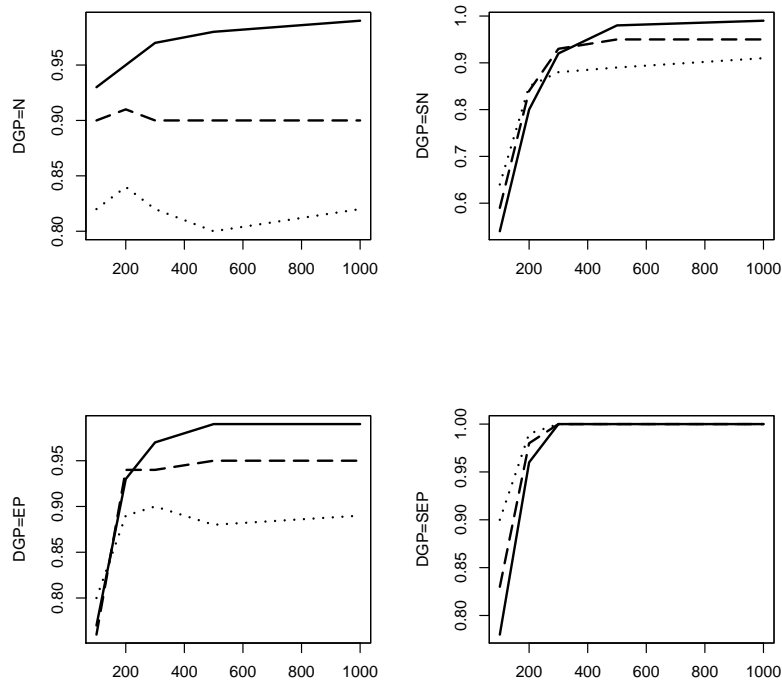


Fig. 1 Probabilities of recovering the DGP by means of an SC information criterion (solid line), and by means of GETS density selection combined with an SC criterion using 5% (dashed line) and 10% (dotted line) significance levels, respectively. All simulations (2000 replications) in R