Testing densities with financial data: an empirical comparison of the Edgeworth–Sargan density to the Student’s $t$

IGNACIO MAULEÓN* and JAVIER PEROTE**

*Universidad Rey Juan Carlos I. **Universidad de Salamanca.

The Edgeworth–Sargan density has been shown capable of capturing salient empirical regularities of financial data in some studies. The main purpose of the reported study is to compare its performance with other densities, most notably to the Student $t$. Both densities can account for thick tails, and asymmetry. One important by-product of the comparison is to test the existence of moments. The comparison of densities is carried out with daily financial observations, spanning 25 years of data from two major world stock markets. Attention is paid to the fitting of other empirical regularities, and especially to the peak, frequently found at the middle of the densities.

Keywords: densities comparison, Edgeworth–Sargan and Student $t$ distributions, financial data, testing moment existence

1. INTRODUCTION

Estimating and testing the densities of econometric models is a major research topic in finance. This may be explained, partly at least, by the following factors: (1) establishing forecasting confidence intervals requires knowledge of the exact density; (2) efficient estimation by maximum likelihood and unbiased estimation; (3) determining the price of derivative products (notably option prices; see, for example, Ball and Torous, 1985; Black and Karasinski, 1991; Heston, 1993; Dupire, 1994; Corrado and Su, 1995; Bates, 1996; Hardle and Hafner, 1997); and (4) evaluating the risk of portfolios (by assessing their exposure to the risk of price changes; Dacorogna et al. 1995; Jorion, 1997; Dowd, 1998).

The problem of fitting the appropriate probability distribution in applied financial modelling has been addressed using several different families of probability densities. These include convolutions of the Poisson and the Normal
(Ball and Roma, 1993; Ball and Torous, 1983; Akgiray and Booth, 1988; Jorion, 1988; Vlaar and Palm, 1993), nonparametric estimation (Silverman, 1986), mixtures of Normals (Hamilton, 1991; Harvey and Zhou, 1993), the Gamma (Nelson, 1991), the Generalized Beta (McDonald and Xu, 1995), and the student’s \( t \) (Praetz, 1972; Blattberg and Gonedes, 1974; Rogalski and Vinso, 1978; Prucha and Kelejian, 1984). In particular, this last density has been the subject of much attention, because it can account for thick tails – a well-known feature of financial data – and does not require the existence of moments of all orders.

Another family of probability densities, which has not been extensively tested in finance, is the Edgeworth–Sargan (henceforth ES) family. This family was brought into econometrics by the early work of Sargan on finite sample distributions (see, for example, Sargan, 1976, 1980; Mauleón, 1983). This density is able to capture salient empirical regularities of financial data, like asymmetry and kurtosis (Gallant and Tauchen, 1989; Bourgoin and Prieul, 1997) and has a number of interesting properties. It provides a systematic way of generalizing the complexity of the density, so that it can be understood as a semi non parametric approach (Gallant and Nychka, 1987). Therefore, it is able to fit adequately a wide variety of density patterns (Gallant and Fenton, 1996). It also allows the introduction of conditional heterogeneity, as a generalization of the time dependence implied by conditional heteroskedasticity (Gallant and Nychka, 1987).

The main purpose of this study is to compare the performance of the ES density to other densities, most notably the student \( t \) distribution. This is because both densities can account for thick tails, possibly the main feature of financial densities (both can account for asymmetry, as well; see Hansen, 1994 for a skewed \( t \) distribution). One important by-product of the comparison is to discuss the existence of moments in financial data. This is because the ES has finite moments of all orders, but the \( t \) does not. If second-order moments do not exist, as suggested by Pagan and Schwertz (1990), and Phillips and Loretan (1994, 1995), then (1) some standard estimation procedures, like the method of moments, would be under jeopardy, and (2) the definition of risk, central to financial analysis, would have to be modified (since it is generally identified with variance, after the work of Markowitz and Sharpe). The problem of testing moment existence can be tackled, alternatively, by estimating the tail index (Phillips and Loretan, 1994, 1995; Dacorogna et al., 1995). This approach being non-parametric, has the advantage of robustness. But it also requires an extremely large set of data, since outliers are treated as stochastic realizations from the same distribution as the remaining observations.

Section 2 reviews the main features of the probability distributions used in the paper. The empirical results (Section 3), are based on a sample of daily observations, spanning approximately 25 years of data from two major world stock markets (the UK and the USA). Conditional densities are considered, and a standard Garch model that captures the main features of the conditional heteroskedasticity of the data is implemented. The main results are summarized in Section 4. Most of the notation used in the paper is that in common use. Where necessary, terms are defined at relevant places in the text.
2. DESCRIPTION OF THE PROBABILITY DISTRIBUTIONS USED IN THE PAPER

This section is devoted to the presentation of the densities implemented in the empirical research reported in Section 3. The densities are the following: (1) Student $t$; (2) expansions based on the Student $t$; (3) the Edgeworth–Sargan; (4) the Gallant–Nychka Transformation; and (5) mixtures of the previous densities. Conditional heteroskedasticity can be added to all these models.

It is convenient, now, to review the main features of densities in financial data. They can be observed by means of simple graphs for almost all major financial markets and instruments. These are the following: (1) thick tails; (2) asymmetry (possibly); (3) conditional heteroskedasticity; and (4) peak at zero. The following comments are in order, regarding these properties: (1) the usual way of tackling this property is to fit a $t$ density; however, the question remains as to whether moments, at least up to the fourth order, exist; besides, the problem is mixed with the existence of outliers; it is one of the purposes of this study, to shed some light on this issue; (2) the ES density can account for asymmetries of various types, although we have found none in the empirical results; (3) as regards conditional heteroskedasticity, we use a standard Garch model to take account of this property, since this is not the main focus of the paper; and, (4) this point is not stressed in empirical research, although we find that it is an overwhelming property of financial series.

The densities, and their main properties, are reviewed next (the notation of the empirical results follows that used here). We start with the $t$ density.

Student $t$

The density function $f_{S,v}(\varepsilon)$ of a random variable, $\varepsilon$, distributed as a student’s $t$ is given by,

$$f_{S,v}(\varepsilon) = \frac{\Gamma[(v + 1)/2]/\Gamma[v/2]}{\sqrt{v\pi}}\left[1 + \frac{\varepsilon^2}{v}\right]^{-\frac{v+1}{2}, \ v > 0} \tag{2.1}$$

where $v$ is a positive number (not necessarily an integer), and $\Gamma(.)$ is the gamma function. This is a standard and well-known density defined almost one century ago (see, for example, Kendall and Stuart, 1977). The scaling can be obtained by defining $\eta = \sigma\varepsilon$, $\sigma > 0$, and the density of $\eta$ is given by, $f_{S,v}(\eta/\sigma)/\sigma$.

The moments of $\eta$ for integer $v$ are given as follows,

$$E(\eta^r) = 0, v > r(r, odd),$$

$$E(\eta^2) = \frac{\sigma^2}{\sqrt{v-2}}, v > 2$$

$$E(\eta^r) = \frac{\sigma^r \sqrt{v}}{B[(r + 1)/2,(v - r)/2]/B[1/2,v/2]} < \infty, v > r(r, even) \tag{2.2}$$

where $B(\ldots)$ is the beta function. For $r \geq v$, the moments do not exist.

In a time series context, we write $\eta_t$ in place of $\eta$ in (2.1) and (2.2). In order to allow for conditional heteroskedasticity, a Garch(1,1) specification can be introduced. Then, $\sigma_t$ takes the place of $\sigma$, as follows,

$$\sigma_t^2 - \alpha_0 = \alpha_1(\sigma_{t-1}^2 - \alpha_0) + \alpha_2 \eta_{t-1}^2 \tag{2.3}$$
We note that if \( v < 2 \), then \( \sigma_i^2 \) cannot be interpreted as a variance, and that conditional on \( t = 1 \), and provided \( v > 2 \), \( E(\eta_i) = \sigma_i^2 \cdot v/(v - 2) \). Hansen (1994) provides a generalization to account for skewness.

Two families of probability densities can be generated from the Student \( t \) distribution. These are new in the literature, and are derived following the same type of expansion that gives rise to the Hermite, or ES, density (Kendall and Stuart, 1977; Sargan, 1976; and (2.10) for the ES density). The motivation behind them is to allow for additional parameters, so that the comparison with the ES density is impartial.

**Expansion 1**

The density derived in the first type of expansion is given by,

\[
f_1(\varepsilon_i) = f_{S,v}(\varepsilon_i).[1 + \sum_{s=1}^{m} \gamma_{1,s}P_s(\varepsilon_i)]
\]

(2.4)

where the polynomial \( P_s(\varepsilon_i) \) is defined by the following identity,

\[
\frac{\partial}{\partial \varepsilon} f_{S,v}(\varepsilon_i) = f_{S,v}(\varepsilon_i).(-1)^sP_s(\varepsilon_i)
\]

(2.5)

and, in particular,

\[
\begin{align*}
P_1(\varepsilon_i) &= \left[\omega/\nu\right] \varepsilon_i \Theta(\varepsilon_i) \\
P_2(\varepsilon_i) &= \left[\left(\nu + 3\right) \omega/\nu^2\right] \varepsilon_i^2 \Theta(\varepsilon_i)^2 - \left[\omega/\nu\right] \Theta(\varepsilon_i) \\
P_3(\varepsilon_i) &= \left[\left(1 + 10/\nu + 23/\nu^2 + 14/\nu^3\right) \varepsilon_i^3 \Theta(\varepsilon_i)^3 - \left[3 + 12/\nu + 9/\nu^2\right] \varepsilon_i \Theta(\varepsilon_i)^2 \\
&\quad - \left[6 \omega^3 + 36 \omega^2 - 48 \omega\right]/\nu^3\varepsilon_i^3 \Theta(\varepsilon_i)^3 + \left[3 \omega \left(\nu - 1\right)/\nu^2\right] \Theta(\varepsilon_i)^2
\end{align*}
\]

(2.6)

where,

\[
\Theta(\varepsilon_i) = 1/(1 + \varepsilon_i^2/\nu) \\
\omega = \nu + 1
\]

(2.7)

It is easily checked that the density integrates up to one without further restrictions. This is because

\[
\int_{-\infty}^{+\infty} P_s(\varepsilon)f_{S,v}(\varepsilon)\,d\varepsilon = 0, \quad s \geq 1
\]

(2.8)

The highest-order finite moment is still the same (i.e. strictly smaller than the number of degrees of freedom). Since \( \varepsilon_i \Theta(\varepsilon_i) \to 0 \), as \( \varepsilon \to \infty \), this density may not necessarily become truncated, even if \( \gamma_{1,m} \) is negative: it will ultimately depend on the value of this coefficient (see Equations 2.16 and 2.17 for further discussion on truncation). It is interesting to note that as \( \nu \to \infty \), this density tends to the density of the ES distribution (defined in (2.10)). It should be remarked, as well, that the polynomials in (2.6) and (2.7) are fairly complex, so that the empirical implementation of this expansion may be difficult, except for low values of \( m \).

**Expansion 2**

The second expansion considered in this paper is
\[ f_2(\varepsilon_t) = f_{S,v}(\varepsilon_t) \left( 1 + \sum_{s=1}^{m} \gamma_{2,s} (\varepsilon_t^s - \mu^s) \right) \]  

(2.9)

where \( \mu^s \) is the \( s \)th order moment of a \( t \) density (with the same number of degrees of freedom). The density is asymmetric if for some odd \( s \), \( \gamma_{2,s} \neq 0 \). This expansion is simpler than expansion 1, and yet, the probability integrates up to one. One property of this expansion, is that the upper limit to the highest-order finite moment is now \((v - m)\) (provided \( \gamma_{2,m} \) is positive). If the coefficient of the highest-order polynomial, \( \gamma_{2,m} \), is negative, then the density becomes truncated, and moments of all orders exist. In this last case, however, the previous expression for the density has to be adjusted by a factor that ensures that the probability integral is one (see the discussion in (2.16) and (2.17)). It could also be possible, following Gallant and Nychka (1987), to modify slightly the expansion to avoid any truncation (see (2.18, 2.19) below).

**The Edgeworth–Sargan (ES) density**

The Edgeworth–Sargan probability density is given by the following specification (Kendall and Stuart, 1977; Sargan, 1976),

\[ f_{ES}(\varepsilon_t) = \phi(\varepsilon_t).\left[ 1 + \sum_{s=1}^{q} \delta_s H_s(\varepsilon_t) \right] \]  

(2.10)

where \( \phi(\varepsilon_t) \) stands for a N(0,1) density, the polynomials \( H_s(\varepsilon_t) \) are defined by means of the identity \( \delta^s \phi(\varepsilon_t) = [(-1)^s \phi(\varepsilon_t) H_s(\varepsilon_t)] \), and the \( \delta_s \) are a set of constants. In particular, the first eight polynomials are the following (only polynomials up to this order will be considered in the empirical application).

\[
\begin{align*}
H_1(\varepsilon_t) &= \varepsilon_t \\
H_2(\varepsilon_t) &= \varepsilon_t^2 - 1 \\
H_3(\varepsilon_t) &= \varepsilon_t^3 - 3 \varepsilon_t \\
H_4(\varepsilon_t) &= \varepsilon_t^4 - 6 \varepsilon_t^2 + 3 \\
H_5(\varepsilon_t) &= \varepsilon_t^5 - 10 \varepsilon_t^3 + 15 \varepsilon_t \\
H_6(\varepsilon_t) &= \varepsilon_t^6 - 15 \varepsilon_t^4 + 45 \varepsilon_t^2 - 15 \\
H_7(\varepsilon_t) &= \varepsilon_t^7 - 21 \varepsilon_t^5 + 105 \varepsilon_t^3 - 105 \varepsilon_t \\
H_8(\varepsilon_t) &= \varepsilon_t^8 - 28 \varepsilon_t^6 + 210 \varepsilon_t^4 - 420 \varepsilon_t^2 + 105
\end{align*}
\]  

(2.11)

The polynomials have a number of interesting properties. First, and since \( \delta^s \phi(\varepsilon_t) \rightarrow 0 \), as \( \varepsilon_t \rightarrow \pm \infty \), then,

\[
\int_{-\infty}^{+\infty} H_s(\varepsilon_t) \phi(\varepsilon_t) d\varepsilon_t = 0, \quad s \geq 1
\]  

(2.12)

Second, the following orthogonality property holds (Kendall and Stuart, 1977),

\[
\int_{-\infty}^{+\infty} H_s(\varepsilon_t) H_r(\varepsilon_t) \phi(\varepsilon_t) d\varepsilon_t = 0, \quad s \neq r
\]

\[= s!, \quad s = r
\]  

(2.13)
Noting now that we can write,
\[
\begin{align*}
\varepsilon^2_1 &= H_2(\varepsilon_r) + 1 \\
\varepsilon^3_1 &= H_3(\varepsilon_r) + 3H_1(\varepsilon_r) \\
\varepsilon^4_1 &= H_4(\varepsilon_r) + 6H_2(\varepsilon_r) + 3
\end{align*}
\]
(2.14)
and from the two properties in (2.12, 2.13), the first four moments of the distribution are immediately given by,
\[
\begin{align*}
E(\varepsilon^2_1) &= \delta_1 \\
E(\varepsilon^3_1) &= 1 + 2\delta_2 \\
E(\varepsilon^4_1) &= 6\delta_3 + 3\delta_1 \\
E(\varepsilon^4_1) &= 24\delta_4 + 12\delta_2 + 3
\end{align*}
\]
(2.15)
Higher-order moments can be obtained similarly. We note, also, that they exist to any order. This is because they can be written as linear combinations of the moments of a Normal density. It is also easily checked that the density integrates up to one (because of (2.12)). If \(\delta_1 = 0\), then the mean is zero, and \(\delta_3\) accounts for asymmetry \((\delta_3 = E(\varepsilon^3_1)/6)\). Also, if \(\delta_2 = 0\), then \(\delta_4 = [E(\varepsilon^4_1) - 3]/24\), so that it provides a direct measure of kurtosis (note that the fourth-order moment of the standard normal distribution is 3). If \(\delta_2 \neq 0\), \(E(\varepsilon^3_1)\) is directly proportional to \(\delta_4\), although kurtosis depends now on both, \(\delta_2\) and \(\delta_4\).

If the coefficient of the highest-order polynomial, \(\delta^q_{\varphi}\), is negative, then the density becomes truncated. In this last case, however, the previous expression for the density has to be adjusted by a factor that ensures that the probability integral is one. More explicitly, we note first, that for large values of \(|\varepsilon_r|\), \(\{\sum_{s=1}^{q} \delta_s H_s(\varepsilon_r)\}\) will be dominated by the highest power on \(\varepsilon_r\) If \(\delta^q_{\varphi}\) is negative, then the largest admissible value for \(\varepsilon_r\), \(\overline{\varepsilon}_r\), will be given by solving
\[
1 + \sum_{s=1}^{q} \delta_s H_s(\overline{\varepsilon}_r) = 0
\]
(2.16)
where, for simplicity, it is assumed that the density is symmetric (the generalization to the asymmetric case is straightforward). Note, also, that negative solutions are prevented, since we assume (2.10) to be a valid density (see (2.18) for further discussion on this point). If there are more than one solution, \(\overline{\varepsilon}_r\), will be the largest.

The density will be defined now, for \(|\varepsilon_r| \leq \overline{\varepsilon}_r\), by,
\[
f_{ES}(\varepsilon_r) = \frac{\phi(\varepsilon_r).\{1 + \sum_{s=1}^{q} \delta_s H_s(\varepsilon_r)\}}{\int_{-\overline{\varepsilon}}^{\overline{\varepsilon}} \phi(\varepsilon_r).\{1 + \sum_{s=1}^{q} \delta_s H_s(\varepsilon_r)\} d\varepsilon_r}
\]
(2.17)
It may also be interesting to note that the sum of polynomials can be written as a sum of powers of \(\varepsilon_r\). Some additional properties of interest of the ES family of probability densities are the following: (1) it can be easily generalized to include more parameters, should they be needed (since \(\delta^q_{\varphi}(\varepsilon_r) = \sum_{s=1}^{q} \delta_s H_s(\varepsilon_r)\))...
[(−1)^i\phi(\varepsilon_i)H_i(\varepsilon_i)], higher-order polynomials H_i(.) can be calculated easily; see, for example, Abramowitz and Stegun (1972), and Kendall and Stuart (1977), for these and other properties; (2) the probability distribution function is easily obtained; (3) it allows the introduction of conditional heterogeneity (besides conditional heteroskedasticity); (4) a multiperiod forecasting stability test can be derived analytically (Mauleón, 1999a); and (5) the multivariate generalization is flexible enough, and provides good empirical fits (Mauleón, 1999b; Perote, 1999).

A generalization to allow for an arbitrary, and time-varying variance, is easily obtained denoting η_i = σ_r\varepsilon_r. The density of η_i is immediately given by, \bar{f}_{ES}(\eta_i/\sigma_i)/\sigma_i. The notation for the Garch(1,1) specification is the same as that used for the t density (see (2.3)). We note, finally, that conditional on t − 1, \(E(\eta_i^2) = (1 + 2.\delta_2).\sigma_i^2\).

**Gallant–Nychka transformation**

One possible problem with the density of (2.10), is that for some values of the \(\delta_s\), it may yield negative results. Restrictions on the coefficients \(\delta_s\) can be imposed, to ensure that the density is always positive. But even for a moderate number of parameters it may be difficult, or even impossible, to work them out explicitly. A transformation to avoid that problem is offered by Gallant and Nychka (1987), as follows,

\[
f_{GN}(\varepsilon_i) = \phi(\varepsilon_i).\{\delta_0 + (\sum_{s=1}^{q} \delta_{GN,s}\varepsilon_i^s)^2\}
\]  

where,

\[
0 < \delta_0 = 1 - \sum_{s=1}^{q} \sum_{n=1}^{q} [\delta_{GN,s}\delta_{GN,n} \int_{+\infty}^{+\infty} \varepsilon_i^{n+s}\phi(\varepsilon_i)d\varepsilon_i]
\]

This ensures that the probability integral is one. Further restrictions can be imposed to obtain zero mean and unit variance, if required. The sum of powers on \(\varepsilon_i^s\) can be reparameterized, so that it becomes a sum of Hermite polynomials. Also, the coefficients \(\delta_{GN,s}\) can be made dependent on lagged values of \(\varepsilon_i\), so that a general form of conditional heterogeneity is allowed. Gallant and Nychka (1987), call this approach semi nonparametric, because they show that by increasing the number of parameters, i.e. the value of q, an arbitrary degree of accuracy fitting any density, can be achieved (see, also, Gallant and Fenton, 1996). They also point out, however, that a parametric interpretation is legitimate, as well. Empirical work with this density is presented in Gallant and Tauchen (1989) and Bourgoin and Prieul (1997).

### 3. EMPIRICAL RESULTS

This section presents the empirical results, organized as follows: first, the criteria used to compare the performance of the probability densities are stated; second, the data set is described; third, the main results are presented and assessed in two tables and three figures; fourth, some details of the empirical
work are discussed, and lastly, the main results of the section are summarized.

The empirical performance of the probability densities is assessed and compared according to the general fit, as given by the value of the likelihood, as the basic criterion (all models have been estimated by maximum likelihood). A graphical analysis, that might detect some misspecification, is also considered. A second criterion is the number of parameters required to obtain an adequate fit, as a measure of complexity. Finally, attention is paid to the highest order finite moment implied by every fitted density. The data set chosen to conduct the estimation is 25 years of daily observations, coming from two of the major world financial markets. These are the UK FTSE index and the USA Dow Jones (6590 observations, from 24/01/1971 to 28/05/1996).

An AR(1) has been applied to the rate of change of raw data in both cases. That is, if $x_t$ is the raw index, the following model has been estimated in both cases,

$$\Delta \log(x_t) = a_0 + a_1 \Delta \log(x_{t-1}) + \eta_t$$

(3.1)

The results shown in Tables 1 and 2 follow this notation. The probability densities have been applied to the residuals of this equation, $\hat{\eta}_t$. A standard Garch(1,1) model to take account of conditional heteroskedasticity has been implemented in all cases, with the notation stated in the previous section, i.e. $\hat{\eta}_t = \sigma_t \hat{\varepsilon}_t$, where $\sigma_t$ is given in (2.3) and $\hat{\varepsilon}_t$ follows either (1) the $t$ density, given in (2.1), (2) the first expansion of the $t$ density given in (2.4), (3) the ES density given in (2.10), (4) a mixture of the $t$ and the Normal, given in (3.2), or (5) a mixture of the ES and the Normal given in (3.3). The notation of the tables matches exactly that of the equations in the main text.

Mixtures of the probability densities are also estimated, as indicated in the previous section. More specifically, results for the following cases are presented: (1) mixture of a Normal and the Student $t$, denoted by $f_{N,S}$, (2) mixture of a Normal and the ES, denoted by $f_{N,ES}$. The remaining notation is given next:

$$f_{N,S} = p_N \phi(\eta_t/\sigma_N)/\sigma_N + (1 - p_N) f_{S,N}(\eta_t/\sigma_t)/\sigma_t$$

(3.2)

$$f_{N,ES} = p_N \phi(\eta_t/\sigma_N)/\sigma_N + (1 - p_N) f_{ES,N}(\eta_t/\sigma_t)/\sigma_t$$

(3.3)

where $p_N$ is the probability attached to the Normal density, and $\sigma_N$ its variance (assumed constant). This notation is followed in Tables 1 and 2 (all remaining symbols have been defined in Section 2).

The main empirical results are summarized in Tables 1 and 2. Figures 1 to 3 illustrate some salient features of the fits obtained. Table 1 gives the results for the Student $t$, columns I and II for the Dow Jones, and columns III and IV for the FTSE. The notation is that of the main text ((2.1), (2.3), and (3.1), (3.2)). Table 2, in turn, gives results for the ES, again, columns I and II for the Dow Jones, and columns III and IV for the FTSE. The notation is that of the main text, as well ((2.3), (2.10) and (3.1), (3.3)). The value attained by the log of the likelihood in every case is denoted by $\log \text{L}$.

We turn now to a discussion of the results themselves. We comment first on the results for the Student $t$ (Table 1, Fig. 1), and secondly on the ES (Table 2,
Table 1. Student t

<table>
<thead>
<tr>
<th>Variable</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>0.00033</td>
<td>0.00031</td>
<td>0.00044</td>
<td>0.00036</td>
</tr>
<tr>
<td></td>
<td>(3.6)</td>
<td>(2.8)</td>
<td>(4.6)</td>
<td>(3.0)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.083</td>
<td>0.088</td>
<td>0.153</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>(6.9)</td>
<td>(7.3)</td>
<td>(12.1)</td>
<td>(13.1)</td>
</tr>
<tr>
<td>$\rho_N$</td>
<td>0.038</td>
<td>0.022</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.8)</td>
<td>(2.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_N$</td>
<td>0.0016</td>
<td>0.00082</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.7)</td>
<td>(2.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>7.4</td>
<td>8.7</td>
<td>8.4</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>(11.4)</td>
<td>(9.0)</td>
<td>(8.7)</td>
<td>(10.1)</td>
</tr>
<tr>
<td>$\gamma_{1,2}$</td>
<td>ns</td>
<td>ns</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.8)</td>
<td>(5.3)</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.0000112</td>
<td>0.0000124</td>
<td>0.0000092</td>
<td>0.0000085</td>
</tr>
<tr>
<td></td>
<td>(6.5)</td>
<td>(6.4)</td>
<td>(7.0)</td>
<td>(7.2)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.967</td>
<td>0.966</td>
<td>0.9</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td>(108)</td>
<td>(107)</td>
<td>(103)</td>
<td>(104)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.028</td>
<td>0.031</td>
<td>0.055</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(7.5)</td>
<td>(7.3)</td>
<td>(9.3)</td>
<td>(9.8)</td>
</tr>
<tr>
<td>$\log LK$</td>
<td>22290</td>
<td>22296</td>
<td>22089</td>
<td>22095</td>
</tr>
</tbody>
</table>

Cols. I, II results for the Dow Jones.
Cols. III, IV results for the FTSE.
ns: nonsignificant.

Fig. 1. Fit of the $t$ density: UK stock market
Figs 2 and 3). Then, the relative performance of the two type of densities is assessed. The results in Table 1 show that the mixture type of density is significant in both markets (II and IV). Inspecting Fig. 1, this was to be expected, since the Normal component of the mixtures has been introduced precisely to capture the peak at the middle (Fig. 1 gives the results for the Student t alone, and the Dow Jones; for the FTSE the picture is very similar). The second result concerns moment existence, and according to the value attained by v, we conclude that moments up to order 7 exist in all cases. Finally, the third interesting result is that the first type of expansion of the Student t given in (2.4) is significant in some cases (γ1,2 in columns III and IV).

The results for the ES are presented in Table 2. As in the previous case, we check that the mixture type of density is significant in both markets (II and IV). Again this was expected, since the fit of the ES alone yields a figure similar to Fig. 1. Figures 2 and 3 provide additional information (the fit of the mixture (3.3) to the FTSE is depicted; the fitted density is simulated with four million replications). We check in Fig. 2, that the general shape of the empirical density is well captured. Tail behaviour is analysed in Fig. 3. We note that the fit is smooth and yields thick tails, as it should. The second point to be noted, is that the Hermite polynomials are significant (see the coefficients δs; highly significant in some cases). Only three polynomials are required, and the fitted density is

<table>
<thead>
<tr>
<th>Table 2. Edgeworth–Sargan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>a₀</td>
</tr>
<tr>
<td>(3.6)</td>
</tr>
<tr>
<td>a₁</td>
</tr>
<tr>
<td>(7.1)</td>
</tr>
<tr>
<td>ρₙ</td>
</tr>
<tr>
<td>(3.7)</td>
</tr>
<tr>
<td>σₙ</td>
</tr>
<tr>
<td>(4.9)</td>
</tr>
<tr>
<td>δ₂</td>
</tr>
<tr>
<td>(4.5)</td>
</tr>
<tr>
<td>δ₄</td>
</tr>
<tr>
<td>(9.1)</td>
</tr>
<tr>
<td>δ₈</td>
</tr>
<tr>
<td>(3.7)</td>
</tr>
<tr>
<td>α₀</td>
</tr>
<tr>
<td>(5.9)</td>
</tr>
<tr>
<td>α₁</td>
</tr>
<tr>
<td>(107)</td>
</tr>
<tr>
<td>α₂</td>
</tr>
<tr>
<td>(6.8)</td>
</tr>
<tr>
<td>log LK</td>
</tr>
</tbody>
</table>

Cols. I, II results for the Dow Jones.
Cols. III, IV results for the FTSE.
symmetric (odd coefficients were not significant in any case). Finally, and as far as moment existence is concerned, we simply note that moments of all orders exist for this family of probability densities, as noted in Section 2.

The relative performance of the two type of densities is assessed next. We focus on the mixture type of density, since that density provides the best fit in all cases, as seen in the previous paragraphs. Results for the Dow are given in column II (Table 1 for the Student \( t \) and Table 2 for the ES). We immediately note that the value attained by the likelihood (log LK) is higher for the ES. This result is also obtained for the FTSE (column IV; Table 1 for the Student \( t \), and Table 2 for the ES). Regarding moment existence, we have seen that the fit provided at the tails by the ES looks adequate (Fig. 3). In fact, the fitted ES yield generally thicker tails than the Student \( t \) (additional figures are omitted to save space). Given that the ES has moments of all orders, this casts doubts on the limit imposed by the Student \( t \) (the highest integer part of \( v \) in Table 1).

A few final remarks concern the optimization procedure. All models presented converge without special difficulties, provided initial values are chosen carefully. As expected, however, when more parameters are involved, the starting point for the optimization algorithm requires a more careful choice of starting values. This is more relevant in the case of mixtures densities, where

**Fig. 2.** Fit of the ES density: UK stock market
preliminary estimates for some parameters can be obtained by fitting a nonlinear model to a smoothed histogram (see Silverman, 1986 for smoothing alternatives; details are available on request). Finally, a few extreme outliers, very far away from the main histogram of data, were deleted from the estimation sample.

Some summarizing comments are in order. First, the ES yields a good fit, and better than the Student $t$ (according to the value of the likelihood). Tail behaviour is adequately captured by the ES density, as well. This casts doubt on the limit on moment existence imposed by the fitted Student $t$. Nevertheless, this is obtained at the cost of some complex models, and the Student $t$ provides, generally, reasonable fits with fewer parameters than the ES. Finally, one expansion of the Student $t$ presented in this paper, is significant in some cases.

4. CONCLUSIONS

Estimating and testing the densities of econometric models is a major research topic in finance, that has been addressed using several different families of probability densities. One of them which has rarely been tested is the Edgeworth–Sargan family of probability densities. The main purpose of this
Testing densities with financial data

study has been to compare the empirical performance of this density with other densities, most notably the Student $t$ distribution.

The empirical results reported are based on daily observations for the UK and the US stock market (6590 observations spanning 25 years, 1971/1996). The main summarizing conclusions that can be drawn from these results are the following: (1) the general shape of the distribution is accurately fitted by a combination of a Normal density with low variance, and another density, either a Student $t$ or the ES, with thick tails and a much larger variance; the ES, nevertheless, provides a better fit, although at the cost of estimating a more complex density with more parameters; (2) the shape of the tails and their thickness is well captured, specially by the ES; as regards moment existence, no strong evidence against the existence of moments up to seventh order has been found; and (3) one of the expansions for the Student $t$ introduced in the paper is statistically significant in some cases; nevertheless, it does not improve the empirical fit dramatically.

On the basis of the results reported in this paper, it may be concluded that the choice of density depends on the purpose of the research: if all that is required is a density simple to implement and easy to compute, the choice should be the Student $t$ density; however, if better fits are required, the most complex choice, that is, the mixture of the ES and a Normal, should be adopted. Finally, it would be interesting to extend the research of this paper to other financial series and other data frequencies.

ACKNOWLEDGEMENTS

We thank the comments received at the ‘Forecasting Financial Markets’ conference (London, 1997), and those from an anonymous referee. Financial support from the Cicyt under project SEC98–1112 is acknowledged. We are solely responsible for any possible remaining errors.

REFERENCES


