Simulation and Estimation of the Meixner Distribution

Journal:	*Communications in Statistics - Simulation and Computation*
Manuscript ID:	LSSP-2008-0176
Manuscript Type:	Original Paper
Date Submitted by the Author:	25-Jul-2008
Complete List of Authors:	Grigoletto, Matteo; University of Padua, Department of Statistical Sciences Provasi, Corrado; University of Padua, Department of Statistical Sciences
Keywords:	Meixner distribution, Johnson translation system, Rejection method, Maximum likelihood, Bootstrap goodness of fit, APARCH model
Abstract:	The Meixner distribution is a special case of the generalized z-distributions. Its properties make it potentially very useful in modeling short-term financial returns. This article proposes an algorithm to simulate the Meixner distribution, and shows how to obtain maximum likelihood estimators of its parameters. A GARCH-type model is then assessed, assuming that the innovation distribution is a standardized Meixner. Goodness of fit properties are investigated for some real financial time series, using bootstrap tests based on the empirical process of the residuals.

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

grigoletto_provasi_revised_2.zip

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Simulation and Estimation of the Meixner Distribution

Matteo Grigoletto and Corrado Provasi

Department of Statistical Sciences, University of Padua, Italy

The Meixner distribution is a special case of the generalized z-distributions. Its properties make it potentially very useful in modeling short-term financial returns. This article proposes an algorithm to simulate the Meixner distribution, and shows how to obtain maximum likelihood estimators of its parameters. A GARCH-type model is then assessed, assuming that the innovation distribution is a standardized Meixner. Goodness of fit properties are investigated for some real financial time series, using bootstrap tests based on the empirical process of the residuals.

Keywords Johnson translation system; Rejection method; Maximum likelihood estimation; Bootstrap goodness of fit; APARCH model.

1.   Introduction

It is well known that financial returns measured on short time intervals, i.e. daily or weekly, are non-Gaussian. In fact, numerous empirical evidences show that their distribution has heavier than Gaussian tails, and is often skewed (see, e.g., Cont, 2001).

One of the most interesting distributions introduced in the literature to interpret financial returns is the normal inverse Gaussian distribution (NIG)

Address correspondence to Matteo Grigoletto, Department of Statistical Sciences, Via Cesare Battisti 241, 35121 Padova, Italy. Email: matteo.grigoletto@stat.unipd.it

1

(Barndorff-Nielsen, 1997), which is a subclass of the generalized hyperbolic distribution (Barndorff-Nielsen, 1977), and is characterized by important theoretical properties, as semi-heavy tails and infinite divisibility. For these reasons the NIG distribution has been often adopted in financial applications, both as conditional distribution in GARCH-type models (see, among others, Jensen and Lunde, 2001, and Forsberg and Bollerslev, 2002) and as marginal return distribution (e.g. Eberlein and Keller, 1995, and Lillestøl, 2000). It should be noted, however, that since the derivatives of the NIG log-likelihood involve the Bessel function, direct likelihood maximization is difficult. As a further complication, the kurtosis parameter is constrained to be larger than the absolute value of the skewness parameter. As a consequence, maximization algorithms suffer from problems like non-convergence and the need for good initial values. This is especially true when the NIG is used within GARCH-type models, and many parameters need to be estimated simultaneously. Karlis (2002) proposed, as a solution to these problems, an EM type algorithm for finding maximum likelihood estimators of the NIG parameters. The algorithm assures convergence, which however might be reached very slowly.

In the present work we will consider, as an alternative approach, the adoption of the Meixner distribution, which has properties similar to those of the NIG. In fact, it has semi-heavy tails and is infinitely divisible. The Meixner distribution was introduced by Schoutens (2002) in the context of stochastic volatility models, and is a subclass of the generalized z-distributions (Grigelionis, 2001). The derivatives of the Meixner log-likelihood are more analytically tractable than those of the NIG, and the Meixner joint parameter space is the Cartesian product of each parameter space; likelihood maximization is therefore much easier to perform. These properties make the Meixner distribution very useful both as conditional distribution in GARCH-type models, and as marginal return distribution.

The paper is organized as follows. In Section 2 the Meixner distribution is defined, while in Section 3 we propose a method to generate pseudo-random

2

samples from it. Section 4 exposes the details of moment and maximum likelihood estimation and investigates, by means of suitable bootstrap tests, goodness of fit to weekly returns of some financial indices. In Section 5 the MXN-APARCH model is introduced, assuming that the conditional distribution for an APARCH model is Meixner, and its goodness of fit to several series of daily returns is analyzed. Concluding remarks are discussed in Section 6.

2. The Meixner distribution

A random variable X is said to follow a Meixner distribution having parameters m, a, b, d, with $a > 0$, $-\pi < b < \pi$, $d > 0$, and $m \in \mathbb{R}$, in symbols $X \sim \mathrm{MXN}(m, a, b, d)$, if its density function, for $x \in \mathbb{R}$, is

$$f_{\mathrm{MXN}}(x; m, a, b, d) = a^{-1} f_{\mathrm{MXN}}\left(\frac{x - m}{a}; 0, 1, b, d\right),$$

where

$$f_{\mathrm{MXN}}(z; 0, 1, b, d) = \frac{(2\cos(b/2))^{2d}}{2\pi\,\Gamma(2d)}\,\mathrm{e}^{bz}\,|\Gamma(d + \imath z)|^2,$$

with $\imath = \sqrt{-1}$ and $\Gamma(\cdot)$ being the gamma function. It should be noted that m and a are location and scale parameters, while b and d decide the shape of the distribution. In other terms, if $X \sim \mathrm{MXN}(m, a, b, d)$, then $Z = (X - m)/a$ is its standardized form, with distribution $\mathrm{MXN}(0, 1, b, d)$.

The characteristic function of X is (Grigelionis, 2001)

$$\phi_{\mathrm{MXN}}(u) = \mathrm{e}^{\imath m u}\left(\frac{2\cos(b/2)}{\cosh\left(\frac{a\,u - ib}{2}\right)}\right)^{2d},$$

so that the Meixner distribution is clearly infinitely divisible. From this definition the mean, variance, skewness and kurtosis of a $\mathrm{MXN}(m, a, b, d)$ distribution

3

can also be immediately derived:

$$
\begin{aligned}
\mathrm{E}[X] &= {}_{\mathrm{MXN}}\mu = m + a\,d\tan(b/2)\,,\\[4pt]
\mathrm{Var}[X] &= {}_{\mathrm{MXN}}\sigma^2 = \frac{a^2\,d}{\cos b + 1},\\[4pt]
\mathrm{Skew}[X] &= {}_{\mathrm{MXN}}\kappa_1 = \sin b\,\sqrt{\frac{1}{d\,(\cos b + 1)}},\\[4pt]
\mathrm{Kurt}[X] &= {}_{\mathrm{MXN}}\kappa_2 = 3 - \frac{\cos b - 2}{d}.
\end{aligned}
\tag{1}
$$

Here, $\mathrm{Skew}[X] = \mathrm{E}[(X - \mathrm{E}[X])^3]/[\mathrm{Var}(X)]^{3/2}$ and $\mathrm{Kurt}[X] = \mathrm{E}[(X - \mathrm{E}[X])^4]/[\mathrm{Var}(X)]^2$. It follows that when $b = 0$ the distribution is symmetric; besides, its kurtosis is always larger than 3, the Gaussian kurtosis.

An important property of the Meixner distribution, that makes it potentially useful in financial applications, is that it has semi-heavy tails (Grigelionis, 2001). Formally, this implies that, for a $\mathrm{MXN}(m, a, b, d)$ distribution, we have

$$
f_{\mathrm{MXN}}(m, a, b, d) \sim C_-|x|^\rho\,\mathrm{e}^{-\sigma_-|x|} \quad \text{as} \quad x \to -\infty,
$$

$$
f_{\mathrm{MXN}}(m, a, b, d) \sim C_+|x|^\rho\,\mathrm{e}^{-\sigma_+|x|} \quad \text{as} \quad x \to +\infty,
$$

for some $\rho \in \mathbb{R}$ and $C_-, C_+, \sigma_-, \sigma_+ \geq 0$, with

$$
\rho = 2d - 1, \quad \sigma_- = \frac{\pi - b}{a} \quad \text{and} \quad \sigma_+ = \frac{\pi + b}{a}.
$$

3. Simulation

Generating random values from a Meixner distribution requires some effort, since the quantile function is not available in closed form and the inversion method is therefore difficult to implement.

As an alternative, since f_{MXN} is unimodal, in the present context we considered the rejection method which, in its most basic form, assumes the existence of a density g and the knowledge of a constant $c \geq 1$ such that (Devroye, 1986)

$$
f(x) \leq c\,g(x)\,, \quad \forall x\,.
\tag{2}
$$

4

Therefore, in order to apply the rejection algorithm, one needs to be able to determine a suitable constant c. Besides, it is necessary to specify and appropriate random variable Y, with density g and cdf G easily invertible. The density g has to dominate f_{MXN} for any value of its parameters (m, a, b, d). In this regard, we propose to approximate the density f_{MXN} with one of the distributions in the Johnson translation system (Johnson, 1949). The Johnson system is highly flexible and, through one of its four functional forms, is able to closely approximate many standard continuous distribution. Also, simulating from this distribution is relatively simple and fast (a random sample from a standardized normal distribution is all that is required). The latter point is especially important since, in order for the Meixner density to be dominated, the factor c needs to grow with the kurtosis. As a consequence, the number of rejections also increases and, if the algorithm to generate from the dominating density is not sufficiently fast, computation becomes slow. An inversion based method was also considered, but was found to be too computationally expensive.

For a generic random variable Y, the transformations that originate the Johnson system have the general form

$$Z = \gamma + \delta \, h\left(\frac{Y - \xi}{\lambda}\right), \tag{3}$$

where Z is a standard normal random variable, $\gamma \in \mathbb{R}$ and $\delta > 0$ are shape parameters, $\lambda > 0$ is a scale parameter and ξ is a location parameter. The nature of function h defines the four families of distributions in the Johnson translation system:

$$\begin{cases} \ln y, & \text{for the } S_L \text{ (lognormal) family,} \\ \sinh^{-1}(y), & \text{for the } S_U \text{ (unbounded) family,} \\ \ln(y/(1-y)), & \text{for the } S_B \text{ (bounded) family,} \\ y, & \text{for the } S_N \text{ (normal) family.} \end{cases}$$

When a parameter vector $(\xi, \lambda, \gamma, \delta)$ is available, computing quantiles or proba-

5

bilities is relatively easy, since these distributions are generated by an increasing transform of the Gaussian distribution and their cdf can be expressed in terms of Φ, the standard Gaussian cdf:

$$F(y; \xi, \lambda, \gamma, \delta) = \Phi\left(\gamma + \delta\, h\!\left(\frac{y - \xi}{\lambda}\right)\right).$$

As we have seen, the Meixner distribution is always characterized by a kurtosis greater than 3. As a consequence, an examination of the diagram by Johnson (1949) (but see also Johnson et al., 1994) shows that the distribution most apt to approximate the Meixner distribution is the S_U, with density function defined by

$$f_{S_U}(y; \xi, \lambda, \gamma, \delta) = \lambda^{-1} f_{S_U}\!\left(\frac{y - \xi}{\lambda}; 0, 1, \gamma, \delta\right),$$

where

$$f_{S_U}(u; 0, 1, \gamma, \delta) = \frac{\delta}{\sqrt{2\pi}\sqrt{u^2 + 1}}\, \exp\!\left(-\frac{1}{2}(\gamma + \delta \sinh^{-1}(u))^2\right), \quad u \in \mathbb{R}\ .$$

The moments of S_U can be obtained by the inverse transformation of (3), and are given by

$$E(Y^r) = \int_{-\infty}^{\infty} \left[\xi + \lambda\, \sinh\!\left(\frac{z - \gamma}{\delta}\right)\right]^r \phi(z)\, dz, \qquad r = 1, 2, \ldots,$$

where ϕ denotes the standard Gaussian density. It can be shown that these moments are available analytically; therefore, by the usual relations among the central moments and the moments about the origin, we have that the mean, the variance and the skewness and kurtosis indices for the S_U distribution are, respectively

$$\mathrm{E}[Y] = {}_{S_U}\mu = \xi - \lambda\, \omega^{1/2} \sinh\theta\ ,$$

$$\mathrm{Var}[Y] = {}_{S_U}\sigma^2 = \frac{\lambda^2}{2}\,(\omega - 1)\,(\omega\, \cosh(2\,\theta) + 1)\ ,$$

$$\mathrm{Skew}[Y] = {}_{S_U}\kappa_1 = -\frac{\omega^{1/2}\,(\omega - 1)^2\,[\omega\,(\omega + 2)\,\sinh(3\,\theta) + 3\,\sinh\theta]}{\sqrt{2}\,[(\omega - 1)\,(\omega\, \cosh(2\,\theta) + 1)]^{3/2}}\ , \qquad (4)$$

$$\mathrm{Kurt}[Y] = {}_{S_U}\kappa_2 = \{\omega^2\,[(\omega^4 + 2\,\omega^3 + 3\,\omega^2 - 3)\cosh(4\,\theta)$$
$$+ 4\,(\omega + 2)\cosh(2\,\theta)] + 3\,(2\,\omega + 1)\}/[2\,(\omega\, \cosh(2\,\theta) + 1)^2]\ ,$$

6

where $\omega = \exp(\delta^{-2})$ and $\theta = \gamma/\delta$. It should be noted that when $\gamma = 0$, $\gamma < 0$ or $\gamma > 0$, the S_U density function is symmetric, right-skewed and left-skewed, respectively.

In order to parametrize the distributions in the Johnson system, numerous moment or quantile-based procedures have been proposed (see, among others, Palmitesta and Provasi, 2000). Here we followed the procedure introduced by Hill et al. (1976), which implies equating the first four moments of the target distribution to those of one of the four distributions in the Johnson system, using the Johnson diagram. In the present case, for approximating the Meixner distribution with the S_U, this procedure amounts to equating $_{\text{MXN}}\mu$, $_{\text{MXN}}\sigma^2$, $_{\text{MXN}}\kappa_1$ and $_{\text{MXN}}\kappa_2$, obtained with a parameter vector (m, a, b, d), to $_{S_U}\mu$, $_{S_U}\sigma^2$, $_{S_U}\kappa_1$ and $_{S_U}\kappa_2$. Then, equations (4) can be solved in $(\xi, \lambda, \gamma, \delta)$.

The procedure can be applied in two steps: first, the values of γ and δ satisfying the last two equations in (4) are found numerically; then, conditionally on these values, the first two equations in (4) are explicitly solved for ξ and λ. Figure 1 shows the densities of some zero mean and unit variance Meixner and S_U distributions, for several values of the skewness and kurtosis coefficients. The last two equations in (4) have been solved employing the Newton algorithm.

Figure 1 clearly indicates that, especially when the kurtosis increases, the density of S_U does not dominate that of the Meixner distribution, especially around the mode. Then, in order to satisfy inequality (2), the constant c can be chosen as

$$c = \sup\left(\frac{f_{\text{MXN}}(x; m, a, b, d)}{f_{S_U}(x; \tilde{\xi}, \tilde{\lambda}, \tilde{\gamma}, \tilde{\delta})} \right), \tag{5}$$

where the parameter vector $(\tilde{\xi}, \tilde{\lambda}, \tilde{\gamma}, \tilde{\delta})$ satisfies the equations in (4), with mean, variance, skewness and kurtosis equal to those of a $\text{MXN}(m, a, b, d)$.

The value of c can be approximated graphically or, if an higher precision is desired, by using numerical algorithms.

To sum up, the procedure to obtain rejection based Monte Carlo samples from a Meixner distribution is composed of the following steps:
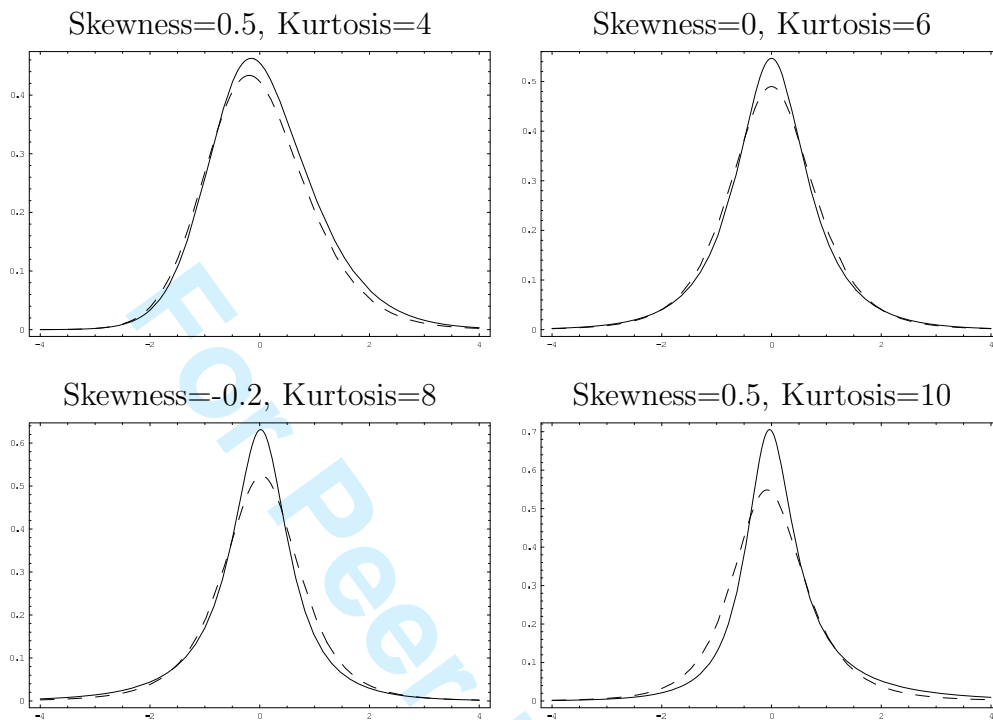
<div align="center">7</div>

Figure 1: Densities of some zero mean and unit variance Meixner (solid lines) and S_U (dashed lines) distributions.

1. Determine the values $\tilde{\xi}, \tilde{\lambda}, \tilde{\gamma}, \tilde{\delta}$ such that the first four moments of the S_U and Meixner distributions coincide.

2. Use (5) to compute c, graphically or with a numerical algorithm.

3. Generate u from a $U(0, 1)$ distribution.

4. Generate x_g from S_U using the transformation

$$x_g = \tilde{\xi} + \tilde{\lambda} \, \sinh((z - \tilde{\gamma})/\tilde{\delta}) \, ,$$

where z was generated from a $\mathcal{N}(0, 1)$ distribution.

5. If

$$u \leq \frac{1}{c} \frac{f_{\text{MXN}}(x_g; m, a, b, d)}{f_{S_U}(x_g; \tilde{\xi}, \tilde{\lambda}, \tilde{\gamma}, \tilde{\delta})} \, ,$$

accept x_g as a random value from $\text{MXN}(m, a, b, d)$; otherwise, return to step 3.

8

As an illustration, Table 1 displays the value of c, the number of rejections and the sample statistics for Monte Carlo samples of size $n = 10000$ from zero mean and unit variance Meixner distributions with several values of the skewness and kurtosis indices. Figure 2 shows the frequency plots comparing the pseudo-random samples and the true Meixner densities.

κ_1	κ_2	c	Rejects	Mean	Var	Skew	Kurt
-0.2	4.0	1.0932	894	-0.0031	1.0195	-0.2600	4.4613
0.0	4.0	1.0772	764	-0.0199	1.0021	0.0009	4.0287
0.5	4.0	1.5021	5021	0.0042	0.9952	0.5250	3.9047
-0.2	6.0	1.2270	2330	0.0050	0.9803	-0.2365	5.8214
0.0	6.0	1.2413	2372	-0.0023	0.9972	0.0368	6.3294
0.5	6.0	1.2048	2068	0.0034	1.0068	0.4831	6.4086
-0.2	8.0	1.3909	3938	-0.0017	1.0059	-0.2540	8.3552
0.0	8.0	1.3677	3739	-0.0043	0.9913	0.0187	7.5852
0.5	8.0	1.4480	4386	-0.0006	1.0015	0.5911	8.0002
-0.2	10.0	1.4959	5008	0.0014	1.0108	-0.3311	9.5562
0.0	10.0	1.4699	4623	-0.0145	0.9815	-0.2505	9.9530
0.5	10.0	1.5527	5573	-0.0001	1.0002	0.5078	10.2164

Table 1: c, number of rejections and sample statistics for Monte Carlo samples of size $n = 10000$ from zero mean and unit variance Meixner distributions with several values of the skewness and kurtosis indices κ_1 and κ_2.

4. Estimation methods

4.1 Moment estimation

The moments of the Meixner distribution, defined in (1), have a form that makes moment based estimation relatively simple. By equating the theoretical moments to their sample counterparts, and solving for the parameters, we
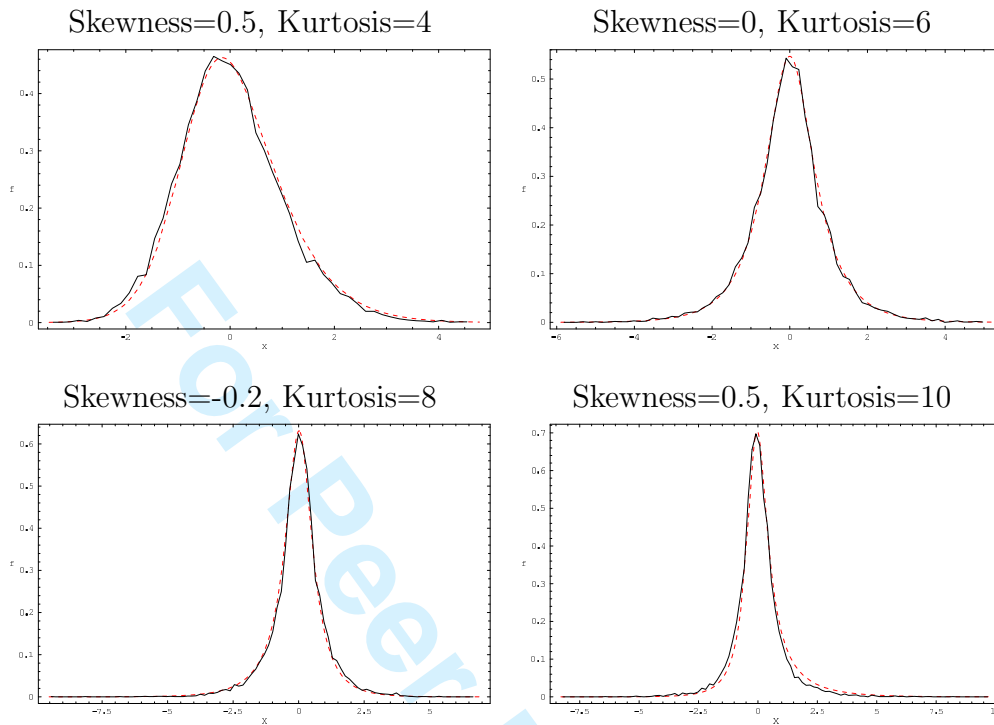
9

Figure 2: Frequency plots for random samples of size 10000 from a Meixner distribution with zero mean, unit variance and several values of the skewness and kurtosis indices. The dashed lines represent the true Meixner densities.

obtain

$$\bar{d} = \frac{1}{\bar{\kappa}_2 - \bar{\kappa}_1^2 - 3} \ ,$$

$$\bar{b} = \text{sign}(\bar{\kappa}_1) \, \cos^{-1}(2 - \bar{d} \, (\bar{\kappa}_2 - 3))$$

$$\bar{a} = s \, \sqrt{\frac{\cos \bar{b} + 1}{\bar{a}}} \ ,$$

$$\bar{m} = \bar{x} - \bar{a} \, \bar{d} \, \tan\big(\bar{b}/2\big) \ ,$$

where $\bar{x}$ and s^2 represent, as usual, the sample mean and variance, respectively, while $\bar{\kappa}_1 = \bar{\mu}_3/\bar{\mu}_2^{3/2}$ and $\bar{\kappa}_2 = \bar{\mu}_4/\bar{\mu}_2^2$, with $\bar{\mu}_k = n^{-1} \sum_{i=1}^{n}(x_i - \bar{x})^k$, are the sample skewness and kurtosis indices. It should be noted that the method of moment estimates do not exist when $\bar{\kappa}_2 < 2\bar{\kappa}_1^2 + 3$.

10

4.2 Maximum likelihood estimation

Suppose that the n observations $x_1, \ldots, x_n$ are a random sample from a Meixner distribution with unknown parameters $\theta = (m, a, b, d)$. The average log-likelihood function is defined as

$$
\ell_n(\theta) = 2\, d\, \log(2 \cos(b/2)) - \log(2\, a\, \pi) - \log(\Gamma(2\, d)) + b\, \bar{z}
$$
$$
+ \frac{1}{n} \sum_{j=1}^{n} \log |\Gamma(d + \imath\, z_j)|^2, \tag{6}
$$

where $z_j = (x_j - m)/a$ and $\bar{z} = \sum_{j=1}^{n} z_i/n$. Then, the MLE of θ, indicated by $\hat{\theta}_{\mathrm{ML}}$, is the solution of the following optimization problem:

$$
\hat{\theta}_{\mathrm{ML}} = \arg \max_{\theta \in \Theta} \ell_n(\theta),
$$

where Θ is the parameter space for θ.

In general it is known that, if θ_0 is the true and unknown parameter vector and under certain regularity conditions (Lehmann and Casella, 1998), the ML estimator $\hat{\theta}_{\mathrm{ML}}$ satisfies

$$
\sqrt{n}\, (\hat{\theta}_{\mathrm{ML}} - \theta_0) \to N(0, I_n(\theta_0)^{-1}) \tag{7}
$$

in distribution as $n \to \infty$, where $I_n(\theta)$ is the information matrix defined as

$$
I_n(\theta) = -E\left[\frac{\partial^2 \ell_n(\theta)}{\partial \theta \partial \theta'} \right] = E\left[\left(\frac{\partial \ell_n(\theta)}{\partial \theta} \right) \left(\frac{\partial \ell_n(\theta)}{\partial \theta} \right)' \right].
$$

In practical situations, we can use

$$
\hat{I}_n(\hat{\theta}_{\mathrm{ML}}) = -\frac{\partial^2 \ell_n(\theta)}{\partial \theta \partial \theta'} \bigg|_{\theta = \hat{\theta}_{\mathrm{ML}}}
$$

or

$$
\hat{I}_n(\hat{\theta}_{\mathrm{ML}}) = \left(\frac{\partial \ell_n(\theta)}{\partial \theta} \right) \left(\frac{\partial \ell_n(\theta)}{\partial \theta} \right)' \bigg|_{\theta = \hat{\theta}_{\mathrm{ML}}}
$$

to obtain an estimate of $I_n(\theta_0)$.

For the Meixner distribution it is possible to compute the matrix $\hat{I}_n(\hat{\theta}_{\mathrm{ML}})$ with high accuracy, since the expressions defining the first and second derivatives of the log-likelihood functions are explicitly available: see Appendix A.

11

These expressions allow to maximize very efficiently the log-likelihood function, using Newton-type algorithms relying on both the first and second derivatives. The method of moments parameter estimates can be used as starting points.

The regularity conditions underlying (7) appear very difficult to prove analytically in the present framework. For this reason we performed a Monte Carlo experiment in order to check the convergence of maximum likelihood estimators to the multivariate normal distribution. In particular, we assessed normality using the multivariate skewness and kurtosis measures by Mardia (1970). Table 2 shows the results obtained with 2000 replications and sample sizes equal to 250, 500, 1000 and 2000. The samples are extracted from a zero mean and unit variance Meixner distribution, characterized by several values of the skewness and kurtosis indexes (κ_1 and κ_2, respectively). The multivariate normality was assessed for the maximum likelihood estimates, standardized according to (7), with the information matrix estimated by $\hat{I}_n(\hat{\theta}_{\mathrm{ML}})$.

It should be noted that, apart from some inconsistency due to sampling errors, convergence of the multivariate skewness and kurtosis indexes to the values characterizing the multivariate normal (i.e. 0 and 24, respectively) is very slow. The speed of convergence is, in particular, inversely related to the kurtosis parameter d. These results suggest that inferential procedures based on the asymptotic properties of the maximum likelihood estimators, when the Meixner distribution is involved, must be carefully evaluated, even when medium-sized samples are considered. The parametric bootstrap can be used for this assessment (see e.g. Shao and Tu, 1996).

It is interesting to remark that the log-likelihood optimization can be simplified on the grounds of the following lemma.

Lemma. The ML estimate of the b parameter characterizing the Meixner distribution can be expressed as a function of the sample mean and of the ML estimates for the remaining parameters:

$$\hat{b} = 2 \tan^{-1}\left(\bar{z}/\hat{d}\right).$$

12

κ_1	κ_2	MultiSkewness				MultiKurtosis			
		250	500	1000	2000	250	500	1000	2000
-0.2	4.0	13.20	9.67	7.67	7.68	44.54	36.53	33.57	35.48
0.0	4.0	11.28	10.47	7.93	6.96	40.66	39.90	35.12	33.77
0.5	4.0	13.21	15.98	12.79	9.33	53.21	49.98	43.54	36.81
-0.2	6.0	8.83	5.90	4.63	2.31	37.49	31.39	32.63	27.22
0.0	6.0	6.56	5.81	3.85	2.03	31.68	32.23	30.07	26.41
0.5	6.0	7.56	6.68	3.32	2.16	33.99	36.05	28.06	27.14
-0.2	8.0	8.11	4.74	2.91	1.48	35.58	31.20	29.17	26.24
0.0	8.0	6.48	5.55	2.91	1.92	32.14	33.55	29.30	27.26
0.5	8.0	6.92	4.27	2.69	1.79	33.42	30.38	27.74	26.21
-0.2	10.0	7.54	4.34	2.92	1.51	35.67	30.40	29.17	25.93
0.0	10.0	7.75	4.26	2.99	1.66	36.26	30.33	28.27	25.89
0.5	10.0	5.82	3.88	3.27	1.53	32.43	28.24	31.17	25.88

Table 2: Multivariate skewness and kurtosis measures for the maximum likelihood estimators. The results concern 2000 Monte Carlo replications and samples (with sizes equal to 250, 500, 1000 and 2000) from zero mean and unit variance Meixner distributions, characterized by several values of the skewness and kurtosis indexes (κ_1 and κ_2, respectively).

Proof. As shown in Appendix A, the derivative of the average log-likelihood function with respect to b is

$$\frac{\partial l_n(\theta)}{\partial b} = \bar{z} - d \tan(b/2) \ .$$

Equating this expression to zero, and solving for b, the lemma is proved.

4.3 Goodness of fit

We are going to assess the ability of the MXN distribution to represent financial data. In particular, misspecification tests based on empirical processes will

13

be considered. For iid observations $x_1, \ldots, x_n$ with cumulative distribution function F_θ, characterized by parameters θ, the empirical process is defined as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{I[x_i \le x] - F_\theta(x)\} \ . \tag{8}$$

For a survey of the contributions of empirical and quantile processes, for iid observations, to the asymptotic theory of goodness-of-fit tests, see del Barrio (2004) and del Barrio et al. (2000).

We will consider a bootstrap implementation of the goodness of fit assessment, which will be based on the estimated empirical process

$$\hat{\nu}_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{I[x_i \le x] - F_{\hat{\theta}}(x)\} \ ,$$

where $\hat{\theta}$ is assumed to be a consistent estimate of θ. Therefore, rather than that of a single MXN distribution, we mean to test the appropriateness of the MXN family as a whole.

The process $\hat{\nu}_n(x)$ can be used for the empirical analysis of several functionals as, for instance, the

- Kolmogorov-Smirnov statistic (KS):

$$T_{\text{KS}}(\hat{\nu}_n) = \max_{1 \le i \le n} |\hat{\nu}_n(x_i)| \ ,$$

  or the

- Cramér-von Mises statistic (CV):

$$T_{\text{CV}}(\hat{\nu}_n) = \int [\hat{\nu}_n(x)]^2 \ dx \ .$$

Le us suppose that $x_1^*, \ldots, x_n^*$ are generated independently from an estimate $\hat{F}_\theta$ of F_θ as, for instance, the empirical distribution function of the observations, or the estimate $F_{\hat{\theta}}$ found by substituting an estimate $\hat{\theta}$ in the definition of F_θ (parametric bootstrap).

14

Then, a new estimate $\hat{\theta}^*$ for θ can be computed, with the same method used to obtain $\hat{\theta}$. The corresponding estimated bootstrap empirical process is defined as

$$\hat{\nu}_n^*(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ I[x_i^* \leq x] - F_{\hat{\theta}^*}(x) \right\} .$$

This process can be used for computing the bootstrap functionals $T_{\mathrm{KS}}(\hat{\nu}_n^*)$ and $T_{\mathrm{CV}}(\hat{\nu}_n^*)$. By repeating the procedure B times, the quantiles of the empirical distribution of the B bootstrap functionals can be used to define suitable critical regions and p-values.

Bootstrap based goodness of fit tests in the presence of estimated parameters have been studied e.g. in Stute et al. (1993), Babu and Rao (2004), Genz and Haeusler (2006) and Szũcs (2008). These results suggest that the present bootstrap resampling scheme is consistent, although we do not attempt a formal proof here.

The bootstrap procedure has been applied to the weekly returns for five international stock indices, in a five year period from January 2002 to December 2006. The percentage returns are defined as $y_t = 100 \left(\log p_t - \log p_{t-1} \right)$, where p_t is the weekly price. The indices are the AEX, Dow Jones, CAC 40, DAX and S&P 500. The goodness of fit results are summarized in Table 3 and show that the Meixner family is well suited for representing the considered data sets.

5. MXN-APARCH Model

The most elaborate GARCH-type model suggested so far seems to be the model presented in Hentschel (1995) or Duan (1997). The conditional heteroskedasticity of a time series y_t, $t = 1, \ldots, T$, is assumed to be generated by the following equations:

$$\begin{aligned} y_t &= \mu_t(\vartheta) + \varepsilon_t , \\ \varepsilon_t &= \sqrt{h_t(\vartheta)} \, z_t , \end{aligned} \tag{9}$$

Source: http://finance.yahoo.com

15

where the $\{z_t\}$ are iid random variables, having zero mean, unit variance and density g_η, with derivative g'_η and cdf G_η, that might depend on unknown shape parameters, henceforth denoted by η. Expressions $\mu_t \equiv \mu_t(\vartheta)$ and $h_t \equiv h_t(\vartheta)$ indicate the mean and variance of y_t, respectively, conditional on the information Φ_{t-1} available at time $t-1$. By $\vartheta \in \mathbb{R}^q$ we represented the vector containing all the parameters in equations (9). It should be noted that this framework includes many GARCH-type models introduced in the literature. Besides, the conditional mean μ_t can depend on exogenous variables, or be a nonlinear function of past observations.

The average log-likelihood function generated by equations (9) can be written as

$$\ell_T(\theta) = -\frac{1}{2\,T}\sum_{t=1}^{T}\log h_t + \frac{1}{T}\sum_{t=1}^{T}\log g(z_t;\eta)\ , \tag{10}$$

where $\theta = (\vartheta, \eta)$, while $z_t = (y_t - \mu_t)/\sqrt{h_t}$ denotes the residuals.

ML estimates can be computed maximizing (10) with respect to the vector of parameters θ. In particular, the gradient of the average log-likelihood function is

$$\frac{\partial \ell_T(\theta)}{\partial \theta} = \begin{cases} \dfrac{\partial \ell_T(\theta)}{\partial \vartheta} = -\dfrac{1}{T}\sum_{t=1}^{T}\left\{\dfrac{1}{\sqrt{h_t}}\dfrac{\partial \mu_t}{\partial \vartheta}\dfrac{g'(z_t;\eta)}{g(z_t;\eta)} + \dfrac{1}{2\,h_t}\dfrac{\partial h_t}{\partial \vartheta}\left(1 + z_t \dfrac{g'(z_t;\eta)}{g(z_t;\eta)}\right)\right\}, \\[2ex] \dfrac{\partial \ell_T(\theta)}{\partial \eta} = \dfrac{1}{T}\sum_{t=1}^{T}\dfrac{1}{g(z_t;\eta)}\dfrac{\partial\, g(z_t;\eta)}{\partial \eta}. \end{cases}$$

In the present context we are going to consider a GARCH-type model very widespread in applications: the $AR(r)$-$APARCH(p,q)$ model (Ding et al., 1993; see also He and Teräsvirta, 1999), further assuming that the data generating process is ruled by the MXN distribution, transformed in order to have zero mean and unit variance. In other terms, the density of z_t is

$$g_{\mathrm{MXN}}(z_t;\eta) = \frac{_{\mathrm{MXN}}\sigma\left(2\cos\left(\frac{b}{2}\right)\right)^{2\,d}}{2\,\pi\,\Gamma(2d)}\exp(b\,(_{\mathrm{MXN}}\mu +_{\mathrm{MXN}}\sigma\,z_t))$$
$$\cdot\,|\Gamma(d + \imath\,(_{\mathrm{MXN}}\mu + _{\mathrm{MXN}}\sigma\,z_t))|^2\,,$$

with $\eta = (b, d)$, while the other parameters are as defined previously.

16

The AR(r)-APARCH(p,q) specifies the conditional means in (9), for $t = 1, \ldots, T$, as

$$\mu_t = \nu + \sum_{i=1}^{r} \phi_i \, y_{t-i} \ ,$$

while conditional variances are determined by

$$h_t^{\delta/2} = \omega + \sum_{i=1}^{p} \alpha_i \left(|\varepsilon_{t-i}| - \gamma_i \, \varepsilon_{t-i}\right)^{\delta} + \sum_{j=1}^{q} \beta_j \, h_{t-j}^{\delta/2} \ , \tag{11}$$

where $\mu \in \mathbb{R}$, $-1 < \phi_i < 1$ $(i = 1, \ldots, r)$, $\omega > 0$, $\alpha_i \geq 0$ $(i = 1, \ldots, p)$, $\beta_j \geq 0$ $(j = 1, \ldots, q)$, $\delta \geq 0$ and $-1 < \gamma_i < 1$. The most relevant characteristics of the model are the presence of a Box-Cox power transformation of the conditional variances and the asymmetric absolute errors. The APARCH model includes as special cases seven other ARCH type models, including the ARCH by Engle (1982), the GARCH by Bollerslev (1986), the NARCH by Higgins and Bera (1992), the log-ARCH by Geweke (1986) and Pentula (1986), and simple asymmetric and threshold GARCH models.

In financial applications the conditional variance h_t of a time series is often assumed to follow an APARCH(1,1) model. In Appendix B we compute the derivatives necessary for maximizing the average log-likelihood when this model is adopted.

5.1 Goodness of fit

Our objective here is to validate the AR-APARCH model, with MXN distribution for the innovations. We will use an approach analogous to that followed in Section 4.3, but applied to the squared residuals.

Horváth et al. (2001) studied the weak convergence of the empirical process of squared residuals of ARCH sequences, defined as

$$\zeta_T(x) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ I[\hat{z}_t^2 \leq x] - Q_\eta(x) \right\} \ , \tag{12}$$

where Q_η is the cumulative distribution function of the squared disturbances. Here we are assuming that a consistent estimate $\hat{\vartheta}$ of ϑ is available, so that the

17

squared residuals for model (9) can be computed by

$$\hat{z}_t^2 = \frac{(y_t - \mu_t(\hat{\vartheta}))^2}{h_t(\hat{\vartheta})}, \quad t = 1, \ldots, T .$$

Horváth et al. (2001) assumed h_t to have an ARCH structure. Extensions to GARCH(p, q) processes and more general ARCH processes are available, respectively, in Berkes and Horváth (2003) and Koul (2002). In any case, the asymptotic distribution of the residual empirical process for GARCH models depends in a complicated way on the model parameters and on the distribution of the innovations. This implies that classical goodness-of-fit tests, like the Cramér-von Mises and Kolmogorov-Smirnov tests, are not readily applicable. However, the existence of a limiting distribution suggests that bootstrap tests based on the empirical process can be expected to detect departures from the postulated distribution of the unobservable innovations (Horváth et al., 2004).

We will consider the estimated empirical process

$$\hat{\zeta}_T(x) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ I[\hat{z}_t^2 \leq x] - Q_{\hat{\eta}}(x) \right\},$$

where the parameter vector η has also been consistently estimated.

This approach is limited in that the quadratic residuals $\{\hat{z}_t^2\}$ are not independent nor identically distributed, and therefore the usual inferential procedures cannot be applied. For this reason, here bootstrap goodness-of-fit procedures for dependent data are considered, as documented, e.g., in Lahiri (2003). More specifically, assume that a sequence of observations $\{y_t\}_{t=1}^{T}$ is available from a process of the kind described in (9). Then, an estimate $\hat{\theta} = (\hat{\vartheta}, \hat{\eta})'$ can be obtained. The replicates of the observed series are generated from

$$y_t^* = \mu_t(\hat{\vartheta}) + \sqrt{h_t(\hat{\vartheta})} \, z_t^*, \tag{13}$$

where the $\{z_t^*\}$ are iid from an estimate $\hat{G}_\eta$ of G_η.

Having generated a bootstrap replicate $\{y_t^*\}_{t=1}^{T}$ from process (9), a new estimate $\hat{\theta}^* = (\hat{\vartheta}^*, \hat{\eta}^*)'$ can be computed, with the same method used to obtain

18

$\hat{\theta}$. As a consequence, the bootstrap residuals

$$\hat{z}_t^{*2} = \frac{(y_t^* - \mu_t(\hat{\vartheta}^*))^2}{h_t(\hat{\vartheta}^*)}, \quad t = 1, \ldots, T \ ,$$

become available. The corresponding estimated bootstrap empirical process is defined as

$$\hat{\zeta}_T^*(x) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ I[\hat{z}_t^{*2} \le x] - Q_{\hat{\eta}^*}(x) \right\} \ .$$

By obtaining B bootstrap replicates $T_{\mathrm{KS}}(\hat{\zeta}_T^*)$ and $T_{\mathrm{CV}}(\hat{\zeta}_T^*)$, critical regions and p-values for the goodness of fit tests can be computed.

The available contributions on the consistency of bootstrap based goodness of fit tests in the presence of estimated parameters assume the observations to be iid (see e.g. Szũcs, 2008). As a consequence, these results do not apply to the present case, where tests concern residuals. The theoretical asymptotic behaviour of bootstrap tests, in this framework, has yet to be studied. However, Horváth et al. (2004) and Grigoletto and Provasi (2008), on the grounds of theoretical results concerning the asymptotic distribution of the relevant empirical processes, applied such tests, finding them to have an overall correct size.

The procedure described above has been applied to the same series of stock indices used in Section 4.3, but considering daily (instead of weekly) returns. The results of the goodness of fit assessment are shown in Table 4. The p-values of the Kolmogorov-Smirnov and Cramér von-Mises tests indicate good agreement of the Meixner APARCH(1,1) model with the data sets analyzed.

6.   Final remarks

The theoretical properties of the Meixner distribution, as semi-heavy tails and infinite divisibility, make it potentially very effective in modeling short-term financial returns. However, the usefulness of a certain distribution depends also on the availability of algorithms to simulate from it, and to estimate its parameters. This article proposes a rejection-based method to generate random

19

values from the Meixner, and shows how maximum likelihood estimators can be obtained. The convergence of these estimators to their asymptotic multivariate normal distribution is assessed through a Monte Carlo study.

In order to study the applicability of the Meixner distribution, a GARCH-type model is then analyzed, assuming a standardized Meixner distribution for the innovations. Some real financial time series are considered to investigate goodness of fit, through bootstrap tests based on the empirical process of the residuals. The results of these tests, together with good computational properties, suggest that the Meixner distribution can be considered a good candidate for representing series of financial returns.

Appendix A

The average log-likelihood function introduced in equation (6) can also be written as

$$\ell_n(\theta) = 2\,d\,\log(2\,\cos(b/2)) - \log(2\,a\,\pi) - \log(\Gamma(2\,d)) + b\,\bar{z}$$

$$+ \frac{1}{n}\sum_{j=1}^{n}(\log(\Gamma(d + \imath\,z_j)) + \log(\Gamma(d - \imath\,z_j))),$$

since $|\Gamma(d + \imath\,z_j)|^2 = \Gamma(d + \imath\,z_j)\Gamma(d - \imath\,z_j)$; hence, the gradient of $\ell_n(\theta)$ is defined as

$$\frac{\partial l_n(\theta)}{\partial m} = -\frac{1}{a}\left(b + \frac{\imath}{n}\sum_{j=1}^{n}\Psi_j^-\right),$$

$$\frac{\partial l_n(\theta)}{\partial a} = -\frac{1}{a}\left(b\,\bar{z} + 1 + \frac{\imath}{n}\sum_{j=1}^{n}z_j\,\Psi_j^-\right),$$

$$\frac{\partial l_n(\theta)}{\partial b} = \bar{z} - d\,\tan(b/2),$$

$$\frac{\partial l_n(\theta)}{\partial d} = 2\,(\log(2\,\cos(b/2)) - \psi(2\,d)) + \frac{1}{n}\sum_{j=1}^{n}\Psi_j^+.$$

On the other hand, the diagonal elements of the Hessian matrix are

$$\frac{\partial^2 l_n(\theta)}{\partial m^2} = -\frac{1}{a^2\,n}\sum_{j=1}^{n}\Upsilon_j^+,$$

$$\frac{\partial^2 l_n(\theta)}{\partial a^2} = \frac{1}{a^2}\left(2\,b\,\bar{z} + 1 + \frac{1}{n}\sum_{j=1}^{n}(2\,\imath\,z_j\,\Psi_j^- - z_j^2\,\Upsilon_j^+)\right),$$

$$\frac{\partial^2 l_n(\theta)}{\partial b^2} = -\frac{d}{1 + \cos(b)},$$

$$\frac{\partial^2 l_n(\theta)}{\partial d^2} = -4\psi'(2\,d) + \frac{1}{n}\sum_{j=1}^{n}\Upsilon_j^+,$$

while the off-diagonal components of the Hessian matrix are defined as

$$\frac{\partial^2 l_n(\theta)}{\partial m\,\partial a} = \frac{1}{a^2}\left(b + \frac{1}{n}\sum_{j=1}^{n}(\imath\,\Psi_j^- - z_j\,\Upsilon_j^+)\right),$$

$$\frac{\partial^2 l_n(\theta)}{\partial m\,\partial b} = -\frac{1}{a},$$

21

$$\frac{\partial^2 l_n(\theta)}{\partial m\,\partial d} = -\frac{\imath}{a\,n}\sum_{j=1}^{n}\Upsilon_j^{-},$$

$$\frac{\partial^2 l_n(\theta)}{\partial a\,\partial b} = -\frac{\bar{z}}{a},$$

$$\frac{\partial^2 l_n(\theta)}{\partial a\,\partial d} = -\frac{\imath}{a\,n}\sum_{j=1}^{n} z_j\,\Upsilon_j^{-},$$

$$\frac{\partial^2 l_n(\theta)}{\partial b\,\partial d} = -\tan(b/2)\,,$$

where $z_j = (x_j - m)/a$, $\bar{z} = \sum_{j=1}^{n} z_j/n$,

$$\Psi_j^{+} = \psi(d + \imath\,z_j) + \psi(d - \imath\,z_j),$$

$$\Psi_j^{-} = \psi(d + \imath\,z_j) - \psi(d - \imath\,z_j),$$

$$\Upsilon_j^{+} = \psi'(d + \imath\,z_j) + \psi'(d - \imath\,z_j),$$

$$\Upsilon_j^{-} = \psi'(d + \imath\,z_j) - \psi'(d - \imath\,z_j),$$

and $\psi(\cdot)$ is the psi function.

Appendix B

When an AR(r)-APARCH(1,1) model is adopted for defining the relations in (9), the derivative of μ_t with respect to ν is 1, while the derivative with respect to ϕ_i is y_{t-i}. Therefore, from equation (11), we have that (Laurent, 2004)

$$\frac{\partial h_t}{\partial(\phi,\omega,\alpha,\gamma,\beta)} = \frac{2h_t}{\delta h_t^{\delta/2}}\,\frac{\partial h_t^{\delta/2}}{\partial(\phi,\omega,\alpha,\gamma,\beta)}$$

and

$$\frac{\partial h_t}{\partial\delta} = \frac{2h_t}{\delta h_t^{\delta/2}}\left(\frac{\partial h_t^{\delta/2}}{\partial\delta} - \frac{h_t^{\delta/2}\log(h_t^{\delta/2})}{\delta}\right).$$

If the sequence of the variances $h_1,\ldots,h_T$ is initialized with $h_1 = \omega^{2/\delta}$, the partial derivatives of the generic term h_t with respect to the parameters of the APARCH(1,1) model defined in equation (11) can be obtained in a recursive

22

form since, for $t > 1$,

$$\frac{\partial h_t^{\delta/2}}{\partial \nu} = \alpha_1 \, \delta \, (|\varepsilon_{t-1}| - \gamma_1 \varepsilon_{t-1})^{\delta-1}(-\mathrm{sign}(\varepsilon_{t-1}) + \gamma_1) + \beta_1 \frac{\partial h_{t-1}^{\delta/2}}{\partial \mu} \; ,$$

$$\frac{\partial h_t^{\delta/2}}{\partial \phi_j} = \alpha_1 \, \delta y_{t-j}(|\varepsilon_{t-1}| - \gamma_1 \varepsilon_{t-1})^{\delta-1}(-\mathrm{sign}(\varepsilon_{t-1}) + \gamma_1) + \beta_1 \frac{\partial h_{t-1}^{\delta/2}}{\partial \phi_j} \; ,$$

$$j = 1, \ldots, r \; ,$$

$$\frac{\partial h_t^{\delta/2}}{\partial \omega} = 1 + \beta_1 \frac{\partial h_{t-1}^{\delta/2}}{\partial \omega} \; ,$$

$$\frac{\partial h_t^{\delta/2}}{\partial \alpha_1} = (|\varepsilon_{t-1}| - \gamma_1 \varepsilon_{t-1})^{\delta} + \beta_1 \frac{\partial h_{t-1}^{\delta/2}}{\partial \alpha_1} \; ,$$

$$\frac{\partial h_t^{\delta/2}}{\partial \gamma_1} = -\alpha_1 \, \delta \, \varepsilon_{t-1} \, (|\varepsilon_{t-1}| - \gamma_1 \varepsilon_{t-1})^{\delta-1} + \beta_1 \frac{\partial h_{t-1}^{\delta/2}}{\partial \gamma_1} \; ,$$

$$\frac{\partial h_t^{\delta/2}}{\partial \beta_1} = h_{t-1}^{\delta/2} + \beta_1 \frac{\partial h_{t-1}^{\delta/2}}{\partial \beta_1} \; ,$$

$$\frac{\partial h_t^{\delta/2}}{\partial \delta} = \alpha_1 \, (|\varepsilon_{t-1}| - \gamma_1 \varepsilon_{t-1})^{\delta} \log(|\varepsilon_{t-1}| + \gamma_1 \varepsilon_{t-1}) + \beta_1 \frac{\partial h_{t-1}^{\delta/2}}{\partial \delta} \; ,$$

where $\varepsilon_{t-1} = y_{t-1} - \nu - \sum_{i=1}^r \phi_i \, y_{t-i-1}$. Obviously, on the basis of the formulations of the APARCH(1,1) model, we must observe that the differentiation of the absolute value of a variable is not defined in zero. Nevertheless, as also observed by Laurent (2004), even if this situation is possible, it is most unlikely in practice.

We also have that

$$\frac{g'_{\mathrm{MXN}}(z_t : \eta)}{g_{\mathrm{MXN}}(z_t; \eta)} = {}_{\mathrm{MXN}}\sigma \, (b + \imath \, (\psi(d + \imath \, w_t) - \psi(d - \imath \, w_t))).$$

The logarithmic derivatives of g_{MXN} with respect to b and d are given by

$$\frac{\partial \log g_{\mathrm{MXN}}(z_t; \eta)}{\partial b} = \frac{{}_{\mathrm{MXN}}\sigma^{(1,0)}}{{}_{\mathrm{MXN}}\sigma} - d \tan(b/2) + w_t$$

$$+ w_t^{(1,0)}(b + \imath(\psi(d + \imath \, w_t) - \psi(d - \imath \, w_t))),$$

$$\frac{\partial \log g_{\mathrm{MXN}}(z_t; \eta)}{\partial d} = \frac{{}_{\mathrm{MXN}}\sigma^{(0,1)}}{{}_{\mathrm{MXN}}\sigma} + 2 \, (\log(2 \cos(b/2)) - \psi(2 \, d))$$

$$+ \psi(d + \imath \, w_t) + \psi(d - \imath \, w_t)$$

$$+ w_t^{(0,1)}(b + \imath(\psi(d + \imath \, w_t) - \psi(d - \imath \, w_t))),$$

23

where $w_t = \mu_t + \sigma_t z_t$, $w_t^{(1,0)} = \mu_t^{(1,0)} + \sigma_t^{(1,0)} z_t$ and $w_t^{(0,1)} = \mu_t^{(0,1)} + \sigma_t^{(0,1)} z_t$, being

$$\mu_t^{(1,0)} = \frac{\partial_{\text{MXN}} \mu}{\partial b} = \frac{d}{2 \cos(b/2)^2} \ ,$$

$$\sigma_t^{(1,0)} = \frac{\partial_{\text{MXN}} \sigma}{\partial b} = \frac{d \sin(b)}{2 \,_{\text{MXN}} \sigma (1 + \cos b)^2},$$

and

$$\mu_t^{(0,1)} = \frac{\partial_{\text{MXN}} \mu}{\partial d} = \tan(b/2) \ ,$$

$$\sigma_t^{(0,1)} = \frac{\partial_{\text{MXN}} \sigma}{\partial d} = \frac{1}{1 + \cos b} \ ,$$

with the notation defined above.

References

Babu, G.J., Rao, C.R. (2004). Goodness-of-fit tests when parameters are estimated. *Sankhyā* 66:63–74.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London, Series A* 353:401–419.

Barndorff-Nielsen, O. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* 24:1–13.

Berkes, I., Horváth, L. (2003). Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Processes and their Applications* 105:279–298.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307–327.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1:223–236.

del Barrio, E. (2004). *Empirical and Quantile Processes in the Asymptotic Theory of Goodness-of-Fit Tests.* Lecture Notes presented at the European Mathematical Society Summer School on Theory and Statistical Applications of Empirical Processes. Laredo, Spain.

del Barrio, E., Cuesta-Albertos, J., Matrán, C. (2000). Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. *Test* 9:1–96.

Devroye, L. (1986). *Non-Uniform Random Variate Generation.* New York: Springer-Verlag Inc. Available at `http://cg.scs.carleton.ca/~luc/rnbookindex.html`.

Ding, Z., Granger, C., Engle, R.F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1:83–106.

Duan, J.C. (1997). Augmented GARCH(p,q) process and its diffusion limit. *Journal of Econometrics* 79:97–127.

Eberlein, E., Keller, U. (1995). Hyperbolic distributions in Finance. *Bernoulli* 1:281–299.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987–1007.

Forsberg, L., Bollerslev, T. (2002). Bridging the gap between the distribution of realized (ecu) volatility and ARCH modelling: the GARCH-NIG model. *Journal of Applied Econometrics* 17:535–548.

Genz, M., Haeusler, E. (2006). Empirical processes with estimated parameters under auxiliary information. *Journal of Computational and Applied Mathematics* 186:191–216.

25

Geweke, J. (1986). Comments on "Modelling the persistence of conditional variances". *Econometric Reviews* 5:57–61.

Grigelionis, N. (2001). Generalized z-distributions and related stochastic processes. *Lithuanian Mathematical Journal* 41:239–251.

Grigoletto, M., Provasi, C. (2008). Misspecification testing for the conditional distribution model in GARCH-type processes. *Econometric Reviews.* To appear.

He, C., Teräsvirta, T. (1999). Statistical properties of the asymmetric power ARCH process. In: Engle, R.F., White, H., eds., *Cointegration, causality and forecasting. Festschrift in Honour of Clive W.J. Granger.* Oxford: Oxford University Press, pp. 462–474.

Hentschel, L. (1995). All in the family: nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics* 39:71–104.

Higgins, M.L., Bera, A.K. (1992). A class of nonlinear ARCH models. *International Economic Review* 33:137–158.

Hill, I., Hill, R., Holder, R. (1976). Fitting Johnson curves by moments. *Applied Statistics* 25:180–189.

Horváth, L., Kokoszka, P., Teyssière, G. (2004). Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals. *Journal of Statistical Computation and Simulation* 74:469–485.

Horváth, L., Kokoszka, P., Teyssière, G. (2001). Empirical process of the squared residuals of an ARCH sequence. *The Annals of Statistics* 29:445–469.

Jensen, M., Lunde, A. (2001). The NIG-S & ARCH model: a fat-tailed stochastic, and autoregressive conditional heteroscedastic volatility model. *Econometrics Journal* 4:319–342.

26

Johnson, N. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* 36:146–176.

Johnson, N., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions, Vol. 1 (2nd Ed.)*. New York: Wiley.

Karlis, D. (2002). An EM type algorithm for maximum likelihood estimation of the normal-inverse Gaussian distribution. *Statistics & Probability Letters* 57:43–52.

Koul, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*. New York: Springer-Verlag Inc.

Lahiri, S. (2003). *Resampling Methods for Dependent Data*. New York: Springer-Verlag Inc.

Laurent, S. (2004). Analytical derivates of the APARCH model. *Computational Economics* 24:51–57.

Lehmann, E., Casella, G. (1998). *Theory of Point Estimation*. New York: Springer.

Lillestøl, J. (2000). Risk analysis and the NIG distribution. *Journal of Risk* 2:41–56.

Mardia, K. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57:519–530.

Palmitesta, P., Provasi, C. (2000). Estimating distribution functions in Johnson translation system by the starship procedure with simulated annealing. *Metron* 58:97–110.

Pentula, S. (1986). Comments on "Modelling the persistence of conditional variances". *Econometric Reviews* 5:71–74.

Schoutens, W. (2002). The Meixner process: theory and applications in Finance. Eurandom Report 2001-004. Eurandom, Eindhoven.

27

Shao, D.J., Tu, D. (1996). *The Jackknife and Bootstrap.* New York: Springer.

Stute, W., Gonzales Manteiga, W., Presedo Quindimil, M. (1993). Bootstrap based goodness-of-fit-tests. *Metrika* 40:243–256.

Szũcs, G. (2008). Parametric bootstrap tests for continuous and discrete distributions. *Metrika* 67:63–81.

	AEX	DJ	CAC 40	DAX	S&P 500
Mean	-0.0033	0.0855	0.0648	0.0828	0.0824
StDev	2.9817	1.9313	2.5136	3.0674	1.9131
Skewness	-0.6427	-0.3161	-0.4143	-0.1237	-0.3290
Kurtosis	4.9723	5.5891	4.2721	4.2031	5.3452
$\hat{m}$	0.9315	0.5989	0.9247	0.8448	0.4183
	1.3334	*0.9494*	*1.7355*	*1.6947*	*0.9253*
$\hat{a}$	6.3637	3.8138	4.1745	5.7616	3.9324
	5.7571	*3.4508*	*4.3937*	*5.5278*	*3.6518*
$\hat{b}$	-0.7225	-0.5556	-0.6020	-0.4794	-0.3714
	1.0668	*1.1206*	*1.1566*	*1.1382*	*1.1272*
$\hat{d}$	0.3888	0.4720	0.6636	0.5411	0.4547
	0.5656	*0.7234*	*0.2171*	*0.9004*	*0.7144*
MALL	-2.4500	-2.0281	-2.3096	-2.5090	-2.0222
$T_{\mathrm{KS}}(\hat{\nu}_n)$	0.0227	0.0311	0.0265	0.0292	0.0310
	0.9445	*0.4735*	*0.7850*	*0.6125*	*0.4630*
$T_{\mathrm{CV}}(\hat{\nu}_n)$	0.0126	0.0371	0.0153	0.0251	0.0267
	0.9890	*0.3330*	*0.9560*	*0.6750*	*0.6000*

Table 3: Bootstrap goodness of fit for weekly returns (2002–2006). The displayed results concern some properties of the observed series, and the estimated parameters of the Meixner distribution, with their estimated standard errors. By MALL we indicate the maximum of the average log-likelihood. The observed values of the Kolmogorov-Smirnov and Cramér-von Mises statistics, and the corresponding bootstrap p-values, are also shown.

29

	AEX	DJ	CAC 40	DAX	S&P 500
$\hat{\mu}$	0.0371	0.0197	0.0362	0.0503	0.0210
	0.1526	*0.1281*	*0.1603*	*0.1745*	*0.1281*
$\hat{\phi}_1$	-0.0302	-0.0594	-0.0782	-0.0597	-0.0698
	0.1791	*0.1818*	*0.1813*	*0.1784*	*0.1812*
$\hat{\omega}$	0.0172	0.0058	0.0239	0.0213	0.0058
	0.0408	*0.0221*	*0.0579*	*0.0645*	*0.0202*
$\hat{\alpha}$	0.0453	0.0243	0.0314	0.0331	0.0159
	0.1419	*0.2621*	*0.1012*	*0.1083*	*0.1030*
$\hat{\gamma}$	0.3893	0.5870	0.3261	0.2079	0.4242
	1.0833	*0.6586*	*0.7528*	*0.5302*	*1.9390*
$\hat{\beta}$	0.8788	0.8880	0.8541	0.8452	0.8984
	0.1171	*0.1328*	*0.1580*	*0.1549*	*0.1209*
$\hat{\delta}$	3.1868	4.3395	4.0319	4.5788	4.2042
	2.6108	*5.0896*	*3.6445*	*4.6931*	*4.4061*
$\hat{b}$	-0.3265	-0.0592	-0.4730	-0.5987	-0.2128
	1.0374	*1.1249*	*1.0966*	*1.2303*	*1.2397*
$\hat{d}$	1.2357	1.5904	1.5101	1.8496	1.9752
	2.7857	*4.3321*	*3.4274*	*4.9910*	*6.3142*
MALL	-1.5024	-1.2289	-1.4866	-1.6126	-1.2470
$T_{\mathrm{KS}}(\hat{\zeta}_T)$	0.0189	0.0256	0.0195	0.0287	0.0248
	0.6160	*0.2880*	*0.6160*	*0.4020*	*0.3560*
$T_{\mathrm{CV}}(\hat{\zeta}_T)$	0.0504	0.1973	0.1122	0.2859	0.1678
	0.7560	*0.2120*	*0.4360*	*0.3280*	*0.3140*

Table 4: Bootstrap goodness of fit results for daily returns (2002–2006). The displayed results concern the estimated parameters of the Meixner AR(1)-APARCH(1,1) model, with their estimated standard errors. By MALL we indicate the maximum of the average log-likelihood. The observed values of the Kolmogorov-Smirnov and Cramér-von Mises statistics, and the corresponding bootstrap *p*-values, are also shown.

30

Reply to the Referee's Report on the revised version of

LSSP-2007-0145 by Grigoletto and Provasi

Simulation and Estimation of the Meixner Distribution

Main concerns

The authors give a satisfactory answer to the first main concern. However, I do not see completely solved the second one. Authors state that there are differences between Meixner and Johnson distributions (that is right), that S_U-distributions have been applied in financial literature and that according to Perez (2004) S_U-distributions may fail to capture financial risk. I have not been able to get this paper and, then, I can say nothing about whether Meixner distributions suffer from the same problem or not.

Anyway, the statement of the authors "the Meixner distribution studied in our paper may prove to be a useful alternative, when the available distributions fail to adequately describe the observations at hand", from me point of view, is weak unless the authors provide some guidelines to see under which conditions are Meixner distributions preferable to the available alternatives.

The paper by Perez (2004) which, among others, highlights the possible shortcomings of S_U-distributions, is available at

http://administracion.itam.mx/workingpapers/hedgefunds.pdf

In the concluding remarks, Perez (2004) states:

"...although the Johnson approximation carries information about skewness and kurtosis, it still fails to reflect entirely higher moment effects. It also fails to detect entirely the left tail risk, specially in the cases where we have a "short option" behavior. In these cases, the left tail of the associated Johnson distribution underestimates the probability of losses. The isolated but large negative returns seem to be too rare to make any difference on the Johnson parameters, and hence they remain undetected."

It would be simple to look for particular observed time series for which the Meixner distribution has better fitting properties than the S_U, but this would not make the message in the paper come through more clearly. We simply suggest that, when fitting models for financial time series, the Meixner distribution deserves to be in the analyst's toolbox.

In our view, it is not possible to find general guidelines suggesting *a priori* when to prefer the Meixner distribution: the suitability of this distribution for representing observed time series is to be found with the usual goodness of fit procedures. On the other hand, there are not even general *a priori* guidelines stating when the S_U distributions is preferable to the available alternatives.

Technical points

Technical points 1 and 3 have been satisfactorily solved. Let us analyze points 2 and 4.
2. I consider that simulations can not replace proofs. Thus, in me opinion, authors do not provide enough justification to use asymptotic distributions. Moreover, if we assume that the results hold, taking into account the slow speed of convergence that the simulations suggest, it seems that the conclusion would be that they are useless.
4. I also consider weak the reasons the authors give to justify the use of the bootstrap here.

It is certainly true that simulations can not replace proofs. However, when formal proofs are very hard to find (as in the present case), simulations can be useful to shed light on the behavior of the procedures under investigation. Slow convergence speed, on the other hand, is to be expected when many parameters are involved. Since no better alternatives seem to be available, maximum likelihood estimation still appears to be the preferable approach in the present framework (and is suggested, among others, by Davison and Hinkley, 1997, p. 148).

The use of bootstrap in the context of parametric likelihood inference has been advocated e.g. by Davison *et al.* (2003), while Boos (2003) applies resampling for a goodness of fit problem based on the use of the Anderson-Darling statistic. In a now classic paper, MacKinnon (2002) states "The astonishing increase in computer performance over the past two decades has made it possible for economists to base many statistical inferences on simulated, or bootstrap, distributions rather than on distributions obtained from asymptotic theory".

References

Boos, D.D. (2003). Introduction to the bootstrap world. *Statistical Science, 18*, 168–174.

Davison, A.C., Hinkley, D.V. (1997). *Bootstrap methods and their application.* Cambridge: Cambridge University Press.

Davison, A.C., Hinkley, D.V., Young, G.A. (2003). Recent developments in bootstrap methodology. *Statistical Science, 18*, 141–157.

MacKinnon, J.G. (2002). Bootstrap inference in Econometrics. *Canadian Journal of Economics, 35*, 615–645.

Pérez, P.G. (2004). An approach to the non-normal behavior of hedge fund indices using Johnson distributions. *Working Paper*, Department of Business and Administration, ITAM.