ELSEVIER

Journal of Economic Dynamics & Control
25 (2001) 1457–1483

JOURNAL OF Economic Dynamics & Control

www.elsevier.nl/locate/econbase

Gram–Charlier densities☆

Eric Jondeau[a], Michael Rockinger[b,*]

[a]*Banque de France, DEER, 41-1391 Centre de Recherche, 31, rue Croix des Petits Champs, 75049 Paris, France*
[b]*HEC-School of Management, Department of Finance, 1, rue de la Libération, 78351 Jouy-en-Josas, France*

Received 5 October 1998; accepted 7 December 1999

Abstract

The Gram–Charlier expansion, where skewness and kurtosis directly appear as parameters, has become popular in Finance as a generalization of the normal density. We show how positivity constraints can be numerically implemented, thereby guaranteeing that the expansion defines a density. The constrained expansion can be referred to as a Gram–Charlier density. First, we apply our method to the estimation of risk neutral densities. Then, we assess the statistical properties of maximum-likelihood estimates of Gram–Charlier densities. Lastly, we apply the framework to the estimation of a GARCH model where the conditional density is a Gram–Charlier density.

JEL classification: C40; C63; G13; F31

Keywords: Hermite expansion; Semi-nonparametric estimation; Risk-neutral density; GARCH model

* Corresponding author. Tel: + 33-01-39-67-72-59.

E-mail address: rockinger@hec.fr (M. Rockinger).

☆Michael Rockinger is also CEPR, and scientific consultant at the Banque de France. He acknowledges help from the HEC Foundation and the European Community TMR Grant: "Financial Market Efficiency and Economic Efficiency". We are grateful to Sanvi Avouyi-Dovi, Sophie Coutant, Ivana Komunjer, and Pierre Sicsic for precious comments. The usual disclaimer applies. The Banque de France does not necessarily endorse the views expressed in this paper.

0165-1889/01/$ - see front matter
PII: S0165-1889(99)00082-2

1. Introduction

Several recent studies in empirical finance have used Gram–Charlier-type expansions as a semi-nonparametric device to overcome the restriction imposed by the usual normality assumption. For instance Knight and Satchell (1997) develop an option pricing model using a Gram–Charlier expansion for the underlying asset. In a similar framework, Abken et al. (1996) end up with a Gram–Charlier expansion to approximate risk neutral densities (RND). Gallant and Tauchen (1989) use Gram–Charlier expansions to describe deviations from normality of innovations in a GARCH framework.

Gram–Charlier expansions allow for additional flexibility over a normal density because they naturally introduce the skewness and kurtosis of the distribution as parameters. However, being polynomial approximations, they have the drawback of yielding negative values for certain parameters. Moreover, there does not seem to be an easy and analytic characterization of those parameters for which the density will take positive values. In a noticeable study by Barton and Dennis (1952) conditions on the parameters guaranteeing positive definiteness of the underlying densities are obtained through a numerical method. In this paper we build on their work and indicate how it is numerically possible to restrict parameters. Once positivity for the expansion gets imposed we may talk of Gram–Charlier densities (GCd).

In this paper we first specialize the method advocated by Barton and Dennis (1992) to characterize the boundary delimiting the domain in the skewness–kurtosis space over which the expansion is positive. We then present a mapping which transforms the constrained estimation problem into an unconstrained one. Since the positivity boundary is only defined as an implicit function, we show how the mapping can be numerically imposed.

In the empirical part of this paper we first show the relevance of our method to estimate risk neutral densities, by extending the work of Abken et al. (1996). Next, we examine the maximum-likelihood properties when GCds are directly fitted to data. We consider the situation where GCds are fitted to GCd distributed data as well as to a mixture of normals. The first simulation allows us to validate our code and to examine estimation properties in situations known to be delicate. Similarly to the estimation of mixtures of normals, for small deviations from normality, we find that it is difficult in that situation to correctly capture the parameters. The second simulation shows possible biases of the estimation when the model is misspecified. Lastly, we indicate how our method improves GARCH estimations when innovations are assumed to be distributed as a Gram–Charlier density rather than a normal one.

This paper is structured in the following manner: In the next section, we provide some properties of Gram–Charlier expansions. In Section 3, we describe our algorithm to implement positivity of the density. In Sections 4 and 5, we show with two examples how it works. We estimate risk neutral densities and

a GARCH model allowing for a conditional density with skewness and kurtosis different from those of a normal distribution.

2. Properties of Gram–Charlier expansions

When the true probability distribution function (pdf) of a random variable z is unknown, yet believed to be similar to a normal one, it is quite natural to approximate it with a pdf of the form

$$g(z) = p_n(z)\phi(z), \tag{1}$$

where $\phi(z)$ is the standard zero mean and unit variance normal density and where $p_n(z)$ is chosen so that $g(z)$ has the same first moments as the pdf of z. Since Hermite polynomials form an orthogonal basis with respect to the scalar product generated by the expectation taken with the normal density ϕ the true density is often approximated using

$$p_n(z) = \sum_{i=0}^{n} c_i He_i(z), \tag{2}$$

where $He_i(z)$ are the Hermite polynomials.[1] The Hermite polynomial of order i is defined by $He_i(z) = (-1)^i(\partial^i\phi/\partial z^i)1/\phi(z)$.[2] When z is standardized, with zero mean and unit variance, two representations are typically adopted in the literature

$$p_4(z) = 1 + \frac{\gamma_1}{6}He_3(z) + \frac{\gamma_2}{24}He_4(z) \tag{3}$$

and

$$p_6(z) = 1 + \frac{\gamma_1}{6}He_3(z) + \frac{\gamma_2}{24}He_4(z) + \frac{\gamma_1^2}{72}He_6(z). \tag{4}$$

These cases correspond, respectively, to the Gram–Charlier type-A and the Edgeworth expansions. The Edgeworth expansion (4) involves one more Hermite polynomial while keeping the number of parameters constant. As shown by Barton and Dennis (1952), the range for γ_1 and γ_2 over which positivity of the approximation is guaranteed is then smaller than for the Gram–Charlier one. For this reason, in this paper we will focus on the first approximation.

[1] For Hermite polynomials we follow the notation of Gradshteyn and Ryzhnik (1994, p. xxxvii).

[2] Straightforward computations yield the following expressions for the first six Hermite polynomials: $He_0(z) = 1, He_1(z) = z$, $He_2(z) = z^2 - 1$, $He_3(z) = z^3 - 3z$, $He_4(z) = z^4 - 6z^2 + 3$, $He_5(z) = z^5 - 10z^3 + 15z$, and $He_6(z) = z^6 - 15z^4 + 45z^2 - 15$.

Property 1. γ_1 *and* γ_2 *correspond, respectively, to the skewness and the excess kurtosis of* $g(z)$.

Proof. Because z is standardized, straightforward but tedious computations show that:[3]

$$\int_{-\infty}^{+\infty} zg(z)\,\mathrm{d}z = 0, \qquad \int_{-\infty}^{+\infty} z^2 g(z)\,\mathrm{d}z = 1,$$
$$\int_{-\infty}^{+\infty} z^3 g(z)\,\mathrm{d}z = \gamma_1, \qquad \int_{-\infty}^{+\infty} z^4 g(z)\,\mathrm{d}z = 3 + \gamma_2.$$

In the following pages, we will, therefore, adopt the notations $s = \gamma_1$ and $k = \gamma_2$ to denote the skewness and the excess kurtosis, respectively. Property 1 partly explains the success of Gram–Charlier expansions in the empirical literature, since the two additional parameters γ_1, γ_2 are directly related to the third and fourth moments. However, Gram–Charlier expansions also have some drawbacks. First, for some (s, k) distant from the normal values (0,3), $g(z)$ can be negative for some z. For other pairs the pdf $g(z)$ may be multimodal.

In this work we focus on implementing numerical conditions so that Gram–Charlier approximations are positive definite. To ensure positivity, Gallant and Tauchen (1989) suggest to square the polynomial part, $p_n(z)$, of Eq. (1). However, by doing so one loses, the interpretation of the various parameters as moments of the density.

Some properties are useful to identify the region $\mathscr{D}$ in the (s, k)-plane for which $g(z)$ is positive definite. For $g(z)$ to be positive definite, we require the polynomial $p_4(z)$ to be positive for every z, that is

$$1 + \frac{s}{6}He_3(z) + \frac{k}{24}He_4(z) \geq 0, \quad \forall z.$$

In order to characterize $\mathscr{D}$ two approaches can be followed. The first direct one consists in establishing general properties of the frontier of $\mathscr{D}$ and then to take z over a large grid and to check if for possible pairs of (s, k) the polynomial $p_4(z)$ is positive.[4] The second one involves notions of analytical geometry. Consider a given value of z. For each such value the equation

$$1 + \frac{s}{6}He_3(z) + \frac{k}{24}He_4(z) = 0 \tag{5}$$

defines a straight line in the (s, k)-plane. A small deviation for z, while holding (s, k) fixed, will then yield a $p_4(z)$ of either positive or negative sign. Thus, it is

[3] See also Johnson et al. (1994).

[4] Such an approach was followed in an earlier version of this paper.

interesting to determine the set of (s, k), as a function of z, such that $p_4(z)$ remains zero for small variations of z since this set will define the requested boundary. This set is determined by the derivative of (5) with respect to z

$$\frac{s}{2}He_2(z) + \frac{k}{6}He_3(z) = 0. \tag{6}$$

The set of (s, k) solving simultaneously (5) and (6), called the *envelope* of $p_4(z)$, yields a parametric representation of the boundary where for a given z the term $p_4(z)$ is zero. Once this boundary is determined it remains to find that subregion delimited by $p_4(z) = 0$ for all z.[5]

Solving the system given by (5) and (6) yields the expression for the skewness and the excess kurtosis as functions of z:

$$s(z) = -24\frac{He_3(z)}{d(z)},$$

$$k(z) = 72\frac{He_2(z)}{d(z)},$$

with $d(z) = 4He_3^2(z) - 3He_2(z)He_4(z)$.

Straightforward computations allow us to rewrite the denominator of both expressions as $d(z) = z^6 - 3z^4 + 9z^2 + 9$. Since its minimum is attained for $z = 0$ where $d(0) = 9$ we obtain that $d(z)$ is always positive.

The sign of $k(z)$ changes with $He_2(z) = z^2 - 1$. It is positive for z between $-\infty$ and -1 and between 1 and $+\infty$. It is negative for $-1 \le z \le 1$. Similarly, the sign of $s(k)$ changes with $He_3(z) = z^3 - 3z$. It is positive for z between $-\infty$ and $-\sqrt{3}$ and between 0 and $\sqrt{3}$. In Fig. 1 we present, the straight lines defined by (5) for various values of z (satisfying $|z| \ge \sqrt{3}$). The thick line delimiting the oval domain is the envelope. Within the envelope $p_4(z)$ will be positive. Similarly, in Fig. 2 we present (5) and its envelope for values of z between $-\sqrt{3}$ and $\sqrt{3}$.

In Fig. 3 we present a summarizing graph: for z between $-\infty$ and $-\sqrt{3}$ one obtains the curve AM_1B; the values $-\sqrt{3} \le z \le 0$ lead to the curve BM_4C; the values $0 \le z \le \sqrt{3}$ lead to the curve CM_3B; lastly when z varies from $\sqrt{3}$ to $+\infty$ one obtains the curve BM_2A. This envelope is clearly symmetric with respect to the horizontal axis. Thus, the region where $g(z)$ is positive for every z is formed by the intersection of the domains delimited in Figs. 1 and 2; that is the curve AM_1BM_2A.

If we concentrate on the envelope where $g(z)$ is positive for all z, we note that the excess kurtosis k is inside the interval [0,4]. Indeed, we find that

[5] This approach has been highlighted by Barton and Dennis (1952) in a slightly different context.

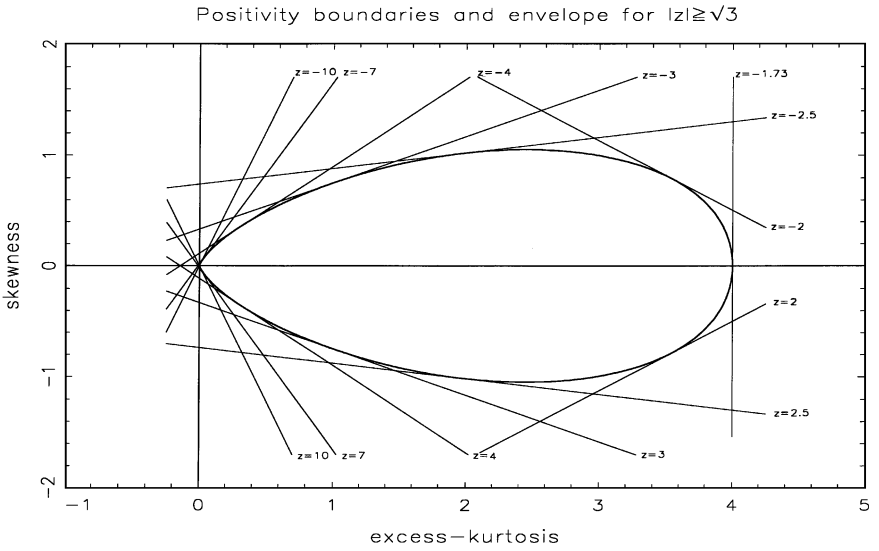

Fig. 1. Represents equations (5) of the text for various values of z with $|z| \geq \sqrt{3}$. We also present the envelope given as a solution to (5) and (6).

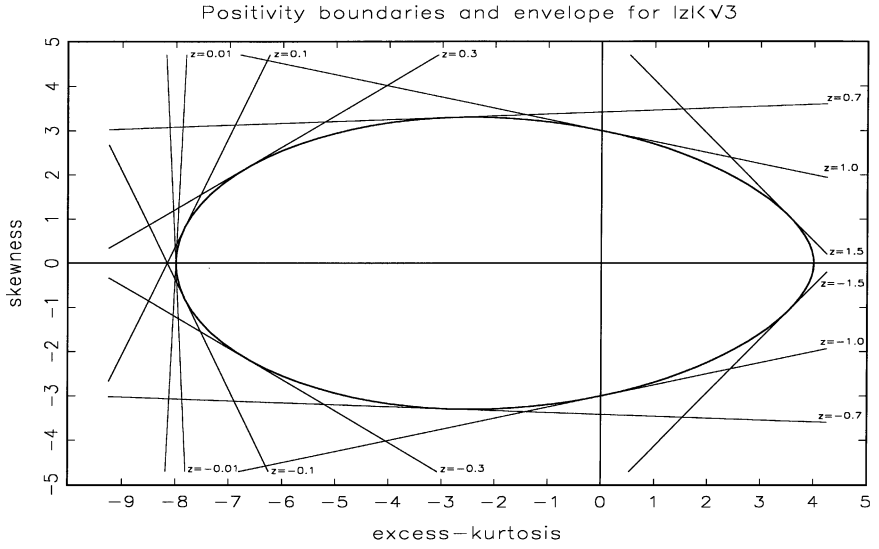

Fig. 2. Represents the same as Fig. 1 but with $|z| < \sqrt{3}$. The scale differs in both figures.

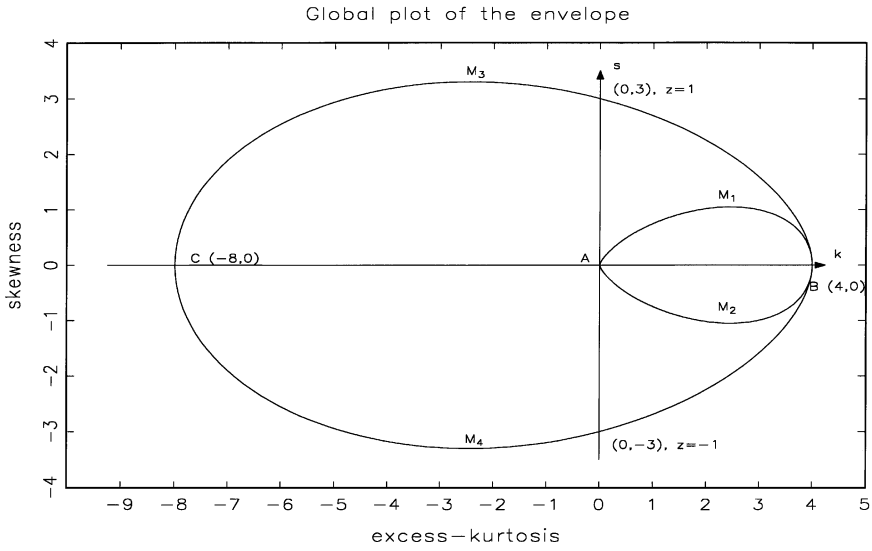

Fig. 3. Presents the plot of the envelope as z varies from $-\infty$ to $+\infty$. For skewness and kurtosis located in the interior of the domain $\mathscr{D}$, delimited by AM_1BM_2A, the Gram–Charlier expansion is actually a density.

$k(\pm\infty)=0$ and $k(\pm\sqrt{3})=4$. The points where the skewness is at a maximum or a minimum are obtained when $s'(z)=z^4-6z^3+6z^2-18z+9=0$. The solutions can be found numerically to be $M_1=(\sqrt{6},\sqrt{6}/\sqrt{3+\sqrt{6}})=(2.4508;1.0493)$ and $M_2=(\sqrt{6},-\sqrt{6}/\sqrt{3+\sqrt{6}})=(2.4508;-1.0493)$.

We notice that the frontier is a steady, continuous, and concave curve. A last remark is that since k is bounded below by 0 the kurtosis of $g(\cdot)$ will always be larger than for a normal density.

3. An algorithm to implement positivity

At this stage we have characterized the domain $\mathscr{D}$ over which the Gram–Charlier approximation is positive. We now wish to indicate how positivity may be numerically implemented in applications where we will have to solve programs such as

$$\max_{(s,k)\in\mathscr{D}} F(s,k),$$

where F is an objective function involving s and k through the Gram–Charlier expansion. F may depend on some other parameters which are unrelated to

s and k.[6] In numerical applications F could be a log-likelihood function or a NLLS problem.[7] F is assumed to be differentiable with respect to all its parameters and we assume that there is a unique optimum.

Since F is not defined for (s,k) outside $\mathscr{D}$ it is necessary to restrict parameters to that region using an ad hoc method. The idea is to transform the constrained optimization into an unconstrained one. This turns out to yield a fast and numerically accurate method.

In the previous section, we derived the parametric equation of the frontier of $\mathscr{D}$. Taking z over a very fine grid, a discretization of the frontier is possible. Alternatively, and this is how we compute the boundary for our empirical work, it is possible to take k over a fine grid. For each value on the grid we know from the previous section that the associated s is bracketed in the interval $[-1.0493, 1.0493]$. Numerically, one can then compute s with a bisection algorithm.[8] Furthermore, for numerical applications it is necessary to have a continuous representation of the boundary, therefore, we substitute the continuous frontier with a piecewise linear one. For each k the corresponding s can then be found with a linear interpolation.

Formally, we proceed in the following manner also illustrated in Fig. 4: we start with a fine grid for kurtosis say k_i, $i = 1, \ldots, N_k$.[9] For each k_i the corresponding s_i is known. We compute and store

$$a_i = \frac{s_i k_{i+1} - k_i s_{i+1}}{k_{i+1} - k_i},$$

$$b_i = \frac{s_{i+1} - s_i}{k_{i+1} - k_i}$$

for $i = 1, \ldots, N_k - 1$. For a given k the maximal, $(s_U(k))$, and minimal, $(s_L(k))$, allowed skewness in $\mathscr{D}$ will be approximated with a linear interpolation by first obtaining the i such that $k_i < k \leq k_{i+1}$, and then computing $s_U = a_i + b_i k$ as well as $s_L = -s_U$.

Now it is possible to introduce an ad hoc mapping transforming the constrained optimization into an unconstrained one. We introduce the logistic map defined by

$$f(x; a, b) = a + (b - a)\frac{1}{1 + \exp(-x)}.$$

[6] To keep notations simple we do not emphasize this possibility.

[7] In this latter case the maximization becomes a minimization.

[8] For more details see Press et al. (1988, p. 259). The simultaneous computation of all s can be easily vectorized yielding a very efficient code.

[9] In numerical applications $N_k = 250$ appears to provide a reasonable compromise between speed and accuracy.

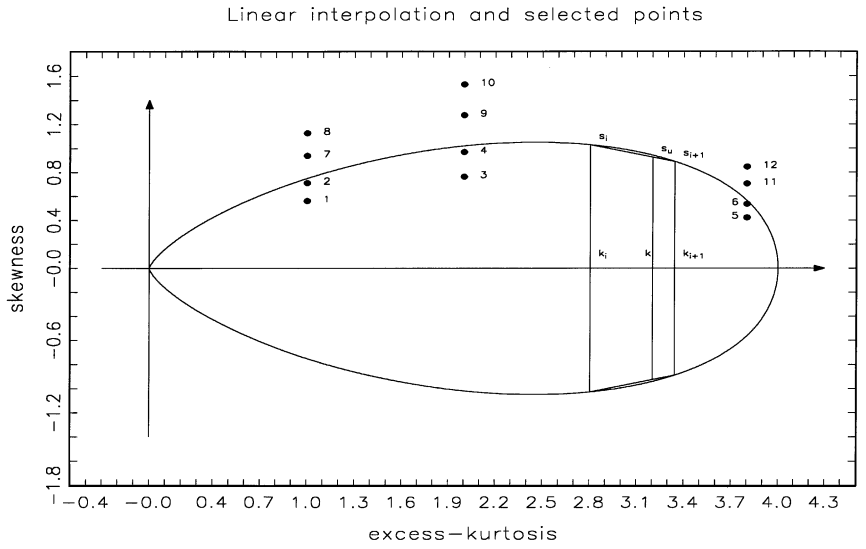

Fig. 4. Indicates how we replace the discrete envelope AM_1BM_2A with segments (not to scale) allowing for a continuous representation of the boundary. The parameters of points 1–6 correspond to rows 1–6 of Tables 2 and 3. Points 7–12, out of the domain, correspond to rows 7–12 of Table 3. For given kurtosis the points correspond to 75, 95, 125 and 150% of the segment $[0, s_u]$.

Let $(\tilde{s}, \tilde{k}) \in \mathscr{R}^2$ be unconstrained values for the skewness and kurtosis. It is easy to see that the map

$$k = f(\tilde{k}; 0, 4) \equiv f_k(\tilde{k}), \tag{7}$$

$$s = f(\tilde{s}; s_L(\tilde{k}), s_U(\tilde{k})) \equiv f_s(\tilde{s}, \tilde{k}) \tag{8}$$

transforms $\mathscr{R}^2$ into $\mathscr{D}$. Given that this mapping involves the logistic function, which is strictly increasing and differentiable, we notice that the first-order conditions of

$$\max_{(\tilde{s},\,\tilde{k}) \in \mathscr{R}^2} G(\tilde{s},\tilde{k}) \equiv F(f_s(\tilde{s}, \tilde{k}), f_k(\tilde{k}))$$

that is

$$\frac{\partial G}{\partial \tilde{s}} = \frac{\partial F}{\partial s}\frac{\partial f_s}{\partial \tilde{s}} = 0,$$

$$\frac{\partial G}{\partial \tilde{k}} = \frac{\partial F}{\partial s}\frac{\partial f_s}{\partial \tilde{k}} + \frac{\partial F}{\partial k}\frac{\partial f_k}{\partial \tilde{k}} = 0$$

imply

$$\frac{\partial F}{\partial s} = 0 \quad \text{and} \quad \frac{\partial F}{\partial k} = 0.$$

Thus, the unconstrained optimum, say $(\tilde{s}^*, \tilde{k}^*)$, is uniquely related to the constrained optimum

$$s^* = f_s(\tilde{s}^*, \tilde{k}^*), \quad k^* = f_k(\tilde{k}^*).$$

Given unicity of the optimum and convexity of our map our restriction is therefore a valid one.

We will henceforth denote by $\mathscr{GC}(\mu, \sigma, s, k)$ the Gram–Charlier density with mean μ and standard deviation σ obtained by imposing positivity constraints on the expansion.

4. The estimation of risk neutral densities

In this section we wish to illustrate the usefulness of our method on a first example dealing with option pricing.

4.1. Theoretical considerations

Let S_t be the price of an asset at time t. We suppose that this asset underlies a European call option with expiration date T and strike price K. Then, at maturity the payoff is $\max(S_T - K, 0)$. In an arbitrage-free economy (see Harrison and Pliska, 1981), it is known that there exists a risk-neutral density (RND), $g(\cdot)$, such that the price of a call option can be written as

$$C_t(K) = \mathrm{e}^{-r(T-t)} \int_K^{+\infty} (S_T - K) g(S_T)\, \mathrm{d}S_T, \tag{9}$$

where $C_t(K)$ is the price at time t of a call option, and r is the continuously compounded interest rate to maturity. The function $C_t(\cdot)$ depends on the parameters r, T, t as well as others characterizing the process followed by S_t. As noted by Breeden and Litzenberger (1978), Leibnitz' rule for differentiating integrals gives:

$$\frac{\mathrm{d}^2 C_t}{\mathrm{d}K^2}(S_T) = \mathrm{e}^{-r(T-t)} g(S_T) \tag{10}$$

which reveals the discounted RND. For the econometrician wishing to estimate $g(S_T)$, formula (10) suggests the use of numerical second derivatives. Numerical computation of second derivatives is, however, a very unstable method. For this reason it is sometimes advantageous to assume some additional structure on

$g(\cdot)$ and to proceed with (9). This is the way suggested by Abken et al. (1996) (AMR for short) who assume that $g(\cdot)$ can be approximated with Hermite polynomials. Inspired by the usual lognormality assumption of the underlying asset they assume first that

$$S_T = S_t \exp((\mu - \tfrac{1}{2}\sigma^2)(T-t) + \sigma(T-t)z), \tag{11}$$

where z is a normal variate with zero mean and unit variance. The parameters μ and σ represent the instantaneous drift and volatility, respectively, of S_T. Last, in the spirit of Section 2 of this paper, they consider that $g(z)$ is given by $g(z) = \lambda(z)\phi(z)$ where $\lambda(z)$ is a perturbation of the normal $\mathcal{N}(0,1)$ density $\phi(\cdot)$. By assuming that $\lambda(z)$ can be approximated by a Hermite expansion they obtain that option prices can be written as

$$C_t(K) = \mathrm{e}^{-r(T-t)} \sum_{k=0}^{4} a_k b_k, \tag{12}$$

where the b_k are parameters to be estimated and where the a_k take the following expressions, denoting $\tilde{\sigma} = \sigma\sqrt{T-t}$:

$$\begin{aligned}
a_0 &= F_t \Phi(d_1) - K\Phi(d_2),\\
a_1 &= F_t(\tilde{\sigma}\Phi(d_1) + \phi(d_1)) - K\phi(d_2),\\
a_2 &= \tfrac{1}{2}[F_t(\tilde{\sigma}^2\Phi(d_1) + 2\tilde{\sigma}n_1 - h_{1,1}) + Kh_{1,2}],\\
a_3 &= \tfrac{1}{6}[F_t(\tilde{\sigma}^3\Phi(d_1) + 3\tilde{\sigma}^2 n_1 - 3\tilde{\sigma}h_{1,1} + h_{2,1}) - Kh_{2,2}],\\
a_4 &= \tfrac{1}{24}[F_t(\tilde{\sigma}^4\Phi(d_1) + 4\tilde{\sigma}^3 n_1 - 6\tilde{\sigma}^2 h_{1,1} + 4\tilde{\sigma}h_{2,1} - h_{3,1}) + Kh_{3,2}],\\
d_1 &= \frac{\ln(S_t/K) + (\mu + \sigma^2/2)(T-t)}{\tilde{\sigma}},\\
d_2 &= d_1 - \tilde{\sigma},\\
F_t &= S_t \exp(r(T-t)),\\
n_1 &= 1/\sqrt{2\pi}\exp(-d_1^2/2),\\
n_2 &= 1/\sqrt{2\pi}\exp(-d_2^2/2),\\
h_{i,j} &= He_i(d_j)n_j.
\end{aligned}$$

In the expression a_0 we recognize up to a discount factor the Black–Scholes–Merton benchmark pricing formula. For the AMR model one can see that option prices are obtained as a perturbation of the benchmark case. F_t is the forward price.

To obtain identifiability and a density for g further assumptions are imposed: $b_0 = 1$ (forcing $g(z)$ to have unit probability mass), $b_1 = 0$ forcing a zero expectation, $b_2 = 0$ imposing a unit variance on z. b_3 and b_4 will control skewness and excess kurtosis.

The risk neutral density eventually becomes

$$g(z) = \left[1 + \frac{b_3}{\sqrt{6}}(z^3 - 3z) + \frac{b_4}{\sqrt{24}}(z^4 - 6z^2 + 3)\right]\phi(z). \tag{13}$$

The inversion of (11) allows one to express the RND with respect to S_T.

To sum up: Given call option prices and their characteristics (such as time to maturity, strike price, value of the underlying asset, interest rates), Eq. (12) can be used to numerically estimate the b_3 and b_4 parameters. Since Eq. (12) also involves in a non-linear manner the volatility σ, it will be necessary to estimate all parameters using a non-linear procedure. The parameter μ can be either estimated (non-linearly) or obtained by imposing the non-arbitrage condition

$$S_t = \mathrm{e}^{-r(T-t)} \int_0^{+\infty} S_T g(S_T)\,\mathrm{d}S_T.$$

Once the non-linear estimation procedure has produced parameter estimates, Eq. (13) can be used to obtain the RND. It should be noted that Eq. (13) is basically the same one as Eq. (1), with $n = 4$, and with $p_n(z)$ defined by Eq. (3). Between the AMR parametrization and the theory presented earlier we obtain the following relations $b_3 = s/\sqrt{6}$ and $b_4 = k/\sqrt{24}$. Furthermore, all earlier developments still bear. Whereas AMR show how parameters can be estimated they do not address the issue that the corresponding (s, k) parameters may not belong to $\mathscr{D}$. As we will show in the next section, by using the algorithm proposed in Sections 2 and 3 this difficulty can be overcome.

4.2. Empirical results

We implement this method with European OTC French Franc to Deutsche Mark options which have been provided to us by a large French bank. For foreign exchange options the Garman and Kohlhagen (1983) model imposes the non-arbitrage condition $\mu = r - r^*$, where r and r^* are the French and German euro-rates, respectively. For illustrative purposes we use data for April 25th, 1997 that is a few days after President Chirac announced the dissolution of the National Assembly leading to snap elections. At this stage the markets were roiling.[10] We obtained option prices for several maturities and strikes as well as the value of the underlying exchange rate. Using formula (12) relating skewness, kurtosis, and non-linearly volatility of the underlying asset with actual option prices we estimate, using the conventional NLLS method, the various parameters without imposing positivity restrictions. The parameters are displayed in Table 1. As shown by those parameters, they lie out of the authorized domain

[10] See also Jondeau and Rockinger (1998) for further details concerning this period.

Table 1
Estimates of the Gram–Charlier expansion without positivity constraints

	Without positivity constraints			
	1 Month	3 Month	6 Month	12 Month
σ	0.0280	0.0275	0.0291	0.0296
s	1.5875	1.7769	1.5724	1.5121
k	0.8836	1.1790	4.7630	4.8310

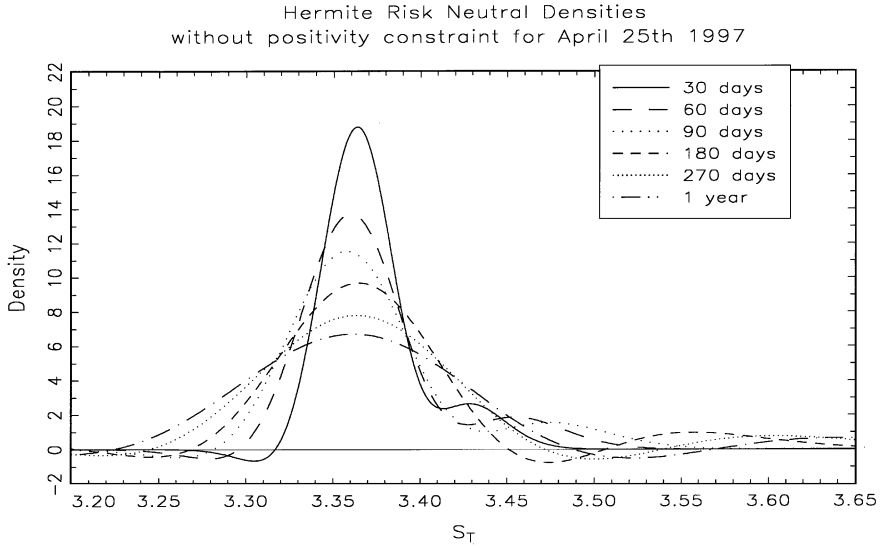

Fig. 5. Represents risk-neutral densities estimated without positivity constraints. The data is FRF/DEM options on April 25th 1997.

represented in Fig. 4 implying that the RND must be negative. Fig. 5 where we represent the RND given by expression (13) shows that this is indeed the case.

The estimation of the model with restricted parameters yields the estimates displayed in Table 2. We notice a significant difference for skewness and kurtosis. It can be checked that the parameters now belong to the authorized domain $\mathscr{D}$. As shown in Fig. 6 the densities are positive.

The observation that the unconstrained estimation yields parameters for which the polynomial approximation is negative suggests that there is a mis-specification in the model. Theoretically one could overcome this difficulty by introducing further terms in the expansion. In practice there are several reasons

Table 2
Estimates of the Gram–Charlier expansion with positivity constraints

	After imposing positivity constraints			
	1 Month	3 Month	6 Month	12 Month
σ	0.0295	0.0291	0.0276	0.0281
s	0.3781	0.9563	0.9808	0.9772
k	2.9920	3.1428	3.0583	3.0720

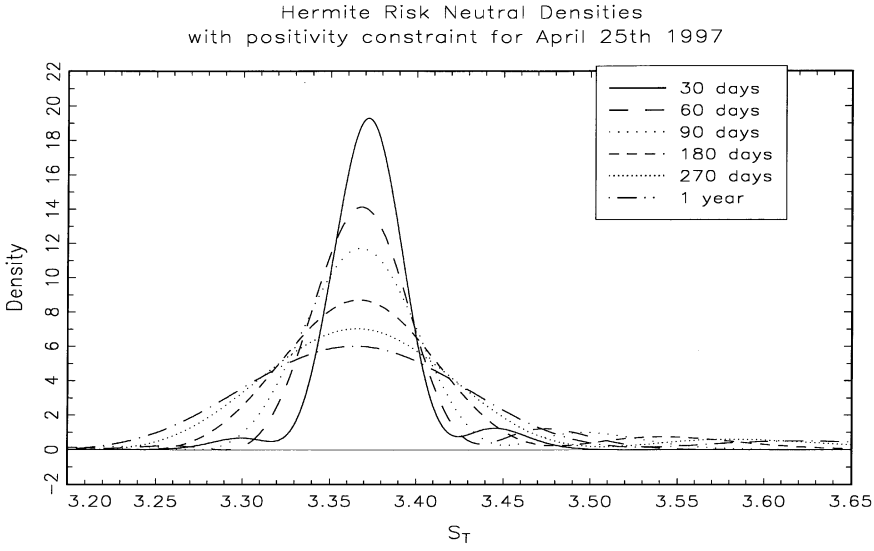

Fig. 6. Represents risk-neutral densities estimated with positivity constraints. The data is FRF/DEM options on April 25th 1997.

why this extension is not fruitful. First, the introduction of additional parameters renders more difficult the research for the domain where the approximation is positive. Second, we have to estimate three parameters but we have only very few option prices (13) for a given maturity. The introduction of further terms in the expansion would yield a numerically unstable problem. Third, as already noticed by Corrado and Su (1996, p. 624) who deal with Jarrow and Rudd's (1982) approximation, if one increases the number of terms in the expansion, one has to deal with multicollinear parameters. The intuition for this comes from the observation that the parameter b_j is related to the jth moment and as a consequence the parameters b_5 and b_6 would turn out to be highly

correlated with b_3 and b_4, respectively. Similarly all parameters where the indices are of same parity are collinear. As Corrado and Su mention: … *adding the terms, b_5 and/or b_6,*[11] *to skewness and kurtosis estimation procedures leads to highly unstable parameter estimates.*

As a consequence, adding more terms to the expansion, beyond the difficulty to characterize the domain where the approximation is positive, raises problems of stability due to a too small sample size and multicollinearity. For those reasons we only focus on moments up to the fourth one.

5. Estimation of Gram–Charlier densities

In this section we wish to investigate the properties of maximum-likelihood estimates when Gram–Charlier densities (GCd) are used in an attempt to directly obtain higher moments that differ from the ones of the normal distribution. For this purpose we investigate how well GCds can be fitted to simulated data. We consider the fit of GCds to data generated with a Gram–Charlier distribution and to data generated with a mixture of normals. Furthermore, in the latter type of simulation we distinguish the situation where parameters for the simulated data are in or out of the restricted domain $\mathscr{D}$. Once the statistical properties are well understood we turn to the estimation of GARCH processes with Gram–Charlier distributed innovations.

5.1. Assessment of statistical properties

5.1.1. Sampling from a Gram–Charlier density

In our first simulation experiment we consider as true data generating process (DGP) random variables distributed according to the Gram–Charlier density. To that data we fit a GCd with maximum-likelihood. We simulate $N = 100$ series of length $T = 2000$ of data $\mathscr{GC}(0, 1, s, k)$.[12] We will retain this type of size for all simulations reported in this work. For excess kurtosis, k, we have arbitrarily chosen the values 1, 2, and 3.8. The first value corresponds to a situation where the tails behave very much like for a normal density, and the third value is close to the upper boundary of excess kurtosis 4. For each value of kurtosis we have chosen values of skewness that correspond to the 75th, and 95th percentile of the $[0, s_u(k)]$ segment. In columns 2–5 of Table 3 we present

[11] We have used our notation b_5, b_6 for the higher moments.

[12] To simulate this density we numerically construct its cumulative distribution function. The inverse of uniformly generated numbers is then distributed as the GCd. For more details see Ripley (1987, p. 59).

Table 3
ML estimation of Gram–Charlier parameters when the true DGP is Gram–Charlier[a]

	Theoretical parameters				ML estimates				STD of ML estimates			
	μ	σ	s	k	μ	σ	s	k	μ	σ	s	k
1	0.00	1.00	0.56	1.00	− 0.0015	0.9906	0.5451	0.8666	0.0240	0.0164	0.1277	0.2384
2	0.00	1.00	0.71	1.00	0.0011	0.9895	0.6940	0.9254	0.0225	0.0200	0.1115	0.1812
3	0.00	1.00	0.76	2.00	0.0015	0.9946	0.7723	1.9814	0.0233	0.0174	0.0621	0.1681
4	0.00	1.00	0.97	2.00	− 0.0042	0.9940	0.9616	1.9850	0.0214	0.0167	0.0459	0.1222
5	0.00	1.00	0.42	3.80	0.0017	0.9976	0.4364	3.7930	0.0210	0.0105	0.0922	0.0904
6	0.00	1.00	0.54	3.80	0.0009	0.9967	0.5414	3.7776	0.0209	0.0128	0.0930	0.0769

[a]This table presents the results of simulations where a GCd is fitted to GCd distributed data. We simulate for each set of parameters 100 samples of length 2000. Columns 2–5 present the theoretical moments chosen. Columns 6–9 present the averages of the ML estimates. This estimation can only be done when positivity is imposed. Columns 10–13 present the standard deviations of the ML estimates.

the various selected parameters (μ, σ, s, k) and in Fig. 4 we represent with dots the associated (s, k) pairs corresponding to the rows 1–6.[13]

As the columns μ and σ for the maximum-likelihood (ML) estimation show, the average of the estimates for the first and second moments are very good. The average of the mean takes values between -0.0042 and 0.0017 which compares with the true value of 0. Turning to skewness and kurtosis we still find that on average the estimates come very close to the theoretical ones. However, we notice that the estimates for kurtosis tend to differ by a larger percentage from the theoretical values than the other moments. In particular, for a given level of theoretical kurtosis, the smaller the skewness, the worse the average of the estimated kurtosis. This suggests that for the situation where the tail-thickness of the density behave like the ones of a normal one, it will be difficult to also allow for a non-zero skewness. In such cases estimation of a GCd is difficult. A similar situation appears in the context of fitting a mixture of densities. Bowman and Shenton (1973) mention that *...there is the paradox that, the nearer to normality the theoretical distribution is, the less likely it is that a normal mixture fit can be found*. Our research suggests that this sentence can be transformed into *..., the nearer kurtosis is to the one of the normal distribution, the less likely it is that a parametric approximation can be found.*

When turning to the dispersion of the parameter estimates, measured with their standard deviation, our earlier observations are corroborated. The estimates of μ and σ vary little. For skewness and kurtosis the dispersion increases. We explain this as resulting from the multicollinearity of the parameters. We also notice that as kurtosis increases the dispersion of the parameters improves. On the other hand, for a given kurtosis, the larger the skewness the better the estimates. This result indicates that the GCd estimation is better the more the tails differ from the normal one.

5.1.2. Sampling from a mixture of normals density

To further assess the ability of the ML estimation of the GCd to correctly capture the moments of the data, we simulate data distributed as a mixture of normals. Formally we assume that the true DGP is given by

$$p\, n_1(r; \mu_1, \sigma_1) + (1 - p)\, n_2(r; \mu_2, \sigma_2),$$

where n_1 and n_2 are normal densities of given mean and standard deviation. The parameter $p \in [0,1]$ indicates the probability of sampling from one or the other distribution. For given moments (up to the fifth moment and belonging to a domain of complex nature), Karl Pearson (1894) showed that the parameters $p, \mu_1, \mu_2, \sigma_1, \sigma_2$ can be obtained as a solution to a *fundamental nonic*, that is

[13] We verified that the moments of the simulated data came on average close to the theoretical ones.

a polynomial of the ninth degree.[14] For five given moments (located in a complicated domain) it is, therefore, possible to infer parameters for the mixture of normals $(p, \mu_1, \sigma_1, \mu_2, \sigma_2)$ yielding precisely those moments.

In Table 4 we present the results for this simulation. Columns 2–6 present the parameters necessary for the mixture to yield the theoretical moments displayed in columns 7–9. In the table d1 and d2 correspond to μ_1 and μ_2, s12 and s22 correspond to σ_1^2 and σ_2^2. In addition to the skewness–kurtosis pairs considered previously (1–6 in Fig. 4), we consider several additional observations, 7–12, laying outside of $\mathscr{D}$ corresponding to a 25 and 50% excess of the segment $[0, s_u(k)]$.[15] In Fig. 7 we represent the graph of the mixture for point 1, that is a mixture of two normals yielding a skewness of 0.562 and an excess kurtosis of 4. The retained fifth moment is then 4.1. We notice the strong deviation from the normal density.

Turning to simulations, we noticed that first and second moments of the simulated data were on average, up to the third decimal, identical with the theoretical ones.[16] Back to Table 4, we notice that the average skewness and excess kurtosis displayed in columns 10 and 11 come very close to the theoretical moments. Our simulation procedure appears to work very well.

On average our ML estimates for the first and second moment come close to the theoretical ones. For the case where our simulated point lies outside $\mathscr{D}$ we notice that the first and second moments are not well estimated. The bias tends, however, to diminish the greater the kurtosis. To summarize our simulation results, for deviations from the true Gram–Charlier DGP, as long as the parameters are within the authorized domain, we have some difficulties to correctly capture skewness and kurtosis. For parameters outside the authorized domain, even the first and second moments are badly estimated. Because of the restrictive shape of the density our parameter estimates will have difficulties to capture the moments. This observation highlights the importance to test if the GC specification is a correct one for the data at hand. Such a test can for instance be performed with a Kolmogorov–Smirnov test. Deviations from the non-conditional moments and the ones obtained with a Gram–Charlier density are suggestive of a model misspecification.

[14] See also Cohen (1967) or Holgersson and Jorner (1979) for a more modern derivation of the formulas. Bowman and Shenton (1973) discuss the space of moments for which moment estimators exist.

[15] Since Pearson's method involves a fifth moment, we first tried to seek a solution for a wide range of this fifth moment. Eventually we chose that 5th moment where p turned out to be closest to 0.25. This ensures that we simulate sufficiently often from both distributions. As Table 4 shows, we often have a boundary solution.

[16] Results not reported here.

Table 4
Estimation of Gram–Charlier parameters when the true DGP is a mixture of normals[a]

	Parameters of mixture distribution					Theoretical moments			Simulated moments		Average ML estimates				STD of ML estimates			
	p	$d1$	$d2$	$s12$	$s22$	s	k	M5	s	k	μ	σ	s	k	μ	σ	s	k
1	0.340	− 0.418	0.215	0.012	1.373	0.56	1.0	4.1	0.572	1.003	− 0.038	0.945	0.397	0.853	0.028	0.030	0.247	0.533
2	0.375	− 0.479	0.288	0.008	1.375	0.71	1.0	5.3	0.716	1.029	− 0.064	0.919	0.566	1.088	0.031	0.033	0.224	0.440
3	0.248	0.601	− 0.199	2.083	0.483	0.76	2.0	10.3	0.760	1.986	− 0.005	0.985	0.581	1.299	0.020	0.024	0.105	0.251
4	0.321	0.632	− 0.299	1.822	0.334	0.97	2.0	11.2	0.975	1.950	− 0.030	0.955	0.781	1.548	0.022	0.021	0.057	0.127
5	0.250	0.219	− 0.073	2.900	0.344	0.42	3.8	6.9	0.416	3.707	− 0.010	0.944	0.251	2.229	0.020	0.027	0.083	0.174
6	0.249	0.280	− 0.093	2.875	0.342	0.54	3.8	8.7	0.561	3.816	− 0.009	0.941	0.326	2.216	0.023	0.027	0.087	0.164
7	0.441	− 0.544	0.429	0.017	1.357	0.94	1.0	7.2	0.931	0.995	− 0.111	0.874	0.786	1.338	0.024	0.026	0.088	0.194
8	0.148	1.958	− 0.341	0.376	0.325	1.12	1.0	8.5	1.127	1.009	− 0.120	0.894	1.027	2.078	0.015	0.013	0.010	0.084
9	0.440	0.596	− 0.468	1.604	0.028	1.27	2.0	11.6	1.278	1.977	− 0.171	0.775	0.928	2.148	0.023	0.034	0.026	0.120
10	0.178	1.815	− 0.394	0.718	0.191	1.53	2.0	13.4	1.526	1.974	− 0.159	0.837	1.045	2.593	0.014	0.016	0.004	0.075
11	0.249	0.374	− 0.124	2.819	0.336	0.70	3.8	11.3	0.710	3.712	− 0.013	0.947	0.426	2.269	0.023	0.026	0.079	0.142
12	0.251	0.453	− 0.152	2.746	0.324	0.85	3.8	13.3	0.849	3.820	− 0.018	0.941	0.514	2.251	0.020	0.023	0.080	0.150

[a]This table presents the results of simulations where a GCd is fitted to data simulated as a mixture of normals. We simulated for each set of parameters 100 samples of length 2000. In columns 2–6 we present those parameters for the mixtures of normals yielding an expectation of 0 and a standard deviation of 1 and further higher moments displayed in columns 7–9. In column 10 and 11 we present averages of the third and fourth moments for the simulated data. (The average for the first two moments are virtually equal to the theoretical ones). Columns 12–19 are similar to columns 6–13 of Table 3. Rows 1–6 correspond to the same theoretical moments as in Table 3. Rows 7–12 correspond to those moments situated out of the positivity domain.

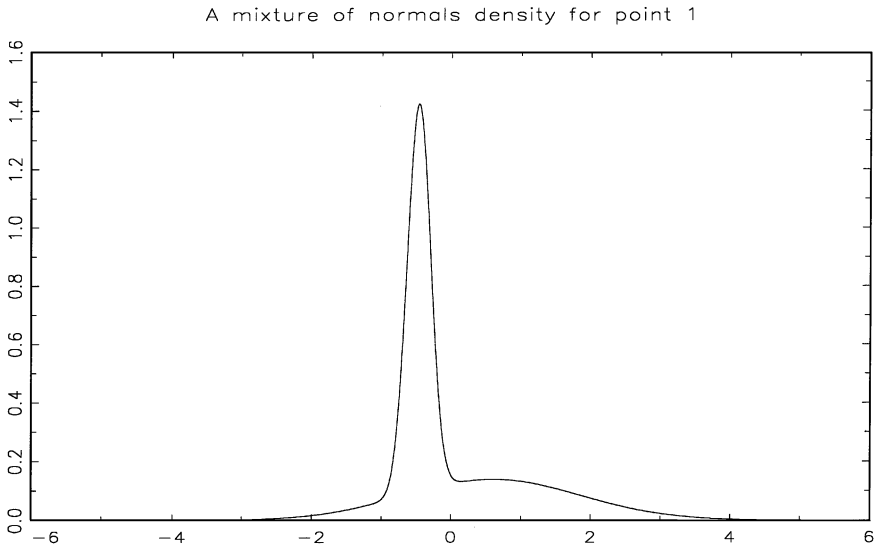

Fig. 7. Displays the density of a mixture of normals corresponding to the parameters of point 1 that is $\mu = 0$, $\sigma = 1$, $s = 0.56$, $k = 1$.

5.2. GARCH models with Gram–Charlier distributed innovations

Let us now turn to a second empirical illustration where our positivity restriction comes handy namely in situations where a Gram–Charlier distribution is used to model innovations in a GARCH model while maintaining the interpretation of the parameters s and k as the skewness and excess kurtosis of the density.

Models based on GARCH-type technology have recognized the possibility of time-changing volatility. First, Engle (1982) proposed his ARCH model. Bollerslev (1986) extended it to GARCH. Time-varying volatility has lead to a significant amount of literature summarized in Bollerslev et al. (1992), as well as in Bera and Higgins (1993). One difficulty with those models is that residuals often remain heavy tailed. Solutions have been proposed to account for this heavy-tailedness such as in Engle and Gonzales-Rivera's (1989) semi-parametric model, or using t-distributions (as in Bollerslev, 1986), or GED distributions (as in Nelson, 1991). In none of those models it is possible to access directly to the skewness and kurtosis parameters.

In the following model we keep the usual GARCH-type parameterization of volatility and for the innovations allow a skewness and kurtosis different from the ones of the normal density. Formally, we assume that S_t is the value of some asset at time t. We assume that the continuously compounded return, defined by

$r_t = 100 \cdot \ln(S_t/S_{t-1})$ may be described by

$$r_t = \mu_t + y_t, \tag{14}$$

$$y_t = \sigma_t z_t, \tag{15}$$

$$\sigma_t^2 = w + a y_{t-1}^2 + b\sigma_{t-1}^2. \tag{16}$$

The term μ_t in (14) corresponds to the conditional mean and y_t to the unexpected part of returns. The variable σ_t is the conditional volatility. In Eq. (16) we allow a GARCH(1,1) representation for the conditional volatility. More complicated processes could be trivially accommodated. In standard GARCH models, it is assumed that the innovation z_t follows a given distribution such as a $\mathcal{N}(0,1)$ or a student t-distribution with v degrees of freedom. Here, we assume that innovations are distributed as a Gram–Charlier density with skewness and excess kurtosis parameters s and k, respectively. Formally, this allows us to complete model (14)–(16) with

$$z_t \sim \mathcal{GC}(0, 1, s, k), \tag{17}$$

$$(s, k) = f(\tilde{s}, \tilde{k}), \tag{18}$$

where f is the mapping from $\mathcal{R}^2$ into $\mathcal{D}$ described in Section 3. The positiveness of $g(z)$ is not only a theoretical problem. Indeed, from a practical point of view, if expression (2) were negative, the log-likelihood is no longer defined and parameters could not be estimated. Therefore, when (s, k) is not in the domain $\mathcal{D}$, the log-likelihood can be actually undefined for some values of z.

5.2.1. The data used

In this study we focus on six foreign exchange rate series with respect to the US dollar: the British Pound (GBP), the Japanese Yen (JPY), the Deutsche Mark (DEM), the French Franc (FRF), and eventually the Canadian dollar (CAD). Our data covers the period from 03/01/1977 to 03/05/1999. We consider weekly data computed with the Friday closing price. The data got extracted from the Datastream service.

In Table 5 we present various descriptive statistics of the data at a daily frequency. For all series we dispose of 5826 observations. We compute, in the spirit of Richardson and Smith (1994), all four moments and associated standard errors with GMM, thus, allowing for possible heteroscedasticity in the data. This also yields a Wald-type test for normality, W, distributed as a χ_2^2. In the table we also present the more traditional Jarque–Bera, JB, test for normality. We notice that the mean return is small in absolute value. The standard deviation of returns is lowest for CAD. The JPY exchange rate had the highest volatility. The DEM and FRF exchange rates have very similar patterns for moments as could be expected. Turning to the skewness and kurtosis we notice for all series that there is a strong non-normality as one can check by looking at

Table 5
Descriptive statistics for the weekly foreign exchange data[a]

	GBP	YEN	DEM	FRF	CAD
mean	0.0047	− 0.0773	− 0.0212	0.0187	0.0324
STD(mean)	0.0427	0.0445	0.0436	0.0425	0. 0184
std	1.4573	1.5189	1.4886	1.4492	0. 6273
STD(std)	0. 0498	0.0477	0.0429	0.0447	0.0189
sk	0.2254	− 0.5962	− 0.1285	0.0063	− 0.0723
STD(sk)	0.2557	0.2115	0.1884	0.2145	0.2173
xku	3.4524	2.6008	1.8638	2.4241	2.2484
STD(xku)	0.8669	1.0382	0.6340	0.6825	0.7970
W	15.86	8.13	8.79	12.85	8.46
p-value	0.0004	0.0172	0.0123	0.0016	0.0146
JB	588.43	397.35	171.84	285.25	246.40
KS	2.12	2.18	1.43	1.40	1.34
Engle 5	88.25	44.13	47.53	40.81	64.59
AR(1)	0.030	0.066	0.040	0.040	− 0.015
AR(2)	0.005	0.112	0.045	0.047	0.025
$Q(5)$	1.30	4.83	1.33	1.50	1.29

[a]The first four moments and their associated standard deviation get estimated with GMM allowing for possible heteroscedasticity. sk and xku correspond to skewness and excess kurtosis. W is a test for normality presented with its p-value. JB and KS are the Jarque–Bera test, respectively, a Kolmogorov–Smirnov test for normality. Engle 5 is the Lagrange multiplier test for heteroscedasticity. AR and Q are the coefficients of autocorrelation and of the Box–Ljiung test for autocorrelation, respectively.

the high values for the Jarque–Bera statistics. Excess kurtosis is in all cases significantly larger than 3 implying that the unconditional density of all series has fatter tails than the normal distribution. The Engle statistic computed with 5 lags indicates strong heteroscedasticity in all the series. The Box–Ljung Q-statistics indicates that weekly returns generally appear to be uncorrelated.

5.2.2. Estimation results

Tables 6 and 7 present estimates of GARCH models with a normal density and a Gram–Charlier density, respectively, for the innovations.

The conditional mean has been estimated separately and is not reported here. Starting with Table 6, the parameter a indicates that subsequent to a large return volatility of the next period remains high. The parameter b indicates that a high volatility is followed by high volatility: As expected volatility is persistent. Furthermore, we estimate the skewness and excess kurtosis for the innovations.

Table 6
GARCH estimates under normality assumption[a]

	GBP	YEN	DEM	FRF	CAD
w	0.0526	0.2824	0.1637	0.0934	0.0358
STE(w)	0.0274	0.1315	0.0849	0.0533	0.0137
a	0.0990	0.0982	0.1186	0.1308	0.1079
STE(a)	0.0258	0.0326	0.0371	0.0408	0.0252
b	0.8814	0.7832	0.8125	0.8354	0.8048
STE(b)	0.0241	0.0729	0.0634	0.0463	0.0424
sk	0.32	− 0.66	− 0.02	0.12	0.13
sk∗	4.50	− 9.18	− 0.29	1.60	1.85
xku	2.38	2.58	1.21	1.53	1.88
xku∗	16.56	17.98	8.41	10.67	13.12
KS	1.53	2.02	1.20	1.22	1.08
Log-Lik	− 2007.50	− 2116.32	− 2079.19	− 2042.77	− 1079.54

[a]Sk* and xku* represent the t-ratios for sk and xku. KS represents the Kolmogorov–Smirnov test for normality. Log-lik is the sum of all log-likelihoods.

We notice for all situations that the kurtosis is significantly different from 0 and incompatible with a normal distribution. Our Kolmogorov–Smirnov statistic, KS, indicates a rejection of the assumption of normality for all series. Those results are well established and indicate that GARCH models should be modeled with distributions for the innovations allowing for unconditional fat-tailedness.

We now turn to the results reported in Table 7 where we have performed the estimations with the Gram–Charlier density. The parameters for w, a, and b are similar to the ones of Table 6. We also report the estimates of skewness and kurtosis. We notice that all the estimated skewness and kurtosis lay in the authorized domain $\mathscr{D}$. Nonetheless, when trying to estimate the likelihood function without the restrictions on skewness and kurtosis, in many situations the algorithm crashed because the likelihood became negative. Residuals are still found to be non-normal. When turning to the Kolmogorov–Smirnov statistics which tests if the residuals have a behavior compatible with the Gram–Charlier density we cannot reject this hypothesis.

For the Deutsche Mark exchange rate series we present in Fig. 8 a plot with a normal density whose moments are matched to the ones of the innovations, a Kernel estimation of the density of the innovations, and the fitted Gram–Charlier density. This figure corroborates our statistical finding that the Gram–Charlier density is an improvement over the normal one.

Table 7
GARCH estimates with Gram–Charlier density[a]

	GBP	YEN	DEM	FRF	CAD
w	0.0529	0.2034	0.1305	0.0620	0.0326
STE(w)	0.0298	0.1083	0.0775	0.0353	0.0118
a	0.1098	0.0832	0.1170	0.1284	0.1143
STE(a)	0.0351	0.0295	0.0385	0.0437	0.0256
b	0.8756	0.8354	0.8321	0.8579	0.8113
STE(b)	0.0303	0.0612	0.0611	0.0406	0.0363
s	0.1885	− 0.3724	− 0.1483	− 0.0410	− 0.0142
STE(s)	0.0880	0.1003	0.0968	0.0994	0.0621
k	0.9811	1.1231	0.7015	0.9639	0.8152
STE(k)	0.2258	0.2201	0.2034	0.2320	0.2183
sk	0.32	− 0.67	− 0.02	0.11	0.14
sk∗	4.41	− 9.33	− 0.31	1.58	1.92
xku	2.38	2.60	1.21	1.55	1.89
xku∗	16.56	18.09	8.44	10.76	13.15
KS(Norm)	1.52	2.02	1.23	1.21	1.08
KS(GC)	0.70	1.05	0.63	0.54	0.61
Log-Lik	− 1986.77	− 2081.40	− 2066.73	− 2023.46	− 1064.50
LRT	41.45	69.84	24.92	38.63	30.08

[a]In addition to the parameters already appearing in Table 6, KS(normal), KS(GC), and Log-Lik represent the Kolmogorov–Smirnov tests for normality, for data to be generated as a Gram–Charlier density, and the log-likelihood value. LRT represents the likelihood-ratio test statistics that the conditional density of residuals is a Gram–Charlier density versus a normal one.

In all situations we reject with a likelihood-ratio test the restriction of a normal density. As a first conclusion we, therefore, notice that the use of the Gram–Charlier density is a success from a statistical point of view. As can be expected, we obtain in general a decrease of the parameters' standard errors. Our estimation is therefore slightly more efficient. On the negative side we notice that the estimates of the skewness and kurtosis parameters for the residuals differ from the ML ones of the GCd. This result, in light of our earlier simulations, suggests that even though the KS statistics does not reject the GCd, our model remains misspecified. In particular it is possible that there remains heteroscedasticity of higher order in the data. It is therefore possible that also skewness and kurtosis need to be made time varying. Within our framework this can be done in a natural way by following Hansen (1994), but is left for further research.

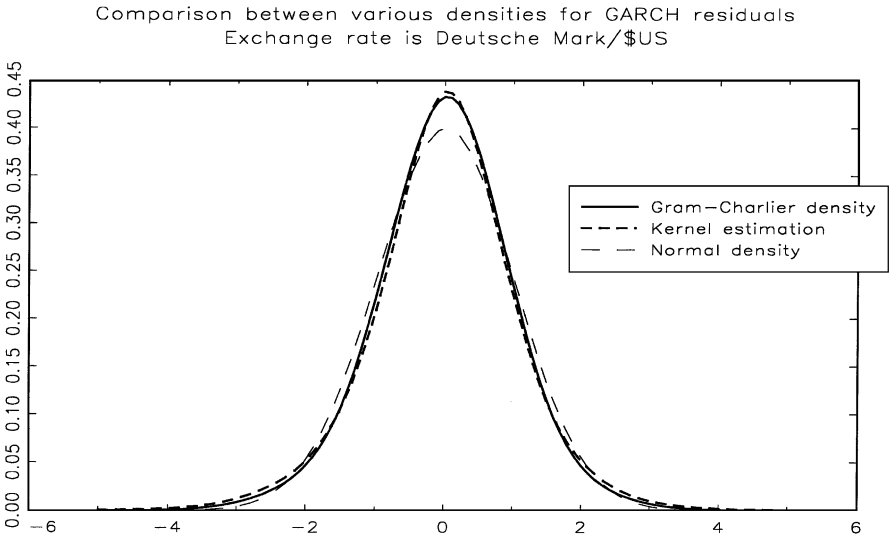

Fig. 8. Displays for the Deutsche Mark to $US series a GARCH regression with a Gram–Charlier density for the innovations. We also present a Kernel estimation of the density of the estimated innovations as well as a normal density with parameters equal to the ones of the innovations.

6. Conclusion

Gram–Charlier expansions are useful to model densities which are deviations from the normal one. In addition to the mean and standard deviation that characterize the normal density, for Gram–Charlier expansions, the third and fourth moments (skewness and kurtosis) are also characterizing elements. In this paper we determine the domain of skewness and kurtosis over which the expansion is positive. Imposing this positivity constraint allows us to talk of Gram–Charlier densities (GCd). We indicated how this constraint can be imposed numerically with a simple mapping and that the unconstrained optimum will be uniquely related to the constrained one.

We apply our method to the estimation of Risk–Neutral densities that arise in an option pricing context and to the estimation of GCds within a GARCH model. In both estimations an unconstrained optimization would have been problematic. Risk–Neutral densities might have been negative and Gram–Charlier densities impossible to fit to GARCH innovations because of the impossibility to compute log-likelihoods. Both types of estimations are very fast and numerically stable once the positivity constraint got imposed.

In the section dealing with the maximum-likelihood estimation of GCds we validate our procedure and notice the following two observations: First, a fit of

a GCd to Gram–Charlier distributed data yields unbiased estimates as long as kurtosis is not too small. Thus, as in other statistical estimations, the fit of a generalization of the normal density becomes more difficult for small deviations from a normal density. Second, when fitting a GCd to data generated with a mixture of moments we notice difficulties in capturing the correct moments. This highlights the importance of testing if the data is compatible with a GCd.

Our GARCH estimation reveals a large improvement in terms of likelihood-ratio tests. Further improvements left for future research could involve a time-varying specification of skewness and kurtosis where those moments would be linked in a GARCH-type specification to the third and fourth moment of innovations.

References

Abken, P., Madan, D.B., Ramamurtie, S., 1996. Estimation of risk-neutral and statistical densities by Hermite polynomial approximation: with an application to eurodollar futures options. Mimeo, Federal Reserve Bank of Atlanta.

Barton, D.E., Dennis, K.E.R., 1952. The conditions under which Gram–Charlier and Edgeworth curves are positive definite and unimodal. Biometrika 39, 425–427.

Bera, A.K., Higgins, M.L., 1993. ARCH models: properties, estimation and testing. Journal of Economic Surveys 7 (4), 305–366.

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 31, 307–328.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH Modeling in Finance. Journal of Econometrics 52, 5–59.

Bowman, K.O., Shenton, L.R., 1973. Space of solutions for a normal mixture. Biometrika 60, 629–636.

Breeden, D., Litzenberger, R., 1978. Prices of state-contingent claims implicit in option prices. Journal of Business 51, 621–651.

Cohen, A.C., 1967. Estimation in mixtures of two normal distributions. Technometrics 9, 15–28.

Corrado, C.J., Su, T., 1996. S&P 500 index option tests of Jarrow and Rudd's approximate option valuation formula. Journal of Futures Markets 6, 611–629.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom Inflation. Econometrica 50, 987–1007.

Gallant, A.R., Tauchen, G., 1989. Semi-nonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. Econometrica 57, 1091–1120.

Garman, M., Kohlhagen, S., 1983. Foreign currency option values. Journal of International Money and Finance 2, 231–238.

Gradshteyn, I.S., Ryzhnik, I.M., 1994. Table of Integrals, Series, and Products, 5th Edition. Academic Press, New York.

Hansen, B., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.

Harrison, J.M., Pliska, S., 1981. Martingales and stochastic integrals in the theory of continuous trading. Stochastic Processes and their Applications 11, 215–260.

Jarrow, R., Rudd, A., 1982. Approximate valuation for arbitrary stochastic processes. Journal of Financial Economics 347–369.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous univariate distribution, 2nd Edition, Vol. 1. Wiley, New York.

Jondeau, E., Rockinger, M., 1998. Reading the smile: the message conveyed by methods which infer risk neutral density. CEPR Discussion Paper no. 2009.

Knight, J., Satchell, S., 1997. Pricing derivatives written on assets with arbitrary skewness and kurtosis. Mimeo, Trinity College.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59 (2), 347–370.

Pearson, K., 1894. Contribution to the mathematical theory of evolution. Proc. Trans. Royal Soc. A 185, 71–110.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1988. Numerical Recipes in C. Cambridge University Press, Cambridge.

Richardson, M., Smith, T., 1994. A direct test of the mixture of distributions hypothesis: measuring the daily flow of information. Journal of Financial and Quantitative Analysis 29, 101–116.

Ripley, B., 1987. Stochastic Simulation. Wiley, New York.