Article:

Abadir, K.M. and Rockinger, M. (2003) Density functionals, with an option-pricing application. Econometric Theory. pp. 778-811. ISSN 0266-4666

https://doi.org/10.1017/S0266466603195047

Econometric Theory, **19**, 2003, 778–811. Printed in the United States of America.
DOI: 10.1017/S0266466603195047

DENSITY FUNCTIONALS, WITH AN OPTION-PRICING APPLICATION

Karim M. Abadir
University of York

Michael Rockinger
HEC Lausanne,
CEPR,
and FAME

We present a method of estimating density-related functionals, without prior knowledge of the density's functional form. The approach revolves around the specification of an explicit formula for a new class of distributions that encompasses many of the known cases in statistics, including the normal, gamma, inverse gamma, and mixtures thereof. The functionals are based on a couple of hypergeometric functions. Their parameters can be estimated, and the estimates then reveal *both* the functional form of the density *and* the parameters that determine centering, scaling, etc. The function to be estimated always leads to a valid density, by design, namely, one that is nonnegative everywhere and integrates to 1. Unlike fully nonparametric methods, our approach can be applied to small datasets. To illustrate our methodology, we apply it to finding risk-neutral densities associated with different types of financial options. We show how our approach fits the data uniformly very well. We also find that our estimated densities' functional forms vary over the dataset, so that existing parametric methods will not do uniformly well.

1. INTRODUCTION

It is often the case that one wishes to estimate a probability density function (p.d.f.) of a variate, without prior knowledge of its functional form. If the variate is directly observable, a number of parametric and nonparametric methods

We thank Hans-Jürg Büttler, Aleš Černý, Tony Culyer, Les Godfrey, David Hendry, Sam Kotz, Steve Lawford, Peter Phillips, Bas Werker, and three anonymous referees for their comments. We also thank for their feedback the participants at the seminars and conferences where this paper has been invited, in particular the 1998 CEPR Finance Network Workshop, the 1998 METU conference, the 1998 FORC (Warwick) conference "Options: Recent Advances," Money Macro & Finance Group, the Swiss National Bank, Imperial College, Tilburg University, Université Libre de Bruxelles, the University of Oxford, Southampton University, and UMIST. The first author acknowledges support from the ESRC (UK) grant R000239538. The second author acknowledges help from the HEC Foundation and the European Community TMR grant "Financial Market Efficiency and Economic Efficiency." This paper was written when the second author was affiliated with HEC-Paris. Address correspondence to: Karim M. Abadir, Department of Mathematics and Department of Economics, University of York, Heslington, York YO10 5DD, England; e-mail: kma4@york.ac.uk.

  0266-4666/03 $12.00

are already available for estimating the density. If, however, the variate is not observable, and its density can only be extracted indirectly, the problem is more complicated. In this paper, we propose a method that can be used in the case of either type of variate. Its advantages are illustrated with the following application.

1.1. A Motivating Problem

Let S_t be the price of an asset at time t. To illustrate, we assume for the moment that the asset does not pay dividends and that it is domestic, i.e., no considerations need to be paid to foreign interest rates. Later on, we will consider an extension to foreign currency options.

Suppose that this asset is underlying a European call option with expiration date T and strike price K. Then, the intrinsic value of this option at expiration is $V_T = \max(S_T - K, 0)$. In an arbitrage-free economy, it is known (see Harrison and Pliska, 1981) that there exists a risk-neutral density (RND) $g(\cdot)$ such that the price of a call option can be written as

$$C_t(K) = \mathrm{e}^{-r(T-t)}\mathrm{E}_t(V_T) = \mathrm{e}^{-r(T-t)} \int_K^{\infty} (S_T - K) g(S_T)\, \mathrm{d}S_T, \tag{1}$$

where $\mathrm{E}_t(\cdot)$ is the expectations operator conditional on all information available at time t, $C_t(\cdot)$ is the price at time t of the call option, and r is the continuously compounded risk-free interest rate. The function $C_t(\cdot)$ depends on the parameters r, T, t and also on others characterizing the process followed by S_t. In an arbitrage-free economy we also have the martingale condition

$$S_t = \mathrm{e}^{-r(T-t)}\mathrm{E}_t(S_T), \tag{2}$$

where the expectation is taken with respect to the RND $g(\cdot)$.

As noticed by Breeden and Litzenberger (1978), differentiating the integral gives

$$\frac{\mathrm{d}}{\mathrm{d}K} C_t(K) = -\mathrm{e}^{-r(T-t)} \int_K^{\infty} g(S_T)\, \mathrm{d}S_T \equiv -\mathrm{e}^{-r(T-t)}[1 - G(K)], \tag{3}$$

where $G(\cdot)$ is the cumulative distribution function (c.d.f.) corresponding to the p.d.f. $g(\cdot)$. The second derivative is given by

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}K^2} C_t(K)\right|_{K=S_T} = \mathrm{e}^{-r(T-t)} g(S_T),$$

which reveals the required density $g(\cdot)$. It is convenient to work with the future value

$$C(K) \equiv \mathrm{e}^{r(T-t)} C_t(K) \tag{4}$$

whose second derivative is the required density function.

If the asset were to return a yield of r^*, or if we were to consider a currency call option with foreign interest rate r^*, the martingale condition of the arbitrage-free economy (2) would be

$$S_t = \mathrm{e}^{-(r-r^*)(T-t)} \mathrm{E}_t(S_T) \tag{2'}$$

instead of (2). Of course, no alterations are necessary in formulas (1), (3), and (4), where r is the unchanged discount factor used to calculate present values.

One way to estimate the density is to postulate a parametric form for $g(\cdot)$, work out analytically the corresponding integrals leading to an explicit $C(\cdot)$, and then fit the observed option prices $C(K)$ to the strike prices K to determine the parameters of $g(\cdot)$. Mixtures of lognormal distributions have been used by Bahra (1996), Melick and Thomas (1997), and Söderlind and Svensson (1997). It is possible to assume that the underlying process is more general than a lognormal diffusion (for jump-diffusion versions, see Bates, 1996a, 1996b; Malz, 1996).

Alternatively, a nonparametric approach may be adopted. Jarrow and Rudd (1982) have developed a method based on an Edgeworth expansion involving a lognormal density. This approach has been implemented by Corrado and Su (1996) to price options. Using a system of Hermite polynomials, Madan and Milne (1994) suggest a method of approximating the underlying RND, and Abken, Madan, and Ramamurtie (1996) provide an application. See also Knight and Satchell (1997) and Jondeau and Rockinger (2001), where Gram–Charlier expansions are used. Stutzer (1996) uses a Bayesian method based on the maximum-entropy principle of Shannon. Finally, in a time series context where large datasets are available, Aït-Sahalia and Lo (1998) fit kernel-based estimates to the RND. See also Bondarenko (2000). For a survey assessing existing methods, see Jondeau and Rockinger (2000). Also, numerical methods can be used to estimate functionals of densities and/or to simulate option prices, but these have not been used to generate RNDs (e.g., see Wilmott, Dewynne, and Howison, 1993).

Improvements over these approaches are possible, because they are either restrictive (lognormals or jump-diffusions), do not always yield positive probabilities (Edgeworth expansions or Hermite polynomials), or require a large number of observations (kernel estimates). Furthermore, the methodology we are about to suggest is general, and it may be applied to other problems not necessarily related to finance or economics.

1.2. The Plan

The initial problem is to design a general function whose second derivative is a class of density functions. We derive such a function, and find that it is based on the confluent hypergeometric ${}_1F_1$ function.[1] Examples of recent uses of this function in economics include the option-pricing approach to investment

(Dixit and Pindyck, 1994) and pricing of callable bonds (Büttler and Waldvogel, 1996). Also, Abadir and Lucas (2000) show that this function is necessary to represent densities associated with the minimal sufficient functionals of Ornstein–Uhlenbeck processes, and all the related processes satisfying an invariance principle such as the functional central limit theorems in Phillips (1987). In these applications, a reason for the success of ${}_1F_1$ is that it includes as special cases the incomplete-gamma and the normal distribution functions, in addition to mixtures of the two. Also, iterated integrals and/or derivatives of ${}_1F_1$ give mixtures of ${}_1F_1$, which makes certain classes of these functions closed under such operations. These features suggest them as a natural tool to model option prices and, more generally, functionals of densities.

For practical purposes, parsimony of the model is important. For example, considering the data in our motivating problem, there are not many strike prices available. Our approach is of a parsimonious semi-nonparametric nature, closest in spirit to fitting a system of orthogonal Hermite polynomials. However, we do not estimate a system but rather the parameters of a couple of functions. The functions we use include Laguerre and Hermite polynomials as special cases, with the added advantage that Abadir's (1993a, 1999) "fractional" polynomials allow for monotonic behavior at the tails of the density and thus do not suffer from the forced oscillatory nature that standard polynomials have for extreme values. Within our model, parameter estimates determine the functional form, in addition to the usual distributional properties (e.g., centering and scaling). This is more efficient than fully nonparametric estimation, which runs into difficulties in small samples. It is also more flexible than parametric methods that restrict functional forms.

We do not restrict the functional form of the density to a single type. The only restriction we impose is that it must belong to some family of densities to be specified in Section 2, which includes exponential and degenerate nonexponential cases. This new class of closed-form densities that we are proposing can have applications in other areas of statistics too and can be used to characterize a broad selection of continuous random variables. For example, this class can be used to provide an alternative approach to fully nonparametric density estimation. Another application would be to model nonnormal densities in generalized autoregressive conditional heteroskedasticity (GARCH) models, etc. Our class of densities does not assume the existence of moments, thus avoiding the problem of basing an estimation procedure on calculated "moments" that may be spurious.

In the next section we introduce our density functional, based on ${}_1F_1$, for the case of double integrals of densities. In Section 3 we illustrate this methodology with two applications in option pricing. We also discuss how the empirical findings can be used for practical purposes. In Section 4, we generalize the methodology to the case of observing any functional (not necessarily double integrals) of densities. We conclude in Section 5. Proofs and lengthy derivations are collected in the Appendix.

We denote the set of natural numbers (which excludes zero) by $\mathbb{N}$ and real numbers by $\mathbb{R}$. The indicator function is written as $1_{\mathcal{K}}$, returning 1 if condition $\mathcal{K}$ is satisfied and 0 otherwise. The (complete) gamma function is denoted by $\Gamma(\nu)$ for $\nu \in \mathbb{R}$, and defining

$$(a)_j \equiv (a)(a+1)\dots(a+j-1) = \frac{\Gamma(a+j)}{\Gamma(a)}$$

leads to the generalized hypergeometric function

$${}_pF_q\left(\begin{matrix}\alpha_1,\dots,\alpha_p;\\ \beta_1,\dots,\beta_q;\end{matrix} z\right) \equiv \sum_{j=0}^{\infty} \frac{\prod_{k=1}^{p}(\alpha_k)_j}{\prod_{k=1}^{q}(\beta_k)_j} \frac{z^j}{j!}, \tag{5}$$

where $-\beta_{\cdot} \notin \mathbb{N} \cup \{0\}$. Special cases that we will be discussing frequently are

$$\gamma(\nu, z) \equiv \int_0^z \mathrm{e}^{-x} x^{\nu-1}\,\mathrm{d}x \equiv \frac{z^\nu}{\nu}\,{}_1F_1(\nu;\nu+1;-z), \qquad -\nu \notin \mathbb{N}\cup\{0\}, \tag{6}$$

$$\begin{aligned}\Phi(z) &\equiv \int_{-\infty}^{z} \mathrm{e}^{-x^2/2}\frac{\mathrm{d}x}{\sqrt{2\pi}} \equiv \frac{1}{2} + \frac{z}{\sqrt{2\pi}}\,{}_1F_1\left(\frac{1}{2};\frac{3}{2};-\frac{z^2}{2}\right)\\ &\equiv \frac{1}{2} + \frac{\operatorname{sgn}(z)}{2\sqrt{\pi}}\,\gamma\left(\frac{1}{2},\frac{z^2}{2}\right),\end{aligned} \tag{7}$$

where $\gamma(\cdot,\cdot)$ is the incomplete-gamma function, $\Phi(z)$ is the standard normal c.d.f., and $\operatorname{sgn}(z)$ is the sign function.

2. THE DESIGN OF DENSITY FUNCTIONALS

The ${}_1F_1$ function can represent a variety of density-related functions, and we will use it in Section 2.1 to propose a generalization of the normal, gamma, and other variates. We will briefly outline how our functional is constructed and how it reduces to some well-known continuous random variables in statistics. Sections 2.2 and 2.3 derive, respectively, the restrictions necessary for the density functional to have a proper underlying density and the moments implied by this function. Section 2.4 discusses strategies for the estimation of our function.

2.1. Specification and Main Special Cases

We need to model integrals of c.d.f.s and possibly mixtures (integrals) thereof. Integrating n times the functions $\gamma(\nu, z)$ of (6)

$$\int\dots\int \gamma(\nu,z)(\mathrm{d}z)^n = \sum_{j=0}^{\infty}\frac{(-1)^j z^{j+\nu+n}}{j!(j+\nu)_{n+1}} = \frac{z^{\nu+n}}{(\nu)_{n+1}}\,{}_1F_1(\nu;\nu+n+1;-z) \tag{8}$$

and $\Phi(z)$ of (7)

$$\int \cdots \int \Phi(z)(\mathrm{d}z)^n = 2^{-(n/2)-1}\left[\frac{z\sqrt{2}}{\Gamma\left(\frac{n+1}{2}\right)}\,{}_1F_1\left(\frac{1-n}{2};\frac{3}{2};-\frac{z^2}{2}\right)\right.$$
$$\left.+\frac{1}{\Gamma\left(\frac{n}{2}+1\right)}\,{}_1F_1\left(-\frac{n}{2};\frac{1}{2};-\frac{z^2}{2}\right)\right]. \tag{9}$$

The latter is obtained by means of the parabolic cylinder function (see Abadir, 1993b, 1999), and integrals of negative orders (i.e., derivatives) yield the well-known Hermite polynomials. Generally, parabolic cylinder functions are linear combinations of ${}_1F_1$ functions that would allow for fractional n in (9) and are closed under differentiation and integration: these operations keep the result within the same class of functions. A weaker version of this property applies to other linear combinations of ${}_1F_1$ too, and this makes them appealing in modeling an arbitrary number of functionals of a class of densities that encompasses the normal, gamma, and others. We defer introducing the case of an arbitrary number of functionals until later in Section 4, because of the added level of difficulty that it poses.

For double integrals of densities, a mixture that extends the cases seen earlier is given by

$$C(z) \equiv c_1 + c_2 z + 1_{z>m_1} a_1 (z-m_1)^{b_1}\,{}_1F_1(a_2;a_3;b_2(z-m_1)^{b_3})$$
$$+ (a_4)\,{}_1F_1(a_5;a_6;b_4(z-m_2)^2), \tag{10}$$

where $-a_3, -a_6 \notin \mathbb{N} \cup \{0\}$, $b_2, b_4 \in \mathbb{R}_-$. The indicator function is required to represent a component of the density with bounded support. It is also sufficient for keeping the function real-valued for general b_1 and b_3.

Not all the parameters in $C(\cdot)$ are free to vary unrelatedly, because some restrictions (at least three in general and at least seven in our motivating example) are needed for the function to be the integral of a c.d.f., and we shall analyze these restrictions in detail in our next section. Now, we analyze the relation of $C(\cdot)$ to familiar distributions.

The first two parameters, c_1 and c_2, drop out when differentiating twice to get the density, but they are needed in general either to represent the constants of indefinite integration or when one of the limits of definite integration is infinite. An example of the need for such constants is in (3), where the integral of the density is over (K,∞) rather than, say, $(-\infty, K)$. If the limits of integration of the functional are an interval that is a strict subset of the support of the density, then one would end up with the difference of two $C(\cdot)$ functions. We shall not consider this case explicitly, because it boils down to using $C(\cdot)$ twice.

The first ${}_1F_1$ function in $C(\cdot)$ covers the gamma and other asymmetric generalizations, whereas the second covers the case of symmetric quadratic-exponentials such as the normal. We do not go further to quartic-exponentials (or higher powers of even order) because they are unlikely to occur in practice unless the variate is almost degenerate. They are nevertheless covered approximately by b_3 in the first ${}_1F_1$ function. Examples of special cases giving integrals of known density functions include

$$\text{gamma: } a_1 = \frac{(-b_2)^{b_1-1}}{\Gamma(b_1+1)}, \qquad a_2 = b_1 - 1, \qquad a_3 = b_1 + 1, \qquad a_4 = 0, \qquad b_3 = 1;$$

$$\text{inverse gamma: } a_1 = \frac{-(-b_2)^{1-b_1}}{b_1\Gamma(2-b_1)}, \qquad a_2 = -b_1, \qquad a_3 = 2 - b_1, \qquad a_4 = 0, \qquad b_3 = -1;$$

$$\text{Weibull: } a_1 = -1, \qquad a_2 = \frac{1}{b_3}, \qquad a_3 = \frac{1}{b_3} + 1, \qquad a_4 = 0, \qquad b_1 = 1;$$

$$\text{normal: } a_1 = 0, \qquad a_4 = \frac{1}{2\sqrt{-b_4\pi}}, \qquad a_5 = -\frac{1}{2}, \qquad a_6 = \frac{1}{2};$$

$$\text{Pareto: } a_1 = \frac{m_1}{a_2}, \qquad a_3 = -m_1 b_2 \quad \text{with } b_2 \to \infty, \qquad a_4 = 0, \qquad b_1 = 0, \qquad b_3 = 1,$$

where standardization (e.g., centering around zero) is not imposed and the constants of integration c_1 and c_2 are to be determined by the problem at hand. For the listing of the normal, recall that the first term in (9) reduces to $z/2$ when $n = 1$. Extreme-value distributions are useful in the field of value-at-risk in finance, and they are a special case of our approach. For further degenerate special cases, see the discussion of confluences in Abadir (1999).[2] We have provided one illustration, the Pareto, which relies on the confluence

$$\lim_{b_2\to\infty} {}_1F_1(a_2; -m_1 b_2; b_2(z - m_1)) = {}_1F_0\left(a_2; 1 - \frac{z}{m_1}\right) \equiv \left(\frac{z}{m_1}\right)^{-a_2} \tag{11}$$

obtained from the expansion in (5). This is a case where the existence of higher order moments will hinge on the magnitude of a_2. By the definition of the Pareto density, we require in addition that $a_2 \in (-1,\infty)$ and $m_1 \in \mathbb{R}_+$. When estimating the function $C(\cdot)$, it is worth testing for its reduction to known simpler cases such as those outlined previously and others in Johnson, Kotz, and Balakrishnan (1994, 1995).

Given the fact that our method encompasses a large class of traditional densities and is flexible, we believe that the potential for misspecification is small. Of course, misspecification is always a possibility, even with the most flexible of methods that assume the least (e.g., fully nonparametric methods). For example, let ζ and ξ be two independent standard normal variates. Then, the product $\eta \equiv \zeta\xi$ has a density that is infinite at the origin and that could be represented by hypergeometric functions of the sort we use (see Abadir, 1999, p. 302). If one were to fit a nonparametric density to data generated for η, then the kernels would smooth (build a "bridge") over the origin in spite of the true density being discontinuous there. Our method avoids such assumptions and encompasses a broad class of statistical densities, including ones with discontinuities.

2.2. Necessary Restrictions on the Parameter Space

The second derivative of the $C(\cdot)$ function is a mixture of the densities mentioned earlier. By differentiating termwise the ${}_1F_1$ series,

$$\begin{aligned} g(z) \equiv{}& \frac{\mathrm{d}^2}{\mathrm{d}z^2} C(z) \\ ={}& 1_{z>m_1} a_1 (z-m_1)^{b_1-2} \Bigg[b_1(b_1-1)\, {}_1F_1(a_2; a_3; b_2(z-m_1)^{b_3}) \\ & + \frac{a_2}{a_3} b_2 b_3 (2b_1 + b_3 - 1)(z-m_1)^{b_3} \\ & \times {}_1F_1(a_2+1; a_3+1; b_2(z-m_1)^{b_3}) \\ & + \frac{a_2(a_2+1)}{a_3(a_3+1)} b_2^2 b_3^2 (z-m_1)^{2b_3} \\ & \times {}_1F_1(a_2+2; a_3+2; b_2(z-m_1)^{b_3}) \Bigg] \\ & + 2a_4 \frac{a_5}{a_6} b_4 \Bigg[{}_1F_1(a_5+1; a_6+1; b_4(z-m_2)^2) \\ & + 2\,\frac{a_5+1}{a_6+1} b_4 (z-m_2)^2\, {}_1F_1(a_5+2; a_6+2; b_4(z-m_2)^2) \Bigg]. \end{aligned} \tag{12}$$

The estimate of $g(z)$ that is implied from the estimate of $C(z)$ should be a density. It should be nonnegative over its support, say, $(z_\ell, z_u) \subseteq \mathbb{R}$, and integrate to 1.

The nonnegativity restriction is hardly ever binding near the optimum parameter estimates, given the design of our function $C(\cdot)$. Nevertheless, it can be imposed on the estimation routine by a Lagrangian penalty function involving $\min(0, g(z))$, if the optimum $g(\cdot)$ is found to be negative. We did not need to do so in our applications.

For the restriction on the integral of the density, we derive

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z}C(z) = c_2 + 1_{z>m_1} a_1 (z-m_1)^{b_1-1}\Big[&(b_1)_1F_1(a_2;a_3;b_2(z-m_1)^{b_3})\\ &+\frac{a_2}{a_3}b_2b_3(z-m_1)^{b_3}\\ &\times {}_1F_1(a_2+1;a_3+1;b_2(z-m_1)^{b_3})\Big]\\ + 2a_4\frac{a_5}{a_6}b_4(z-m_2)_1F_1(a_5+1;a_6+1;b_4(z-m_2)^2).\end{aligned} \tag{13}$$

This expression is linear in, at least, c_2 and a_4 so that the two restrictions that ensure the proper choice of constants of integration, namely,

$$\int_{z_\ell}^{z_\ell} g(z)\,\mathrm{d}z = 0,$$

$$\int_{z_\ell}^{z_u} g(z)\,\mathrm{d}z = 1,$$

give rise to an explicit constraint on each of c_2 and a_4. These can be directly substituted into our $C(z)$ and there are two fewer parameters to estimate. The two restrictions will be illustrated in the Appendix.

We have mentioned restrictions on the p.d.f. (positivity) and then two more on the c.d.f. It is also natural that the integral of the c.d.f. will necessitate one more restriction, namely, one to do with the constant of integration, which is application-specific. It takes the form of a boundary condition on $C(\cdot)$, e.g., $C(\infty)$ equals some fixed value. Such conditions are often known in economics (especially in growth theory) as transversality conditions and will be illustrated in the Appendix. Because $C(\cdot)$ is linear in c_1, this condition implies an explicit restriction on c_1 that can be substituted directly into our function $C(z)$.

2.3. Explicit Characterization of the Moments

Finally, it is useful to characterize the moments of the density. One of the reasons for doing so could be the desire to investigate and/or impose restrictions on the moments of the density if some theory (e.g., arbitrage pricing theory) requires them. Another reason may be the desire to estimate our function $C(\cdot)$

by fitting the data to its analytical moments, which is discussed later in Section 2.4.

Subject to these functions being nondegenerate (i.e., the existence condition for the moments),

$$\begin{aligned}
\mathrm{E}(z^n) &= \int_{z_\ell}^{z_u} z^n \,\mathrm{d}G(z) \\
&= \int_{z_\ell}^{z_u} z^n \,\mathrm{d}\,\frac{\mathrm{d}C(z)}{\mathrm{d}z} \\
&= \left(z^n \frac{\mathrm{d}C(z)}{\mathrm{d}z} - n z^{n-1} C(z) \right) \Bigg|_{z_\ell}^{z_u} + n(n-1) \int_{z_\ell}^{z_u} z^{n-2} C(z)\,\mathrm{d}z
\end{aligned} \tag{14}$$

by integrating by parts two times. For $n = 1$ and assuming $m_1 \in (z_\ell, z_u)$, this gives

$$\begin{aligned}
\mathrm{E}(z) &= \left(z \frac{\mathrm{d}C(z)}{\mathrm{d}z} - C(z) \right) \Bigg|_{z_\ell}^{z_u} \\
&= a_1 (z_u - m_1)^{b_1 - 1} \Bigg[((b_1 - 1) z_u + m_1)\, {}_1F_1(a_2; a_3; b_2 (z_u - m_1)^{b_3}) \\
&\qquad + \frac{a_2}{a_3} b_2 b_3 z_u (z_u - m_1)^{b_3} \\
&\qquad \times {}_1F_1(a_2 + 1; a_3 + 1; b_2 (z_u - m_1)^{b_3}) \Bigg] \\
&\quad - a_4 \Bigg[{}_1F_1(a_5; a_6; b_4 (z - m_2)^2) \\
&\qquad - 2 \frac{a_5}{a_6} b_4 z (z - m_2)\, {}_1F_1(a_5 + 1; a_6 + 1; b_4 (z - m_2)^2) \Bigg] \Bigg|_{z_\ell}^{z_u},
\end{aligned} \tag{15}$$

by substituting from (13) and (10), respectively. This expression is linear in a_1 and can be used to reduce further the number of parameters to estimate, if there are reasons to believe that $\mathrm{E}(z)$ needs to be restricted, e.g., as a result of the no-arbitrage condition (2) in our example. This, along with a simplification of the expression, will be done in the Appendix.

For $n > 1$, (14) requires us to work out an explicit formula for the integral. Substituting for $C(z)$ from (10) and integrating termwise,

$$
\begin{aligned}
&\int_{z_\ell}^{z_u} z^{n-2} C(z)\,\mathrm{d}z \\
&\quad = \frac{c_1}{n-1}\left(z_u^{n-1} - z_\ell^{n-1}\right) + \frac{c_2}{n}\left(z_u^n - z_\ell^n\right) \\
&\qquad + a_1 \int_{m_1}^{z_u} z^{n-2}(z-m_1)^{b_1}\,{}_1F_1(a_2;a_3;b_2(z-m_1)^{b_3})\,\mathrm{d}z \\
&\qquad + a_4 \int_{z_\ell}^{z_u} z^{n-2}\,{}_1F_1(a_5;a_6;b_4(z-m_2)^2)\,\mathrm{d}z \\
&\quad = \frac{c_1}{n-1}\left(z_u^{n-1} - z_\ell^{n-1}\right) + \frac{c_2}{n}\left(z_u^n - z_\ell^n\right) \\
&\qquad + a_1 \sum_{j=0}^{n-2}\binom{n-2}{j} m_1^{n-2-j} \sum_{k=0}^{\infty} \frac{(a_2)_k b_2^k}{(a_3)_k k!} \int_{m_1}^{z_u} (z-m_1)^{b_1+j+b_3 k}\,\mathrm{d}z \\
&\qquad + a_4 \sum_{j=0}^{n-2}\binom{n-2}{j} m_2^{n-2-j} \sum_{k=0}^{\infty} \frac{(a_5)_k b_4^k}{(a_6)_k k!} \int_{z_\ell}^{z_u} (z-m_2)^{j+2k}\,\mathrm{d}z \\
&\quad = \frac{c_1}{n-1}\left(z_u^{n-1} - z_\ell^{n-1}\right) + \frac{c_2}{n}\left(z_u^n - z_\ell^n\right) \\
&\qquad + a_1 \sum_{j=0}^{n-2}\binom{n-2}{j} \frac{m_1^{n-2-j}(z_u-m_1)^{b_1+j+1}}{j+b_1+1} \\
&\qquad \times {}_2F_2\left(\begin{matrix} a_2, \dfrac{j+b_1+1}{b_3}; \\ a_3, \dfrac{j+b_1+1}{b_3}+1; \end{matrix}\; b_2(z_u-m_1)^{b_3}\right) \\
&\qquad + a_4 \sum_{j=0}^{n-2}\binom{n-2}{j} \frac{m_2^{n-2-j}(z-m_2)^{j+1}}{j+1}\,{}_2F_2\left(\begin{matrix} a_5, \dfrac{j+1}{2}; \\ a_6, \dfrac{j+3}{2}; \end{matrix}\; b_4(z-m_2)^2\right)\Bigg|_{z_\ell}^{z_u}.
\end{aligned}
$$

Asymptotic expansions of ${}_pF_q$ can be used to simplify the expression further, for extreme values of z_ℓ or z_u, as is illustrated in the Appendix.

2.4. Strategies for Estimation and Inference

There is more than one possible method for estimating our function $C(\cdot)$. These include maximum likelihood (ML), generalized least squares (GLS), generalized method of moments (GMM), and other approaches. We only discuss these three. Other attractive estimation methods, such as M-estimators, can be analyzed in the same way if desired.

The pseudo-likelihood approach is one of the most appealing. In the simplest case where the errors of estimating our functional $C(\cdot)$ are spherical normals, we can fit the functional to the observed data by nonlinear least squares (LS). In general, however, the errors may be nonspherically distributed. If there was evidence of nonspherical errors, one could specify a covariance structure for the residuals. Keeping the assumption of a pseudo-Gaussian likelihood, one could then estimate the variance-covariance matrix, say, $\boldsymbol{\Omega}$, by an iterative (e.g., two-step) procedure. An analogous idea would apply to GLS estimation, albeit with a difference of a factor of $\frac{1}{2}\log(\det(\boldsymbol{\Omega}))$ for the objective function to be optimized.

Another potentially appealing approach would rely on the explicit formulas for the moments, which we have derived in the previous section. GMM estimation would be based on these expressions. However, we have an important concern with applying this method here. We would only recommend it if the user were willing to restrict estimation to a class of our underlying densities where the moments do exist. We have given an example of our class of densities, the Pareto, where some moments do not necessarily exist. Subject to this proviso, the appeal of GMM would be in the ease of obtaining an estimate of the covariance matrix for the moment restrictions.

Two theorems will now be derived for the case of a random sampling of N observed values of $C(z)$ and the nonlinear LS estimation of the model

$$C(z) = \hat{C}(z) + \hat{\varepsilon},$$

where estimates are denoted by a hat, ε has zero mean and variance $\omega^2 < \infty$. The conditions on ε are sufficient but not necessary for the theorems to hold. Moreover, when the sampling scheme is not independent and identically distributed (i.i.d.) then, for the theorems to be valid, we assume that the other methods of estimation discussed earlier are used with the correct covariance structure. Let $\boldsymbol{\theta}$ denote the vector of parameters of the function $C(z)$, which are defined in (10) subject to the (exclusion) restrictions outlined there and which we denote by $\boldsymbol{\theta} \in \Theta$. Then, we have the following result.

THEOREM 1. *The nonlinear* LS *estimators of* $\boldsymbol{\theta} \in \Theta$ *are consistent for any of the parameters that have a nonzero impact on the function* $C(z)$.

The last part of the theorem's statement refers to the situation when some component of $C(z)$ drops out, in which case some (irrelevant) parameters will not be estimated consistently. For example, if $a_1 = 0$, the parameters $m_1, a_2, a_3, b_1, b_2, b_3$ in (10) are neither identified nor estimated consistently. In such cases, however, these parameters have no effect at all on $C(z)$ and are of no interest.

We now turn to inference regarding the estimated parameters. For this, we restrict $\boldsymbol{\theta} \in \Theta$ further to $\boldsymbol{\theta} \in \tilde{\Theta} \subset \Theta$ by excluding the cases where either of a_1, a_2, b_2, a_5 is zero,[3] to keep standard asymptotic Gaussian inference. Otherwise, when a hypothesis on some parameters causes other parameters to disappear from the model, then nonstandard asymptotics arise (e.g., see Andrews and Ploberger, 1994; Hansen, 1996; and references therein). The case of nonstandard asymptotics is dealt with by Lawford (2001) in the context of general hypergeometric functions and is not treated here.

THEOREM 2. *The nonlinear* LS *estimators of* $\boldsymbol{\theta} \in \tilde{\Theta}$ *are asymptotically normal with mean* $\boldsymbol{\theta}$ *and covariance matrix*

$$\omega^2 \left(N \mathrm{E} \left(\frac{\partial C}{\partial \boldsymbol{\theta}} \frac{\partial C}{\partial \boldsymbol{\theta}'} \right) \right)^{-1},$$

where N is the sample size.

We need to qualify inferences in our applications of the next section with a warning, because small datasets are involved and asymptotic inference may be of limited relevance there. The issue of finite-sample inference within a general context is, as ever, an unresolved problem. Various possibilities for attempting to improve such inference exist, but we do not pursue them in this paper, and we rely instead on the well-documented robustness of F-tests (but not of t-ratios) (e.g., see Ali and Sharma, 1996; Godfrey and Orme, 2001; and their reference lists). Examples that can help improve finite-sample performance include resampling methods such as the bootstrap. The difficulties with bootstrapping in the current setup are that it would be computationally very expensive and that establishing analytically its validity for ${}_1F_1$ functions is beyond the scope of this paper. For examples on the inconsistency of the bootstrap, see Basawa, Mallik, McCormick, Reeves, and Taylor (1991), Young (1994), and Andrews (2000). Of particular relevance is the discussion in Young (1994, especially p. 385) about its failure in environments where moments need not exist (e.g., stable laws), which can be the case when our function collapses to the Pareto density (11).

3. APPLICATIONS TO OPTION PRICING AND RISK-NEUTRAL DENSITIES

3.1. The Empirical Setup

We illustrate our methodology with foreign exchange options and with S&P500 index options. Both types of options are known to exhibit deviations from lognormality and are likely to reveal an interesting pattern in terms of densities (for foreign exchange options, see Bates, 1996a, 1996b; for S&P500 options, see Rubinstein, 1994). Because our only goal in this part is to illustrate the usefulness of our method, rather than to describe of all available options, we represent the results for a few dates and maturities.

First, for foreign exchange data, we use over the counter (OTC) options. Even though there exist options listed on currency exchanges, the vast majority of currency options are traded in the OTC market. According to the Bank of International Settlement, in June 1999, the notional value of outstanding OTC currency derivatives was about 200 times larger than that of currency exchanges. This implies that foreign currency OTC markets have greater liquidity. The drawback is that OTC option prices are usually unavailable to the public. We were able to collect OTC data of European French franc/Deutsche mark (FF/DM) rate options, in addition to the current exchange rate, from a large French bank for one randomly selected date (17 May 1996) during which no particular event happened and also for another date occurring a few days after President Chirac announced a snap election (28 April 1997) that eventually led to a landslide victory of the then opposition party.

For a given time of the day, the bank that provided us with the data requested quotes from the other dealers in the market. Dealers are compelled to give a bid-ask quote. The bank then retained the best bid and ask quote for various strike prices. The fact that the bank retained dealer's quotations guarantees that the option prices are determined simultaneously. The situation would have been quite different if only prices for traded options were available. In such a situation reported prices might have been obtained at very different times of the day, possibly, for very different values of the underlying asset, in which case prices would have been stale. Also, for many strike prices there may not have been transactions. Similarly, we were given the exchange rate from the same source. For exchange-rate quotes, the mechanism is similar to OTC options in that dealers must quote a bid and ask price. This means that for the foreign currency options, the magnitude of the nonsynchronicity bias is small. The drawbacks of this market are that prices are not publicly available and that we were able to obtain data only for a few days.

For the first and second dates, we were able to obtain the bid and ask prices for options corresponding to ($N =$) 13 and 11 different strike prices, respectively. Even though one could, for the purpose of option pricing, compute RNDs for the bid or the ask side, we decided to follow initially the convention of using the average of the bid and ask price for each strike price. Taking the mid-spread is compatible with the use made by central bankers, who are mainly interested in the evolution of the density and its shape (e.g., see Campa, Chang, and Reider, 1997). Nevertheless, we check in Section 3.3 the importance of using the mid-spread rather than the bid or the ask prices. We report results on options with two different maturities, 1 and 3 months.

Second, for the S&P500 options, we downloaded the prefiltered data used and described in Aït-Sahalia and Lo (1998). The options are European, cover 1993, and for each date there are various maturities available. Unlike our foreign exchange options, these index options are characterized by many strikes and a high liquidity. We selected three dates randomly and chose a maturity with a large number of strikes.

The benchmark option pricing model assumes that the underlying asset's price process follows a lognormal diffusion with constant volatility σ. In this case, European-type call options are traditionally priced with

$$C_t(K) = \mathrm{e}^{-r^*(T-t)} S_t \Phi(d_1) - \mathrm{e}^{-r(T-t)} K \Phi(d_2),$$

where

$$d_1 \equiv \frac{\log(S_t/K) + \left(r - r^* + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}},$$

$$d_2 \equiv \frac{\log(S_t/K) + \left(r - r^* - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}.$$

Similar formulas apply to put options. If r^* is a foreign risk-free interest rate, then we obtain the Garman and Kohlhagen (1983) formula for foreign exchange options. If, instead, r^* is the continuous dividend yield, then we obtain the formula for index options of Black and Scholes (1973) and Merton (1973). In the numerical applications for FF/DM options, we take for r and r^* the domestic (French) and foreign (German) euro-interest rates chosen to match the expiration of the options. We obtain data on these rates and transform them into their continuously compounded equivalents.

If one is willing to assume that volatility is constant across all strike prices, then it is possible to estimate the single volatility parameter σ^2. We follow the literature in doing this by minimizing the quadratic distance between actual and fitted prices, i.e., by nonlinear LS. Table 1 presents the estimates of such volatilities and also the associated goodness of fit measures. For both dates of

TABLE 1. Parameter estimates for lognormal distributions

	FF/DM options				S&P500 options		
	17.05.1996		28.04.1997		3.05.1993	11.06.1993	20.10.1993
$T-t$	30 days	90 days	30 days	90 days	46 days	98 days	58 days
σ	0.0208	0.0234	0.0265	0.0256	1.5347	1.5426	1.3891
σ_A	(0.33%)	(0.37%)	(0.42%)	(0.41%)	(24.28%)	(24.48%)	(22.05%)
N	13	13	11	11	15	17	20
R^2	0.996841	0.994207	0.975545	0.966069	0.951508	0.928570	0.980671

the FF/DM options, we notice a slightly better fit for the first maturity than for the second maturity. For the S&P500 options, we notice a better fit than for FF/DM options, with the fit being similar across the three dates.

The annualized volatility is denoted by σ_A^2. For FF/DM, σ_A ranges between 0.33% for the 30 days-to-maturity options on 17 May 1996 up to 0.42% for similar options on 28 April 1997. These figures are slightly smaller than those reported by Malz (1996), who considers the DM/£ in 1992, a series that is known to behave more erratically than the FF/DM. The annualized volatility of the S&P 500 ranges between 22.05% for the option on 20 October 1993 to 24.48% for the option on 11 June 1993. The difference in magnitude between the volatilities of exchange rates and stock returns is large. However, the data are of very different nature, and the market structures also differ. For instance, the FF/DM market was subject to tight trading bands within the European Monetary System.

To get an overall feel for the inaccuracy of these benchmark fits, we present in Figure 1 a typical plot of actual and fitted prices taken from the 3-months to maturity options of 28 April 1997. We notice that, for options with a high strike price, the fit is bad. This implies that a model with constant volatility across different strike prices is incomplete. In other words, volatility is dependent on K (a phenomenon known as the volatility "smile" because of the U-shaped relation), and further information that is not extracted by the lognormal is con-

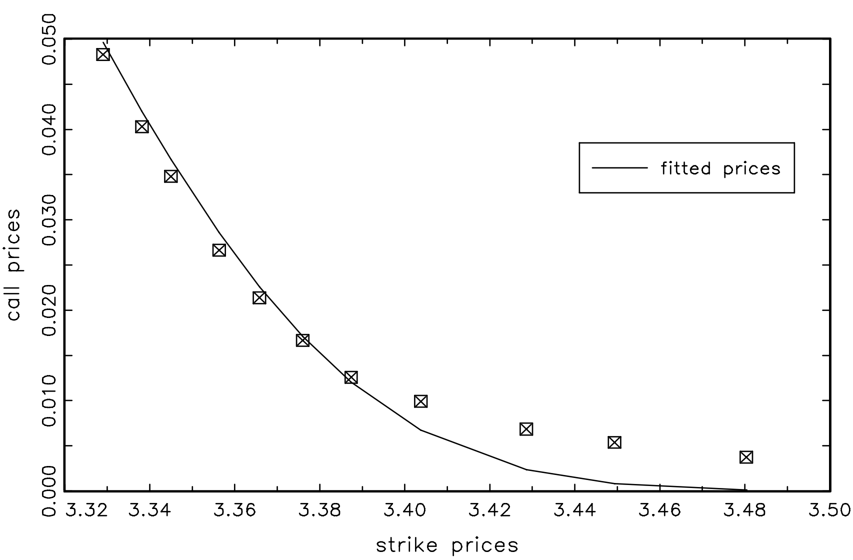

FIGURE 1. Original and fitted Garman–Kohlhagen call option prices, 28 April 1997, 3-months to maturity.

tained in the available option prices. This further information can be captured with a hypergeometric-based RND, as we will show.

3.2. Numerical Aspects of Estimating Our Nonlinear Function $C(\cdot)$

We performed all optimizations with the GAUSS program and the OPTMUM module. We used nonlinear LS as our fitting criterion for $C(\cdot)$, as we did with the lognormal. In general, though, one should consider other methods based on GLS or a likelihood criterion, as we have discussed in Section 2.4. We do not elaborate on this point in this section, given how uniformly good the fit turned out to be and given that the sample we have for the observed $C(\cdot)$ is random. This is because the data are given over a grid of options' "deltas," the term for $\partial C/\partial S$. This derivative is a constant plus the distribution function $G(\cdot)$ (e.g., see (A.2) in the Appendix). Because the distribution function of a continuous variate is uniformly distributed when the function's argument is random, taking a grid over the option's deltas is equivalent to stratified sampling.

In nonlinear estimation, it is crucial for numerical stability to obtain estimates of comparable magnitudes. To help the program find such estimates, we rescale the strike prices K by defining the linear transform $z \equiv \alpha + \beta K$. Because $C(\cdot)$ is tied to the density of z, it is also necessary to rescale the option prices by the Jacobian of the transform, according to $\beta C(K) = C(z)$. Because the bulk of standardized density masses tends to lie in the interval $[-3,3]$, the mapping of strike prices achieving approximately $z \in [-3,3]$ will yield a numerically stable estimation procedure.

Another important element in numerical optimization is the choice of initial values. The lognormal benchmark has a shape that is not too different from a normal. We therefore chose for the optimization to start with parameter values that arise from a normal RND. From the discussion following (10),

$$a_5 = -\frac{1}{2}, \qquad a_6 = \frac{1}{2}, \qquad b_4 = -\frac{1}{2 \times \text{variance of } z}, \qquad m_2 = \text{mean of } z,$$

and the component starting with a_1 is set to zero. Setting $a_1 = 0$ initially does not guarantee a well-behaved function at the next iteration. For example, care should be taken not to let $-a_3 \in \mathbb{N} \cup \{0\}$, as a division by zero may occur in computing the ${}_1F_1$ function (see the expansion in (5) and the subsequent exclusion restrictions). To make the program run smoothly, the starting values of the initially omitted component (the one beginning with a_1) may be those of a restricted gamma, e.g.,

$$b_1 = 1 + a_2 b_3, \qquad b_3 = 1, \qquad a_3 = a_2 + 2, \qquad a_2 = 4, \qquad m_1 = m_2,$$

because of the shape of RNDs known to arise in the literature (see the references cited earlier).

We saw in Section 2.2 that explicit restrictions on c_2, a_4, and c_1 exist and must be implemented in the estimation routine, and in the Appendix we simplify them for the setup of our problem. Additionally, we found in our numerical applications that the reductions implied by the joint hypothesis $b_1 = 1 + a_2 b_3$, $a_5 = -\frac{1}{2}$, $a_6 = \frac{1}{2}$ cannot be rejected by an F-test. This simplifies the former restrictions further to

$$c_1 = -c_2 m_2,$$

$$c_2 = -1 + a_4\sqrt{-b_4\pi},$$

$$a_4 = \frac{1}{2\sqrt{-b_4\pi}}\left[1 - a_1(-b_2)^{-a_2}\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}\right], \tag{16}$$

as shown in the Appendix. Finally, for the no-arbitrage condition to be checked, (2) or (15) simplifies to

$$\mathrm{E}(z) = a_1\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}(-b_2)^{-a_2}(m_1 - m_2) + m_2, \tag{17}$$

which is naturally a weighted average of m_1 and m_2. In both our applications, we found that the condition was satisfied by the data. We now summarize the empirical results.

3.3. Summary of the Results

Following the estimation strategy outlined previously, we obtain the estimates displayed in Table 2 for the various dates and maturities. Because R^2 is invari-

Table 2. Estimation of our $C(\cdot)$ function

	FF/DM options				S&P500 options		
	17.05.1996		28.04.1997		3.05.1993	11.06.1993	20.10.1993
$T - t$	30 days	90 days	30 days	90 days	46 days	98 days	58 days
a_2	3.2375	5.2241	3.4870	2.1929	3.7889	2.8683	2.0763
a_3	5.2462	7.0280	5.6126	6.5079	6.3536	6.2076	4.7731
b_2	−0.9250	−1.3249	−0.8918	−0.7344	−1.6314	−0.9381	−3.9701
b_3	1.4641	1.1114	1.5385	2.0968	1.7809	1.7482	0.5589
b_4	−0.5907	−0.1979	−0.3806	−0.1131	−0.2023	−0.1837	−0.6288
m_1	−1.5888	−2.2587	−2.3913	−0.8451	−0.6736	−1.8040	−1.4815
m_2	0.0128	0.2263	−0.9261	−0.8509	0.2864	−0.1961	1.8474
N	13	13	11	11	15	17	20
R^2	0.999955	0.999957	0.999945	0.999899	0.999922	0.999911	0.999661
$\bar{R}^2$	**0.999909**	**0.999913**	**0.999863**	**0.999749**	**0.999864**	**0.999857**	**0.999504**

ant to affine transformations of the data, a comparison of this statistic with the one in Table 1 is feasible. We notice that across all estimations, the R^2 and the R^2 adjusted for degrees of freedom (i.e., $\bar{R}^2$) for our fit are larger by a sizable magnitude than the ones obtained for the lognormal distribution. For DM/FF options we display in Figure 2 the plot of actual and fitted prices taken from 28 April 1997 (3-months maturity), which is representative of the fit at all the dates we tried. We notice an excellent fit for all options. Also, a direct computation of option prices involving formula (1) with a numerical routine corroborated that our method gives a very good fit.

We obtain a discretization of the RND by evaluating $g(\cdot)$ over a grid consisting of 1,000 points, and the plot for two dates and two maturities is in Figures 3–6. The link between the RND and actual (or objective) probabilities involves the degree of risk aversion of investors, and an interpretation of the RND as if it concerned the probability of the financial asset belonging to a certain range is a rough approximation. Aït-Sahalia and Lo (2000) and also Jackwerth (2000) show how a measure of risk aversion may be inferred from RNDs and the actual probabilities. Conversely, if one had a model describing risk aversion one could make a statement about actual probabilities. These investigations are beyond the scope of our paper, and central bankers are typically more interested in how RNDs change from day to day than in measuring risk aversion.

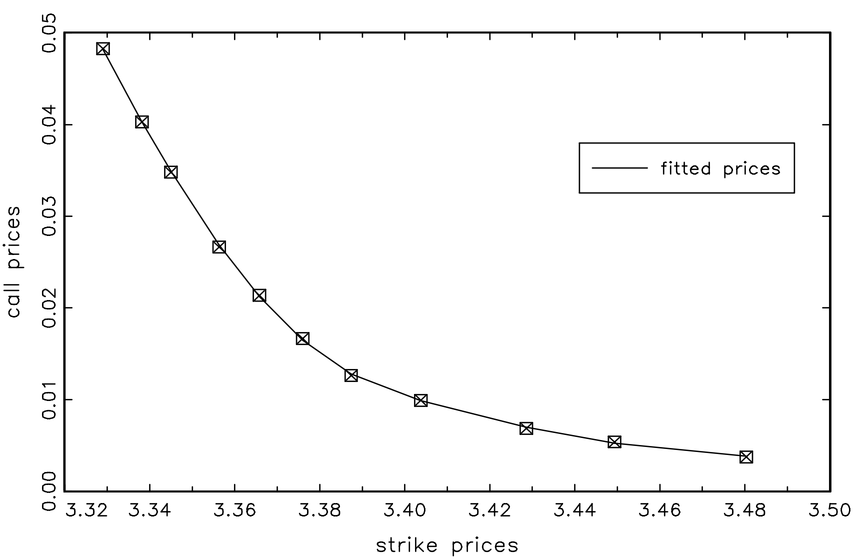

FIGURE 2. Original data and hypergeometric-based fit, 28 April 1997, 3-months to maturity.

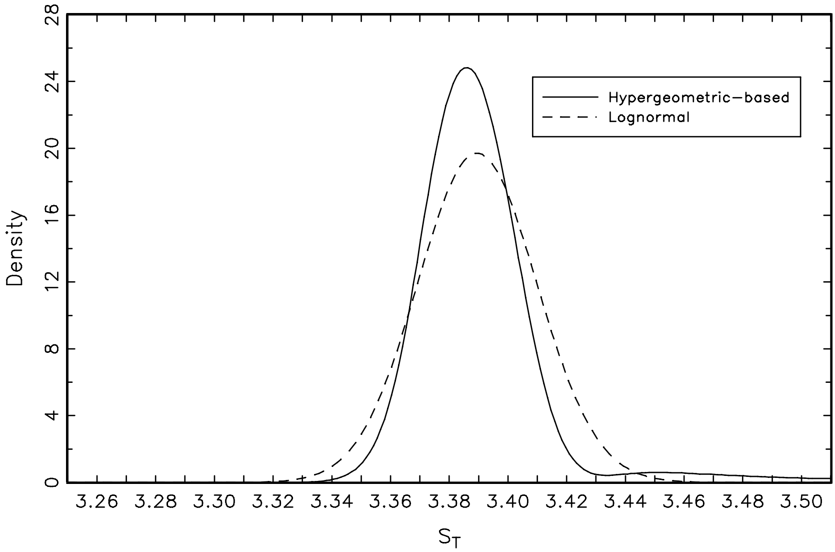

FIGURE 3. Lognormal and hypergeometric-based risk-neutral densities, 17 May 1996, 1 month maturity.

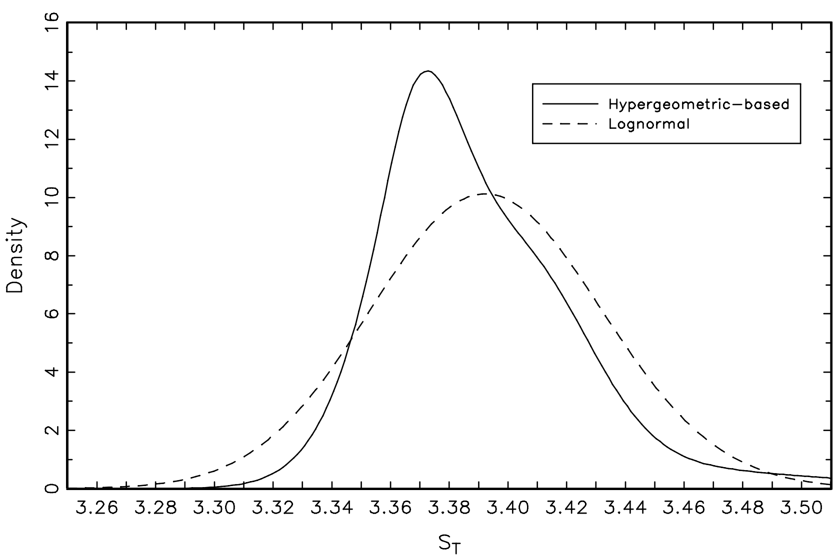

FIGURE 4. Lognormal and hypergeometric-based risk-neutral densities, 17 May 1996, 3-months to maturity.

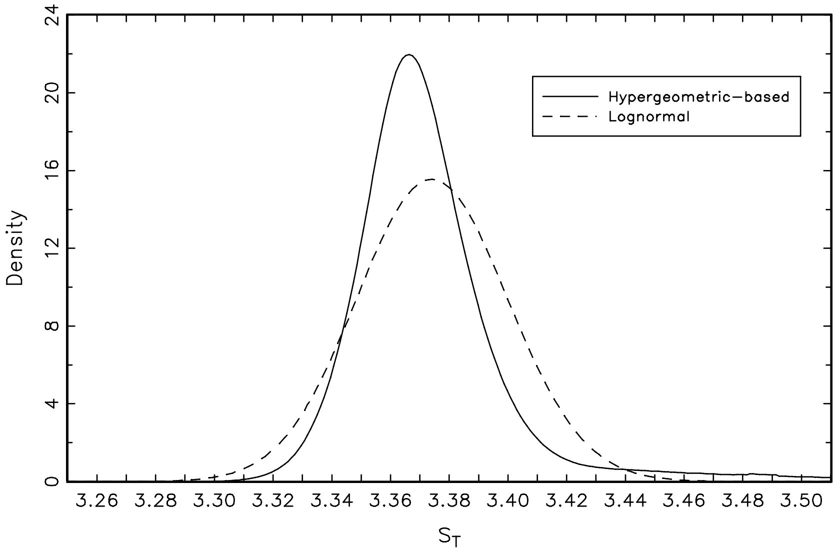

FIGURE 5. Lognormal and hypergeometric-based risk-neutral densities, 28 April 1997, 1 month maturity.

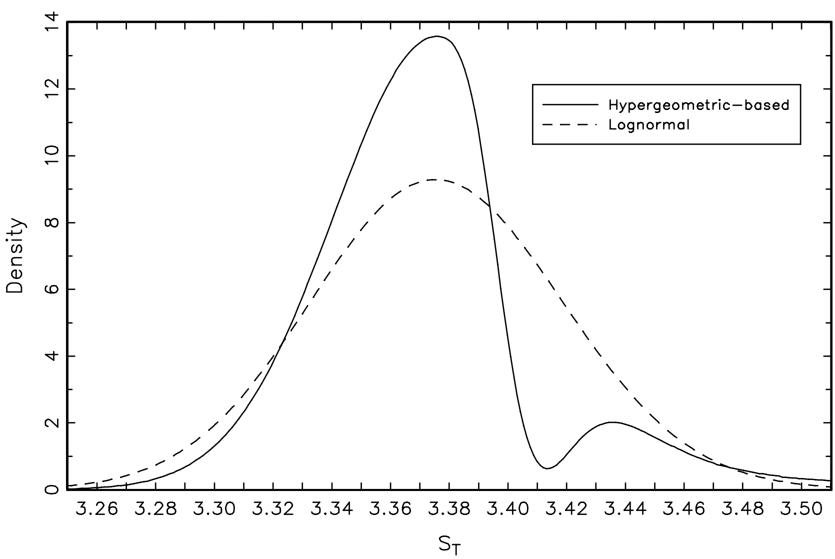

FIGURE 6. Lognormal and hypergeometric-based risk-neutral densities, 28 April 1997, 3-months to maturity.

Interesting comparative statics can be drawn from analyzing RNDs for different dates and/or maturities. The influence of major events is reflected in their changing shapes. In our first application, graphical inspection of the RNDs confirms that the first date was a rather quiet one, whereas the second date was amid a period of great agitation. Figures 3 and 4 display the RND for FF/DM options at 17 May 1996, for the two maturities. We notice that for the higher maturity the spread is much larger, translating the fact that market participants associate more uncertainty with the longer run. Also, in comparison with the lognormal case, we notice that our method reveals a heavier right tail. This translates the fact that investors, even on a rather normal day, are paying a premium in the anticipation of a latent devaluation of the FF (the so-called peso problem).

When we turn to the second date, there is an overall shift to the left for all distributions coming from the fact that, at that date, the FF had appreciated with respect to the DM. When we compare the distributions with those of the first date, there is a larger spread for the p.d.f. at all maturities. In other words, for the second date the global uncertainty is much larger. In addition, the right tail is more slowly decaying for the second date, reflecting the fact that market participants were contemplating the possibility of a large subsequent depreciation of the FF, with a nonnegligible probability.

It should be mentioned that the bimodality of the RND, which has arisen in the literature cited earlier, is also found to varying extents in Figures 3–6. In our function, we have not forced this bimodality (or any other oscillatory-tail features), as is clear from looking at our graphs. Our earlier talk of "two" components in our $C(\cdot)$ should not give the erroneous impression that our $C(\cdot)$ is based on just a mixture of two densities with at most two modes (see (12) for analytical details). Our graphs illustrate that this is not the case, some displaying multimodality but others not. See also the discussion after (A.8) in the Appendix, where we could not reduce our estimated functions to just mixtures of normal and gamma.

The multimodality, when it appears, could be due to a number of reasons. For exchange rates, the most plausible explanation of the right hump is the expectation of a widening of the target band. The center of the target zone was set at 3.35 FF/DM and the width was set at $\pm 2.5\%$, implying a band ranging from 3.2746 up to 3.4254. The right hump, being outside this band, suggests that market participants put a nonnegligible probability on a devaluation of the FF with respect to the DM.

To check the robustness of our estimated RNDs, say, $\hat{g}(\cdot)$, and the corresponding distributions $\hat{G}(\cdot)$, we perform a sensitivity analysis. We estimate the parameters of the hypergeometric-based $C(\cdot)$ when omitting one different strike price at a time, obtaining N estimates of RNDs, each based on $N-1$ strikes. Denote each of these estimated densities by $\hat{g}_j(\cdot)$ and their corresponding distributions by $\hat{G}_j(\cdot)$, where $j = 1,2,\ldots N$. To illustrate, we choose our "worst"-fitting $C(\cdot)$, namely, the FF/DM options at 28 April 1997, for the second maturity. We find that the maximal absolute deviation between $\hat{G}(S_T)$ and each of the 11 c.d.f.s $\hat{G}_j(S_T)$, for any j or S_T, is

$$\max_{j,S_T}\{|\hat{G}(S_T) - \hat{G}_j(S_T)|\} = 0.0056, \tag{18}$$

which is about half a percentage point, thus indicating a very stable estimate of the c.d.f. This observation is corroborated by Figure 7. There, the dotted line represents our density estimate $\hat{g}(S_T)$, whereas the narrow band around it is obtained pointwise (i.e., for each S_T) as the minimal and maximal value over the N estimates $\hat{g}_j(S_T)$. This indicates that our estimated RNDs are robust to a number of possible concerns raised in the finance literature, in addition to the traditional econometric ones. For example, our RNDs are robust to dropping the observation with the largest bid-ask spread, or the observation with the largest (or smallest) strike price K, etc.

We check further the robustness of the estimated shapes to bid-ask spreads, because this is a particularly important issue. To do so, we reestimate the densities by first assuming that all option prices were at the ask side and then by assuming that they were all on the bid side. We perform these estimations for 28 April 1997, a day when markets were rather agitated, and for the relatively illiquid 3 months to maturity option. For these options, the spreads tend to be

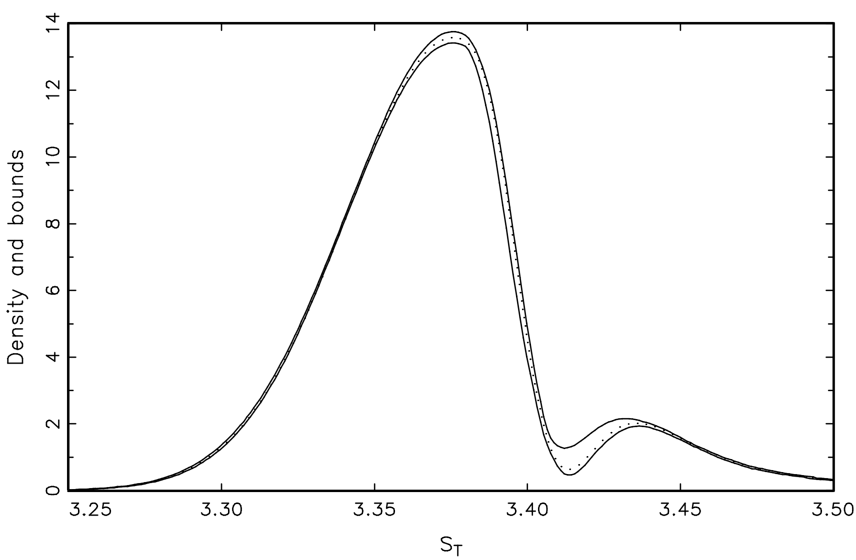

FIGURE 7. Sensitivity bounds for the hypergeometric-based density with one strike deleted at a time.

the largest of all the options considered in this study. The results of the estimation are displayed in Figure 8. As can be seen, the overall shape of the densities is the same. In particular, the bimodality of the RNDs for this date appears to be numerically stable and not simply due to an error in the data. Considering the implied c.d.f.s, we find that the maximal discrepancy between the mid-spread and the bid-only (respectively, ask-only) prices is 0.058 (respectively, 0.047). Though the shapes are analogous, the discrepancy of roughly 5% is larger than the one found in (18) because here we take the sensitivity analysis to the extreme of changing all prices simultaneously. Even so, the maximal discrepancy between the distributions occurs very close to the peak (first mode) of the density, namely, at 3.37 and 3.35, respectively, but not in its tails, which are of most interest as explained earlier in connection with Central Bank policy-making.

We now turn to the estimation results pertaining to fitting our $C(\cdot)$ function to the S&P500 options. Inspection of the second part of Table 2 confirms our earlier results: the hypergeometric-based $C(\cdot)$ function allows for a very close fit, of comparable quality to the one obtained for the FF/DM options. For 3 May 1993, Figure 9 represents the RND obtained with the lognormal model and the hypergeometric one. We notice now, in line with the figures represented in Aït-Sahalia and Lo (1998), that the RND is negatively skewed for Index options. This pattern translates the idea that investors fear a crash. A lognormal cannot represent a negatively skewed density, and thus seems to be

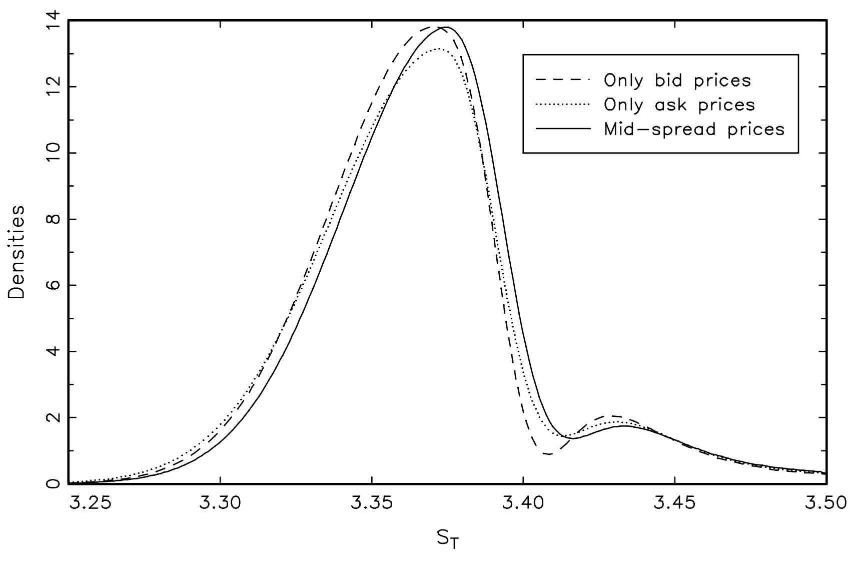

FIGURE 8. Hypergeometric-based estimated densities with ask, with bid, and with mid-spread option prices.

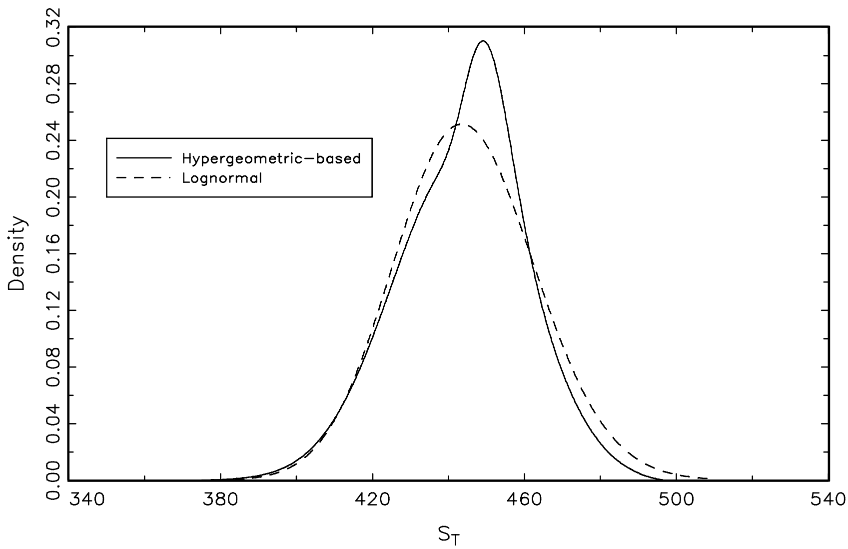

FIGURE 9. Lognormal and hypergeometric-based risk-neutral densities, S&P500 index options, 3 May 1993, 46-days to maturity.

incompatible with the data, whereas our hypergeometric-based densities allow for either type of skewness.

Finally, we compare our results to the fits obtained by other methods. We do not fit the entropy-based model, because it has as many parameters as data points and is therefore going to lead to $R^2 = 1$ and $\bar{R}^2 = 0$ everywhere. As for the others, Table 3 reports the results. The label "Hermite" refers to the method used by Jondeau and Rockinger (2001) to enforce nonnegativity constraints on

TABLE 3. A comparison of R^2 and $\bar{R}^2$ (in bold) for other methods

	FF/DM options				S&P500 options		
	17.05.1996		28.04.1997		3.05.1993	11.06.1993	20.10.1993
$T - t$	30 days	90 days	30 days	90 days	46 days	98 days	58 days
Hermite	0.999669	0.999878	0.990926	0.991626	0.997214	0.984918	0.993403
	[**0.999603**]	[**0.999853**]	[**0.988658**]	[**0.989532**]	[**0.996750**]	[**0.982764**]	[**0.992627**]
Jumps	0.999676	0.999926	0.999850	0.999804	0.997926	0.991013	0.995244
	[**0.999611**]	[**0.999911**]	[**0.999812**]	[**0.999755**]	[**0.997580**]	[**0.989729**]	[**0.994685**]
Mixtures	0.999947	0.999878	0.999842	0.999835	0.998267	0.990682	0.996039
	[**0.999906**]	[**0.999817**]	[**0.999812**]	[**0.999725**]	[**0.997573**]	[**0.987577**]	[**0.994983**]

the estimation of Gram–Charlier densities by means of Hermite polynomials. "Jumps" refers to the Malz-type jump-diffusion model discussed in the introduction, whereas "Mixtures" is the generalization of the benchmark into lognormal mixtures. It is clear from the table that our method is always best in terms of R^2 and is only once beaten by "Jumps" in terms of an $\bar{R}^2$ that differs from ours by less than a unit in the fifth decimal place! Our method does very well and uniformly so for

(1) volatile days, in addition to quiet ones;
(2) different maturities; and
(3) options on very different underlying assets.

Our estimated densities' functional forms vary over the dates and maturities, as can be seen in Table 2 from the parameter estimates of a_2, a_3, b_3, the other parameters playing mainly centering and scaling roles. Therefore, parametric methods that restrict functional forms (e.g., mixtures of lognormals) will not do uniformly well. Notice also that Table 2 indicates that, in all cases studied, the tails of the RNDs decay exponentially fast. For the markets studied, this implies that all moments of finite order exist and that the perceived probability of extreme events declines at exponential rates. The estimated parameters of the asymmetric component of (10) were such that it ended up representing extreme events by means of long tails, especially because b_3 is mostly smaller than the exponential power 2 of the other component.

4. THE DENSITY FUNCTIONAL FOR REPEATED INTEGRALS

Section 2 gave a $C(\cdot)$ function based on generalizing (8) and (9) for the case of double-integrals of p.d.f.'s. Here, we extend this approach to the situation where the pth integral of a p.d.f. is observed, where $p \in \mathbb{N}$, by taking

$$\begin{aligned} C(z) \equiv\ & c_1 + c_2 z + \cdots + c_p z^{p-1} \\ & + 1_{z>m_1} a_1 (z-m_1)^{b_1} {}_1F_1(a_2; a_3; b_2 (z-m_1)^{b_3}) \\ & + (a_4)\,{}_1F_1(a_5; a_6; b_4 (z-m_2)^2). \end{aligned} \tag{19}$$

The parameters are not generally (i.e., for $p \neq 2$) in correspondence with our earlier $C(z)$, but the modeling methodology is the same and so is the number of free parameters. The case $p = 1$ is one where the variate z is directly observable, and its empirical c.d.f. is available to compare to our fitted c.d.f. $C(z)$.

Three types of restrictions on the parameters of $C(z)$ are needed:

Type I:

A: restrictions on the c.d.f. given by (up to arbitrary constants) $\mathrm{d}^{p-1}C(z)/\mathrm{d}z^{p-1}$ for the edges of its support, namely, z_ℓ and z_u;

B: nonnegativity restrictions on the p.d.f. given by $\mathrm{d}^p C(z)/\mathrm{d}z^p$ over its support (z_ℓ, z_u);

C: other specifics of the problem (in our previous applications, the transversality and no-arbitrage conditions).

Type II: a priori information.

Type III: hypotheses of reduction to known simpler forms.

In this section, without a specific application in mind, one can only implement restrictions of Types IA (up to a constant) and IB and provide formulas (such as expressions for the moments) that are likely to be required for Type IC. Types II and III are application-specific and are to be determined by the user.

For Types IA and IB, we need the $(p-1)$th and pth derivatives of $C(z)$. They are obtained by expanding the ${}_1F_1$ functions in (19) and differentiating termwise. The result is, for either $q = p-1$ or $q = p$,

$$\begin{aligned}\frac{\mathrm{d}^q}{\mathrm{d}z^q} C(z) = {} & 1_{q=p-1}\, c_p (p-1)! \\ & + 1_{z>m_1}(-1)^q a_1 \sum_{j=0}^{\infty} \frac{(a_2)_j b_2^j}{(a_3)_j j!} (-b_3 j - b_1)_q (z-m_1)^{b_1 - q + b_3 j} \\ & + a_4 (-1)^q \sum_{j=q_i}^{\infty} \frac{(a_5)_j b_4^j}{(a_6)_j j!} (-2j)_q (z-m_2)^{2j-q},\end{aligned}$$

where q_i is the integer part of $(q+1)/2$. There are two equivalent ways of expressing these sums. The first is the one we have seen earlier, and it is the sum of ${}_1F_1$ functions. For general p, the expression would be quite cumbersome, so we resort to the second possible method. It makes use of the generalized hypergeometric function in (5) to write

$$\begin{aligned}\frac{\mathrm{d}^q}{\mathrm{d}z^q} C(z) = {} & 1_{q=p-1}\, c_p (p-1)! \\ & + 1_{z>m_1}(-1)^q a_1 (-b_1)_q (z-m_1)^{b_1-q} \\ & \times {}_{q+1}F_{q+1}\left(\begin{matrix} a_2, \frac{b_1}{b_3}+1, \ldots, \frac{b_1-q+1}{b_3}+1; \\ a_3, \frac{b_1}{b_3}, \ldots, \frac{b_1-q+1}{b_3}; \end{matrix} \; b_2 (z-m_1)^{b_3} \right) \\ & + \frac{a_4}{\sqrt{\pi}} \frac{(a_5)_{q_i} \Gamma\left(q_i + \frac{1}{2}\right)}{(a_6)_{q_i} \Gamma(2q_i - q + 1)} (4b_4)^{q_i} (z-m_2)^{2q_i - q} \\ & \times {}_3F_3\left(\begin{matrix} a_5 + q_i, 1, \frac{1}{2} + q_i; \\ a_6 + q_i, q_i - \frac{q}{2} + 1, q_i - \frac{q}{2} + \frac{1}{2}; \end{matrix} \; b_4 (z-m_2)^2 \right),\end{aligned} \qquad \textbf{(20)}$$

where the duplication formula

$$\sqrt{\pi}2^{q-2q_i}\Gamma(2q_i - q + 1) = \Gamma\left(q_i - \frac{q}{2} + 1\right)\Gamma\left(q_i - \frac{q}{2} + \frac{1}{2}\right)$$

has been used.

For restrictions of Type IC, it can be helpful to derive an expression for the moments of z. By letting $q = p$ and $p_i \equiv q_i$ (= integer part of $(p + 1)/2$) in (20), we get

$$\begin{aligned}
\mathrm{E}(z^n) &= \int_{z_\ell}^{z_u} z^n \frac{\mathrm{d}^p C(z)}{\mathrm{d}z^p}\,\mathrm{d}z \\
&= (-1)^p a_1(-b_1)_p \\
&\quad \times \int_{m_1}^{z_u} z^n (z - m_1)^{b_1 - p}\,{}_{p+1}F_{p+1}\left(\begin{matrix} a_2, \dfrac{b_1}{b_3} + 1, \ldots, \dfrac{b_1 - p + 1}{b_3} + 1; \\ a_3, \dfrac{b_1}{b_3}, \ldots, \dfrac{b_1 - p + 1}{b_3}; \end{matrix}\, b_2 (z - m_1)^{b_3}\right)\mathrm{d}z \\
&\quad + \frac{a_4}{\sqrt{\pi}} \frac{(a_5)_{p_i}\Gamma\left(p_i + \frac{1}{2}\right)}{(a_6)_{p_i}\Gamma(2p_i - p + 1)} (4b_4)^{p_i} \\
&\quad \times \int_{z_\ell}^{z_u} z^n (z - m_2)^{2p_i - p}\,{}_3F_3\left(\begin{matrix} a_5 + p_i, 1, \dfrac{1}{2} + p_i; \\ a_6 + p_i, p_i - \dfrac{p}{2} + 1, p_i - \dfrac{p}{2} + \dfrac{1}{2}; \end{matrix}\, b_4 (z - m_2)^2\right)\mathrm{d}z \\
&= (-1)^p a_1(-b_1)_p \sum_{j=0}^{n} \binom{n}{j} \frac{m_1^{n-j}(z_u - m_1)^{j + b_1 - p + 1}}{j + b_1 - p + 1} \\
&\quad \times {}_{p+2}F_{p+2}\left(\begin{matrix} a_2, \dfrac{b_1}{b_3} + 1, \ldots, \dfrac{b_1 - p + 1}{b_3} + 1, \dfrac{j + b_1 - p + 1}{b_3}; \\ a_3, \dfrac{b_1}{b_3}, \ldots, \dfrac{b_1 - p + 1}{b_3}, \dfrac{j + b_1 - p + 1}{b_3} + 1; \end{matrix}\, b_2 (z_u - m_1)^{b_3}\right) \\
&\quad + \frac{a_4}{\sqrt{\pi}} \frac{(a_5)_{p_i}\Gamma\left(p_i + \frac{1}{2}\right)}{(a_6)_{p_i}\Gamma(2p_i - p + 1)} (4b_4)^{p_i} \sum_{j=0}^{n} \binom{n}{j} \frac{m_2^{n-j}(z - m_2)^{j + 2p_i - p + 1}}{j + 2p_i - p + 1} \\
&\quad \times {}_4F_4\left(\begin{matrix} a_5 + p_i, 1, \dfrac{1}{2} + p_i, \dfrac{j + 2p_i - p + 1}{2}; \\ a_6 + p_i, p_i - \dfrac{p}{2} + 1, p_i - \dfrac{p}{2} + \dfrac{1}{2}, \dfrac{j + 2p_i - p + 3}{2}; \end{matrix}\, b_4 (z - m_2)^2\right)\Bigg|_{z_\ell}^{z_u}.
\end{aligned}$$

This completes the explicit formulas mentioned in connection with restrictions of Type I.

5. CONCLUDING COMMENTS

We have provided a new methodology for estimating density-related functionals, without prior knowledge of the density's functional form. In this endeavor, we have been originally motivated by a problem in the area of contingent-claim valuation. We have shown in our applications that our method did uniformly very well, thus illustrating the analytical justifications for our new approach. However, the new general methodology we propose is equally applicable to different problems. Future work could apply our method to different areas of research. What this paper has done is to lay down the technical foundations necessary for subsequent applications.

NOTES

1. There are two different types of confluent hypergeometric functions, Kummer's ${}_1F_1$ and Tricomi's Ψ (see Abadir, 1999, and references therein). There are also two related confluent hypergeometric functions due to Whittaker, M and W. We use only ${}_1F_1$ in this paper.

2. The function can also mimic the basic properties of a lognormal. For this, use page 68 of Abramowitz and Stegun (1972): the leading term is picked up by the symmetric quadratic component (latter one) of $C(\cdot)$, whereas the asymmetry is represented by the rest. See also Johnson et al. (1994, 1995) for approximating the lognormal density by normal, gamma, and/or Weibull ones.

3. The parameter a_4 is not free to vary unrestrictedly (see the discussion of (13)). To some extent, m_2 is also not free to vary, if one believes in arbitrage pricing (see (15) or more clearly (17)).

REFERENCES

Abadir, K.M. (1993a) The limiting distribution of the autocorrelation coefficient under a unit root. *Annals of Statistics* 21, 1058–1070.

Abadir, K.M. (1993b) Expansions for some confluent hypergeometric functions. *Journal of Physics A* 26, 4059–4066 [Corrigendum for printing error (1993) 7663].

Abadir, K.M. (1999) An introduction to hypergeometric functions for economists. *Econometric Reviews* 18, 287–330.

Abadir, K.M. & A. Lucas (2000) A comparison of minimum MSE and maximum power for the nearly integrated non-Gaussian model. University of York Discussion paper in economics, 2000/21. Forthcoming in *Journal of Econometrics*.

Abken, P., D.B. Madan, & S. Ramamurtie (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.

Abramowitz, M. & I. Stegun (1972) *Handbook of Mathematical Functions*. New York: Dover Publications.

Aït-Sahalia, Y. & A.W. Lo (1998) Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance* 53, 499–547.

Aït-Sahalia, Y. & A.W. Lo (2000) Nonparametric risk management and implied risk aversion. *Journal of Econometrics* 94, 9–51.

Ali, M.M. & S.C. Sharma (1996) Robustness to nonnormality of regression *F*-tests. *Journal of Econometrics* 71, 175–205.

Amemiya, T. (1985) *Advanced Econometrics*. Oxford: Basil Blackwell.

Andrews, D.W.K. (2000) Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 68, 399–405.

Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.

Bahra, B. (1996) Probability distributions of future asset prices implied by option prices. *Bank of England Quarterly Bulletin*, August, 299–311.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves, & R.L. Taylor (1991) Bootstrapping unstable first-order autoregressive processes. *Annals of Statistics* 19, 1098–1101.

Bates, D.S. (1996a) Dollar jump fears, 1984–1992: Distributional abnormalities implicit in currency futures options. *Journal of International Money and Finance* 15, 65–93.

Bates, D.S. (1996b) Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. *Journal of Financial Studies* 9, 69–107.

Black, F. & M. Scholes (1973) The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–659.

Bondarenko, O.P. (2000) Recovering Risk-Neutral Densities: A New Nonparametric Approach. Mimeo, Department of Finance, University of Illinois at Chicago.

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

Büttler, H.-J. & J. Waldvogel (1996) Pricing callable bonds by means of Green's function. *Mathematical Finance* 6, 53–88.

Campa, J.C., P.H.K. Chang, & R.L. Reider (1997) ERM bandwidths for EMU and after: Evidence from foreign exchange options. *Economic Policy* 24, 55–87.

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

Dixit, A.K. & R.S. Pindyck (1994) *Investment under Uncertainty*. Princeton: Princeton University Press.

Erdélyi, A. (ed.) (1953) *Higher Transcendental Functions*, vols. 1 and 2. New York: McGraw-Hill.

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

Godfrey, L.G. & C.D. Orme (2001) Significance Levels of Heteroskedasticity-Robust Tests for Specification and Misspecification. Mimeo, Department of Economics, University of York.

Gouriéroux, C. & A. Monfort (1995) *Statistics and Econometric Models*, vol. 1. Cambridge: Cambridge University Press.

Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

Harrison, J.M. & S.R. Pliska (1981) Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications* 11, 215–260.

Jackwerth, J.C. (2000) Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* 13, 433–451.

Jarrow, R. & A. Rudd (1982) Approximate valuation for arbitrary stochastic processes. *Journal of Financial Economics* 10, 347–369.

Johnson, N.L., S. Kotz, & N. Balakrishnan (1994) *Continuous Univariate Distributions*, 2nd ed., vol. 1. New York: Wiley.

Johnson, N.L., S. Kotz, & N. Balakrishnan (1995) *Continuous Univariate Distributions*, 2nd ed., vol. 2. New York: Wiley.

Jondeau, E. & M. Rockinger (2000) Reading the smile: The message conveyed by methods which infer risk neutral densities. *Journal of International Money and Finance* 19, 885–915.

Jondeau, E. & M. Rockinger (2001) Gram-Charlier densities. *Journal of Economic Dynamics and Control* 25, 1457–1483.

Knight, J. & S. Satchell (1997) Pricing Derivatives Written on Assets with Arbitrary Skewness and Kurtosis. Mimeo, Trinity College, University of Cambridge.

Lawford, S.D.C. (2001) *Improved Modelling in Finite-Sample and Nonlinear Frameworks*. DPhil thesis, University of York.

Madan, D.B. & F. Milne (1994) Contingent claims valued and hedged by pricing and investing in a basis. *Mathematical Finance* 4, 223–245.

Malz, A.M. (1996) Using option prices to estimate realignment probabilities in the European monetary system. *Journal of International Money and Finance* 15, 717–748.

Melick, W.R. & C.P. Thomas (1997) Recovering an asset's implied PDF from options prices: An application to crude oil during the Gulf crisis. *Journal of Financial and Quantitative Analysis* 32, 91–116.

Merton, R. (1973) Rational theory of option pricing. *Bell Journal of Economics and Management Science* 4, 141–183.

Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.

Rubinstein, M. (1994) Implied binomial trees. *Journal of Finance* 49, 771–818.

Söderlind P. & L.E.O. Svensson (1997) New techniques to extract market expectations from financial instruments. *Journal of Monetary Economics* 40, 383–429.

Stutzer, M. (1996) A simple nonparametric approach to derivative security valuation. *Journal of Finance* 51, 1633–1652.

Wilmott, P., J. Dewynne, & S. Howison (1993) *Option Pricing: Mathematical Models and Computation*. Oxford: Oxford Financial Press.

Young, G.A. (1994) Bootstrap: More than a stab in the dark? *Statistical Science* 9, 382–415 (with discussion).

APPENDIX

Proof of Theorem 1. The proof follows by checking the standard conditions for consistency (e.g., see Amemiya, 1985, Ch. 4; Gouriéroux and Monfort, 1995, Ch. 8). It draws on the properties of continuity and differentiability of hypergeometric functions detailed in, e.g., Erdélyi (1953). First, by means of the polygamma function, the function $C(\cdot)$ is differentiable an arbitrary number of times with respect to the parameters $\boldsymbol{\theta}$ when $\boldsymbol{\theta} \in \Theta$, except possibly (depending on b_1) at the point $m_1 = z$ that has probability measure zero for the continuous variate z. Second, the covariance of $C(\cdot)$ for different values of $\boldsymbol{\theta} \in \Theta$ is finite, this being the integral of a distribution function that is bounded by definition. Finally, two nontrivial (nonconstant) ${}_1F_1$ functions will be identical for all z, if and only if their parameters are identical. The values of randomly sampled $C(z)$ are therefore distinct for different $\boldsymbol{\theta} \in \Theta$. ∎

Proof of Theorem 2. The nonlinear LS estimators of $\boldsymbol{\theta} \in \tilde{\Theta} \subset \Theta$ are consistent, by Theorem 1, so we now check the two additional conditions for asymptotic normality. First, as mentioned in the proof of Theorem 1, the function $C(\cdot)$ is arbitrarily differentiable with respect to $\boldsymbol{\theta}$ when $\boldsymbol{\theta} \in \Theta$ and, a fortiori, when $\boldsymbol{\theta} \in \tilde{\Theta}$. Second, the matrix of second derivatives of $C(\cdot)$ with respect to $\boldsymbol{\theta}$ is finite with probability 1 as seen in the proof of Theorem 1, and it has a nonsingular expectation for $\boldsymbol{\theta}$ restricted further to $\boldsymbol{\theta} \in \tilde{\Theta}$ by the identifiability of all the parameters. ∎

Derivations for Section 3.2. Letting $O(z^\nu)$ denote terms of order of magnitude of at most z^ν,

$$ {}_1F_1(a;c;z) = \begin{cases} \dfrac{\Gamma(c)}{\Gamma(c-a)}\,|z|^{-a}\left(1 + O\left(\dfrac{1}{z}\right)\right), & \text{as } z \to -\infty \\ \dfrac{\Gamma(c)}{\Gamma(a)}\,z^{a-c}e^{z}\left(1 + O\left(\dfrac{1}{z}\right)\right), & \text{as } z \to \infty \end{cases} \tag{A.1} $$

for $z \in \mathbb{R}$. We saw in Section 2.2 that explicit restrictions on c_2, a_4, and c_1 exist. Here, we specialize them to the setup of our applications.

First, consider the restrictions on c_2 and a_4. In the context of (3) and (4)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z} C(z)\bigg|_{z=z_\ell} &= G(z_\ell) - 1 = -1,\\ \frac{\mathrm{d}}{\mathrm{d}z} C(z)\bigg|_{z=z_u} &= G(z_u) - 1 = 0,\end{aligned} \tag{A.2}$$

where $\mathrm{d}C(z)/\mathrm{d}z$ is given by (13). On the assumption that $z_\ell \le m_1$, the first constraint reduces to

$$c_2 = -1 - 2a_4 \frac{a_5}{a_6} b_4(z_\ell - m_2)\,{}_1F_1(a_5 + 1; a_6 + 1; b_4(z_\ell - m_2)^2) \tag{A.3}$$

and the second is

$$\begin{aligned} c_2 = &-a_1(z_u - m_1)^{b_1 - 1}\bigg[(b_1)\,{}_1F_1(a_2; a_3; b_2(z_u - m_1)^{b_3})\\ &\qquad + \frac{a_2}{a_3} b_2 b_3 (z_u - m_1)^{b_3}\,{}_1F_1(a_2 + 1; a_3 + 1; b_2(z_u - m_1)^{b_3})\bigg]\\ &- 2a_4 \frac{a_5}{a_6} b_4(z_u - m_2)\,{}_1F_1(a_5 + 1; a_6 + 1; b_4(z_u - m_2)^2).\end{aligned}$$

The first of these is linear in c_2, and the second may be reformulated by combining the two as

$$\begin{aligned} a_4 = 1 &- a_1(z_u - m_1)^{b_1 - 1}\bigg[(b_1)\,{}_1F_1(a_2; a_3; b_2(z_u - m_1)^{b_3})\\ &\qquad + \frac{a_2}{a_3} b_2 b_3 (z_u - m_1)^{b_3}\,{}_1F_1(a_2 + 1; a_3 + 1; b_2(z_u - m_1)^{b_3})\bigg]\\ &\div \bigg\{2 \frac{a_5}{a_6} b_4[(z_u - m_2)\,{}_1F_1(a_5 + 1; a_6 + 1; b_4(z_u - m_2)^2)\\ &\qquad - (z_\ell - m_2)\,{}_1F_1(a_5 + 1; a_6 + 1; b_4(z_\ell - m_2)^2)]\bigg\}\end{aligned} \tag{A.4}$$

which is linear in a_4. In our two applications, we found that the reductions implied by the joint hypothesis $b_1 = 1 + a_2 b_3$, $a_5 = -\frac{1}{2}$, $a_6 = \frac{1}{2}$ cannot be rejected by an F-test. As a result, (A.3) and (A.4) simplify further to

$$\begin{aligned} c_2 &= -1 - 2a_4 \frac{a_5}{a_6} (-b_4)^{-a_5} \frac{\Gamma(a_6 + 1)}{\Gamma(a_6 - a_5)} \lim_{z_\ell \to -\infty} (m_2 - z_\ell)^{-2a_5 - 1}\\ &= -1 + a_4\sqrt{-b_4 \pi}\end{aligned} \tag{A.5}$$

and

$$a_4 = \left[1 - a_1(-b_2)^{-a_2}\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}[b_1 - a_2 b_3]\lim_{z_u\to\infty}(z_u - m_1)^{b_1 - a_2 b_3 - 1}\right]$$
$$\div\left[-2\frac{a_5}{a_6}(-b_4)^{-a_5}\frac{\Gamma(a_6+1)}{\Gamma(a_6 - a_5)}\right.$$
$$\left.\times\left(\lim_{z_u\to\infty}(z_u - m_2)^{-2a_5-1} + \lim_{z_\ell\to-\infty}(m_2 - z_\ell)^{-2a_5-1}\right)\right]$$
$$= \frac{1}{2\sqrt{-b_4\pi}}\left(1 - a_1(-b_2)^{-a_2}\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}\right), \tag{A.6}$$

respectively, by means of formula (A.1) for extreme values and by substitution for b_1, a_5, a_6. As explained in Abadir (1999), the asymptotic formula (A.1) does not require $|z| \to \infty$ but requires only that z reach some extreme values for the standardized distribution.

Second, the restriction implied for c_1 by the boundary condition of our applications can be stated explicitly here. The reader will have noticed from (1) and (4) that $C_t(\infty) = C(\infty) = 0$. For our function $C(z)$ of (10), this transversality condition translates into

$$c_1 = \lim_{z\to\infty}\left(-c_2 z - a_1(-b_2)^{-a_2}(z - m_1)^{b_1 - a_2 b_3}\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}\right.$$
$$\left. - a_4(-b_4)^{-a_5}\frac{\Gamma(a_6)}{\Gamma(a_6 - a_5)}(z - m_2)^{-2a_5}\right) \tag{A.7}$$

by means of (A.1). Similarly to the simplification of (A.3) and (A.4) into (A.5) and (A.6), respectively, we may simplify (A.7) further for our applications by substituting for b_1, a_5, a_6 as

$$c_1 = \lim_{z\to\infty}\left(-c_2 z - \left(a_1(-b_2)^{-a_2}\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)} + a_4\sqrt{-b_4\pi}\right)(z - m_2)\right)$$
$$= \lim_{z\to\infty}((1 - a_4\sqrt{-b_4\pi})z - (1 - 2a_4\sqrt{-b_4\pi} + a_4\sqrt{-b_4\pi})(z - m_2))$$
$$= \lim_{z\to\infty}((1 - a_4\sqrt{-b_4\pi})m_2)$$
$$= (1 - a_4\sqrt{-b_4\pi})m_2$$
$$= -c_2 m_2. \tag{A.8}$$

Note that for a gamma component to exist in $C(\cdot)$, one would require the further simplifications $b_3 = 1$ and $a_3 = a_2 + 2$. These were not supported by our data here (not even at the 1% significance level) and were therefore not included in (A.5)–(A.8). In other words, the first three terms of the formula for the density given in (12) do not simplify to a gamma, but the latter two terms do reduce to a normal component.

Finally, to implement the no-arbitrage condition (2) of our example, we use (A.1) to simplify $\mathrm{E}(z)$ of (15) as

$$
\begin{aligned}
\mathrm{E}(z) = {} & a_1 \frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)} (-b_2)^{-a_2} \lim_{z_u \to \infty} (z_u - m_1)^{b_1 - a_2 b_3 - 1}((b_1 - a_2 b_3 - 1) z_u + m_1) \\
& - a_4 \frac{\Gamma(a_6)}{\Gamma(a_6 - a_5)} (-b_4)^{-a_5} \lim_{-z_\ell, z_u \to \infty} \left(\left(|z - m_2|^{-2a_5} \left(1 + 2a_5 \frac{z}{z - m_2} \right) \right) \Big|_{z_\ell}^{z_u} \right) \\
= {} & a_1 \frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)} (-b_2)^{-a_2} \lim_{z_u \to \infty} (m_1) \\
& - a_4 \sqrt{-b_4 \pi} \lim_{-z_\ell, z_u \to \infty} \left(\left(|z - m_2| \left(1 - \frac{z}{z - m_2} \right) \right) \Big|_{z_\ell}^{z_u} \right) \\
= {} & a_1 \frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)} (-b_2)^{-a_2} m_1 - a_4 \sqrt{-b_4 \pi} \lim_{-z_\ell, z_u \to \infty} (-m_2 - m_2) \\
= {} & a_1 \frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)} (-b_2)^{-a_2} m_1 + 2a_4 (\sqrt{-b_4 \pi}) m_2,
\end{aligned}
$$

which can also be written as in (17) by substituting for a_4 from (A.6). ■