

# Estimation of risk-neutral densities using positive convolution approximation

Oleg Bondarenko\*

*Department of Finance, University of Illinois at Chicago, 601 S. Morgan St.,  
Chicago, IL 60607, USA*

---

## Abstract

This paper proposes a new nonparametric method for estimating the conditional risk-neutral density (RND) from a cross-section of option prices. The idea of the method is to fit option prices by finding the optimal density in a special admissible set. The admissible set consists of functions, each of which may be represented as a convolution of a positive kernel with another density. The method is termed the positive convolution approximation (PCA). The important properties of PCA are that it (1) is completely agnostic about the data generating process, (2) controls against overfitting while allowing for small samples, (3) always produces arbitrage-free estimators, and (4) is computationally simple.

In a Monte-Carlo experiment, PCA is compared to several popular methods: mixtures of log-normals (with one, two, and three lognormals), Hermite polynomials, two regularization methods (for the RND and for implied volatilities), and sigma shape polynomials. PCA is found to be a promising alternative to the competitors.

© 2003 Elsevier B.V. All rights reserved.

*JEL classification:* C13; C14; G13

*Keywords:* Risk-neutral density; Nonparametric estimation; Option valuation

---

## 1. Introduction

Exchange-traded options are typically available for a number of different strikes and maturities. Collectively, their prices contain rich information about the underlying asset price process. Implicit in option prices is the so-called *risk-neutral density* (RND), which plays a key role in the risk-neutral valuation of Ross (1976); and

---

\* Tel.: +1-312-996-2362; fax: +1-312-413-7948.

*E-mail address:* [olegb@uic.edu](mailto:olegb@uic.edu) (O. Bondarenko).

Cox and Ross (1976). The power of the RND is due to the fact that it allows securities prices to be expressed as the present value of their expected payoffs computed under the RND.

In a dynamically complete arbitrage-free market, the RND can be recovered from a set of European option prices using the relationship proposed in Ross (1976), Breeden and Litzenberger (1978), and Banz and Miller (1978). The relationship states that the RND is proportional to the second derivative of the call (or put) pricing function with respect to the strike. Recently, the better availability of option data has created a substantial interest in developing methods that recover the implied RND from prices of traded options. The literature on the RND estimation is extensive and is reviewed in the next section.

A number of papers have also demonstrated the importance of the RND for various applications. These include:

- (1) pricing exotic derivatives from prices of the standard ones,
- (2) assessing the market beliefs about important political and economic events—Campa and Chang (1996), Melick and Thomas (1997), Leahy and Thomas (1996), Söderlind and Swensson (1997), Söderlind (2000),
- (3) estimating parameters of the underlying stochastic process—Bates (1996),
- (4) testing market rationality—Bondarenko (1997, 2002),
- (5) estimating investors risk preferences—Aït-Sahalia and Lo (2000), Jackwerth (2000), Rosenberg and Engle (2002),
- (6) managing risk—Aït-Sahalia and Lo (2000).

In this paper, we propose a new nonparametric approach for estimating the conditional RND from a cross-section of option prices. The underlying idea of the approach is to construct a special set of admissible densities from which the optimal density is selected. The optimal density is the one that provides the best fit to the option prices. The admissible set consists of functions that can be represented as a convolution of a fixed positive kernel and another density. The choice of the kernel controls the “smoothness” of the admissible densities. By construction, the admissible set is rich and flexible. It has a continuum of degrees of freedom and can accommodate a wide variety of shapes. At the same time, the admissible set only includes smooth, well-behaved densities. The new method is termed the *positive convolution approximation* (PCA).

We show that, in the PCA method, the infinite-dimensional optimization can be replaced by its finite-dimensional counterpart, which is a standard quadratic program. The important properties of the PCA method are that it (1) is completely agnostic about the data generating process, (2) controls against overfitting while allowing for small samples, (3) always produces arbitrage-free estimators, and (4) is computationally simple.

To study the performance of the PCA method, we conduct a Monte-Carlo experiment. In this experiment, PCA is contrasted to 7 other RND methods recently proposed in the literature. Those include both nonparametric and flexible parametric methods: mixtures of lognormals (with one, two, and three lognormals), as in Bahra (1997),

Melick and Thomas (1997); Hermite polynomials/Edgeworth expansion, as in Madan and Milne (1994), Jarrow and Rudd (1982); the regularization method for the RND, as in Jackwerth and Rubinstein (1996); the regularization method for implied volatilities, as in Campa et al. (1998), Jackwerth (2000); and sigma shape polynomials, as in Rosenberg (1998). We evaluate the methods on the basis of the RMISE criterion and find that PCA performs considerably better than the competitors. The striking finding is that PCA, which is a fully *nonparametric* method, outperforms the mixture of three lognormals, which is designated in the experiment as the correct *parametric* model.

Additional results are reported in Bondarenko (2000). They indicate that PCA might also be useful in other frameworks, such as estimating statistical densities from historical returns. Bondarenko (2000) also presents an empirical application, where PCA is implemented to recover the conditional RND implicit in the S&P 500 Index options.

The rest of the paper is organized as follows. Section 2, discusses the theoretical relationship option prices and the RND and reviews the literature on the RND estimation. Section 3 introduces the new nonparametric method. Section 4 reports the results of the Monte-Carlo experiment. Sections 5 and 6 discuss and summarize the paper. Appendix A and B contains proofs of the results and the details of the Monte-Carlo experiment.

## 2. Option prices and risk-neutral densities

Let  $S_t$  denote value of an underlying asset on *trading* date  $t$ . Consider a general European option with payoff  $Z(S_T)$  on *maturity* date  $T$ . In a dynamically complete arbitrage-free market, the option's price  $Z_t$  can be found using risk-neutral valuation (see Ross, 1976; Cox and Ross, 1976). In this approach, the option's price is given by the expected present value of the payoff computed under the *risk-neutral density* (RND). Formally,

$$Z_t = \frac{1}{R_{t,T}} \int_0^\infty Z(S_T) f_t(S_T) dS_T, \quad R_{t,T} := e^{\int_t^T r_s ds}, \quad (1)$$

where  $f_t(S_T) = f(S_T, T; S_t, t)$  is the conditional risk-neutral density, and  $r_s$  is the risk-free rate. Consider European call and put options with payoffs  $C_T(x) = \max(S_T - x, 0)$  and  $P_T(x) = \max(x - S_T, 0)$ , where  $x$  is the strike price. Then the RND is related to  $C_t(x)$  and  $P_t(x)$  as follows:

$$f_t(S_T) = \frac{1}{R_{t,T}} \left. \frac{\partial^2 C_t(x)}{\partial x^2} \right|_{x=S_T} = \frac{1}{R_{t,T}} \left. \frac{\partial^2 P_t(x)}{\partial x^2} \right|_{x=S_T}. \quad (2)$$

The relationship in (2) has first been proposed in Ross (1976), Breeden and Litzenberger (1978), Banz and Miller (1978). It states that the RND is proportional to the second derivative of the call (put) pricing function with respect to the strike. In what follows we focus on equation (2) expressed in terms of the put price  $P_t(x)$ .

To simplify exposition, we also assume that the asset pays no dividends and that the risk-free rate is zero, i.e.,  $R_{t,T} = 1$ .<sup>1</sup>

The RND has the following intuitive interpretation: quantity  $f_t(x) dx$  is the price of an elementary security that pays \$1 if the asset's price  $S_T$  falls between  $x$  and  $x + dx$ , and nothing otherwise.<sup>2</sup> The RND contains information regarding the market perceptions about the distribution of the future returns as well as investors' risk preferences. Armed with the RND, one can value any European derivative written on  $S_t$  by simply computing the discounted expectation in (1). In this respect, the RND summarizes what is relevant about the market conditions from a pricing perspective.

The enormous information content of the RND and the better availability of option data have recently created a substantial interest in developing methods that recover the RND from market prices of options. In view of the relationship in (2), to estimate the RND one needs to estimate the second derivative of the put price  $P_t(x)$  with respect to the strike  $x$ . Apparently, this estimation would be a simple matter if the function  $P_t(x)$  were accurately known for all strikes  $x$  from zero to infinity. Under such ideal circumstances, the RND would be uniquely determined.

### 2.1. Data limitations

In practice, there are a number of challenges which make the RND estimation a nontrivial task. The estimation is complicated by the following data limitations:

- (1) Strikes of available options do not span the entire support of  $S_T$ . Due to the rules that the exchange uses to introduce new option contracts, very low and high strikes are usually unavailable.
- (2) Strikes are set at discrete intervals. For example, strikes for the S&P 500 Index options are usually spaced \$5 apart.
- (3) Option prices are recorded with substantial measurement errors. The main sources of measurement errors are nonsynchronous trading, price discreteness, and the bid-ask bounce.

In a typical application, the conditional RND is estimated from a cross-section of 10 to 30 option prices (for a given maturity). For such tiny samples, two additional issues must be addressed. The first one is that the problem of estimating a function's second derivative is *ill-posed*. This basically means that the quality of an estimator for the function's derivative will be much worse than the quality of an estimator for the function itself, the phenomenon often referred to as the *curse of differentiation*. The second issue is that an RND estimator  $\hat{f}(x)$  must satisfy the no-arbitrage constraints.

<sup>1</sup> This is without loss of generality. In the presence of the nonzero interest rates and dividends, the asset and option prices should be viewed as *forward* prices. It is then straightforward to convert back to *spot* prices. Dumas et al. (1998) describe this procedure in the context of the S&P 500 Index options.

<sup>2</sup> Another term for the risk-neutral density is the *state-price density*. The RND is also closely related to the *equivalent martingale measure*. The equivalent martingale measure always exists if the market is arbitrage-free (Harrison and Kreps, 1979) and is unique if the market is complete (Harrison and Pliska, 1981).

This means that the estimator must be a proper *probability density*, i.e., a nonnegative function that integrates to one.<sup>3</sup>

## 2.2. Literature on the RND estimation

Research on the estimation of risk neutral densities from option prices has been very active in recent years. Several dozens of alternative methods appear in the literature. In this subsection, we mention only a few of many papers devoted to the RND estimation and refer to Jackwerth (1999) for additional references.

Generally, all methods can be divided into *parametric* and *nonparametric* ones. Parametric methods are highly structured techniques that rely on particular assumptions on the data generating process. Three related approaches exist in the literature:

- (1) Models for the risk-neutral density
  - Expansion methods—Jarrow and Rudd (1982), Madan and Milne (1994), Longstaff (1995), Abken et al. (1996), Abadir and Rockinger (1998), Rubinstein (1998).
  - Generalized distributions—Aparicio and Hodges (1998), Posner and Milevsky (1998), Rosenberg (1998).
  - Mixture methods—Ritchey (1990), Bahra (1997), Melick and Thomas (1997).
- (2) Models for the implied volatility function—Shimko (1993), Malz (1997), Dumas et al. (1998).
- (3) Models for the stochastic process—Black and Scholes (1973), Cox and Ross (1976), Hull and White (1987), Goldenberger (1991), Heston (1993), Bates (1996).

The advantage of parametric models is that they are parsimonious so that a small number of parameters need to be estimated. These models yield explicit expressions for option prices and the hedge ratios. The most famous one is, of course, the Black–Scholes model where the underlying asset's price is assumed to follow the geometric Brownian motion. A problem with parametric techniques is that their success critically depends on whether the model is specified correctly. Unfortunately, despite the theoretical advances, popular parametric models seem unable to fully explain the data.

In contrast, nonparametric methods are flexible, data-driven methods. They make no strong assumptions about the data generating process. These methods can be divided into:

- (1) Implied trees—Derman and Kani (1994), Dupire (1994), Rubinstein (1994).
- (2) Smoothing techniques:
  - kernel methods—Aït-Sahalia and Lo (1998), Aït-Sahalia and Duarte (2000).
  - regularization methods for the RND—Jackwerth and Rubinstein (1996).

<sup>3</sup> See Bondarenko (2000) for the detailed analysis of the difficulties of the RND estimation and their implications for empirical work.

- regularization methods for implied volatility—Aparicio and Hodges (1998), Campa et al. (1998), Bliss and Panigirtzoglou (2002), Jackwerth (2000).
- (3) Maximum Entropy methods—Buchen and Kelly (1996), Stutzer (1996).
- (4) Neural networks—Garcia and Gencay (2000), Gottschling et al. (2000).

The disadvantage of nonparametric techniques is that they are usually data-intensive and are not effective in small samples. In the following sections, we introduce a new nonparametric method for the conditional RND estimation and argue that it might be a promising alternative to the existing methods.

### 3. Positive convolution approximation

#### 3.1. Basic idea

Let  $L^d$  denote the set of all probability densities, i.e., nonnegative functions from  $L_1(-\infty, \infty)$  that integrate to one. We start by fixing a *basis density* or *kernel function*  $\phi(x) \in L^d$ . Similar to the kernel methods, we can rescale  $\phi(x)$  with the bandwidth  $h$  to form a new density  $\phi_h(x) := 1/h\phi(\frac{x}{h})$ .

For fixed  $\phi_h(x)$ , we introduce the approximating set  $\mathcal{W}_h = \mathcal{W}_{\phi_h}$  of functions  $g$  that can be represented as a convolution of  $\phi_h$  and another density. Formally,

$$\mathcal{W}_h := \{g \in L^d \mid g = \phi_h * u, \text{ for some } u \in L^d\},$$

where for integrable functions  $f$  and  $g$ ,

$$f * g := \int_{-\infty}^{\infty} f(x - y)g(y) \, dy.$$

The set  $\mathcal{W}_h$  represents all admissible or candidate densities. In this set, we search for the *optimal* density, which provides the best fit to a given cross-section of options. Specifically, suppose that one observes a cross-section of put prices  $\{P_i\}$  with strikes  $x_1 < \dots < x_n$  corresponding to the RND  $f(x)$ . Then, an estimator of the RND is the function  $\hat{f}(x) \in \mathcal{W}_h$  that solves the following minimization problem:

$$\text{Minimize}_{\hat{f} \in \mathcal{W}_h} \sum_{i=1}^n (P_i - D^{-2} \hat{f}(x_i))^2, \tag{3}$$

where  $D^{-2}g(x) := \int_{-\infty}^x (\int_{-\infty}^y g(z) \, dz) \, dy$  denotes the second integral of  $g(x)$ . We refer to the solution of the above problem as the *positive convolution approximation* (PCA) of  $f(x)$ .<sup>4</sup>

---

<sup>4</sup> The fact that  $\mathcal{W}_h$  is constructed by taking convolutions of  $\phi_h$  with *positive* functions  $u$  is important. Without the positivity constraint, almost *any* function can be represented as convolution of  $\phi_h$  and some  $u \in L_1$ .

The motivation for PCA is the following. To construct the admissible set  $\mathcal{W}_h$ , densities from  $L^d$  are smoothed with the basis density  $\phi_h(x)$ . While the space  $L^d$  contains very general densities (including discontinuous, economically implausible ones), the set  $\mathcal{W}_h$  is comprised of only “well-behaved” densities. As with the kernel methods, the bandwidth  $h$  controls the smoothness of densities in the admissible set. A good choice for the bandwidth is critical for the end result. In contrast, a specific choice of the basis density  $\phi(x)$  is less important, with different choices for  $\phi(x)$  producing very similar estimators.<sup>5</sup>

To simplify exposition, in what follows we assume that the basis density is the standard normal one:

$$\phi(x) = n(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

To investigate the approximating properties of the set  $\mathcal{W}_h$ , we define the distance of density  $f$  to  $\mathcal{W}_h$  as

$$\rho(f, \mathcal{W}_h) := \inf_{g \in \mathcal{W}_h} \|f - g\| = \min_{g \in \mathcal{W}_h} \|f - g\|,$$

where  $\|\cdot\| = \|\cdot\|_2$  is the  $L_2$ -norm, and the last equality follows from the completeness of  $\mathcal{W}_h$ . Intuitively, the distance  $\rho(f, \mathcal{W}_h)$  measures how closely the density  $f$  can be approximated by another density in the admissible set. The distance satisfies the following basic properties, which are proved in Appendix A.

**Proposition 1.** *Let  $f_1, f_2 \in L^d$  and  $\alpha \in [0, 1]$ . Then*

$$\rho(\alpha f_1 + (1 - \alpha)f_2, \mathcal{W}_h) \leq \alpha \rho(f_1, \mathcal{W}_h) + (1 - \alpha)\rho(f_2, \mathcal{W}_h).$$

In words, if two densities can be closely approximated in the set  $\mathcal{W}_h$  then so will be their convex combination.

**Proposition 2.** *Let  $h_1 < h_2$ . Then*

$$\rho(f, \mathcal{W}_{h_1}) \leq \rho(f, \mathcal{W}_{h_2}).$$

That is, the distance  $\rho(f, \mathcal{W}_h)$  is nondecreasing in the bandwidth  $h$ . This is intuitive: an approximation in an admissible set with a smaller bandwidth is always no worse than an approximation in an admissible set with a larger bandwidth.

**Proposition 3.** *When the basis density  $\phi(x)$  is normal,*

$$\rho(\phi_\sigma, \mathcal{W}_h) = \begin{cases} 0, & \text{if } \sigma \geq h, \\ \sqrt{\frac{1}{2\sqrt{\pi}} \left( \frac{1}{\sigma} + \frac{1}{h} - \frac{2\sqrt{2}}{\sqrt{\sigma^2 + h^2}} \right)}, & \text{if } \sigma < h. \end{cases}$$

<sup>5</sup> Note, however, that densities in  $\mathcal{W}_h$  inherit the regularity of the basis density  $\phi$  in the following sense: if  $\phi(x)$  is  $m$ -times continuously differentiable then so will be each function in  $\mathcal{W}_h$ .

From Proposition 3, any normal density  $\phi_\sigma(x - \mu)$  with  $\sigma \geq h$  belongs to the admissible set  $\mathcal{W}_h$ . Moreover, it follows from Proposition 1 that any mixture of such normal densities also belongs to  $\mathcal{W}_h$ . Specifically, let

$$f(x) = \sum_{j=1}^J a_j \phi_{\sigma_j}(x - \mu_j), \quad a_j \geq 0, \quad \sum_{j=1}^J a_j = 1.$$

Then for all  $\mu_j$  and  $\sigma_j \geq h$ , the density  $f$  belongs to  $\mathcal{W}_h$ . This suggests that the set  $\mathcal{W}_h$  is rich and flexible. The admissible set is infinitely-dimensional and can accommodate very general shapes of densities.

At the same time,  $\mathcal{W}_h$  includes only smooth, well-behaved densities. If a density  $f \in \mathcal{W}_h$ , then all its derivatives exist and are bounded. Specifically, the absolute value of the  $m$ -th derivative is bounded by that of the basis density  $\phi_h(x)$ :

$$\max_x |f^{(m)}(x)| \leq \max_x |\phi_h^{(m)}(x)| = \frac{1}{h^m} \max_x |\phi^{(m)}(x)|.$$

In particular,

$$\max_x |f'(x)| \leq \frac{1}{h} \frac{1}{\sqrt{2\pi}}.$$

In other words, by changing the bandwidth  $h$ , one can control the smoothness of the admissible densities.

### 3.2. Probabilistic interpretation

Densities in the admissible set  $\mathcal{W}_h$  have a simple probabilistic interpretation. Consider independent random variables  $\tilde{x}_1$  and  $\tilde{x}_2$  with the probability densities  $\phi_h$  and  $u$ , respectively. Then their sum  $\tilde{x} = \tilde{x}_1 + \tilde{x}_2$  is distributed with the probability density  $f = \phi_h * u$ . Therefore,  $f \in \mathcal{W}_h$ .

### 3.3. Computational issues

The optimization problem in (3) has a continuum of degrees of freedom. In practice, we will solve it numerically by discretizing the admissible set. Specifically, consider an equally-spaced grid on the real line, defined by points  $z_j = j\Delta z$ , where  $j = 0, \pm 1, \dots$ , and  $\Delta z$  is the grid step. Define the discrete admissible set as

$$\mathcal{W}_h^{\Delta z} := \left\{ g \in L^d \mid g(x) = \sum_{j=-\infty}^{\infty} a_j \phi_h(x - z_j), \quad a_j \geq 0, \quad \sum_{j=-\infty}^{\infty} a_j = 1 \right\}.$$

The set  $\mathcal{W}_h^{\Delta z}$  is a subset of  $\mathcal{W}_h$ . Moreover, the two can be made arbitrarily close by selecting a sufficiently small  $\Delta z$ . A density in  $\mathcal{W}_h^{\Delta z}$  can be represented as a convolution of the basis density with a mixture of Dirac delta functions on the equally-spaced grid:

$$f = \phi_h * u, \quad \text{where } u(x) = \sum_{j=-\infty}^{\infty} a_j \delta(x - z_j), \quad a_j \geq 0, \quad \sum_{j=-\infty}^{\infty} a_j = 1.$$

We observe that the distance of  $f$  to  $\mathcal{W}_h^{\Delta z}$  can be bounded as

$$\rho(f; \mathcal{W}_h^{\Delta z}) \leq \rho(f; \mathcal{W}_{h_0}) + \rho(\mathcal{W}_{h_0}; \mathcal{W}_h^{\Delta z}), \quad \text{for all } h_0, \tag{4}$$

where

$$\rho(\mathcal{W}_{h_0}, \mathcal{W}_h^{\Delta z}) := \max_{g \in \mathcal{W}_{h_0}} \rho(g, \mathcal{W}_h^{\Delta z})$$

is the distance between sets  $\mathcal{W}_{h_0}$  and  $\mathcal{W}_h^{\Delta z}$ . This distance shows that any function from  $\mathcal{W}_{h_0}$  can be approximated by a function from  $\mathcal{W}_h^{\Delta z}$  with the accuracy no worse than  $\rho(\mathcal{W}_{h_0}, \mathcal{W}_h^{\Delta z})$ .<sup>6</sup> Intuitively, the inequality in (4) states that, if we need to approximate  $f$  in  $\mathcal{W}_h^{\Delta z}$ , then we can always first approximate  $f$  by some density  $f_{h_0}$  in the continuous set  $\mathcal{W}_{h_0}$  and then approximate  $f_{h_0}$  in the discrete set  $\mathcal{W}_h^{\Delta z}$ .

Suppose now that  $f \in \mathcal{W}_{h_0}$ . Then the first term in (4) is zero. When choosing  $h$  and  $\Delta z$ , our goal is to make sure that the loss in the approximating properties due to discretization is negligible, i.e., that the distance  $\rho(\mathcal{W}_{h_0}, \mathcal{W}_h^{\Delta z})$  is very small. It is important to emphasize that the choice of  $h$  and  $\Delta z$  is at complete discretion of an econometrician. In particular, the choice does not depend on  $f$  or on the available data. The econometrician can select these parameters to achieve any required degree of closeness between sets  $\mathcal{W}_{h_0}$  and  $\mathcal{W}_h^{\Delta z}$ . As a theoretical matter, the following proposition establishes an intuitive result that, as  $\Delta z$  becomes small, the difference between the discrete and continuous admissible sets vanishes.

**Proposition 4.** *For any density  $f \in L^d$ , let  $f_h$  and  $f_h^{\Delta z}$  denote its approximants in the continuous set  $\mathcal{W}_h$  and the discrete set  $\mathcal{W}_h^{\Delta z}$ , respectively:*

$$f_h = \arg \min_{g \in \mathcal{W}_h} \|f - g\|, \quad f_h^{\Delta z} = \arg \min_{g \in \mathcal{W}_h^{\Delta z}} \|f - g\|.$$

Then

$$(1) \quad \lim_{\Delta z \rightarrow 0} f_h^{\Delta z}(x) = f_h(x), \quad \text{for all } x.$$

Moreover,

$$(2) \quad \lim_{\Delta z \rightarrow 0} \rho(\mathcal{W}_{h_0}, \mathcal{W}_h^{\Delta z}) = 0, \quad \text{for } h_0 \geq h.$$

So how should we choose the pair  $(h, \Delta z)$  in practice? To answer this question, the left panel of Fig. 1 plots the distance  $\rho(\mathcal{W}_{h_0}, \mathcal{W}_h^{\Delta z})$  (computed numerically) as a function of the grid step  $\Delta z/h$  for different values of  $h/h_0$ . Based on this figure, we propose to set

$$h = 0.95h_0, \quad \Delta z = 0.5h. \tag{5}$$

For these parameters, the discrete set  $\mathcal{W}_h^{\Delta z}$  approximates the continuous set  $\mathcal{W}_{h_0}$  almost perfectly: the distance  $\rho(\mathcal{W}_{h_0}, \mathcal{W}_h^{\Delta z})$  is  $1.0 \times 10^{-5}$ . Such an approximation accuracy is more than enough for the applications that we have in mind. Still, we want to

<sup>6</sup> In the definition of the distance, we can use *max* (as opposed to *sup*) because the two sets are complete.

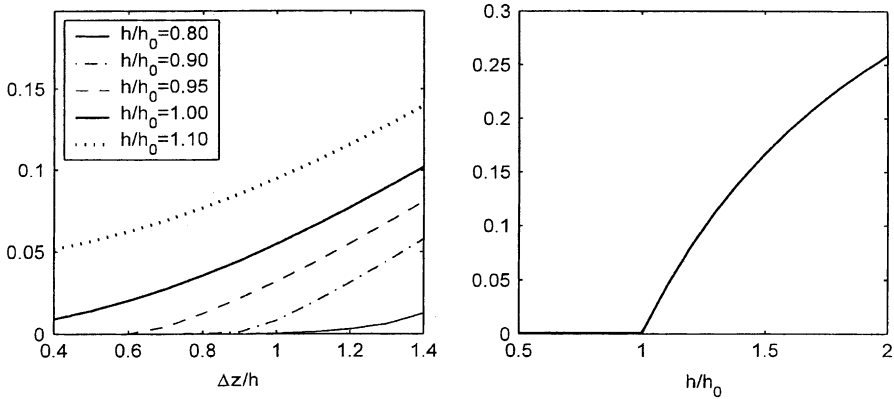


Fig. 1. The left panel plots the distance  $\rho(\mathcal{W}_{h_0}, \mathcal{W}_h^{\Delta z})$  between the continuous and discrete admissible sets  $\mathcal{W}_{h_0}$  and  $\mathcal{W}_h^{\Delta z}$  as a function of grid step  $\Delta z/h$ , for different values of  $h/h_0$ . The right panel plots the distance between two continuous admissible sets  $\rho(\mathcal{W}_{h_0}, \mathcal{W}_h)$  as a function of  $h/h_0$ . Plots for both panels are computed numerically for the case  $h_0 = 1$ .

reiterate that by reducing  $\Delta z$ , the distance can be made arbitrarily small. For instance, for the choice  $h=0.95h_0$  and  $\Delta z=0.25h$ , the distance  $\rho(\mathcal{W}_{h_0}, \mathcal{W}_h^{\Delta z})$  is only  $1.6 \times 10^{-15}$ .

A clarification on the terminology may be needed. In this paper, we say an “almost perfect” fit and a “negligible” approximation error to indicate that the error is much smaller than the required accuracy. Admittedly, the notion of the “required accuracy” would be inappropriate if one is interested in the asymptotic results. However, our focus is on *small* sample problems. In such problems, the presence of the sampling noise implies that a measure of an estimator’s quality (say, the *root mean integrated squared error*, or RMISE), is bounded from zero (even for the true parametric model). This bound determines the *required accuracy*. For example, in a typical estimation of the conditional RND, the best attainable RMISE is at least three orders of magnitude larger than the proposed approximation accuracy of  $10^{-5}$ .

Why do we recommend the choice of  $h$  and  $\Delta z$  as in (5)? Clearly, many other pairs will achieve the desired approximation accuracy, with a smaller bandwidth  $h$  implying a larger grid step  $\Delta z$ . In this tradeoff, we prefer to select a larger bandwidth at the cost of having a finer grid. For larger bandwidths, noise in the data is smoothed out more effectively. Smaller grid steps  $\Delta z$  imply that more parameters  $a_i$  need to be estimated. However, this is not a concern because the resulting minimization problem is computationally simple.

In the right panel of Fig. 1, we illustrate Proposition 2 by plotting the distance between two continuous sets  $\rho(\mathcal{W}_{h_0}, \mathcal{W}_h)$  as a function of  $h/h_0$ . This plot demonstrates that functions in  $\mathcal{W}_{h_0}$  can be approximated *exactly* in  $\mathcal{W}_h$  as long as  $h \leq h_0$ . When  $h$  exceeds  $h_0$ , however, the quality of approximation starts to deteriorate very rapidly.

In practice, we construct the PCA estimator by approximating the true density  $f$  on a large but finite interval  $[v, w]$  so that the number of parameters to be estimated is finite. Specifically, let  $\mathcal{W}_h^{\Delta z}([v, w])$  denotes the “truncated” discrete set, for which

$a_j = 0$  if  $z_j$  is outside the interval  $[v, w]$ :

$$\mathcal{W}_h^{\Delta z}([v, w]) := \left\{ g(x) = \sum_j a_j \phi_h(x - z_j) \in \mathcal{W}_h^{\Delta z} \mid a_j = 0, \text{ if } z_j < v \text{ or } z_j > w \right\}.$$

Thus, instead of the problem in (3) we solve its finite-dimensional approximation:

$$\underset{\hat{f} \in \mathcal{W}_h^{\Delta z}([v, w])}{\text{Minimize}} \quad \sum_{i=1}^n (P_i - D^{-2} \hat{f}(x_i))^2, \tag{6}$$

with  $\Delta z = 0.5h$ . The optimization problem in (6) is a standard quadratic program, for which efficient computational algorithms are available. Typically, the number of parameters to be estimated is from 10 to 30 and finding the solution is very fast.

#### 4. Monte-Carlo experiment

In this section, we perform a Monte-Carlo experiment in which the conditional RND is estimated from a cross-section of option prices using the PCA method and a number of other methods recently proposed in the literature.

Suppose that the actual RND  $f(x)$  is specified as a mixture of three lognormals

$$f(x) = \pi_1 LN(x; \eta_1, \sigma_1) + \pi_2 LN(x; \eta_2, \sigma_2) + \pi_3 LN(x; \eta_3, \sigma_3), \tag{7}$$

$$\pi_1 + \pi_2 + \pi_3 = 1,$$

and  $LN(x; \eta, \sigma)$  is a lognormal density:

$$LN(x; \eta, \sigma) := \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{(\ln \frac{x}{\eta} - \frac{1}{2}\sigma^2)^2}{2\sigma^2}}.$$

The parameters of the lognormals and the mixing probabilities are chosen to describe a typical cross-section of the S&P 500 Index options traded at the Chicago Board Options Exchange (CBOE) with 1 month to maturity. To calibrate the RND  $f(x)$ , we use the closing prices on March 21, 1995 of the S&P 500 options with the maturity date on April 21, 1995. The fitted parameters are

---


$$\begin{array}{lll} \pi_1 = 0.1194 & \pi_2 = 0.8505 & \pi_3 = 0.0301 \\ \eta_1 = 475.59 & \eta_2 = 498.17 & \eta_3 = 524.91 \\ \sigma_1 = 0.0550 & \sigma_2 = 0.0206 & \sigma_3 = 0.0146 \end{array} \tag{8}$$


---

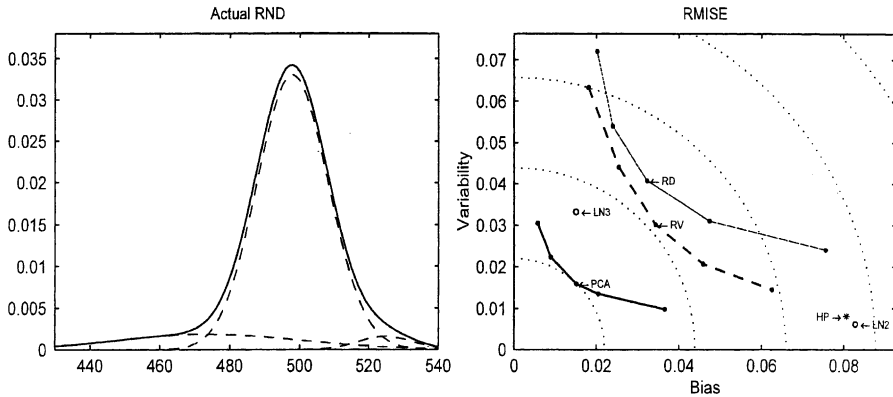


Fig. 2. Monte-Carlo experiment. The left panel plots the actual RND modeled as a mixture of 3 lognormals with parameters given in (8). The lognormals (times weights) are shown by the dashed lines. The actual RND is used to generate 500 sets of (noisy) put prices with 23 strikes ( $x = 430, 435, \dots, 540$ ). The right panel compares performance of 8 RND methods: mixtures of lognormals (LN1, LN2, and LN3), Hermite polynomials (HP), regularization method for density (RD), regularization method for implied volatility (RV), sigma shape polynomials (SSP), and PCA. By design, the LN3 method is the correct parametric model. For each estimator, variability ( $\mu_2$ ) is plotted versus bias ( $\mu_1$ ). The dotted lines are levels of constant RMISE. Two methods (LN1 and SSP) are clipped out because of very large biases. The nonparametric methods are shown for a range of possible values of the penalty parameter  $\lambda$  (RD and RV) or the bandwidth  $h$  (PCA).

The resulting RND  $f(x)$  is shown in the left panel of Fig. 2 along with the 3 lognormal components. It should be pointed out that  $f(x)$  is not in the approximating set  $\mathcal{W}_h$  for any  $h$ , because a *lognormal* density cannot be represented as a convolution with the *normal* basis density.

We assume that we observe 23 strikes  $x_i = 430, 435, \dots, 540$ . To create simulated put prices  $P(x_i)$ , we add noise  $\varepsilon_i$  to the theoretical prices computed from  $f(x)$ . Noise  $\varepsilon_i$  is introduced to model observational errors that arise from nonsynchronicity, bid-ask spread, and other market imperfections. We assume that  $\varepsilon_i$  is independently and uniformly distributed with  $E[\varepsilon_i] = 0$ . The details of our specification for  $\varepsilon_i$  are explained in Appendix B.

For each set of simulated prices, the RND is estimated with 8 different methods, which include both parametric and nonparametric ones:

- Mixtures of lognormals, as in Bahra (1997), Melick and Thomas (1997):
  - LN1—one lognormal (or, the Black–Scholes model),
  - LN2—two lognormals,
  - LN3—three lognormals.
- Hermite polynomials/Edgeworth expansion, as in Madan and Milne (1994), Jarrow and Rudd (1982), Longstaff (1995), Abken et al. (1996)—HP.
- Regularization method for density, as in Jackwerth and Rubinstein (1996)—RD.

- Regularization method for implied volatility, as in Campa et al. (1998), Bliss and Panigirtzoglou (2002), Jackwerth (2000)—RV.
- Sigma shape polynomials, as in Rosenberg (1998)—SSP.
- Positive convolution approximation—PCA.

The competitors of PCA are selected because of three main reasons. First, the above methods are sufficiently flexible to accommodate general shapes of densities. Second, they represent rather distinct ideas to the RND estimation. Three, the methods have been extensively studied in the literature and have been used in a number of important empirical applications. (See Jackwerth (2000) for the references.)

Of the 8 methods, three are nonparametric (RD, RV, and PCA)<sup>7</sup> while the others are parametric. The parametric methods (mixtures of lognormals, HP, and SSP) can be viewed as flexible, because the number of parameters in these methods can be increased to allow for more general shapes of densities. Five methods (LN1, LN2, LN3, SSP, and PCA) produce estimators that are always arbitrage-free. HP, RD, and RV may produce estimators with negative probabilities. Note also that the simulation design is biased in favor of the LN3 method, which is designated as the true parametric model. The details of the implementation of the 8 methods are explained in Appendix B. The computer code is available at [www.uic.edu/~olegb/](http://www.uic.edu/~olegb/).

To compare different estimators, we focus on the RMISE criterion. If  $\hat{f}$  is an estimator of  $f$ , then the (normalized) RMISE is defined as

$$\begin{aligned} \mu(\hat{f}) &:= \text{RMISE}(\hat{f}) = \frac{1}{\|f\|} (E[\|\hat{f} - f\|^2])^{1/2} \\ &= \frac{1}{\|f\|} \left( E \left[ \int_{-\infty}^{\infty} (\hat{f}(x) - f(x))^2 dx \right] \right)^{1/2}. \end{aligned}$$

For our purposes, it is convenient to represent RMISE as

$$\begin{aligned} \mu^2(\hat{f}) &= \mu_1^2(\hat{f}) + \mu_2^2(\hat{f}), \\ \mu_1(\hat{f}) &:= \text{RISB}(\hat{f}) = \frac{1}{\|f\|} \left( \int_{-\infty}^{\infty} (E[\hat{f}(x)] - f(x))^2 dx \right)^{1/2}, \\ \mu_2(\hat{f}) &:= \text{RIV}(\hat{f}) = \frac{1}{\|f\|} \left( \int_{-\infty}^{\infty} E[(\hat{f}(x) - E[\hat{f}(x)])^2] dx \right)^{1/2}, \end{aligned} \tag{9}$$

where RISB is the (normalized) *root integrated squared bias* and RIV is the (normalized) *root integrated variance*. Intuitively, RMISE is a measure of the overall quality of the estimator, RISB is a measure of the accuracy, and RIV is a measure of the stability. The representation in (9) allows us to study the relative contributions of the bias  $\mu_1$  and the variability  $\mu_2$  to RMISE of different methods.

<sup>7</sup> More precisely, the regularization method for implied volatility (RV) should be qualified as *semi-nonparametric*.

The results of simulations for the 8 methods are presented in the right panel of Fig. 2, where the variability  $\mu_2$  is plotted against the bias  $\mu_1$ . The dotted lines show constant levels of RMISE. Because nonparametric methods require specifying the penalty parameter  $\lambda$  (RD and RV) or the bandwidth  $h$  (PCA), their estimators are constructed for ranges of values of  $\lambda$  ( $h$ ). Therefore, these methods are shown on the plot with “curves”. The downward sloping curves reflect the tradeoff between the variability and the bias. For each of the three methods, we search for the optimal parameter ( $\lambda$  or  $h$ ) that minimizes RMISE. These optimal parameters are

$$\lambda_{RD}^* = 9.8 \times 10^4, \quad \lambda_{RV}^* = 3.5 \times 10^5, \quad h^* = 10.5.$$

Since there is only one estimator corresponding to each of the parametric methods, they are shown with single “points”. Note that, because LN1 and SSP have very large bias terms, these methods are not shown on the plot. The bias, the variability, and RMISE for the 8 methods are given in the following table:

	LN1	LN2	LN3	HP	RD*	RV*	SSP	PCA*
Bias ( $\mu_1$ )	0.229	0.083	0.015	0.081	0.032	0.034	0.206	0.015
Variability ( $\mu_2$ )	0.003	0.006	0.033	0.008	0.041	0.030	0.012	0.016
RMISE ( $\mu$ )	0.229	0.083	0.036	0.081	0.052	0.046	0.207	0.022

In Fig. 3, the estimators of the 8 methods are constructed for a single simulation. For the nonparametric methods RD, RV, and PCA, the estimators are shown for the optimal value of  $\lambda^*$  ( $h^*$ ).

As seen from Fig. 2 and the above table, PCA performs considerably better than the competitors. PCA is both very accurate (as measured by the bias) and stable (as measured by the variability). Less flexible methods LN1, LN2, HP, and SSP have better stability than PCA but at the expense of producing much larger biases. Biases are particularly large for LN1 and SSP.<sup>8</sup> The other methods (LN3, RD, and RV) are both less stable and less accurate than PCA. In particular, RMISE for RD and RV are 137% and 108% larger than that for PCA.

The striking feature of Fig. 2 is that PCA, which is a fully *nonparametric* method, outperforms the correct *parametric* method LN3. The difference in the performances of the two methods is substantial: RMISE for LN3 is 66% larger than that for PCA.

The explanation for this counterintuitive result is as follows. It is well known in the econometrics literature that, in small samples, a complicated true model may underperform overly smooth false model. The reason is that the variance of estimating the many parameters of the true model can be much larger than the bias of the smooth false model. In large samples, however, the true model can be expected to outperform any other method.

The above table reveals that PCA has much smaller variability term  $\mu_2$  than LN3 (while the two methods have the same bias term  $\mu_1$ ). In the Monte-Carlo experiment,

<sup>8</sup> A very large bias of LN1 is, of course, expected. It is more surprising, however, that SSP, which is a five-parameter model, is also a rather inflexible method.

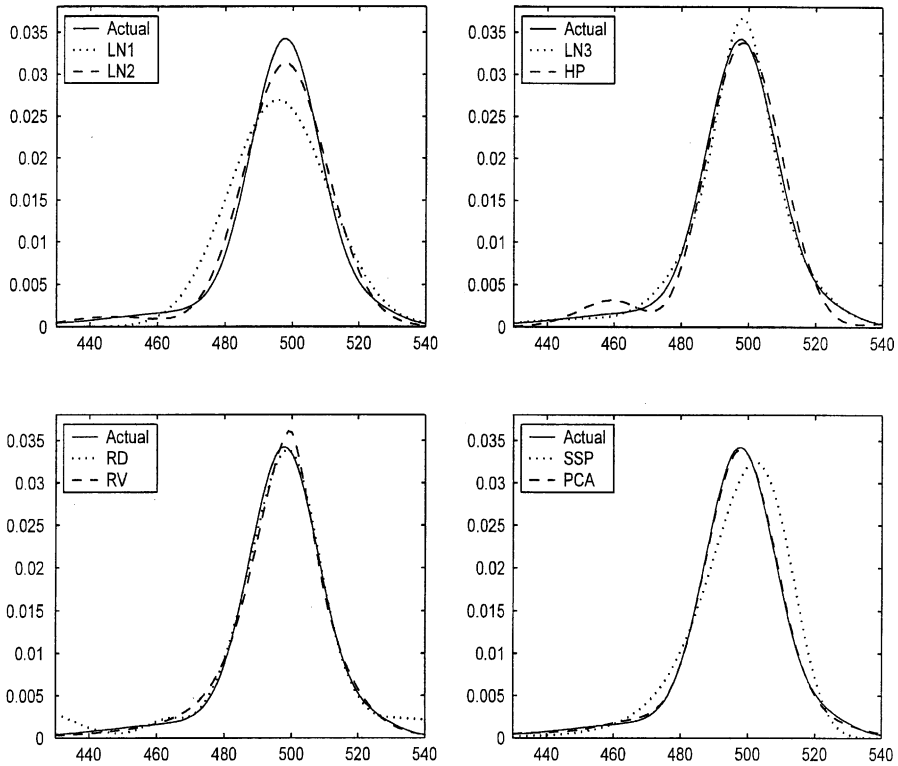


Fig. 3. Monte-Carlo experiment. This figure plots the estimators of the 8 methods for a single simulation. See also Fig. 2. For the nonparametric methods RD, RV, and PCA, the estimators are shown for the optimal penalty parameters ( $\lambda_{RD}^* = 9.8 \times 10^4$ ,  $\lambda_{RV}^* = 3.5 \times 10^5$ ) and bandwidth ( $h^* = 10.5$ ).

the variability of LN3 is relatively large because its 9 parameters (subject to two equality constraints) must be estimated from a tiny sample of only 23 option prices. In contrast, the variability term  $\mu_2$  for PCA is low because PCA is constrained to produce smooth estimators.<sup>9</sup>

In terms of computation time, the methods can be divided into slow (LN3 and SSP) and fast (the rest). For the fast methods, computation time is not an issue at all. The ranking of the methods (from the fastest to the slowest) is RD, LN1, PCA, HP, LN2, RV, and then SSP, and LN3.<sup>10</sup>

<sup>9</sup>Note that PCA involves a larger number of independent parameters than LN3 (in the experiment, the number of coefficients  $\{a_j\}$  is more than 20). Section 5 provides more discussion as to why PCA is constrained from overfitting.

<sup>10</sup>For comparison, LN3 is about 30 times slower than PCA in our simulations. Note also that, among the fast methods, the two regularization methods RD and RV are the fastest and the slowest ones. RV is slower than RD only because RV involves computing implied volatilities for all strikes.

Finally, we have repeated the Monte-Carlo analysis for a number of different specifications. In particular, when fitting the mixture of lognormals, we have used the option prices for several other trading days. We have also tried different specification for noise  $\varepsilon_i$ . In all these experiments, the relative performance of the 8 methods has been very similar. Invariably, PCA has significantly outperformed the alternatives. As with any Monte-Carlo analysis, our design can be criticized for specific assumptions. It is important to emphasize, however, that the primary goal of our experiment is to compare the *relative* performance of the alternative methods and, therefore, the design does not have to be completely realistic as long as it does not favor some methods over others. In this respect, the design serves the purpose.

To illustrate the robustness of the Monte-Carlo analysis, we report the results for one modification. We use the same specification as before, except we *double* the level of noise  $\varepsilon_i$ . For the case of high noise, the bias, the variability, and RMISE are given in the following table:

	LN1	LN2	LN3	HP	RD*	RV*	SSP	PCA*
Bias ( $\mu_1$ )	0.229	0.083	0.025	0.080	0.048	0.046	0.205	0.021
Variability ( $\mu_2$ )	0.006	0.014	0.066	0.016	0.062	0.042	0.020	0.029
RMISE ( $\mu$ )	0.229	0.084	0.070	0.082	0.079	0.062	0.206	0.035

The table shows that the relative ranking remains essentially the same (the only exception is that RV now slightly outperforms LN3). PCA still performs much better than the competitors. In particular, RMISE for LN3, RD, and RV are 99%, 123%, and 76% larger than that for PCA.<sup>11</sup> In fact, the PCA estimator for the *high* noise still outperforms the LN3 estimator for the *low* noise.

It may also be instructive to compare how the optimal parameters of the nonparametric methods change with the level of noise. These are now

$$\lambda_{RD}^* = 2.4 \times 10^5, \quad \lambda_{RV}^* = 7.0 \times 10^5, \quad h^* = 10.5.$$

Predictably, the optimal parameters for the two regularization methods increase, reflecting that more smoothing is needed when noise is high. For example, for the RD method,  $\lambda_{RD}^*$  increases 2.5 times from the previous case. In contrast, the optimal bandwidth for PCA does not change (at least within 0.1 precision that we employ when searching for the optimal bandwidth). This is because  $h^*$  is mainly determined by the shape of the estimated density  $f$ . Unlike the optimal parameters for the regularization methods,  $h^*$  is insensitive to the level of noise. (It is also possible to demonstrate that  $h^*$  is rather insensitive to the sample size, again, in sharp contrast to the regularization methods.)

<sup>11</sup> Typically, RMISE increases with the level of noise. However, the increase is smaller for less flexible methods than for more flexible ones. Somewhat surprisingly, for the two inflexible methods LN1 and SSP, RMISE does not increase at all.

## 5. Discussion

Section 2.2 discusses the existing methods for the RND estimation and classifies them into several categories. It is interesting to point out that PCA does not fall in any of the categories, although it does share some similarities with two groups of methods: parametric mixture methods and nonparametric smoothing techniques.

### 5.1. PCA versus mixture methods

In mixture methods, an estimator is constructed as a weighted sum of several simple densities. In the context of the RND estimation, Bahra (1997) and Melick and Thomas (1997) have proposed to use mixtures with 2 and 3 lognormal densities, respectively.<sup>12</sup>

PCA may be viewed as a mixture of a large number of simple densities (in this paper, normal densities) with *fixed* parameters. Specifically, the PCA estimator is a mixture of  $k$  normal densities which have the same bandwidth  $h$  and whose centers are placed on an equally-spaced grid. The PCA minimization problem, therefore, amounts to finding only  $k$  mixing probabilities  $\{a_i\}$ . This problem is computationally simple and can easily be solved for very large  $k$  (say, 100 or more). In the Monte-Carlo experiment of the previous section, for example, the number of normals  $k$  depends on the selected bandwidth  $h$  and varies from 21 to 25.

Consider now the standard mixture of normals (MN) with  $k$  components. In MN, one needs to estimate  $3k$  parameters:  $k$  centers of densities,  $k$  standard deviations, and  $k$  mixing probabilities. One difficulty with MN is that the corresponding minimization problem is highly nonlinear and computationally very challenging. As a practical matter, this problem cannot be solved satisfactory when  $k > 3$ . This is because the objective function is not globally convex, there are numerous local minima, and the search procedure can converge to a corner solution. Another problem is that MN is prone to overfitting in small samples. For example, when a mixture of 3 normals (or lognormals) is fitted to only about 20 option prices, the estimator often has spurious spikes due to sampling noise.

In contrast, the important advantage of PCA is that it is constrained from overfitting. By this we mean the following. Recall that in the Monte-Carlo experiment, we use 23 option prices to estimate 21 to 25 mixing probabilities  $\{a_i\}$ . At first glance, this would indicate severe overfitting. In reality, this is not the case. Even with so many independent parameters, the PCA estimator does not fit the data exactly.

To illustrate this property, consider again the setup of the Monte-Carlo experiment. Suppose that we keep constant the PCA bandwidth  $h$  and the interval  $[u, v]$  of the truncated discrete set  $\mathcal{W}_h^{\Delta z}([v, w])$  in the problem (6). By reducing the grid size  $\Delta z \leq 0.5h$ , we can increase the number of normal components  $k$  in the PCA estimator. It turns out, however, that, as  $k$  increases to 30, 100, or even more, the goodness-of-fit of the

<sup>12</sup> Mixtures of lognormals have probably been the most popular method for the RND estimation, among both academics and practitioners. The method is used in applications by Leahy and Thomas (1996), Campa et al. (1998), and many others. A close variant of this method is when the distribution of the *log*-return is approximated by a mixture of *normals*, as in Ritchey (1990), Söderlind and Swensson (1997), and Söderlind (2000).

problem (6) improves only negligibly and the shape of the corresponding estimators remain virtually the same. (Note that even such an extreme overparameterization presents no problem for PCA. Having too many parameters does not improve the performance much, but it does not hurt it either.)

The fact that all normals in PCA have the same bandwidth means that the PCA estimator is always a smooth, well-behaved density—the estimator cannot have spikes or wiggles. Intuitively, for a given  $h$ , it is impossible to generate fine details (those that  $\ll h$ ), even if a very large number of normals is used. An additional insight is provided by Proposition 4, which states that the discrete set  $\mathcal{W}_h^{\Delta z}$  approaches the continuous set  $\mathcal{W}_h$  as the grid step  $\Delta z$  goes to zero (thus,  $k$  goes to infinity). When  $k$  is sufficiently large, the difference between the two sets becomes negligible. At that point, the estimator from the discrete set is practically the same as the estimator selected from the continuous set. No further improvement can be achieved by increasing  $k$  after that.

This property of PCA is in sharp contrast to mixture methods (and many other flexible methods, such as polynomial methods, orthogonal series expansions, and certain spline methods), for which the *perfect* fit is achieved when the degree of freedom reaches the number of observations.<sup>13</sup> The fact that PCA is constrained from overfitting is particularly useful for ill-posed, small-sample problems.

## 5.2. PCA versus smoothing techniques

Smoothing techniques are a broad group of related nonparametric methods which include kernel methods, regularization methods, spline methods, orthogonal series expansions, and others. In all these methods, there is a *smoothing parameter* that controls the tradeoff between the goodness-of-fit and some form of smoothness. (For the aforementioned methods, the smoothing parameter is the bandwidth  $h$ , the penalty parameter  $\lambda$ , the number of splines, and the number of orthogonal functions, respectively.) Smoothing techniques are sometimes called *linear smoothers*, because their estimators are linear combinations of the observed responses. Specifically, these estimators can be reexpressed as

$$\hat{r}(x) = \sum_{i=1}^n w_i(x) y_i, \quad (10)$$

where the estimated model is  $y_i = r(x_i) + \varepsilon_i$ , and the weight function  $w_i(x)$  generally depends on all datapoints  $\{x_i\}$ , that is,  $w_i(x) = w_i(x; x_1, \dots, x_n)$ . (See Scott, 1992, Chapter 8.)

Even though PCA may appear similar to kernel methods, it is *not* a smoothing technique and cannot be reformulated as in (10).<sup>14</sup> This fact explains why PCA exhibits different properties from smoothing techniques.

In particular, let us contrast the asymptotic behavior of the optimal bandwidths in PCA and kernel methods, as the number of observations grows to infinity. Suppose

<sup>13</sup> Of course, the perfect fit is undesirable in the presence of sampling noise, because it leads to a huge variability term  $\mu_2$ .

<sup>14</sup> Unlike kernel methods, PCA requires solving a constrained optimization problem and its estimator is not linear in the responses  $\{y_i\}$ .

that the estimated density  $f$  belongs to the admissible set  $\mathcal{W}_{h_0}$ . As the sample size  $n$  increases to infinity, the optimal bandwidth for kernel methods will approach zero at the asymptotic rate that is a power of  $n$ . (Similarly, if regularization methods are used, then the optimal penalty parameter  $\lambda$  will also converge to zero as  $n \rightarrow \infty$ .) In contrast, the bandwidth for PCA will never fall below  $h_0$ .

In general, the optimal smoothing parameter of smoothing techniques depends not only on the shape of the estimated density  $f$ , but also on the sample size, statistical properties of noise, and other characteristics of the problem. Because of this, choosing the appropriate smoothing parameter is a difficult task in real applications.<sup>15</sup> In contrast, the bandwidth in PCA is mainly determined by  $f$  and is relatively insensitive to the other factors. In particular, the Monte-Carlo experiment of Section 4 has illustrated that  $h^*$  is insensitive to the level of noise. Using these insights, Bondarenko (2000) proposes an effective data-driven procedure for estimating the PCA bandwidth. Briefly, the procedure first constructs a preliminary, *undersmoothed* estimator for a small bandwidth. Then, the preliminary estimator is used to estimate  $h^*$ . In simulations, this procedure is found to select the bandwidth that is very close to the optimal value.

## 6. Conclusion

In this paper, we introduce a new nonparametric method for recovering the conditional RND from prices of traded options. The method is termed the *positive convolution approximation* (PCA). The attractive properties of the method are that it (1) is flexible and completely agnostic about the true underlying process, (2) controls against overfitting while allowing for small samples, (3) always produces arbitrage-free estimators, and (4) is computationally simple. The Monte-Carlo experiment suggests that PCA might be a promising alternative to the methods proposed in the literature.

Several important topics are left for future research. One interesting extension of our approach is estimation of the *multivariate* RND. Multivariate contingent claims are becoming more common as the use of derivatives increases. Examples include spread options, exchange options, cross currency options. Consider, for instance, the Yen-Pound cross currency options traded on the Philadelphia Stock Exchange. These options, together with the Dollar-Yen and Dollar-Pound currency options, can be used to estimate the *joint* bivariate RND  $f(y_T, d_T)$  of the Dollar-Yen and Dollar-Pound exchange returns  $y_T$  and  $d_T$ . The bivariate RND will then provide valuable new information about the interaction of the two exchange returns, not present in the two *marginal* RNDs of  $y_T$  and  $d_T$ . (See also Rosenberg, 1998.)

We believe that PCA will be useful in this application because of two reasons. First, it is straightforward to extend PCA to two variables. One only needs to modify

<sup>15</sup> Recall that the Monte-Carlo experiment in Section 4 does not address the issue of how to select the optimal parameters ( $\lambda$  and  $h$ ) of the nonparametric methods RD, RV, and PCA when the actual density  $f$  is *not* provided. In the experiment, the optimal parameters are found by minimizing RMISE, which requires the knowledge of  $f$ . Since in real applications this information is not available, the optimal parameters must be estimated.

the definition of the admissible set by replacing one-dimensional convolutions with two-dimensional ones. The method will again produce smooth, arbitrage-free estimators and be computationally efficient. Second, the fact that PCA is constrained from overfitting will be important for this application. In small samples, PCA handles well the *curse of differentiation* and it should also be able to handle well the *curse of dimensionality*.

Another important issue is estimation of statistical densities from historical returns. Nonparametric estimation of statistical densities is most commonly done with kernel methods. However, Bondarenko (2000) provides an example that suggests that, for this problem, PCA might be a promising alternative to the canonical kernel method. We leave this investigation for another paper.

### Acknowledgements

I would like to thank for helpful comments Gib Bassett, Federico Bandi, Luca Benzoni, Per Mykland, Ser Huang Poon, Pietro Veronesi, Rafal Wojakowski, and participants at University of Chicago Financial Engineering Workshop, European Finance Association Meeting 2000, CBOT European Research Symposium 2000, and Western Finance Association Meeting 2001. The paper has benefited from suggestions of Eric Ghysels and George Tauchen (the editors) and two referees. Any remaining errors are mine. Financial support from UIC CRB and CBOT selected paper award are gratefully acknowledged. For a longer version of this paper which includes an empirical application, see Bondarenko (2000).

### Appendix A. Proofs

**Proof of Proposition 1.** Consider two densities  $g_1$  and  $g_2$  from  $\mathcal{W}_h$  that optimally approximate  $f_1$  and  $f_2$ . That is,  $g_i = \phi_h * u_i$  for some  $u_i \in L^d$  and  $\rho(f_i, \mathcal{W}_h) = \|f_i - g_i\|$ , for  $i = 1, 2$ . (Recall that the space  $\mathcal{W}_h$  is complete in  $L_2$  and, thus, such  $g_1$  and  $g_2$  always exist.)

For  $\alpha \in [0, 1]$ , let  $f := \alpha f_1 + (1 - \alpha)f_2$  and  $g := \alpha g_1 + (1 - \alpha)g_2 = \phi_h * (\alpha u_1 + (1 - \alpha)u_2)$ . Since  $g$  belongs to  $\mathcal{W}_h$ , we obtain

$$\rho(f, \mathcal{W}_h) = \min_{g' \in \mathcal{W}_h} \|f - g'\| \leq \|f - g\| \leq \alpha \|f_1 - g_1\| + (1 - \alpha) \|f_2 - g_2\|. \quad \square$$

**Proof of Proposition 2.** Suppose that  $h_1 < h_2$  and let  $h = \sqrt{h_2^2 - h_1^2}$ . Consider the density  $g_2 = \phi_{h_2} * u_2$  (for some  $u_2 \in L^d$ ) that optimally approximates  $f$  in  $\mathcal{W}_{h_2}$ . We now argue that since the density  $g_2$  belongs to  $\mathcal{W}_{h_2}$ , it must also belong to a *less smooth* admissible set  $\mathcal{W}_{h_1}$ . Indeed, for the normal basis densities,  $\phi_{h_2} = \phi_{h_1} * \phi_h$ . Using properties of a convolution, we can represent  $g_2$  as

$$g_2 = \phi_{h_1} * u_1, \quad \text{where } u_1 := \phi_h * u_2.$$

Proposition 2 now follows immediately since

$$\rho(f, \mathcal{W}_{h_1}) = \min_{g \in \mathcal{W}_{h_1}} \|f - g\| \leq \|f - g_2\| = \rho(f, \mathcal{W}_{h_2}). \quad \square$$

**Proof of Proposition 3.** In view of Proposition 2, if  $\sigma \geq h$ , we have that  $\rho(\phi_\sigma, \mathcal{W}_h) \leq \rho(\phi_\sigma, \mathcal{W}_\sigma) = 0$  (since  $\phi_\sigma$  belongs to  $\mathcal{W}_\sigma$ ). If  $\sigma < h$ , then

$$\rho(\phi_\sigma, \mathcal{W}_h) = \min_{g \in \mathcal{W}_h} \|\phi_\sigma - g\| = \|\phi_\sigma - \phi_h\|.$$

By direct computation,

$$\begin{aligned} \|\phi_\sigma - \phi_h\|^2 &= \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2h^2}} \right)^2 dx \\ &= \frac{1}{2\sqrt{\pi}} \left( \frac{1}{\sigma} + \frac{1}{h} - \frac{2\sqrt{2}}{\sqrt{\sigma^2 + h^2}} \right). \quad \square \end{aligned}$$

**Proof of Proposition 4.** Let  $f_h = \phi_h * u$ , where  $u \in L^d$ . Define  $g_h^{\Delta z}$  as

$$g_h^{\Delta z}(x) = \sum_{j=-\infty}^{\infty} a_j^{\Delta z} \phi_h(x - z_j), \quad \text{where } a_j^{\Delta z} = \int_{-0.5\Delta z}^{+0.5\Delta z} u(z_j + y) dy.$$

The function  $g_h^{\Delta z}$  belongs to the set  $\in \mathcal{W}_h^{\Delta z}$ . Moreover,  $g_h^{\Delta z}$  converges to  $f_h$  as the grid step  $\Delta z$  approaches zero. In particular, it is easy to check that  $|f_h(x) - g_h^{\Delta z}(x)| \leq Const (\Delta z)^2$ , and

$$\lim_{\Delta z \rightarrow 0} \|f_h - g_h^{\Delta z}\| = 0.$$

We use the following series of inequalities:

$$\|f - f_h\| \leq \|f - f_h^{\Delta z}\| \leq \|f - g_h^{\Delta z}\| \leq \|f - f_h\| + \|f_h - g_h^{\Delta z}\|.$$

The first inequality obtains because  $f_h$  optimally approximates  $f$  in  $\mathcal{W}_h$  and because  $f_h^{\Delta z} \in \mathcal{W}_h$ . The second inequality obtains because  $f_h^{\Delta z}$  optimally approximates  $f$  in  $\mathcal{W}_h^{\Delta z}$  and because  $g_h^{\Delta z} \in \mathcal{W}_h^{\Delta z}$ . In the limit,

$$\lim_{\Delta z \rightarrow 0} \|f - f_h^{\Delta z}\| = \|f - f_h\|. \tag{A.1}$$

Because  $f_h$  is the optimal approximant of  $f$  in  $\mathcal{W}_h$ , we must have

$$(f - f_h, f_h^{\Delta z} - f_h) \leq 0,$$

where  $(\cdot, \cdot)$  denotes the inner product in  $L_2$ , i.e.,  $(g_1, g_2) = \int_{-\infty}^{\infty} g_1(y)g_2(y) dy$ . (The inequality must hold because otherwise, for a sufficiently small  $\alpha$ , there exists a convex combination  $g_\alpha = (1 - \alpha)f_h + \alpha f_h^{\Delta z}$  such that  $\|f - g_\alpha\| < \|f - f_h\|$ .) It follows from the above inequality that

$$\begin{aligned} \|f - f_h^{\Delta z}\|^2 &= \|f - f_h\|^2 + \|f_h^{\Delta z} - f_h\|^2 - 2(f - f_h, f_h^{\Delta z} - f_h) \\ &\geq \|f - f_h\|^2 + \|f_h^{\Delta z} - f_h\|^2. \end{aligned}$$

Taking the limit and using (A.1), we obtain

$$\lim_{\Delta z \rightarrow 0} \|f_h^{\Delta z} - f_h\| = 0. \tag{A.2}$$

The two claims in Proposition 4 follow from (A.2). Specifically, the first claim obtains from (A.2) and the fact that  $f_h$  and  $f_h^{\Delta z}$  have bounded first derivatives. To prove the second claim, we observe that for  $h_0 \geq h$  and any  $g \in \mathcal{W}_{h_0}$ , the distance  $\rho(g, \mathcal{W}_h^{\Delta z}) = 0$ . Therefore,

$$\begin{aligned} \lim_{\Delta z \rightarrow 0} \rho(\mathcal{W}_{h_0}, \mathcal{W}_h^{\Delta z}) &= \lim_{\Delta z \rightarrow 0} \max_{g \in \mathcal{W}_{h_0}} \rho(g, \mathcal{W}_h^{\Delta z}) \\ &= \max_{g \in \mathcal{W}_{h_0}} \lim_{\Delta z \rightarrow 0} \rho(g, \mathcal{W}_h^{\Delta z}) \\ &= \max_{g \in \mathcal{W}_{h_0}} \rho(g, \mathcal{W}_h) = 0. \quad \square \end{aligned}$$

**Appendix B. Monte-Carlo experiment**

The computer code (Matlab) for the Monte-Carlo analysis is available at [www.uic.edu/~olegb/](http://www.uic.edu/~olegb/). Below we provide the details of the experiment. When evaluating the performance of the 8 estimation methods, it is important to ensure their implementations are as comparable as possible. This sometimes requires making slight adjustments to the original procedures proposed in the cited references.

*B.1. Methods*

For the 5 parametric methods (LN1, LN2, LN3, HP, SSP), the estimator  $\hat{f}(x) = \hat{f}(x; \theta)$  is constructed by minimizing the goodness-of-fit

$$\text{Minimize}_{\theta} \sum_{i=1}^n (P_i - D^{-2} \hat{f}(x_i; \theta))^2, \tag{B.1}$$

with respect to a general vector of parameters  $\theta$ . For example,  $\theta$  consists of 4 parameters for HP and of 9 parameters for LN3. We require that the estimator satisfy the following two constraints:

$$\int_0^{\infty} \hat{f}(x) dx = 1, \tag{B.2}$$

$$\int_0^{\infty} x \hat{f}(x) dx = S_t, \tag{B.3}$$

where  $S_t$  is the underlying asset price on the trading date- $t$ . For some methods, constraint (B.2) is satisfied automatically.

*LN1, LN2, and LN3 methods:* Conceptually, estimations for these models are straightforward. Consider, for example, the LN3 method. In this case, the 9 parameter

model in (7) is estimated with  $\theta = (\pi_1, \eta_1, \sigma_1, \pi_2, \eta_2, \sigma_2, \pi_3, \eta_3, \sigma_3)$ , subject to two constraints:

$$\pi_1 + \pi_2 + \pi_3 = 1,$$

$$\pi_1 v_1 + \pi_2 v_2 + \pi_3 v_3 = S_t.$$

Computationally, however, LN3 is difficult. The estimation requires solving a constrained nonlinear optimization problem with a large number of parameters. Since the objective function is not globally convex, the numerical procedure may converge to a local minima as opposed to the global one. Another difficulty is that the numerical procedure may converge to a corner solution or fail to converge at all. Moreover, the final solution is sensitive to starting values of the parameters: when the starting values are chosen very close to the true ones, LN3 performs well; however, the performance is much worse when the starting values are far from the true ones.

To deal with these problems, we follow an “evolutionary” approach whereby a problem with a smaller number of variables is solved first. Specifically, we estimate LN1 and use its estimator to form a starting guess for LN2. Then we estimate LN2 and use its estimator to form a starting guess for LN3. This guarantees that LN3 always converges to a sensible minimum and that the goodness of fit for LN3 is always better than that for LN2, which in turn is better than that for LN1. Furthermore, to facilitate the numerical convergence, we impose the lower and upper bounds on possible values of the parameters.

*HP method:* This estimation method has been derived using two related theoretical techniques: (1) the Edgeworth expansion as in Jarrow and Rudd (1982), Longstaff (1995), and (2) a more general Hermite polynomial expansion as in Madan and Milne (1994). In this approach, the RND estimator has a four-parameter representation:

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{-\frac{z^2}{2}} \left( 1 + \frac{\eta_3}{6}(z^3 - 3z) + \frac{\eta_4}{24}(z^4 - 6z^2 + 3) \right),$$

where

$$z = z(x) = \frac{\ln(x/\eta_1) - 0.5\sigma^2}{\sigma}.$$

The four parameters  $\theta = (\sigma, \eta_1, \eta_2, \eta_3)$  are estimated by solving the problem in (13) subject to constraint (15). Constraint (14) is satisfied automatically by the above specification. The estimator  $\hat{f}(x)$  may result in negative probabilities.

*Sigma shape polynomials (SSP):* This estimation technique has been proposed in Rosenberg (1998). In this approach, the RND is represented by a lognormal density with the parameter  $\sigma$  being a function of the strike  $x$ . Specifically,

$$\hat{f}(x) = \frac{\lambda}{\sqrt{2\pi} \sigma(x)x} \exp\left(-\frac{(\ln \frac{x}{\eta} - \frac{1}{2}\sigma^2(x))^2}{2\sigma^2(x)}\right),$$

where  $\ln \sigma(x)$  is a quadratic polynomial of  $\ln x$ :

$$\ln \sigma(x) = \beta_0 + \beta_1 \ln x + \beta_2 (\ln x)^2.$$

In the above representation,  $\lambda$  is a scaling factor which ensures that the density integrates to one. Put prices are computed from  $\hat{f}(x)$  via numerical integration. The five

parameters  $\theta = (\lambda, \eta, \beta_0, \beta_1, \beta_2)$  are then estimated in the nonlinear minimization problem, subject to the constraints (B.2)–(B.3). Computationally, SSP is difficult because it involves numerical integration and because the gradient of the objective function cannot be computed analytically, making the search algorithm much less efficient. As with LN2 and LN3, we facilitate the numerical convergence by specifying good starting values for parameters and by imposing the lower and upper bounds on possible values of the parameters.

*RD method:* When implementing the RD method, we closely follow Jackwerth and Rubinstein (1996). The estimator of the RND is assumed to have a large but finite support. The support is taken to be  $[0, 2S_t]$ . It is sampled at discrete points  $z_j$  ( $z_j = j\Delta$ ,  $j=0, 1, \dots, J$  with the increment  $\Delta z = \$5$ ) so that the grid includes available strikes  $\{x_i\}$ . Then Jackwerth and Rubinstein search for a discrete function  $\hat{P}(z_j)$  which achieves the optimal tradeoff between the goodness of fit to the available puts  $P_i$  at strike  $\{x_i\}$  and the smoothness as measured by a finite-difference approximation of the fourth derivative. Specifically, the following minimization problem is solved:

$$\text{Minimize}_{\hat{P}(z_j)} \sum_{i=1}^n (P_i - \hat{P}(x_i))^2 + \frac{\lambda}{\Delta z^8} \sum_{j=1}^J (\hat{P}(z_{j+2}) - 4\hat{P}(z_{j+1}) + 6\hat{P}(z_j) - 4\hat{P}(z_{j-1}) + \hat{P}(z_{j-2}))^2,$$

with the boundary conditions

$$\hat{P}(z_{-1}) = 0, \quad \hat{P}(z_0) = 0, \quad \hat{P}(z_{J+1}) = z_{J+1} - S_t, \quad \hat{P}(z_{J+2}) = z_{J+2} - S_t.$$

In the above expression, the first sum is computed over the available strikes  $\{x_i\}$  while the second sum is computed over the whole support  $\{z_j\}$ . As before, the *penalty parameter*  $\lambda$  controls the tradeoff between the goodness of fit and the smoothness. After the solution is found, it is numerically differentiated two times to obtain the RND estimator  $\hat{f}_j = \hat{f}(z_j)$ . For  $x \neq z_j$ , values  $\hat{f}(x)$  are interpolated with a cubic spline. Note that the boundary conditions ensure that constraints (B.2)–(B.3) are satisfied. Still, the estimator  $\hat{f}(x)$  may result in negative probabilities. Note also that we do not employ the “clamping down” procedure of Jackwerth and Rubinstein (1996).

*RV method:* Several papers have proposed to estimate the RND by first smoothing option implied volatilities with splines or related methods. See, for example, Campa et al. (1998), Jackwerth (2000), Bliss and Panigirtzoglou (2002). We implement the RV method as in Jackwerth (2000), which is a modification of the regularization method of Jackwerth and Rubinstein (1996) applied to the implied volatility function. In this approach, the put price  $P_i$  for the strike  $x_i$  is transformed into the implied volatility  $\sigma_i^{imp} = \sigma_i$ . The implied volatility is found by inverting the Black–Scholes formula:

$$P_i = S_t(N(d_1) - 1) - x_i(N(d_2) - 1), \quad d_{1,2} = \frac{\ln(S_t/x_i)}{\sigma_i\sqrt{T-t}} \pm \frac{1}{2}\sigma_i\sqrt{T-t},$$

where  $N(\cdot)$  is the standard normal cdf. The estimator of the implied volatility function  $\hat{\sigma}(z_j)$  is constructed on the equally-spaced grid  $\{z_j\}$  ( $j = 1, \dots, J$ ) on interval  $[x_1, x_n]$ , such that  $z_1 = x_1$ ,  $z_{j+1} = z_j + \Delta z$  with  $\Delta z = 2.5$ , and  $J = 2n - 1$ . The estimator  $\hat{\sigma}(z_j)$

is found by minimizing the tradeoff between the *weighted* goodness of fit to  $\sigma_i$  and the smoothness as measured by a finite-difference approximation of the second derivative:

$$\text{Minimize}_{\hat{\sigma}(z_j)} \sum_{i=1}^n \left( \frac{\sigma_i - \hat{\sigma}(x_i)}{\omega_i} \right)^2 + \frac{\lambda}{\Delta z^4} \sum_{j=2}^{J-1} (\hat{\sigma}(z_{j+1}) - 2\hat{\sigma}(z_j) + \hat{\sigma}(z_{j-1}))^2,$$

where  $\omega_i = \partial P_i / \partial \sigma_i$  is the option vega. This particular weighting ensures that the results of the RV method are comparable with those for the other methods. The inverse vega weighting when fitting implied volatilities gives the same weights to the observations as the equal weighting when fitting put prices in (B.1). See also Bliss and Panigirtzoglou (2002) for the discussion on different weighting schemes. (It should be pointed out, however, that other weighting schemes for the RV method produce very similar results.)

After the solution is found, it is interpolated for all  $x \in [x_1, x_n]$  with a cubic spline. Given the estimator  $\hat{\sigma}(x)$ , the estimator for the RND  $\hat{f}(x)$  is computed analytically from  $\hat{\sigma}(x)$  and its first and second derivatives. The RV method may produce negative probabilities.

*PCA method:* The PCA method is implemented as described in Section 3.3. We solve the problem in (6) for  $\mathcal{W}_h^{\Delta z}([v, w])$  with  $[v, w] = [430, 540]$  and  $\Delta z = 0.5h$ . The number of parameters  $\{a_j\}$  depends on the bandwidth  $h$  and varies from 21 to 25. Constraint (B.2) is satisfied automatically by all functions in  $\mathcal{W}_h^{\Delta z}([v, w])$ . Constraint (B.3) becomes

$$\sum a_j z_j = S_t.$$

When solving the problem in (6), we express the second integral of  $\hat{f}(x)$  as

$$D^{-2} \hat{f}(x) = \sum a_j D^{-2} \phi_h(x - z_j),$$

and use the relationship  $D^{-2} \phi(x) = \phi(x) + xN(x)$ , where  $N(\cdot)$  is the standard normal cdf.

### B.2. Error specification

In the Monte-Carlo experiment, noise terms  $\varepsilon_i$  are added to the theoretical put prices. We have used several specifications for the measurement errors and have found that the relative performance of the estimators is qualitatively very similar. Our first approach (not reported) is to simply model  $\varepsilon_i$ 's as i.i.d. random variables, uniformly distributed on  $[-c, c]$  for some constant  $c$ . In this naive approach, the level of noise does not depend on the strike  $x_i$ . In particular, the level of noise is the same for low-priced out-of-the-money (OTM) options and for high-priced in-the-money (ITM) options.

The second approach is more realistic but also more complex. We assume that an empiricist observes concurrent bid and ask quotes  $(q_i^b, q_i^a)$  for puts on the S&P 500 Index with strikes  $\{x_i\}$ . The empiricist uses the midpoint quote  $P_i = 0.5(q_i^b + q_i^a)$  as an approximation to the true price. The introduced measurement error  $\varepsilon_i$  is assumed to be uniformly distributed on  $[-0.5s_i, 0.5s_i]$ , where  $s_i$  is the spread, i.e.,  $s_i := q_i^a - q_i^b$ . The value of the spread  $s_i$  depends on the strike  $x_i$ ; the spread is larger for the ITM options and smaller for the OTM options.

We assume that  $s_i$  is proportional to the maximum spread permitted by the exchange. The CBOE rules state that the maximum bid-ask spread is  $\frac{1}{4}$  for options with bid quote  $q^b$  below than \$2,  $\frac{3}{8}$  for bid quotes between \$2 and \$5,  $\frac{1}{2}$  for bid quotes between \$5 and \$10,  $\frac{3}{4}$  for bid quotes between \$10 and \$20, and 1 for bid quotes above \$20. To reflect the CBOE rules, we construct function  $M(q)$  to represent the maximum spread for the quote  $q$ . Specifically, let

$$M(0) = \frac{1}{8}, \quad M(2) = \frac{1}{4}, \quad M(5) = \frac{3}{8}, \quad M(10) = \frac{1}{2}, \quad M(20) = \frac{3}{4},$$

$$M(q) = 1, \quad q \geq 50,$$

and  $M(q)$  is linearly interpolated for all other  $q \in [0, 50]$ . The maximum spread  $M(q)$  is about 19% when  $q = \$1$  and about 4% when  $q = \$20$ . Next, we assume that the empiricist observes quotes not only for puts but for calls as well. In practice, the OTM options are more liquid (and have smaller spreads) than the ITM options. Thus, we assume that, between puts and calls, the empiricist uses for estimation the more accurate options. In particular, for large strikes, she employs the put-call parity relation to convert more accurate OTM call prices into the corresponding put prices. In other words, for the strike  $x_i$ , the relevant spread is the minimum of the spreads for put  $P_i$  and call  $C_i = S_t + P_i - x_i$ . Thus, we set

$$s_i = c \min(M(P_i), M(S_t + P_i - x_i)),$$

where  $c$  is a constant. By varying  $c$  the level of noise can be increased or decreased across all strikes. We report the results for two cases: (1)  $c = 0.5$  (the observed spreads are half of the exchange allowed maximum), and (2)  $c = 1$  (the observed spreads equal the allowed maximum).<sup>16</sup> The advantages of the proposed specification for  $\varepsilon_i$  are that (1) noise is smaller in the *absolute* terms for far-from-the-money strikes, (2) noise is larger in the *relative* terms for far-from-the-money strikes, and (3) simulated option prices are always nonnegative.

The third approach (not reported) is similar to the second one. We still assume that  $\varepsilon_i$ 's are independently and uniformly distributed on  $[-0.5s_i, 0.5s_i]$ . However, the spread  $s_i$  is now set as a percentage of the option price, with fixed minimum and maximum.

## References

- Abadir, K., Rockinger, M., 1998. Density-embedded functions. Working paper, HEC.
- Abken, P., Madan, D., Ramamurtie, S., 1996. Estimation of risk-neutral and statistical densities by Hermite polynomial approximation: with an application to Eurodollar futures options. Working paper, Federal Reserve Bank of Atlanta.
- Ait-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance* 53, 499–547.
- Ait-Sahalia, Y., Lo, A., 2000. Nonparametric risk management and implied risk aversion. *Journal of Econometrics* 94, 9–51.

<sup>16</sup> Dumas et al. (1998) studies the S&P 500 Index options over period from 1988 to 1993. They find that the spread behavior is consistent with CBOE's maximum spread rules and report that the average bid-ask spread for the S&P 500 options is approximately 47 cents.

- Aït-Sahalia, Y., Duarte, J., 2000. Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, forthcoming.
- Aparicio, S., Hodges, S., 1998. Implied risk-neutral distribution: a comparison of estimation methods. Working paper, Warwick University.
- Bahra, B., 1997. Implied risk-neutral probability density functions from option prices: theory and application. Working paper, Bank of England.
- Banz, R., Miller, M., 1978. Prices for state-contingent claims: some estimates and applications. *Journal of Business* 51, 653–672.
- Bates, D., 1996. Jumps and stochastic volatility: exchange rate process implicit in deutsch mark options. *Review of Financial Studies* 9, 69–107.
- Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–654.
- Bliss, R., Panigirtzoglou, N., 2002. Testing the stability of implied probability density functions. *Journal of Banking and Finance* 26, 381–422.
- Bondarenko, O., 1997. Testing rationality of financial markets: an application to S&P 500 Index options. Working paper, Caltech.
- Bondarenko, O., 2000. Recovering risk-neutral densities: a new nonparametric approach. Working paper, University of Illinois at Chicago.
- Bondarenko, O., 2002. Statistical arbitrage and securities prices. *Review of Financial Studies*, forthcoming.
- Breedon, D., Litzenberger, R., 1978. Prices of State Contingent Claims Implicit in option Prices. *Journal of Business* 51, 621–652.
- Buchen, P., Kelly, M., 1996. The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis* 31, 143–159.
- Campa, J., Chang, K., 1996. Arbitrage based tests of target zone credibility: evidence from ERM cross-rate options. *American Economic Review* 86, 726–740.
- Campa, J., Chang, K., Reider, R., 1998. Implied exchange rate distributions: evidence from OTC option markets. *Journal of International Money and Finance* 17, 117–160.
- Cox, J., Ross, S., 1976. The valuation of options for alternative stochastic processes. *Journal of Financial Economics* 3, 145–166.
- Derman, E., Kani, I., 1994. The volatility smile and its implied tree. *Quantitative Strategies Research Notes*. Goldman Sachs, New York.
- Dumas, B., Fleming, J., Whaley, R., 1998. Implied volatility functions: empirical tests. *Journal of Finance* 53, 2059–2106.
- Dupire, B., 1994. Pricing with a smile. *Risk* 7, 18–20.
- Garcia, R., Gencay, R., 2000. Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics* 94, 93–115.
- Goldenberger, D., 1991. A unified method for pricing options on diffusion processes. *Journal of Financial Economics* 29, 3–34.
- Gottschling, A., Haefke, C., White, H., 2000. Closed form integration of artificial neural networks with some applications in finance. Working paper, UC San Diego.
- Harrison, M., Kreps, D., 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20, 381–408.
- Harrison, M., Pliska, S., 1981. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Process and Their Applications* 11, 215–260.
- Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6, 327–343.
- Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42, 281–300.
- Jackwerth, J., 1999. Option implied risk-neutral distributions and implied binomial trees: a literature review. *Journal of Derivatives* 7, 66–82.
- Jackwerth, J., 2000. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* 13, 433–451.
- Jackwerth, J., Rubinstein, M., 1996. Recovering probability distributions from option prices. *Journal of Finance* 51, 1611–1631.

- Jarrow, R., Rudd, A., 1982. Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics* 10, 347–369.
- Leahy, M., Thomas, C., 1996. The sovereignty option: the Quebec referendum and market views on the Canadian dollar. Working paper, Federal Reserve Board.
- Longstaff, F., 1995. Option pricing and the martingale restriction. *Review of Financial Studies* 8, 1091–1124.
- Madan, D., Milne, F., 1994. Contingent claims valued and hedged by pricing and investing in a basis. *Mathematical Finance* 4, 223–245.
- Malz, A., 1997. Estimating the probability distribution of the future exchange rate from option prices. *Journal of Derivatives* 5 (2), 18–36.
- Melick, W., Thomas, C., 1997. Recovering an asset's implied PDF from option prices: an application to crude oil during the Gulf crisis. *Journal of Financial and Quantitative Analysis* 32, 91–115.
- Posner, S., Milevsky, M., 1998. Valuing exotic options by approximating the SPD with higher moments. *Journal of Financial Engineering* 7, 109–125.
- Ritchey, R., 1990. Call option valuation for discrete normal mixtures. *Journal of Financial Research* 13, 285–295.
- Ross, S., 1976. Options and efficiency. *Quarterly Journal of Economics* 90, 75–89.
- Rosenberg, J., 1998. Pricing multivariate contingent claims using estimated risk-neutral density functions. *Journal of International Money and Finance* 17, 229–247.
- Rosenberg, J., Engle, R., 2002. Empirical pricing kernels. *Journal of Financial Economics* 64, 341–372.
- Rubinstein, M., 1994. Implied binomial trees. *Journal of Finance* 49, 771–818.
- Rubinstein, M., 1998. Edgeworth binomial trees. *Journal of Derivatives* 5 (3), 20–27.
- Scott, D., 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York.
- Shimko, D., 1993. Bounds of probability. *RISK* 6, 33–37.
- Söderlind, P., 2000. Market expectations in the UK before and after the ERM crisis. *Economica* 67, 1–18.
- Söderlind, P., Swensson, L., 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.