Recovering Risk-Neutral Probability Density Functions from Options Prices using Cubic Splines

Ana Margarida Monteiro* Reha H. Tütüncü† Luís N. Vicente‡

Abstract

We present a new approach to estimate the risk-neutral probability density function (pdf) of the future prices of an underlying asset from the prices of options written on the asset. The estimation is carried out in the space of cubic spline functions, yielding appropriate smoothness. The resulting optimization problem, used to invert the data and determine the corresponding density function, is a convex quadratic or semidefinite programming problem, depending on the formulation. Both of these problems can be efficiently solved by numerical optimization software.

In the quadratic programming formulation the positivity of the risk-neutral pdf is heuristically handled by posing linear inequality constraints at the spline nodes. In the other approach, this property of the risk-neutral pdf is rigorously ensured by using a semidefinite programming characterization of nonnegativity for polynomial functions.

We tested our approach using data simulated from Black-Scholes option prices and using market data for options on the S&P 500 Index. The numerical results we present show the effectiveness of our methodology for estimating the risk-neutral probability density function.

Keywords: option pricing, risk-neutral density estimation, cubic splines, quadratic programming, semidefinite programming.

*Faculdade de Economia, Universidade de Coimbra, Av. Dias da Silva, 165, 3004-512 Coimbra, Portugal (`amonteiro@fe.uc.pt`).

†Department of Mathematical Sciences, 6113 Wean Hall, Carnegie Mellon University, Pittsburgh, PA 15213, USA (`reha@cmu.edu`). Support for this author was provided by the National Science Foundation under grants CCR-9875559 and DMS-0139911.

‡Departamento de Matemática, Universidade de Coimbra, 3001-454 Coimbra, Portugal (`lnv@mat.uc.pt`). Support for this author was provided by Centro de Matemática da Universidade de Coimbra, by FCT under grant POCTI/35059/MAT/2000, by the European Union under grant IST-2000-26063, and by Fundação Calouste Gulbenkian. The author would also like to thank the IBM T.J. Watson Research Center and the Institute for Mathematics and Its Applications for their local support.

1 Introduction

The risk-neutral probability measure is a fundamental concept in arbitrage pricing theory. By definition, a risk-neutral probability measure (RNPM) is a measure under which the current price of each security in the economy is equal to the present value of the discounted expected value of its future payoffs given a risk-free interest rate. Fundamental theorems of asset pricing indicate that RNPMs are guaranteed to exist under an assumption of no arbitrage.

If a unique RNPM on the space of future states of an economy is given, we can price any security for which we can determine the future payoffs for each state in the state space. Therefore, a fundamental problem in asset pricing is the identification of a risk neutral probability measure. While the dynamics of an economy and the parameters for its stochastic models are not directly observable, one can infer some information about these dynamics from the current prices of the securities in this economy. In particular, one can extract one or more implied risk-neutral densities of the future price of a security that are consistent with the prices of options written on that security. When there are multiple RNPMs consistent with the observed prices, one may try to choose the "best" one, according to some criterion. We address this problem in this article using optimization models.

For a stock or index, the set of possible future states can be represented as an interval or ray, discretized if appropriate or necessary. In most cases the number of states in this state space is much larger than the number of observed prices, resulting in a problem with many more variables than equations. This underdetermined problem has many potential solutions and we can not obtain an unique or sensible solution without imposing some additional structure into the risk neutral probability measure we are looking for.

The type of additional structure imposed has been the differentiating feature of the existing approaches to the problem of identifying implied RNPMs. These approaches can be broadly classified as parametric and non-parametric techniques and are reviewed by Jackwerth [13], see also Section 2 below. Parametric methods choose a distribution family (or a mixture of distributions) and then try to identify the parameters for these distributions that are consistent with the observed prices [3, 16]. In non-parametric techniques, one achieves more flexibility by allowing general functional forms and structure is introduced either using prior distributions or smoothness restrictions. Our approach fits into this last category and we ensure the desired smoothness of the RNPM using spline functions.

Spline functions are piecewise polynomial functions that assume a predetermined value at certain points (*knots*) and satisfy certain smoothness properties. Other authors have also used spline fitting techniques in the context of risk-neutral density estimation, see [1, 8]. In contrast to existing techniques, we allow the displacement of spline knots in a superset of the set of points corresponding to option strikes. The additional set of knots makes our model flexible and we use this flexibility to optimize the fit of the spline function to the observed prices. The basic formulation, without requiring the nonnegativity of the risk-neutral probability density function (pdf), is a convex quadratic programming (QP) problem.

Two strategies to impose the nonnegativity of the RNPM are presented and discussed in this paper. The first and the simpler strategy is to require the estimated pdf to remain nonnegative at the spline nodes. This scheme keeps the structure of the problem since it brings only linear inequality constraints to the basic formulation. However, there is no guarantee of nonnegativity between the spline nodes. Our second approach replaces the basic QP formulation with a semidefinite programming (SDP) formulation but rigorously ensures the nonnegativity of the estimated pdf in its entire domain. It is based on an SDP characterization of nonnegative polynomial functions due to Bertsimas and Popescu [2] and requires additional variables and linear equality constraints as well as semidefiniteness constraints on some matrix variables. To our knowledge, this is the first spline function approach to risk-neutral density estimation with a positivity guarantee.

The rest of this paper is organized as follows: In Section 2, we provide the definition of RNPMs and briefly discuss some of the existing approaches. In Section 3, we discuss our spline approximation approach to RNPMs and develop our basic QP optimization model. The treatment of nonnegativity is given in Section 4. Section 5 is devoted to a numerical study of our approach both with simulated and market data. We provide a brief conclusion in Section 6.

2 Risk-neutral probability measures and existing approaches

We consider the following one-period economy: There are n securities whose current prices are given by s_0^i for $i = 1, \ldots, n$. At the end of the current period, the economy will be in one of the states from the state space Ω. If the economy reaches state $\omega \in \Omega$ at the end of the current period, security i will have the payoff $s_1^i(\omega)$. We assume that we know all s_0^i's and $s_1^i(\omega)$'s but do not know the particular terminal state ω, which will be determined randomly.

As an example of the set-up explained in the previous paragraph, we consider a particular security (stock, index, *etc.*) and let the n securities be financial options written on this stock. Here, Ω denotes the state space for the terminal price of the underlying stock and $s_1^i(\omega)$ denotes the payoff of the option i when the underlying stock price is ω at termination. For example, if option i is a European call with strike price K_i to be exercised at the end of the current period, we would have $s_1^i(\omega) = (\omega - K_i)^+$.

Next, we give a definition of RNPMs:

Definition 1 *Consider the economy described above. Let r denote the one-period (risk-free) interest rate. A risk neutral probability measure in the*

- **discrete case** *and on the state space $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ is a vector of positive numbers $p_1, p_2, \dots, p_m$ such that*

  1. $\sum_{j=1}^m p_j = 1$,
  2. $s_0^i = \frac{1}{1+r} \sum_{j=1}^m p_j s_1^i(\omega_j), \quad i = 1, \dots, n$;

- **continuous case** *and on the state space $\Omega = (a, b)$ is a density function $p : \Omega \to \mathbb{R}_+$ such that*

  1. $\int_a^b p(\omega) d\omega = 1$,
  2. $s_0^i = \frac{1}{1+r} \int_a^b p(\omega) s_1^i(\omega) d\omega, \quad i = 1, \dots, n$.

It is well known that the existence of a risk-neutral probability measure is strongly related to the absence of arbitrage opportunities as expressed in the *First Fundamental Theorem of Asset Pricing* (see [10]). We first give an informal definition of arbitrage and then state this theorem:

Definition 2 *An* arbitrage *is a trading strategy*

- *that has a positive initial cash flow and has no risk of a loss later, or*
- *that requires no initial cash input, has no risk of a loss, and a positive probability of making profits in the future.*

Theorem 1 *A risk-neutral probability measure exists if and only if there are no arbitrage opportunities.*

As we argued in the Introduction, since the payoffs of the derivatives depend on the future values of the underlying asset, we can use the prices of these derivatives to get information about the probability distribution of the future values of the underlying. We can say that the prices of option contracts contain some information about the market expectations, namely a possible correspondence between the price of the underlying and its strike.

There are several approaches, reported in the literature, to derive risk-neutral probabilities from options prices (see the surveys in [1], [3], [5], [13], and [19]).

Among the methods developed to estimate the risk-neutral probability measure we can specify: approximation function methods applied to the probability density function, stochastic process methods for the underlying asset, finite difference methods, approximating function methods applied to the volatility smile, and implied binomial tree methods. In the next paragraphs we provide a brief description of these methods. As we will see, some of them assume a specific parametrized form for the density function on the underlying asset and then try to identify the optimal parameters. Others try to fit the data by a risk-neutral probability density function (pdf) with unprescribed shape. Parametric methods derive the risk-neutral pdf's from a set of statistical distributions and the set of observational data. Non-parametric methods infer those densities solely from the set of observational data.

Approximating function methods applied to the probability density functions assume that the risk-neutral density function has a predefined form, such as a mixture of lognormals (see Bahra [3] and Mellick and Thomas [16]). These methods use the option pricing formula (see Cox and Ross [9]), which shows that the price of a call option is the discounted risk-neutral expected value of the payoffs

$$C\left(t,T,K\right)=e^{-r(T-t)}\int_K^\infty p(\omega)\left(\omega-K\right)\,d\omega. \qquad (1)$$

For put options we have

$$P\left(t,T,K\right)=e^{-r(T-t)}\int_{-\infty}^K p(\omega)\left(K-\omega\right)\,d\omega. \qquad (2)$$

Here, $C\left(t,T,K\right)$ and $P\left(t,T,K\right)$ are the prices of European calls and puts at time t, respectively, with striking price K and expiring time T, r is the risk-free interest rate, and $p\left(\omega\right)$ is the risk-neutral pdf for the value ω of the underlying asset at time T. After replacing $p\left(\omega\right)$ by some predefined form, the risk-neutral pdf can be estimated by minimizing the distance between the observed option prices and the prices produced by the formulas (1) and (2).

Rather than assuming a parametric form for the risk-neutral pdf one can consider a particular *stochastic process for the prices of the underlying asset.* The analytical formula of the risk-neutral pdf is then derived from the parameters of the stochastic process. The canonical example is the Black-Scholes model [4] in which the geometric Brownian motion followed by the underlying asset price implies a lognormal risk-neutral pdf.

It is shown in [6] that if one could obtain prices of puts and calls, with the same expiration but different strike prices varying in $\mathbb{R}$, then one can determine the risk-neutral distribution uniquely, since the second derivative of the call function (1) with respect to the strike K is related to the probability density function by:

$$\frac{\partial^2 C\left(t, T, K\right)}{\partial K^2} = e^{-r(T-t)} p\left(K\right). \tag{3}$$

Breeden and Litzenberger [6] applied *finite difference methods* to approximate the second derivative in the left hand side, as a way to approximate the risk-neutral pdf that appears in the right hand side.

Approximating function methods applied to the volatility smile try to fit the implied volatility curves. This method was developed by Shimko [18]. First, the author used the Black-Scholes option pricing formula to obtain implied volatilities from a set of observed option prices. Then a continuous implied volatility function is fitted. The implied volatility function, given by the Black-Scholes model, is used to derive a continuous option pricing function. Finally, using (3) a probability density function is obtained. Shimko [18] used a polynomial smoothing function for fitting the implied volatility curves. Brunner and Hafner [7] first fit a curve to the smile between available strikes to obtain the corresponding portion of the pdf and then extrapolate the tails of the pdf using mixtures of two log-normal distributions. Other authors like Campa *et al.* [8] or Anagnou *et al.* [1] have used splines. Despite the use of the Black-Scholes model these methods do not explicitly assume a lognormal risk-neutral pdf.

Implied binomial tree methods were used by Rubinstein [17]. First a prior guess of the risk-neutral pdf for all possible states $j = 1, ..., m$ is established using binomial trees. These prior guesses p_j^ℓ are set according to a lognormal distribution. The prices calculated by this process must fit correctly the observed option prices. Rubinstein [17] achieved this goal by minimizing the sum of the squared deviations between the probabilities p_j

that are being sought, and the priors p_j^ℓ:

$$\min \sum_{j=1}^{m} \left(p_j - p_j^\ell\right)^2 \quad \text{(least squares fitting)}.$$

Jackwerth and Rubinstein [14] proposed different objective functions, such as:

$$\sum_{j=1}^{m} \frac{\left(p_j - p_j^\ell\right)^2}{p_j^\ell} \qquad \text{(goodness of fit)},$$

$$\sum_{j=1}^{m} \left|p_j - p_j^\ell\right| \qquad (\ell_1 \text{ fitting}),$$

$$-\sum_{j=1}^{m} p_j \log\left(\frac{p_j}{p_j^\ell}\right) \quad \text{(maximum entropy)},$$

and

$$\sum_{j=2}^{m-1} \left(p_{j-1} - 2p_j + p_{j+1}\right)^2.$$

It was observed by Jackwerth and Rubinstein [14] that these criteria, as the number of strikes increases, lead to similar risk-neutral pdf's independently of the values of the priors p_j^ℓ. Note also that the last criterion does not assume a prior but instead it searches for a discrete approximation of a risk-neutral pdf by minimizing an approximation to its second-order derivative with respect to the underlying asset level (see the details in [14]).

3 The basic formulation using splines

As discussed in the Introduction, one of the desired structural properties of a RNPM estimate is smoothness. The strategy developed in this section guarantees appropriate smoothness of the risk-neutral pdf by estimating it using cubic splines. The estimation is carried out by the solution of an optimization problem where the optimization variables are the parameters of the spline functions.

3.1 Splines

In this subsection, we recall the definition of spline functions. Consider a function $f : [a, b] \to \mathbb{R}$ to be estimated by using its values $f(x_s)$ given on a set of points x_s, $s = 1, \ldots, n_s + 1$. It is assumed that $x_1 = a$ and $x_{n_s+1} = b$.

Definition 3 *A spline function, or spline, is a piecewise polynomial approximation $S(x)$ to the function f such that the approximation agrees with f on each node x_s, i.e., $S(x_s) = f(x_s)$, $s = 1, \dots, n_s + 1$.*

The graph of a spline function S contains the data points $(x_s, f(x_s))$ (called *knots*) and is continuous on $[a, b]$. A spline on $[a, b]$ is of order q if (i) its first $q - 1$ derivatives exist on each interior knot, (ii) the highest degree for the polynomials defining the spline function is q.

A cubic (third order) spline uses cubic polynomials of the form $f_s(x) = \alpha_s x^3 + \beta_s x^2 + \gamma_s x + \delta_s$ to estimate the function in each interval $[x_s, x_{s+1}]$ for $s = 1, \dots, n_s$. A cubic spline can be constructed in such a way that it has second-order derivatives at each node. For $n_s + 1$ knots $(x_1, \dots, x_{n_s+1})$ there are n_s intervals and, therefore, $4n_s$ unknown constants to evaluate. To determine these $4n_s$ constants we use the following conditions:

$$f_s(x_s) = f(x_s), \ s = 1, \dots, n_s, \ \text{and} \ f_{n_s}(x_{n_s+1}) = f(x_{n_s+1}), \tag{4}$$
$$f_{s-1}(x_s) = f_s(x_s), \ s = 2, \dots, n_s, \tag{5}$$
$$f'_{s-1}(x_s) = f'_s(x_s), \ s = 2, \dots, n_s, \tag{6}$$
$$f''_{s-1}(x_s) = f''_s(x_s), \ s = 2, \dots, n_s, \tag{7}$$
$$f''_1(x_1) = 0 \text{ and } f''_{n_s}(x_{n_s+1}) = 0. \tag{8}$$

The last condition leads to a so-called *natural* spline.

3.2 The Quadratic Programing Formulation

We now formulate an optimization problem with the objective of finding a risk-neutral pdf described by cubic splines for future values of an underlying security that provides a best fit with the observed option prices on this security.

For the security under consideration, we fix an exercise date, a range $[a, b]$ for possible terminal values of the price of the underlying security at the exercise date of the options, and an interest rate r for the period between now and the exercise date. The other inputs to our optimization problem are market prices C_K of call options and P_K for put options on the chosen underlying security, with strike price K and the chosen expiration date. Let $\mathcal{C}$ and $\mathcal{P}$, respectively, denote the set of strike prices K for which reliable market prices C_K and P_K are available. For example, $\mathcal{C}$ may denote the strike prices of call options that were traded on the day that the problem is formulated.

Next, we consider a super-structure for the spline approximation to the risk-neutral pdf, meaning that we choose how many knots to use, where to place the knots and what kind of polynomial (quadratic, cubic, *etc.*) functions to use. For example, one may decide to use cubic splines as we do in this paper and $n_s + 1$ equally spaced knots. The parameters of the polynomial functions that comprise the spline function will be the variables of the optimization problem we are formulating. For cubic splines with $n_s + 1$ knots, we will have $4n_s$ variables $(\alpha_s, \beta_s, \gamma_s, \delta_s)$ for $s = 1, \ldots, n_s$. Collectively, we will represent these variables by $y \in \mathbb{R}^{4n_s}$. For all y chosen so that the corresponding polynomial functions f_s satisfy the systems (5)-(8) of the previous section, we will have a particular (natural) spline function defined on the interval $[a, b]$. Let $p_y(\omega)$ denote this function. Note that we do not impose the constraints given in (4) because the values of the pdf we are approximating are unknown and will be the result of the solution of the optimization problem.

By imposing the following additional restrictions we make sure that p_y is a probability density function:

$$p_y(\omega) \geq 0, \ \forall \omega \in [a, b], \tag{9}$$

$$\int_a^b p_y(\omega) d\omega = 1. \tag{10}$$

In practice the requirement (10) is easily imposed by including the following constraint in the optimization problem:

$$\sum_{s=1}^{n_s} \int_{x_s}^{x_{s+1}} f_s(\omega) d\omega = 1. \tag{11}$$

One can easily see that this is a linear constraint in the components $(\alpha_s, \beta_s, \gamma_s, \delta_s)$ of the optimization variable y. The treatment of (9) is postponed to the next section and is ignored until the end of this section.

Next, we define the discounted expected value of the terminal value of each option using p_y as the risk-neutral probability density function:

$$C_K(y) = \frac{1}{1+r} \int_a^b p_y(\omega)(\omega - K)^+ d\omega, \tag{12}$$

$$P_K(y) = \frac{1}{1+r} \int_a^b p_y(\omega)(K - \omega)^+ d\omega. \tag{13}$$

If p_y was the actual risk-neutral probability density function, the quantities $C_K(y)$ and $P_K(y)$ would be the fair values of the call and put options with strikes K. The quantity

$$(C_K - C_K(y))^2$$

measures the squared difference between the observed value and discounted expected value considering p_y as the risk-neutral pdf. Now consider the overall residual least squares function for a given y:

$$E(y) = \sum_{K \in \mathcal{C}} (C_K - C_K(y))^2 + \sum_{K \in \mathcal{P}} (P_K - P_K(y))^2. \tag{14}$$

The objective now is to choose y such that $E(y)$ is minimized subject to the constraints already mentioned. The resulting optimization problem is a convex quadratic programming problem corresponding to the following formulation:

$$\min_y \;\; E(y) \;\; \text{s.t.} \;\; (5), (6), (7), (8), (11). \tag{15}$$

3.3 Functions $C_K(y)$ and $P_K(y)$

We now look at the structure of problem (15) in more detail. In particular, we evaluate the function $C_K(y)$. Consider a call option with strike K such that $x_\ell \leq K < x_{\ell+1}$. Recall that y denotes a collection of variables $(\alpha_s, \beta_s, \gamma_s, \delta_s)$ for $s = 1, \ldots, n_s$ and that $x_1 = a, x_2, \ldots, x_{n_s}, x_{n_s+1} = b$ represent the locations of the knots. The formula for $C_K(y)$ can be derived as follows:

$$\begin{aligned}
&(1+r)C_K(y) \\
&= \int_a^b p_y(\omega)(\omega - K)^+ d\omega \\
&= \sum_{s=\ell}^{n_s} \int_{x_s}^{x_{s+1}} p_y(\omega)(\omega - K)^+ d\omega \\
&= \int_K^{x_{\ell+1}} p_y(\omega)(\omega - K) d\omega + \sum_{s=\ell+1}^{n_s} \int_{x_s}^{x_{s+1}} p_y(\omega)(\omega - K) d\omega \\
&= \int_K^{x_{\ell+1}} \left(\alpha_\ell \omega^3 + \beta_\ell \omega^2 + \gamma_\ell \omega + \delta_\ell\right) (\omega - K) d\omega \\
&\quad + \sum_{s=\ell+1}^{n_s} \int_{x_s}^{x_{s+1}} (\alpha_s \omega^3 + \beta_s \omega^2 + \gamma_s \omega + \delta_s)(\omega - K) d\omega.
\end{aligned}$$

One can easily see that this expression for $C_K(y)$ is linear in the components $(\alpha_s, \beta_s, \gamma_s, \delta_s)$ of the optimization variable y. A similar formula can be derived for $P_K(y)$. Another relevant aspect that should be pointed out is that the formula for $C_K(y)$ will involve coefficients of the type x_s^5 which can, and in fact does, make the Hessian matrix of the QP problem (15) severely ill-conditioned.

4 Guaranteeing nonnegativity

The simplest way to deal with the requirement of nonnegativity of the risk-neutral pdf is to weaken condition (9), requiring the cubic spline approximation to be nonnegative only at the knots:

$$f_s(x_s) \geq 0,\ s = 1, \dots, n_s \quad \text{and} \quad f_{n_s}(x_{n_s+1}) \geq 0. \tag{16}$$

Then, the basic QP formulation changes to:

$$\min_y \ E(y) \ \text{ s.t. } (5), (6), (7), (8), (11), (16). \tag{17}$$

One can easily see that problem (17) is still a convex quadratic programming problem, since (16) are linear inequalities in the optimization variables. The drawback of this strategy is the lack of guarantee of nonnegativity of the spline functions between the spline knots. This heuristic strategy proved sufficient to obtain nonnegative pdf estimates in most of our experiments some of which are reported in Section 5. However, in some instances pdf estimates assumed negative values between knots. Since our aim is to estimate a probability density function, estimates with negative values are not acceptable.

In what follows, we develop an alternative optimization model where the nonnegativity of the resulting risk-neutral pdf estimate is rigorously guaranteed. The cost we must pay for this guarantee is an increase in the complexity of the optimization problem. Indeed, the new model involves semidefiniteness restrictions on some matrices related to new optimization variables. While the resulting problem is still a convex optimization problem and can be solved with standard conic and semidefinite optimization software (see, *e.g.*, [20]), it is also more expensive to solve than a convex QP.

The model we consider is based on necessary and sufficient conditions for ensuring the nonnegativity of a single variable polynomial in intervals, as well as on rays and on the whole real line. This characterization is due to Bertsimas and Popescu [2] and is stated in the next proposition.

Proposition 1 (Proposition 1 (d),[2]) *The polynomial $g(x) = \sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \geq 0$ for all $x \in [a, b]$ if and only if there exists a positive semidefinite matrix $X = [x_{ij}]_{i,j=0,\dots,k}$ such that*

$$\sum_{i,j:i+j=2\ell-1} x_{ij} \ = \ 0, \quad \ell = 1, \dots, k, \tag{18}$$

$$\sum_{i,j:i+j=2\ell} x_{ij} = \sum_{m=0}^{\ell} \sum_{r=m}^{k+m-\ell} y_r \binom{r}{m} \binom{k-r}{\ell-m} a^{r-m} b^m, \tag{19}$$

$$\ell = 0, \ldots, k, \tag{20}$$

$$X \succeq 0. \tag{21}$$

In the statement of the proposition above, the notation $\binom{r}{m}$ stands for $\frac{r!}{m!(r-m)!}$ and $X \succeq 0$ indicates that the matrix X is symmetric and positive semidefinite. For the cubic polynomial $f_s(x) = \alpha_s x^3 + \beta_s x^2 + \gamma_s x + \delta_s$ we have the following corollary:

Corollary 1 *The polynomial $f_s(x) = \alpha_s x^3 + \beta_s x^2 + \gamma_s x + \delta_s$ satisfies $f_s(x) \geq 0$ for all $x \in [x_s, x_{s+1}]$ if and only if there exists a 4×4 matrix $X^s = [x^s_{ij}]_{i,j=0,\ldots,3}$ such that*

$$\begin{aligned}
x^s_{ij} &= 0, \text{ if } i+j \text{ is 1 or 5}, \\
x^s_{03} + x^s_{12} + x^s_{21} + x^s_{30} &= 0, \\
x^s_{00} &= \alpha_s x_s^3 + \beta_s x_s^2 + \gamma_s x_s + \delta_s, \\
x^s_{02} + x^s_{11} + x^s_{20} &= 3\alpha_s x_s^2 x_{s+1} + \beta_s(2x_s x_{s+1} + x_s^2) \\
&\quad + \gamma_s(x_{s+1} + 2x_s) + 3\delta_s, \\
x^s_{13} + x^s_{22} + x^s_{31} &= 3\alpha_s x_s x_{s+1}^2 + \beta_s(2x_s x_{s+1} + x_{s+1}^2) \\
&\quad + \gamma_s(x_s + 2x_{s+1}) + 3\delta_s, \\
x^s_{33} &= \alpha_s x_{s+1}^3 + \beta_s x_{s+1}^2 + \gamma_s x_{s+1} + \delta_s, \\
X^s &\succeq 0.
\end{aligned} \tag{22}$$

Observe that the positive semidefiniteness of the matrix X^s implies that the first diagonal entry x^s_{00} is nonnegative, which corresponds to our earlier requirement $f_s(x_s) \geq 0$. In light of Corollary 1, we see that introducing the additional variables X^s and the constraints (22), for $s = 1, \ldots, n_s$, into the earlier quadratic programming problem (15), we obtain a new optimization problem which necessarily leads to a risk-neutral pdf that is nonnegative in its entire domain. The new formulation has the following form:

$$\min_{y, X^1, \ldots, X^{n_s}} E(y) \text{ s.t. } (5), (6), (7), (8), (11), [(22), s = 1, \ldots, n_s]. \tag{23}$$

All constraints in (23), with the exception of the positive semidefiniteness constraints $X^s \succeq 0$, $s = 1, \ldots, n_s$, are linear in the optimization variables $(\alpha_s, \beta_s, \gamma_s, \delta_s)$ and X^s, $s = 1, \ldots, n_s$. The positive semidefiniteness constraints are convex constraints and thus the resulting problem can be reformulated as a (convex) semidefinite programming problem with a quadratic objective function.

For appropriate choices of the vectors c, f_i, g_k^s, and matrices Q and H_k^s, we can rewrite problem (23) in the following equivalent form:

$$
\begin{array}{rl}
\min_{y,X^1,\ldots,X^{n_s}} & c^\top y + \frac{1}{2} y^\top Q y \\
\text{s.t.} & f_i^\top y = b_i,\ i = 1,\ldots,3n_s, \\
& H_k^s \bullet X^s = 0,\ k = 1,2,\ s = 1,\ldots,n_s, \\
& (g_k^s)^\top y + H_k^s \bullet X^s = 0,\ k = 3,4,5,6,\ s = 1,\ldots,n_s, \\
& X^s \succeq 0,\ s = 1,\ldots,n_s,
\end{array}
\tag{24}
$$

where $\bullet$ denotes the trace matrix inner product.

We should note that standard semidefinite optimization software such as SDPT3 [20] can solve only problems with linear objective functions. Since the objective function of (24) is quadratic in y a reformulation is necessary to solve this problem using SDPT3 or other SDP solvers. We replace the objective function with min t where t is a new artificial variable and impose the constraint $t \geq c^\top y + \frac{1}{2} y^\top Q y$. This new constraint can be expressed as a second-order cone constraint after a simple change of variables; see, *e.g.*, [15]. This final formulation is a standard form *conic optimization* problem — a class of problems that contain semidefinite programming and second-order cone programming as special classes. Since SDPT3 can solve standard form conic optimization problems we used this formulation in our numerical experiments.

5 Numerical experiments

In this section, we report some numerical experiments obtained with the methodologies introduced in this paper to estimate the risk-neutral pdf, namely the approaches that led to the formulation of problems (17) and (23). We have applied the active set method provided by MATLAB to solve the convex QP problem (17) and the MATLAB-based interior-point code SDPT3 [20] to solve the SDP problem (23) (more precisely its reformulation described at the end of the last section). The performance of these two approaches is illustrated with two different data sets, one generated from a Black-Scholes model and the other extracted from the S&P 500 Index.

In the problem formulations, we chose the number of knots not much bigger than the number of strikes. The first knot a is smaller than the first strike and the last knot b is bigger than the last strike. This assignment guarantees that the range of the possible terminal values for the underlying asset at maturity includes all strikes.

Numerically, we solved scaled versions of both the QP problem (17) and the SDP problem (24). The need for scaling the data of these problems results from the fact that the Hessian matrix in (15), which appears in both problems, is highly ill-conditioned, as we have already pointed out in Section 3.3. Since the magnitude of ω plays a relevant role in the size of the entries of this Hessian matrix, we used as our reference scaling factor the average value of the components of the vector of the knots. Let us call this average value x_{avg}. Then each knot x_s, $s = 1, \ldots, n_s + 1$, is scaled by x_{avg} and replaced by $x'_s = x_s / x_{avg}$. Such a scaling amounts at the end to scale the variables $\alpha_s, \beta_s, \gamma_s, \delta_s$ corresponding to the spline coefficients by, respectively, a, b, c, d, whose values depend on x_{avg} as well as on the expressions for the integrations given in Section 3.3. The problem is then solved in the scaled variables $\alpha'_s, \beta'_s, \gamma'_s, \delta'_s$, $s = 1, \ldots, n_s$. We also multiply each term of the objective function in (15) by $1/x_{avg}^2$. The unscaled solution is recovered by the formulas $(\alpha_s, \beta_s, \gamma_s, \delta_s) = (a\alpha'_s, b\beta'_s, c\gamma'_s, d\delta'_s)$, $s = 1, \ldots, n_s$.

5.1 Black-Scholes data

The first example corresponds to Black-Scholes options data generated using the function BLSPRICE provided by the Financial Toolbox of MATLAB. This function computes the value of the call or put option in agreement with the Black-Scholes formula. To generate the data we must supply the current value of the underlying asset, the exercise price, the risk-free interest rate, the time to maturity of the option, the volatility, and the dividend rate.

The call and put option prices were generated considering 50 as the current price for the underlying asset, 0.1 as the risk-free interest rate, a time to maturity of 0.5, a volatility of 0.2, and no dividend rate. We considered 129 call options and 129 put options with strikes varying from 1 to 129 with increment 1. The number of knots was set to 131 and the knots were equally spaced between 0.01 and 130. The risk-neutral pdf corresponding to the prices generated from this data is known to be the following lognormal density function

$$p(\omega) \;=\; \frac{1}{\omega\sigma\sqrt{2\pi\,(T-t)}} e^{-\frac{\left(\ln(\omega/S_0)-(r-\sigma^2/2)(T-t)\right)^2}{2\sigma^2(T-t)}},$$

where $r = 0.1$, $\sigma = 0.2$, $T - t = 0.5$, and $S_0 = 50$. This function is depicted in solid lines in all the four plots of Figure 1.

We solved the scaled instances of problems (17) and (24) defined by the Black-Scholes data and scaling reported above. The plots of the recovered probability density functions are depicted in Figure 1 (left) for both problems.

In our formulations, the Hessian matrix is known to be positive semidefinite. However, it is also highly rank-deficient and, due to round-off errors, it exhibits small negative eigenvalues, around -10^{-18}. These negative eigenvalues proved to be troublesome for MATLAB's active set QP. The scaling reduced significantly the ill-conditioning of this matrix, allowing a relatively accurate eigenvalue computation. We have modified the Hessian matrix, by adding a multiple ξ of the identity to the scaled Hessian matrix, using as coefficient $\xi = (3/5)|\lambda_{min}|10^4$. Under this modification, the modified scaled Hessian becomes numerically positive definite. This choice for ξ approximately provided the best fit to the lognormal shape.

In both QP and SDP cases, the recovered pdf obtained with Hessian modification approximately exhibited the desired lognormality property. It can be seen from both plots that the pdf computed is slightly less positively skewed than the lognormal one. We also observe at the ends that the recovered pdf's started deviating from the lognormal flatness. Finally, we point out that the expected prices of the call options computed using the recovered risk-neutral pdf adjusted relatively well to the Black-Scholes prices (see right plots of Figure 1).

5.2 S&P 500 data

The other data was obtained from publicly available market data. We collected information related to European call and put options on S&P 500 Index traded in the Chicago Board of Options Exchange (CBOE) on April 29, 2003 with maturity on May 17 (data set 1), on March 24, 2004 with maturity on April 17 (data set 2), and on March 24, 2004 with maturity on June 17 (data set 3). We chose this market because it is one of the most dynamic and liquid options markets in the world.

The interest rate was obtained from the Federal Reserve Bank of New York. We considered a Treasury Bill with time to expiration as closest as possible to the time of expiration of the options.

5.2.1 Preprocessing the data

As indicated in Section 2, a risk-neutral probability measure exists if and only if there are no arbitrage opportunities. It is possible, however, to observe arbitrage opportunities in the prices of illiquid derivative securities. These prices do not reflect true arbitrage opportunities — once these secu-

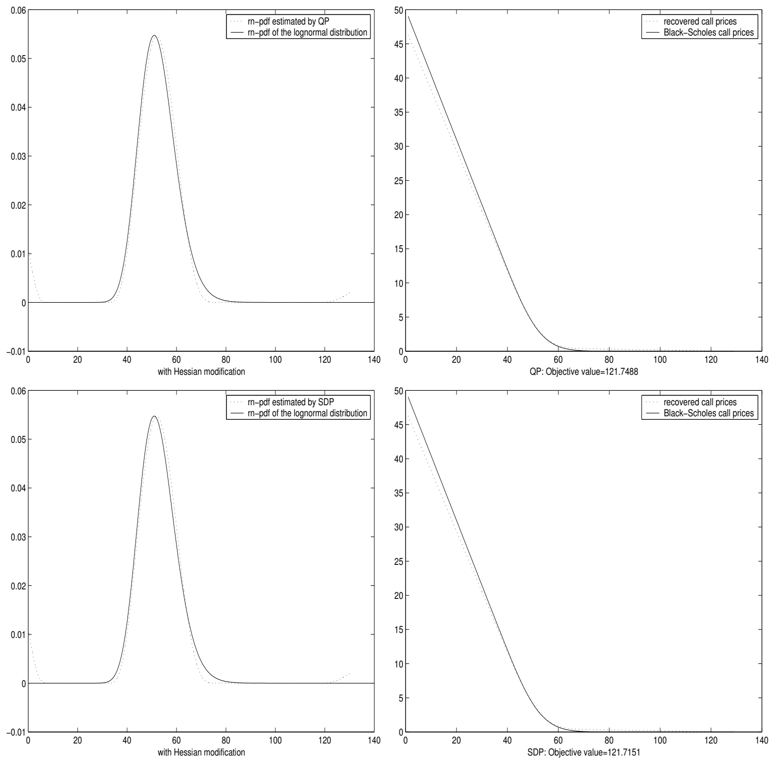

Figure 1: Recovered probability density functions from data generated by a Black-Scholes model using QP and SDP approaches (left plots). Fitted recovered expected prices for both approaches (right plots).

rities start trading, their prices will be corrected and arbitrage will not be realized.

Still, in order to have meaningful solutions for the optimization problems that we formulated in the previous sections, it is necessary to use prices in these optimization models which contain no arbitrage opportunities. Thus, before solving these problems we need to eliminate prices with arbitrage violations such as absence of monotonicity. The following theorem establishes necessary and sufficient conditions for the absence of arbitrage in the prices of European call options with concurrent expiration dates:

Theorem 2 (Herzel [12]) *Let $K_1 < K_2 < \cdots < K_n$ denote the strike prices of European call options written on the same underlying security with the same maturity, and let C_i denote the current prices of these options.*

These securities do not contain any arbitrage opportunities if and only if the prices C_i satisfy the following conditions:

1. $C_i > 0,\ i = 1, \ldots, n.$
2. $C_i > C_{i+1},\ i = 1, \ldots, n-1.$
3. *The piecewise linear function $C(K)$ with break-points at K_i's and satisfying $C(K_i) = C_i,\ i = 1, \ldots, n,$ is strictly convex in $[K_1, K_n]$.*

Theorem 2 provides us with a simple mechanism to eliminate "artificial" arbitrage opportunities from the prices we use. In our numerical experiments, after gathering price data for call and put options from the S&P 500 Index, we first eliminated options whose prices were outside the ask-bid interval, and then we generated call option prices from each one of the put option prices using the put-call parity. In cases where there was already a call option with a matching strike price, in the event that the price of the traded call option did not coincide with the price obtained from the put option price using put-call parity, we used the price corresponding to the option with the higher trading volume. After obtaining a fairly large set of call option prices in this manner, we tested for monotonicity and strict convexity in these call prices as indicated by Theorem 2. After the prices that violated these conditions had been removed, we formulated and solved the optimization problems as outlined in Section 4.

In order to guarantee the quality of the data we collected another piece of information related to the market options: the trading volume (see [11]). It is known that end-of-day settlement prices can contain options that are not very liquid and these prices may not reflect the true market prices. Inaccurate prices are usually related to less traded options. In contrast, options with higher volume represent better the "market sentiment" and the investors expectations. We experimented to incorporate the trading volume in our problem formulation by modifying the objective function of problems (17) and (24) in the following way:

$$\sum_{K \in \mathcal{C}} \theta_K [(C_K - C_K(y))]^2 + \sum_{K \in \mathcal{P}} \mu_K [(P_K - P_K(y))]^2.$$

Here θ_K is the ratio between the trading volume for the option C_K and the trading volume for all options:

$$\theta_K = \frac{\text{trading volume for } C_K}{\text{trading volume for all call options}}.$$

The weight μ_K is defined similarly for put options. Note that options with zero volume have a weight equal to zero. However, we observed that the effect of incorporating this type of weighting after eliminating arbitrage was relatively minor.

5.2.2 Results

The results are presented for the three data sets mentioned before, in a manner similar to the Black-Scholes case. In the first data set (Figure 2) the original number of calls and puts was 40 each. After eliminating arbitrage opportunities we reduced the problem dimension to 24 calls for which we considered 36 knots. In the second data set (Figure 3) the original number of calls and puts was 38 each. After eliminating arbitrage opportunities we reduced the problem dimension to 24 calls for which we considered 32 knots. Finally, in the third data set (Figure 4) the original number of calls and puts was 29 each. After eliminating arbitrage opportunities we reduced the problem dimension to 14 calls for which we considered 23 knots.

The upper plots of Figures 2, 3, and 4 correspond to the QP approach whereas the lower ones were obtained by SDP. The Hessian modification has been done by adding ξI to the scaled Hessian matrix, choosing the reference value $\xi = (3/5)|\lambda_{min}|10^4$ adjusted for the Black-Scholes data.

The recovered probability density functions are slightly negatively skewed, as opposed to what happened in the Black-Scholes case. This behavior is expected according to some authors and to what is known about the behavior of the risk-neutral pdf after the crash of 1987 (see [14]).

We have observed that the pdf estimated using the QP model and the Hessian modification assumes small negative values at the higher tail of the distribution, roughly between 1050 and 1100 (Figure 2), between 890 and 925 (Figure 3), and between 1380 and 1480 (Figure 4). As prescribed, the semidefinite optimization model corrects this behavior and obtains a nonnegative pdf estimate.

Finally, we point out that the expected prices of the call options computed using the recovered risk-neutral pdf adjusted relatively well to the S&P 500 prices (see right plots of Figures 2, 3, and 4).

6 Concluding remarks

We have developed and tested a new way of recovering the risk-neutral probability density function (pdf) of an underlying asset from its corresponding option prices. Our approach is nonparametric and uses cubic splines. The

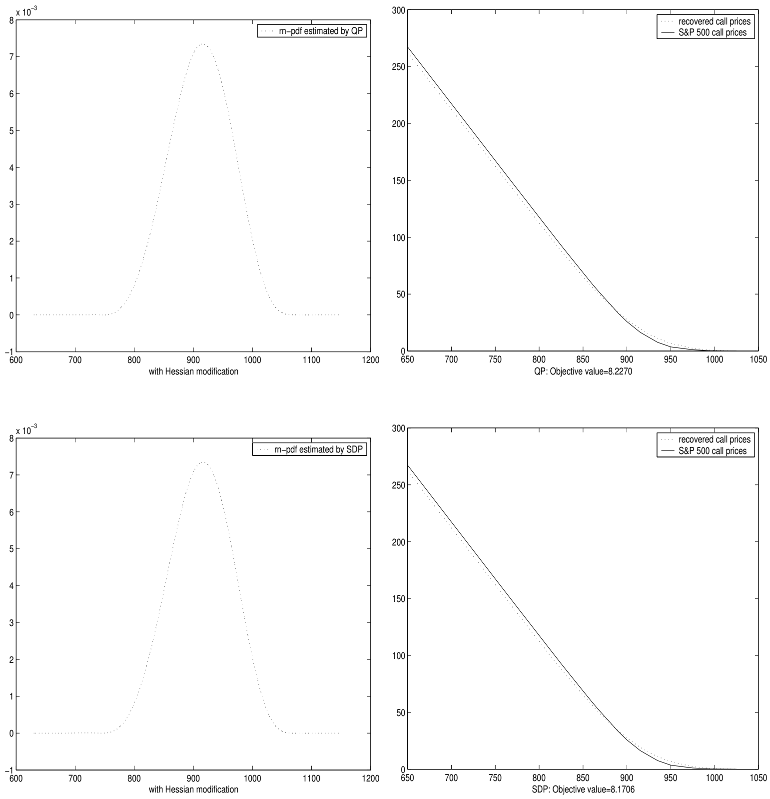

Figure 2: Recovered probability density functions from S&P 500 Index data using QP and SDP approaches (left plots). Fitted recovered expected prices for both approaches (right plots). Data set 1.

core inversion problem is a quadratic programming (QP) problem with a convex objective function and linear equality constraints.

To guarantee the nonnegativity of the inverted risk-neutral pdf we followed two alternatives. In the first one we kept the QP structure of the core problem, adding linear inequalities that reflect only the nonnegativity of this pdf at the spline nodes. The second one extends the nonnegativity requirement to the entire domain of the recovered pdf by imposing appropriate semidefinite constraints. In the examples tested, we observed that the QP approach is less sensitive to scaling than the semidefinite programming (SDP) approach. While the simpler QP approach is generally sufficient to recover an appropriate risk-neutral pdf both with simulated and market

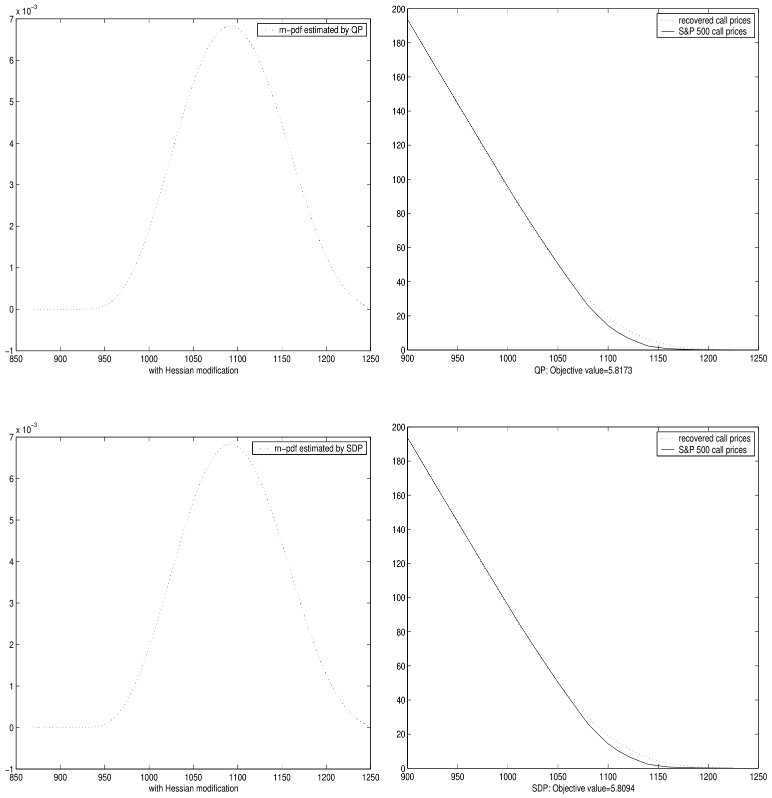

Figure 3: Recovered probability density functions from S&P 500 Index data using QP and SDP approaches (left plots). Fitted recovered expected prices for both approaches (right plots). Data set 2.

data, there are instances where the solution of the more difficult SDP model is necessary to obtain a nonnegative pdf estimate.

We plan to investigate the numerical estimation of the volatility based on the knowledge of the previously estimated risk-neutral pdf. Another topic of future research is to consider uncertainty in the data and to study the robust solution of the corresponding uncertain QP and SDP problems.

References

[1] I. Agnanou, M. Bedendo, S. Hodges, and R. Thompkins. The relation between implied and realised probability density functions. Techni-

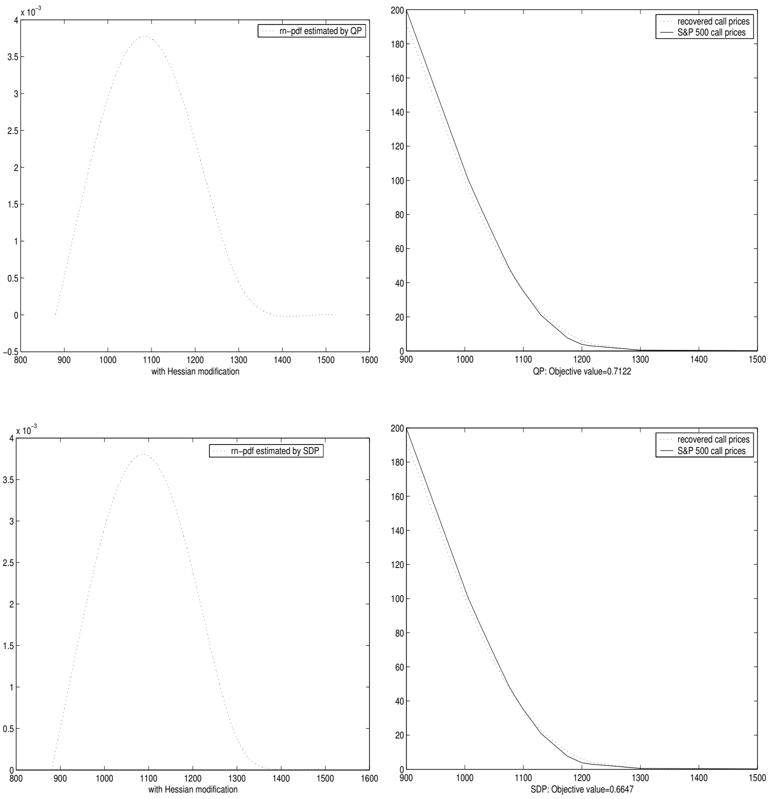

Figure 4: Recovered probability density functions from S&P 500 Index data using QP and SDP approaches (left plots). Fitted recovered expected prices for both approaches (right plots). Data set 3.

cal report, Finantial Options Research Centre, University of Warwick, 2002.

[2] D. Bertsimas and I. Popescu. On the relation between option and stock prices: A convex programming approach. *Oper. Res.*, 50:358–374, 2002.

[3] B. Bhara. Implied risk-neutral probability density functions from options prices: Theory and application. Technical report, Bank of England, 1997.

[4] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–659, 1973.

[5] R. R. Bliss and N. Panigirtzoglou. Testing the stability of implied probability density functions. Technical report, Bank of England, 2000.

[6] D. Breeden and R. Litzenberger. Prices of state-contingent claims implicit in options prices. *Journal of Business*, 51:621–651, 1978.

[7] B. Brunner and R. Hafner. Arbitrage-free estimation of the risk-neutral density from implied volatility smile. *The Journal of Computational Finance*, 7(1):75–106, 2003.

[8] J. Campa, K. Chang, and R. Reider. Implied exchange rate distributions: Evidence from OTC option markets. *Journal of International Money and Finance*, 17:117–160, 1998.

[9] J. Cox and S. Ross. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3:145–166, 1976.

[10] D. Duffie. *Dynamic Asset Pricing Theory.* Princeton University Press, Princeton, NJ, 1992.

[11] D. Y. Dupont. Extracting risk-neutral probability distributions from options prices using trading volume as a filter. Economic Series 104, Institute for Advanced Studies, Vienna, 2001.

[12] S. Herzel. Arbitrage opportunities on derivatives: A linear programming approach. Technical report, Dipartimento di Economia, Universit di Perugia, 2000.

[13] J. C. Jackwerth. Option-implied risk-neutral distributions and implied binomial trees: A literature review. *The Journal of Derivatives*, 7:66–82, 1999.

[14] J. C. Jackwerth and M. Rubinstein. Recovering probabilities distributions from options prices. *The Journal of Finance*, 51:1611–1631, 1996.

[15] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and Its Applications*, 284:193–228, 1998.

[16] W. R. Mellick and C. Thomas. Recovering an asset's implied pdf from options prices: An application to crude oil during the Gulf Crisis. *Journal of Financial and Quantitative Analysis*, 32:91–115, 1997.

[17] M. Rubinstein. Implied binomial trees. *The Journal of Finance*, 49:771–818, 1994.

[18] D. Shimko. Bounds of probability. *Risk*, 6:33–37, 1993.

[19] A. S. Tay and K. F. Wallis. Density forecasting: a survey. *Journal of Forecasting*, 19:235–254, 2000.

[20] R. H. Tütüncü, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Math. Programming*, 95:189–217, 2003.