Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 116 (2003) 9–47
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Nonparametric option pricing under shape restrictions

Yacine Aït-Sahalia[a,*], Jefferson Duarte[b]

[a] *Department of Economics, Princeton University and NBER, Princeton, NJ 08544-1021, USA*
[b] *Department of Finance and Business Economics, University of Washington, 267 MacKenzie Hall, Box 353200, Seattle, WA 98195-3200, USA*

Abstract

Frequently, economic theory places shape restrictions on functional relationships between economic variables. This paper develops a method to constrain the values of the first and second derivatives of nonparametric locally polynomial estimators. We apply this technique to estimate the state price density (SPD), or risk-neutral density, implicit in the market prices of options. The option pricing function must be monotonic and convex. Simulations demonstrate that nonparametric estimates can be quite feasible in the small samples relevant for day-to-day option pricing, once appropriate theory-motivated shape restrictions are imposed. Using S&P 500 option prices, we show that unconstrained nonparametric estimators violate the constraints during more than half the trading days in 1999, unlike the constrained estimator we propose.

JEL classification: C22; G12

Keywords: State price density; Kernel; Local polynomials; Regression; Constraints; Monotonicity; Convexity

1. Introduction

In many settings, economic theory only restricts the direction of the relationship between variables, not the particular functional form of their relationship. Typically, a theory would predict that some economic variable Y should increase when some other variable X increases. Beyond that, the typical economic theory is often not very restrictive about the specific nature of the relationship between Y and X, and if it is, it is often as a result of choosing a particularly tractable model which the theorist

* Corresponding author. Tel.: +1-609-258-4015; fax: +1-609-258-0719.
E-mail addresses: yacine@princeton.edu (Y. Aït-Sahalia), jduarte@u.washington.edu (J. Duarte).

0304-4076/03/$ - see front matter
doi:10.1016/S0304-4076(03)00102-7

understands to be for illustrative purposes only. Sometimes, economic theories manage to put additional restrictions on the shape of the function that links X to Y. For instance, the relationship may be predicted by the theory to be not only monotonic, but also concave. Or it may satisfy some other inequality restrictions on the function and/or its derivatives. Or the function may be homogenous of some degree, or homothetic (i.e., a positive monotonic transformation of the function is homogenous of degree one).

Examples of this nature abound in economics (see for example Matzkin, 1991, 1992, Matzkin and Richter, 1991, and Varian, 1982, 1983, 1984). The cost function of a standard perfectly competitive firm must be increasing and convex. For such a firm, the production function linking its inputs and outputs must be increasing and concave. The utility function of a typical economic agent must be increasing and concave. In fact, the most specific result in this literature, Afriat's Theorem, states that a utility function can be found to rationalize a set of observations on prices and quantities if and only if it is nonsatiated, continuous, concave and monotonic (see Afriat, 1967). No specific functional form can be deduced from the axioms of utility theory, yet one would often parametrize the utility function as an exponential function, a power function, a logarithmic function or rely on more complex functional forms.

Of course, stringent parametric assumptions are very useful for a variety of reasons. First, they allow extrapolation beyond the support of the observed data. Many economic policy questions require that hypothetical experiments be performed in the context of the model (what would the effect of a tax cut be on consumption and investment?). Strategic decisions made by firms also require extrapolation (how would profits be affected if prices were raised further?). Second, it is easy to specify a functional form that will necessarily satisfy the theory-determined restrictions (for example, $Y = Ln(X)$ will always be increasing and concave). Indeed, the common approach in empirical work, for example in microeconometrics, has been to specify parametric functional forms which satisfy the necessary shape restrictions (see e.g., Diewert, 1973). Third, more general parametric models can be built and tested against nested models that satisfy the restrictions imposed by the theory to see if these restrictions are valid. For instance, if the function is predicted to be increasing and concave and the adopted model is $Y = X^{\rho}$, an estimate of ρ can be readily used to test the concavity restriction, i.e., $0 < \rho < 1$. Fourth, the theoretical restrictions can be imposed and result in a decrease in the variance of the estimated parameters.

Despite all their advantages, parametric assumptions have their drawbacks. First, any specification error will typically lead to inconsistent estimates. Second, any test of the theory such as that described above is a joint test of the theory and the (essentially arbitrary) parametric model. Changing the parametric specification of the model will produce different answers. As a result, nonparametric methods are often used in empirical work, at least as a first step in the analysis of the data useful to guide the specification effort. With nonparametric methods, it becomes possible to examine say, whether Y increases with X, without assuming a particular model for the conditional expectation of Y given X. Unfortunately, nonparametric estimators pay for

their robustness to specification errors in other ways. They converge more slowly than their parametric counterparts, thereby requiring a larger sample size to achieve the same degree of accuracy—often, but not always, a small price to pay for the elimination of misspecification risk. Moreover, their rate of convergence deteriorates even further when derivatives of the function are estimated. Consequently, *in small samples*, the estimated first and second derivatives of the function of interest can often fail to satisfy the restrictions that the theory imposes, simply because of sampling noise.

It is therefore quite natural for the literature to have evolved towards estimates that are nonparametric in nature, yet satisfy whatever theory-motivated properties are appropriate. The main body of literature deals with the use of monotone restrictions to estimate a nonparametric regression (see Barlow et al., 1972; Robertson et al., 1988 and Matzkin, 1994 for an excellent survey). A common model is $Y = m(X) + \varepsilon$, where either the expected value or the median of ε given X is zero and $m(\cdot)$ is estimated by minimizing the least squares or least absolute deviations of the residuals, under the constraint that it be monotonous. Brunk (1970) and Hanson et al. (1973) proved the consistency of the estimator under different assumptions.

The rate of convergence of the least squares estimator is available (see Wright, 1981). The estimation of concave regression functions (same context as above except that $m(\cdot)$ is known to be concave) has also been extensively considered (see e.g., Hildreth, 1954 and Hanson and Pledger, 1976) and its distribution is known in the least squares case (see Wang, 1993). Finally, algorithms that extend Hildreth's to estimate a regression curve under inequality restrictions have been proposed by Dykstra (1983) and Ruud (1997), again in the constrained least squares context.

Rather than attempt to solve the least squares (or least absolute deviations) problem, we propose in this paper a method to impose shape restrictions as a simple modification of nonparametric locally polynomial estimators. The standard Nadaraya–Watson kernel regression estimator is a special case of a locally polynomial estimator, corresponding to a "locally constant" specification, i.e., a polynomial of order zero. By modifying locally polynomial estimators, instead of attempting to devise a new type of constrained nonparametric estimator, we can rely on a well-understood set of tools in the unconstrained regression case (see e.g., Fan and Gijbels, 1996). Moreover, our estimators are smooth like any other kernel-type regression estimator, unlike for instance the estimator produced by solving the constrained least squares problem. Our constrained nonparametric estimators satisfy, by construction, the restrictions imposed by economic theory. We focus on locally linear estimators and on the case where inequality constraints are imposed on the first two derivatives of the regression function.

As is often the case, and the estimation of option-implied densities in finance is no exception, there are many different ways to smooth a curve—Nadaraya–Watson kernel regression as in Aït-Sahalia and Lo (1998), splines with a penalty for lack of smoothness (Mammen and Thomas-Agnan, 1999), constrained splines (Dole, 1999 and Bates, 2000), flexible parametric functional forms (in the context of SPDs, see for example Abadir and Rockinger, 1998), neural networks (see Garcia and Gencay, 2000 and Haefke et al., 2000), etc. Bates's paper in particular considers cubic splines

estimated under the same constraints as ours, while Bondarenko (1997) considers the same constrained least squares problem we start with. Nonparametric methods have been applied to other asset-pricing contexts (see Aït-Sahalia, 1996a, 1996b).

We focus on a particular method, locally polynomial regression. In our view, locally polynomial estimators present a few advantages, some of which are shared by the other possible choices. First, they are truly nonparametric. Second, they have well-documented good small sample behavior (see e.g., Fan and Gijbels, 1996), especially relative to Nadaraya–Watson kernel regression estimators. Third, we are able to implement the method in such a way that the locally polynomial estimator will *always* produce estimates satisfying the constraints, which is also possible with some of the other methods, but in our case turns out to require no modification to the estimator, only its application to some transformed data. This said, we do not mean to suggest that local polynomials are necessarily a dominating alternative to everything else nonparametric (otherwise there would not be such a long list of available methods!), but rather our objective is to add to the nonparametric toolkit by showing how this particular method can be amended to reflect shape constraints, especially those that are of interest in derivative pricing. This is achieved in our main theoretical result, Proposition 1, which we hope will be of independent interest beyond our application to the estimation of state-price densities.

Our estimator extends the results of Mammen (1991). Mammen introduced a two-step kernel regression that results in monotonic estimates. We extend Mammen's results in two directions. First, we incorporate restrictions in the first *and* in the second derivatives, which is empirically relevant in a large number of economic contexts. Second, we work with locally polynomial estimators (locally linear in our specific context) as opposed to the Nadaraya–Watson kernel regression estimator used by Mammen, which is a locally constant polynomial estimator.

The remainder of the paper is organized as follows. We start in Section 2 by describing the main example that motivates this paper, the kernel estimation of the state-price density implicit in the market prices of traded options. In Section 3 we introduce our estimator and compare it to the unconstrained Nadaraya–Watson and locally linear nonparametric estimators. We show in particular that our estimator will satisfy the constraints imposed in sample and not just asymptotically. The results of a Monte-Carlo analysis of these three estimators are presented in Section 4. In Section 5, we apply our methodology to option pricing. Section 6 concludes. Technical proofs and results are in the Appendix.

2. Monotonicity and convexity of option pricing functions

The motivation for our empirical work is the theory-imposed restriction that the price of a call option must be a decreasing and convex function of the option's strike price. Assuming that markets are dynamically complete, the absence of arbitrage opportunities implies the pricing operator is linear. Continuity and linearity of the pricing operator implies by the Riesz representation theorem the existence of a state-price density (SPD),

which we denote by $p^*(S_T|S_t,\tau,r_{t,\tau},\delta_{t,\tau})$.[1] The call pricing function at time t is then given by:

$$C(S_t,X,\tau,r_{t,\tau},\delta_{t,\tau}) = \mathrm{e}^{-r_{t,\tau}\tau}\int_0^{+\infty}\max(S_T-X,0)\,p^*(S_T|S_t,\tau,r_{t,\tau},\delta_{t,\tau})\,\mathrm{d}S_T \tag{2.1}$$

where S_t is the underlying asset price at date t, X the strike price, τ the time-to-expiration, $T=t+\tau$ the expiration date, $r_{t,\tau}$ the deterministic risk free interest rate for that maturity, and $\delta_{t,\tau}$ the corresponding dividend yield of the asset. In what follows, we will leave the conditioning information implicit, and write $p^*(S_T)$ for $p^*(S_T|S_t,\tau,r_{t,\tau},\delta_{t,\tau})$.

In order to rule out arbitrage opportunities, C must be a decreasing function of X and the first derivative of C with respect to X must be greater than $-\mathrm{e}^{-r_{t,\tau}\tau}$. This follows from (2.1) since

$$\frac{\partial C(S_t,X,\tau,r_{t,\tau},\delta_{t,\tau})}{\partial X} = -\mathrm{e}^{-r_{t,\tau}\tau}\int_X^{+\infty} p^*(S_T)\,\mathrm{d}S_T \tag{2.2}$$

thus from the positivity of the density and its integrability to one

$$-\mathrm{e}^{-r_{t,\tau}\tau} \leqslant \frac{\partial C(S_t,X,\tau,r_{t,\tau},\delta_{t,\tau})}{\partial X} \leqslant 0. \tag{2.3}$$

By differentiating the call price function twice with respect to the strike price, one obtains, as in Breeden and Litzenberger (1978) and Banz and Miller (1978):

$$\frac{\partial^2 C(S_t,X,\tau,r_{t,\tau},\delta_{t,\tau})}{\partial X^2} = \mathrm{e}^{-r_{t,\tau}\tau}\,p^*(X) \geqslant 0 \tag{2.4}$$

i.e., $\partial^2 C(\cdot)/\partial X^2$ is proportional to a probability density function and hence must be positive. Any local non-convexity of the call pricing function implies negative state prices, which constitute a violation of the no arbitrage principle.

Thus the first two derivatives of the "cross-sectional" option pricing function $X \mapsto C_{t,\tau}(X) \equiv C(S_t,X,\tau,r_{t,\tau},\delta_{t,\tau})$ for given $(S_t,X,\tau,r_{t,\tau},\delta_{t,\tau})$, i.e., at each point in time t and for each maturity τ, must satisfy the set of inequality constraints

$$\begin{aligned} &-\mathrm{e}^{-r_{t,\tau}\tau} \leqslant C'_{t,\tau}(X) \leqslant 0,\\ &C''_{t,\tau}(X) \geqslant 0. \end{aligned} \tag{2.5}$$

[1] The existence and characterization of an SPD can be obtained either in preference-based equilibrium models, e.g., Lucas (1978), Rubinstein (1976), or in the arbitrage-based models by Black and Scholes (1973) and Merton (1973). In the equilibrium framework, the SPD can be expressed in terms of a *stochastic discount factor* or *pricing kernel* such that asset prices are martingales under the actual distribution of aggregate consumption after multiplication by the stochastic discount factor.

Among the no-arbitrage models, the SPD is often called the *risk-neutral density* based on the analysis of Cox and Ross (1976) who observed that the Black–Scholes formula can be obtained by assuming that all investors are risk neutral and, consequently, all assets in such a world must yield an expected return equal to the risk-free rate of interest. The SPD also uniquely characterizes the *equivalent martingale measure* under which all asset prices discounted at the risk-free rate of interest are martingales (see Harrison and Kreps, 1979), and the *state-price deflator* (see Duffie, 1996). Finally, information about market efficiency can be gleaned by comparing the SPD estimated in complete markets from the cross-section of option prices to the SPD inferred from the time series of the underlying asset (see Aït-Sahalia et al., 2001).

The theory also imposes no arbitrage bounds for the call option pricing function itself:

$$\max(0, S_t e^{-\delta_{t,\tau}\tau} - X e^{-r_{t,\tau}\tau}) \leqslant C_{t,\tau}(X) \leqslant S_t e^{-\delta_{t,\tau}\tau}. \tag{2.6}$$

Note first that it follows from (2.1) and (2.4) that $C''_{t,\tau}(X) \geqslant 0$ implies $C_{t,\tau}(X) \geqslant 0$. Secondly, if the forward price $F_{t,\tau}$ at t for delivery of the underlying asset at date $T = t + \tau$ is observable, then by no arbitrage

$$\begin{aligned} F_{t,\tau} &= \int_0^{+\infty} S_T p^*(S_T)\, dS_T \\ &= S_t \exp((r_{t,\tau} - \delta_{t,\tau})\tau). \end{aligned} \tag{2.7}$$

In this case, it follows from the fact that $S_T - X \leqslant \max(S_T - X, 0) \leqslant S_T$ and from (2.1) and (2.7) that $C_{t,\tau}(X) \leqslant S_t e^{-\delta_{t,\tau}\tau}$. It also follows from these equations and the fact that p^* is a density that $C_{t,\tau}(X) \geqslant S_t \exp(-\delta_{t,\tau}\tau) - X\exp(-r_{t,\tau}\tau)$. Indeed,

$$\begin{aligned} & e^{r_{t,\tau}\tau}\{C_{t,\tau}(X) - S_t e^{-\delta_{t,\tau}\tau} + X e^{-r_{t,\tau}\tau}\} \\ &\quad = \int_X^{+\infty} (S_T - X) p^*(S_T)\, dS_T - \int_0^{+\infty} S_T p^*(S_T)\, dS_T + X \\ &\quad = \int_0^X (X - S_T) p^*(S_T)\, dS_T \\ &\quad \geqslant 0. \end{aligned}$$

These restrictions can be expressed as restrictions on $C''_{t,\tau}(X)$, by writing them in the form

$$\int_0^{+\infty} C''_{t,\tau}(X)\, dX = e^{-r_{t,\tau}\tau}, \tag{2.8}$$

$$\int_0^{+\infty} X C''_{t,\tau}(X)\, dX = e^{-r_{t,\tau}\tau} F_{t,\tau}. \tag{2.9}$$

Therefore, the constraints imposed by the theory can all be summarized in terms of the functions $C'_{t,\tau}(X)$ and $C''_{t,\tau}(X)$, and our primary objective in this paper will be to construct nonparametric estimators of the functions $X \mapsto C'_{t,\tau}(X)$ and $C''_{t,\tau}(X)$ that satisfy the constraints (2.5), (2.8) and (2.9).

Aït-Sahalia and Lo (1998) proposed to estimate the SPD nonparametrically by using market prices to estimate an option-pricing formula $\hat{C}(\cdot)$ nonparametrically, then differentiate this estimator twice with respect to X to obtain $\partial^2\hat{C}(\cdot)/\partial X^2$. Under suitable regularity conditions, the convergence (in probability) of $\hat{C}(\cdot)$ to the true option-pricing formula $C(\cdot)$ implies that $\partial^2\hat{C}(\cdot)/\partial X^2$ will converge to $\partial^2 C(\cdot)/\partial X^2$. Consequently, to arrive at the SPD from (2.4) it is sufficient to estimate the second derivative of the call price function in relation to the strike price. Without any restrictions on the full

nonparametric regression of call prices of stock value, strike, time-to-maturity, interest rate and dividend yield, the estimates are too variable to be useful in practice. Therefore Aït-Sahalia and Lo (1998) reduced the dimensionality of the regression function by using a semiparametric specification. Suppose that the call pricing function is given by the parametric Black–Scholes formula

$$C_{BS}(F_{t,\tau}, X, \tau, r_{t,\tau}; \sigma) = \mathrm{e}^{-r_{t,\tau}\tau}\{F_{t,\tau}\Phi(d_1) - X\Phi(d_2)\} \tag{2.10}$$

where $F_{t,\tau} = S_t \exp((r_{t,\tau} - \delta_{t,\tau})\tau)$ is the forward price for delivery of the underlying asset at date T and

$$d_1 \equiv \frac{\ln(F_{t,\tau}/X) + (\sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \quad d_2 \equiv d_1 - \sigma\sqrt{\tau} \tag{2.11}$$

except that the volatility parameter for that option is a nonparametric function $\sigma(X/F_{t,\tau}, \tau)$ of the option's moneyness $M_{t,\tau} \equiv X/F_{t,\tau}$ and time-to-maturity τ:

$$C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = C_{BS}(F_{t,\tau}, X, \tau, r_{t,\tau}; \sigma(X/F_{t,\tau}, \tau)). \tag{2.12}$$

In this semiparametric model, they only need to compute the lower-dimensional kernel regression of implied volatilities on moneyness $F_{t,\tau}$, X and τ to estimate $\hat{\sigma}(\cdot)$. The rest of the call pricing function $C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})$ is parametric, thereby substantially reducing the sample size of options required to achieve the same degree of accuracy as the full nonparametric estimator. This approach nevertheless has its own drawbacks. First, it is not fully nonparametric. Second, it still requires a fairly large sample size to be effective. In a typical cross-section of options at one point in time, one often observes the prices of 20 to 50 options with different strike prices (for a given maturity). This limitation of the traded strikes is a consequence of a deliberate strategy on the part of the options exchanges to insure that the market for each one of them remains sufficiently liquid. Enlarging the sample by gathering data from different dates is useful for data description purposes but opens the door to potential nonstationarity and regime shift issues. Moreover, the inputs of interest, such as the underlying assets price, its volatility or the interest rate, can be volatile enough to preclude aggregating data from different days.

Finally, it is possible for the implied volatility smile function $\sigma(X/F_{t,\tau}, \tau)$ to have sufficiently large derivatives with respect to the option's moneyness $M_{t,\tau}$ for the resulting semiparametric SPD to violate the nonnegativity constraint, especially for long-term options. That is, differentiating (2.12) yields

$$\frac{\partial C}{\partial X} = \frac{\partial C_{BS}}{\partial X} + \frac{1}{F}\frac{\partial \sigma}{\partial M}\frac{\partial C_{BS}}{\partial \sigma},$$

$$\frac{\partial^2 C}{\partial X^2} = \frac{\partial^2 C_{BS}}{\partial X^2} + \frac{2}{F}\frac{\partial \sigma}{\partial M}\frac{\partial^2 C_{BS}}{\partial X \partial \sigma} + \frac{1}{F^2}\left(\frac{\partial \sigma}{\partial M}\right)^2 \frac{\partial^2 C_{BS}}{\partial \sigma^2} + \frac{1}{F^2}\frac{\partial^2 \sigma}{\partial M^2}\frac{\partial C_{BS}}{\partial \sigma}$$

and the right hand sides of these expressions need not satisfy the respective constraints that their left hand sides should satisfy.

Non- and semiparametric estimators of the call pricing function will satisfy the restrictions in the first and second derivatives only when the sample is large enough, and the true function verifies them. This follows simply from the pointwise convergence of nonparametric regression estimators and their derivatives. As in all the other examples from economic theory discussed above, nonparametric estimates may violate the theory-imposed convexity restriction, but parametric estimates can misspecify interesting properties of the SPD (such as its skewness and kurtosis patterns) because they are overly rigid.

As a result, the estimation of the SPD is an empirical problem where the sample size is small, and where economic theory places no restrictions on the function *other than* the restrictions (2.5), (2.8) and (2.9). Because of the potential risk involved in misspecifying the SPD, it is desirable not to impose tight parametric restrictions on the density. And the constraints imposed by the theory provide no guidance whatsoever in terms of specifying a parametric model for the SPD. In fact, as long as the candidate parametric SPD is a proper density function, no matter how it is specified parametrically, the constraints will be satisfied. Moreover, only when sufficiently strong assumptions are made on the underlying asset-price dynamics can the SPD be obtained in closed form. For example, if asset prices follow geometric Brownian motion and the riskfree rate is constant, the SPD is log-normal—this is the Black–Scholes/Merton case. For more complex stochastic processes, the SPD cannot be computed in closed-form and must be approximated by numerically intensive methods. So this is a typical situation where we need a nonparametric estimator that can be constrained to satisfy given shape restrictions.

3. Constrained nonparametric estimation

To obtain a nonparametric estimator satisfying the required shape properties, we use a combination of constrained least squares regression and smoothing.

3.1. Constrained least squares regression

The problem of constrained least squares regression consists in finding the closest values m_i, in the sense of least squares, to a set of n observations $y_1, y_2, \ldots, y_n$ satisfying a set of constraints. The constraints involve n observations on an explanatory variable, $x_1, x_2, \ldots, x_n$. In our case, y_i is the price of the call option with strike x_i. Without loss of generality assume that the observations on the explanatory variable have been ordered, i.e., $x_i \geqslant x_j$ for $i > j$, $i, j \in \{1, 2, \ldots, n\}$.

The constrained least squares regression consists in finding the vector m that solves, for the observation vector y:

$$\min_{m \in R^n} \sum_{i=1}^{n} (m_i - y_i)^2 = \min_{m \in R^n} \|m - y\|^2 \tag{3.1}$$

subject to the slope and convexity constraints:

$$-e^{-r_{t,\tau}\tau} \leqslant \frac{m_{i+1}-m_i}{x_{i+1}-x_i} \leqslant 0 \quad \text{for all } i=1,\ldots,n-1,$$

$$\frac{m_{i+2}-m_{i+1}}{x_{i+2}-x_{i+1}} \geqslant \frac{m_{i+1}-m_i}{x_{i+1}-x_i} \quad \text{for all } i=1,\ldots,n-2. \tag{3.2}$$

If we were only imposing monotonicity of the pricing function, then this would reduce to the classical isotonic regression (see e.g., Barlow et al., 1972). We can eliminate some constraints that are redundant. The convexity constraints insure that the slopes $M_{i+1,i} \equiv (m_{i+1}-m_i)/(x_{i+1}-x_i)$ are nondecreasing. Therefore the inequality constraints on the interior slopes ($i=2,\ldots,n-2$) are redundant and only the boundary slope constraints (lower bound for $i=1$ and upper bound for $i=n-1$) matter. Therefore the constraints (3.2) can be rewritten as

$$\frac{m_2-m_1}{x_2-x_1} \geqslant -e^{-r_{t,\tau}\tau} \quad \text{and} \quad m_{n-1}-m_n \geqslant 0,$$

$$\frac{m_{i+2}-m_{i+1}}{x_{i+2}-x_{i+1}} \geqslant \frac{m_{i+1}-m_i}{x_{i+1}-x_i} \quad \text{for all } i=1,2,\ldots,n-2. \tag{3.3}$$

This reduces the total number of constraints from $2n-3$ to n, which has computational implications when n is moderately large.

Note that the price constraint corresponding to (2.6) can be imposed as

$$\max(0, S_t e^{-\delta_{t,\tau}\tau} - x_i e^{-r_{t,\tau}\tau}) \leqslant m_i \leqslant S_t e^{-\delta_{t,\tau}\tau} \quad \text{for all } i=1,\ldots,n.$$

In light of the monotonicity constraints already present, these n constraints can be reduced to

$$S_t e^{-\delta_{t,\tau}\tau} - x_1 e^{-r_{t,\tau}\tau} \leqslant m_1 \leqslant S_t e^{-\delta_{t,\tau}\tau} \quad \text{and} \quad m_n \geqslant 0 \tag{3.4}$$

(with the call at strike x_1 in the money and that at strike x_n out of the money). In any event, the three additional constraints (3.4) need not be implemented at this stage. As we discuss later in Section 3.6, we will obtain an estimator of the pricing function $C_{t,\tau}$ directly from the SPD estimator, i.e., from $C''_{t,\tau}$ up to discounting. Provided the SPD estimator satisfies constraints (2.8)–(2.9), which we will ensure, our price function estimator will satisfy the constraints (2.6).

When the strike prices are equally spaced, $x_{i+1}-x_i = \Delta x$ for all i, which is the case in most if not all options markets, the second constraint in (3.3) becomes

$$m_{i+2} + m_i - 2m_{i+1} \geqslant 0 \tag{3.5}$$

which says that the butterfly portfolio constructed by buying a call struck at x_{i+2}, one struck at x_i and selling two calls struck at x_{i+1} must have a nonnegative price.

When solving the constrained minimization problem, we are effectively "cleaning" the data y_i in a non-arbitrary manner. Of course, we mean to apply this step after obvious data recording errors (such as a price recorded as 0, etc.) have been corrected.

Solving this problem can be contrasted to the commonly used practice of simply deleting from the sample the recalcitrant observations—those that fail to satisfy the arbitrage restrictions—under the rationale that they must be the result of unacceptable measurement errors. Besides being questionable as a general practice, deleting observations can be quite damaging when the sample is tiny to start with.

Naturally, in cases where the constraints are satisfied by the original option prices, the solution is simply $m_i = y_i$ for all $i = 1, 2, \ldots, n$. But how often is this not the case empirically? Based on the full year 1999, violations of the constraints (3.3) occurred 24% of the time in the raw high frequency S&P 500 index option data from the Chicago Board Options Exchange (lower frequency observations have lower violation occurrences). Hentschel (2001) provides more evidence regarding how noisy the raw option data are.

Finally, the least squares criterion function (3.1) can be weighted as in

$$\min_{m \in R^n} \sum_{i=1}^{n} (m_i - y_i)^2 \omega_i \tag{3.6}$$

to reflect the relative liquidity of different options. In this framework, more actively traded options would receive a higher weight ω_i than those less actively traded. Readily available data can be used for that purpose. In transaction-level data, the actual weights can be determined on the basis of the size and time of the most recent transaction and the bid-ask spread. In closing prices, the open interest and the bid-ask spread can be used to proxy for liquidity.

Solving the constrained least squares problem has a long history. Von Neumann (1950) originally proposed to solve it using alternative projections. While this insight remains at the heart of the more modern algorithms, Von Neumann's approach was limited in the possible set of constraints. Hildreth (1954), then Dykstra (1983) progressively extended the set of possible constraints to convex cones (a cone is such that if the solution vector m belongs to it then λm also belongs to it for any constant λ). This would suit our purposes, except that the lower bound constraint on the slopes in (3.3) make that constraint affine (a convex set) instead of linear (a convex cone). We show in Appendix A that we can first transform it to one with conic constraints, to which we can then apply Dykstra's algorithm. We also describe Dykstra's algorithm, applied to the transformed problem, in Appendix A.

3.2. Locally polynomial kernel smoothing

We now have the transformed data m_i. The transformed data (not y_i) then serve as inputs to the next and last step in our procedure. This step involves smoothing the transformed data m_i and we wish to do so in a way that preserves the constraints that were enforced in the previous step.

Let us now turn to a brief description of locally polynomial regression, which allows us also to introduce some notation. Suppose that the regression function $m(z) \equiv E[Y|Z=z]$ is to be approximated locally for z in a neighborhood of a given state value

x by Taylor's formula up to order p

$$m(z) \approx \sum_{k=0}^{p} \beta_k(x) \times (z-x)^k \tag{3.7}$$

with $\beta_k(x) \equiv m^{(k)}(x)/k!$. This representation of the function m suggests modeling $m(z)$ around x by a polynomial in z, and to use the regression of $m(z)$ on powers of $(z-x)$ to estimate the coefficients β_k. To insure that the estimated coefficients reflect the local nature of the representation, we should intuitively use a weighted regression putting more weights on points close to x. A natural way to achieve this is to introduce a kernel function $K(.)$, a bandwidth h and to use as weights $K_h(x_i - x) \equiv K((x_i - x)/h)/h$. This leads to the estimates of the coefficients $\hat{\beta}_k(x)$ as the minimizers of

$$\sum_{i=1}^{n} \left\{ m_i - \sum_{k=0}^{p} \beta_{k,p}(x) \times (x_i - x)^k \right\}^2 K_h(x_i - x) \tag{3.8}$$

which is, at each fixed point x, a generalized least squares regression of the $m_i's$ on powers of the $(x_i - x)'s$ with diagonal weight matrix formed by the weights $K_h(x_i - x)$. This regression is "local" in the sense that the regression coefficients in equation are only valid in a neighborhood of each point x.

The estimates of the regression function (and its successive derivatives) are then given by

$$\hat{m}^{(k)}(x) \equiv \hat{m}_{k,p}(x) = k!\hat{\beta}_{k,p}(x). \tag{3.9}$$

In particular, $\hat{m}(x) \equiv \hat{\beta}_{0,p}(x)$ is the coefficient of the constant term in the polynomial regression of degree p. In this framework, the classical Nadaraya–Watson kernel regression corresponds to the special case of a "locally constant" estimator where the polynomial is reduced to a constant term, i.e., $p = 0$. Indeed,

$$\hat{m}_{0,0}(x) = \frac{\sum_{i=1}^{n} K_h(x_i - x) m_i}{\sum_{i=1}^{n} K_h(x_i - x)} = \frac{\sum_{i=1}^{n} k_i m_i}{\sum_{i=1}^{n} k_i}, \tag{3.10}$$

where the heteroskedastic weights are $k_i = K_h(x_i - x)$, is the generalized least squares (GLS) regression coefficient of the m_i's on a constant.

More generally, the GLS estimator $\hat{\beta}_p = (\hat{\beta}_{0,p}, \hat{\beta}_{1,p}, \ldots, \hat{\beta}_{p,p})'$ can be written as

$$\hat{\beta}_p = \begin{pmatrix} S_{n,0} & S_{n,1} & \cdots & S_{n,p} \\ S_{n,1} & S_{n,2} & \cdots & S_{n,p+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,p} & S_{n,p+1} & \cdots & S_{n,2p} \end{pmatrix}^{-1} \begin{pmatrix} T_{n,0} \\ T_{n,1} \\ \vdots \\ T_{n,p} \end{pmatrix}, \tag{3.11}$$

where

$$S_{n,j} = \sum_{i=1}^{n} (x_i - x)^j k_i \quad \text{and} \quad T_{n,j} = \sum_{i=1}^{n} (x_i - x)^j m_i k_i. \tag{3.12}$$

The sums $S_{n,j}$ and $T_{n,j}$ depend on x, but we leave that dependence implicit to keep the notation simple. In particular if $p=0$ (Nadaraya–Watson case), $\hat{m}_{0,0}(x)=T_{n,0}/S_{n,0}$, while if $p=1$ (locally linear regression), we have

$$\hat{m}_{0,1}(x)=\hat{\beta}_{0,1}=\frac{S_{n,2}T_{n,0}-S_{n,1}T_{n,1}}{S_{n,2}S_{n,0}-S_{n,1}^2} \tag{3.13}$$

which can be rewritten in the form

$$\hat{m}_{0,1}(x)=\frac{\sum_{i=1}^n w_i m_i}{\sum_{i=1}^n w_i}$$

where the regression weights are $w_i \equiv k_i\{S_{n,2}-(x_i-x)S_{n,1}\}$ compared to k_i in the Nadaraya–Watson case of (3.10). Therefore the locally linear estimator assigns weights that are asymmetric, whereas the Nadaraya–Watson weights are always symmetric. This turns out to be a critical improvement especially when x is near the boundaries of the support, i.e., in the tails of the distribution. There, the locally polynomial regression assigns weights that adjust for the relative scarcity of the data, unlike those assigned by the locally constant Nadaraya–Watson estimator.

3.3. Estimation of derivatives

To estimate the derivative of order k of the regression function m, we can simply set $p=k+1$ and use the estimator $\hat{m}_{k,p}$ obtained from (3.9). For instance, a locally linear regression serves to estimate the regression function $\hat{m}_{0,1}$, a locally quadratic regression for the first derivative $\hat{m}_{1,2}$ and a locally cubic regression for the second derivative $\hat{m}_{2,3}$. This is generally the optimal choice on the basis of asymptotics (see (3.17) below). But alternatives are available, and they may outperform the asymptotic optimum in small samples. The Nadaraya–Watson estimator in (3.10) can easily be differentiated to yield an estimator of the partial derivative of $m(x)$ with respect to x.

$$\hat{m}'_{0,0}(x)=\frac{(\sum_{i=1}^n k'_i m_i)}{(\sum_{i=1}^n k_i)}-\frac{(\sum_{i=1}^n k_i m_i)(\sum_{i=1}^n k'_i)}{(\sum_{i=1}^n k_i)^2}, \tag{3.14}$$

where $k'_i=(1/h)K'((x-x_i)/h)$. Further differentiation of (3.10) will produce an estimator of the second derivative $m''_{0,0}(x)$.

We can also consider the estimators $\hat{m}_{0,1}$ for the regression function, $\hat{m}_{1,1}$ for its first derivative and $\hat{m}'_{1,1}$ for the second derivative. In this case,

$$\hat{m}_{1,1}(x)=\hat{\beta}_{1,1}=\frac{S_{n,0}T_{n,1}-S_{n,1}T_{n,0}}{S_{n,2}S_{n,0}-S_{n,1}^2}=\frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i-x_j)(m_i-m_j)k_i k_j}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i-x_j)^2 k_i k_j} \tag{3.15}$$

from which $\hat{m}'_{1,1}$ follows. Our shape-constrained estimator is based on applying the latter estimators to the transformed data m_i rather than the original data y_i. We show below that this insures that the desired shape restrictions are satisfied in sample, not just

asymptotically. For comparison purposes, we also consider the unconstrained estimators $\hat{m}_{k,2}$ for $k=0,1,2$, corresponding to a locally quadratic regression, and $\hat{m}_{k,3}$ for $k=0,1,2$, corresponding to a locally cubic regression.

3.4. A word on asymptotics

Under standard regularity conditions, both $\hat{m}(x)$ and its derivatives converge pointwise to their true values, as the sample size n goes to infinity. Assume that the conditional expectation $m(x)$ admits q continuous derivatives. The best achievable asymptotic rate of convergence of the estimator $\hat{m}^{(k)}(x)$ of the kth derivative of $m(x)$—in the integrated mean-squared error sense—is given by:

$$n^{(q-k)/(1+2q)}. \tag{3.16}$$

This is actually the best rate of convergence that can be achieved by any nonparametric estimator (see Stone, 1983). The fact that the rate of convergence in (3.16) slows down as the order k of the derivative to be estimated increases is often referred to as the curse of differentiation. This rate is achieved for instance by the Nadaraya–Watson kernel regression when the bandwidth satisfies $h=\mathrm{O}(n^{1/(1+2q)})$. In the case of locally polynomial estimators, the optimal choice of polynomial order p on the basis of asymptotics is given by

$$p=k+1 \tag{3.17}$$

(see Fan and Gijbels (1996, Section 3.3)).

In theory, all the estimators we discussed so far have desirable asymptotic properties. In empirical work, however, the slow rate of convergence of the derivative estimators can be a major hindrance. In our empirical application, the object of interest is the second derivative of the call option pricing function, $C_{t,\tau}(\cdot)$, with respect to the options strike price, X, when the sample size is of the order of 20 to 50 observations. The asymptotic guidance given by (3.17) would lead to locally quadratic estimators to estimate $C'_{t,\tau}$ and locally cubic ones for $C''_{t,\tau}$. We compare below these unconstrained (but asymptotically optimal) estimators to our constrained locally linear procedure. Monte Carlo simulations immediately reveal that the asymptotics are a poor guide in terms of predicting the behavior of the estimators for such small sample sizes and hence as a guide to selecting them. Moreover, as we illustrate in Figs. 1–3, the constraints are quite often violated by the unconstrained nonparametric estimators with these sample sizes. In addition, we would ideally like an increase in the sample size n to correspond to an increase in the number of strike prices for which prices are observed rather than additional prices obtained at a different point of time for the same strikes. The latter could potentially introduce nonstationarity, with prices at a different instant drawn from a different state-price density. But then collecting data for additional strikes requires going to the over-the-counter market where quotes can be obtained beyond and between the Exchange's limited traded strikes. Liquidity issues can be substantial. For all these reasons, we are interested in constructing estimators that will be nonparametric in nature, yet will not require large sample sizes to satisfy the constraints—we want them to satisfy the desired constraints *in sample*, rather than just asymptotically.

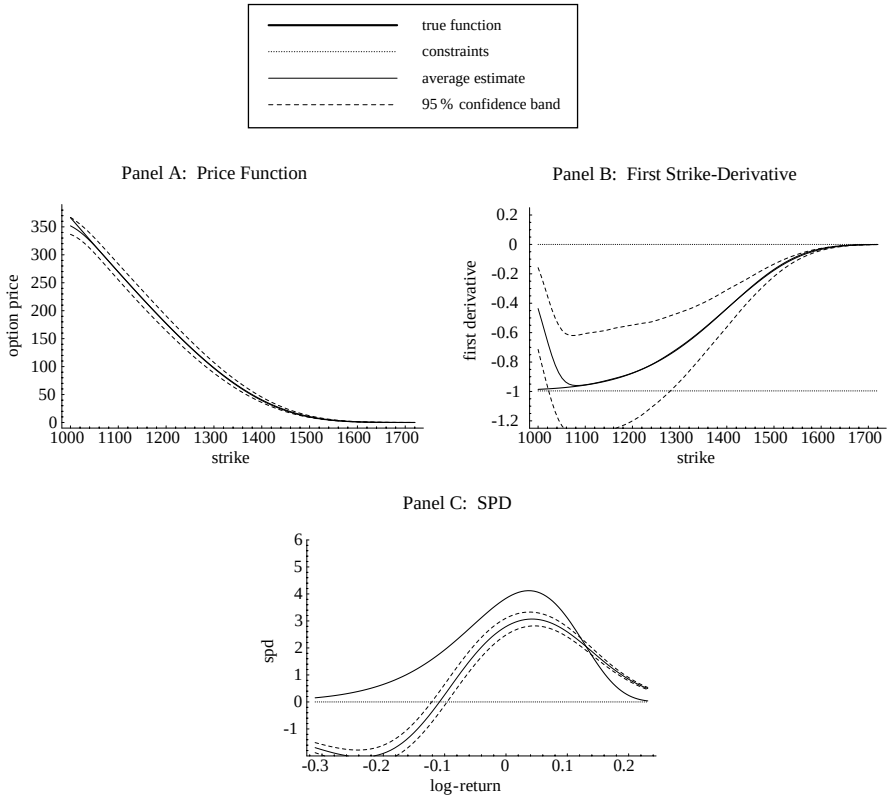

Fig. 1. Nadaraya–Watson estimator.

3.5. Bandwidth selection

A bandwidth of $h = 0$ results in interpolating each data point (the most complex model), whereas a bandwidth of infinity results in a single global polynomial fit of degree p throughout the sample (the simplest model). How to choose the bandwidth is therefore equivalent to choosing the model's complexity. Hence it is highly desirable to rely on automatic procedures that remove any potential arbitrariness in the bandwidth's choice. By minimizing the conditional mean-squared error at x

$$\{E[\hat{m}^{(k)}(x)|x] - m^{(k)}(x)\}^2 + Var[\hat{m}^{(k)}(x)|x] \tag{3.18}$$

the optimal *local* (i.e., variable with x) bandwidth is (see e.g., Fan and Gijbels, 1996):

$$h_{\text{local}}(x) = C_{k,p}\left[\frac{v(x)}{\{m^{(p+1)}(x)\}^2\pi(x)} \times \frac{1}{n}\right]^{1/(2p+3)}, \tag{3.19}$$

where $\pi(x)$ is the marginal density of the regressors and $v(x)$ their variance.

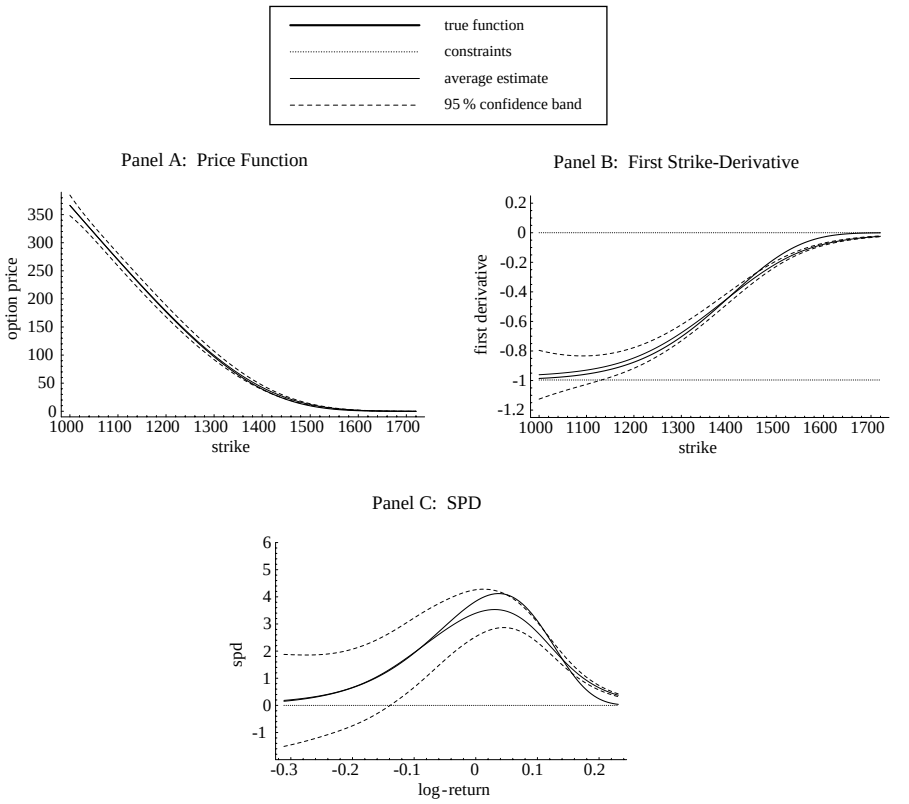

Fig. 2. Locally linear estimator.

If we are interested in a global bandwidth (i.e., one that is independent of x), minimizing the weighted mean integrated squared error with weight function $\omega(x)$

$$\int \{\{E[\hat{m}^{(k)}(x)|x] - m^{(k)}(x)\}^2 + Var[\hat{m}^{(k)}(x)|x]\}\omega(x)\,\mathrm{d}x \tag{3.20}$$

produces the optimal bandwidth

$$h_{\text{global}} = C_{k,p}\left[\frac{\int v(x)\omega(x)/\pi(x)\,\mathrm{d}x}{\int\{m^{(p+1)}(x)\}^2\omega(x)\,\mathrm{d}x} \times \frac{1}{n}\right]^{1/(2p+3)}. \tag{3.21}$$

The constants $C_{k,p}$ depend upon the choice of the kernel. For example, for the Gaussian kernel $K(u)=\exp(-u^2/2)/\sqrt{2\pi}$, the relevant constants are $C_{0,1}=0.776$, $C_{0,3}=1.161$, $C_{1,2}=0.884$ and $C_{2,3}=1.006$. The bandwidth expressions involve unknown quantities:

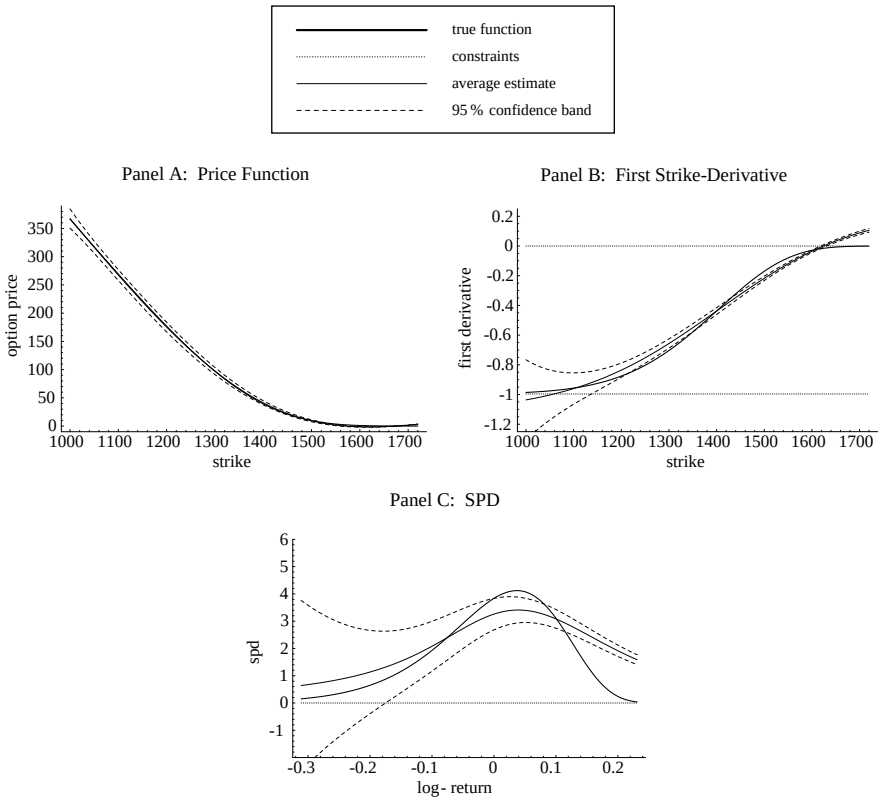

Fig. 3. Locally quadratic estimator.

$\pi(x)$, $v(x)$ and $m^{(p+1)}(x)$, which all need to be estimated prior to the calculation of the optimal bandwidth. A simple way to do so is by fitting a polynomial of order $p+3$ globally to $m(x)$, i.e., $m(x)=\sum_{k=0}^{p+3}\alpha_k x^k$ estimate the parameters α_k by ordinary least squares, v by the sum of squares of residuals (so that the estimator is independent of x), and $m^{(p+1)}(x)$ as the second order polynomial obtained by differentiation of the polynomial fit of order $p+3$ of $m(x)$, i.e.,

$$m^{(p+1)}(x)=\sum_{k=p+1}^{p+3}\alpha_k k(k-1)\ldots(k-p+1)x^{k-(p+1)}. \tag{3.22}$$

For the global optimal bandwidth, a typical choice of weighting function would be $\omega(x)=\omega_0(x)f(x)$ where $\omega_0(x)$ is a fixed function (for instance $\omega_0(x)$ is 1 for all x between the mean of the x_i's minus 1.5 times the standard deviation of the x_i's and the mean plus 1.5 times the standard deviation, and 0 for x outside this interval). In this case, $\int \omega_0(x)\,\mathrm{d}x=3\sqrt{Var(X)}$, estimated by replacing $Var(X)$ by the sample moment.

The estimated optimal global bandwidth is then

$$\hat{h}_{\text{global}} = C_{k,p}\left[\frac{ssr \times \int \omega_0(x)\,\mathrm{d}x}{\sum_{i=1}^{n}\{m^{(p+1)}(X_i)\}^2\omega_0(X_i)} \times \frac{1}{n}\right]^{1/(2p+3)}, \tag{3.23}$$

where *ssr* is the sum of squares of residuals from the regression (3.22).

3.6. The result: Estimation under inequality constraints

We now show that the two-step procedure we proposed, namely constrained least square regression of the data followed by a locally linear estimation using the transformed data, results in an estimator satisfying the constraints. The following proposition states our result. The shape-constrained estimator we described will always satisfy the constraints for every sample size, not just asymptotically:

Proposition 1. *Consider a set of n observations on the dependent variables,* $y_1, y_2, \ldots, y_n$ *and the corresponding independent variable values* $x_1, x_2, \ldots, x_n$. *Without loss of generality, let* $x_i \geqslant x_j$ *for* $i > j$, $i, j \in \{1, 2, \ldots, n\}$. *Assume that the transformed data* m_i *result from applying the constrained least squares algorithm to the original data* y_i. *Then the locally linear estimator obtained from the transformed data and a log-concave kernel function satisfies the required constraints in sample*: $-\mathrm{e}^{-r_{t,\tau}\tau} \leqslant \hat{m}^{(1)}(x) \leqslant 0$, *and* $\hat{m}^{(2)}(x) \geqslant 0$.

Proof. See Appendix B.

The last two constraints (2.8)–(2.9) on the function $\hat{m}^{(2)}(x)$ are easily satisfied. Restriction (2.8) is a scaling constraint: replacing $\hat{m}^{(2)}(x)$ by $\exp(-r_{t,\tau}\tau)\hat{m}^{(2)}(x)/\int_0^{+\infty}\hat{m}^{(2)}(z)\,\mathrm{d}z$ produces the desired result. Note that $\hat{m}^{(2)}$ is an estimator of $C''_{t,\tau}$; the corresponding estimator of the SPD $p^*(x)$ is $\exp(r_{t,\tau}\tau)\hat{m}^{(2)}(x)$: recall (2.4). Restriction (2.9) amounts to a fixed translation of the estimated density $\exp(r_{t,\tau}\tau)\hat{m}^{(2)}(x)$ to achieve the desired expected value $F_{t,\tau}$: replace $\hat{m}^{(2)}(x)$ by the shifted function $\hat{m}^{(2)}(x-z)$ with the fixed shift amount z determined by setting the expected value of the resulting function to the desired level $\exp(-r_{t,\tau}\tau)F_{t,\tau}$. As we show in Section 4 below, these two adjustments have very little effect on the estimator in practice.

We then define the estimator $\hat{m}^{(0)}(x)$ of the call pricing function from the SPD estimator by

$$\hat{m}^{(0)}(x) \equiv \int_0^{+\infty} \max(z-x, 0)\hat{m}^{(2)}(z)\,\mathrm{d}z \tag{3.24}$$

(with an obvious generalization if we wish to price another European-style payoff: just replace $\max(z-x,0)$ by that contingent claim's payoff function). The estimator $\hat{m}^{(0)}(x)$ will automatically satisfy the no-arbitrage bounds (2.6) satisfied by the call pricing function. In effect, having a proper SPD estimator in the form of $\exp(r_{t,\tau}\tau)\hat{m}^{(2)}(x)$ will automatically result in the price function satisfying the arbitrage bounds appropriate for its payoff structure (in particular, (2.6) for a call option). In the case of American-style

payoffs, this would include adding to the right hand side of (3.24) a supremum over the dates over which exercise may occur.

Note that the price function estimator $\hat{m}_{0,1}$ for the regression function will not necessarily satisfy the constraints (2.6) in sample, whereas $\hat{m}^{(0)}$ defined in (3.24) always will. When it does, however, $\hat{m}_{0,1}$ is a straightforward estimator to use for the purpose of estimating the call pricing function, and one which avoids the computation of the SPD. This is worth keeping in mind because in practice the constraints (2.6) on the price function will only be violated in extreme circumstances, whereas the constraints on the higher derivatives are more likely to be violated. This works only for calls and puts, by put-call-parity. More complicated payoffs need to be priced via the SPD. As for the first derivative, it can be estimated either by $\hat{m}_{1,1}$ or by $-\int_x^{+\infty} \hat{m}^{(2)}(z)\,\mathrm{d}z$. Both estimators satisfy in sample the constraints $-\mathrm{e}^{-r_{t,\tau}\tau} \leqslant \hat{m}^{(1)}(x) \leqslant 0$. It is logical to use $\hat{m}_{1,1}$ when $\hat{m}_{0,1}$ is used, and $-\int_x^{+\infty} \hat{m}^{(2)}(z)\,\mathrm{d}z$ when $\int_0^{+\infty} \max(z-x,0)\hat{m}^{(2)}(z)\,\mathrm{d}z$ is used.

Finally, while we are motivated by the problem of constraining our locally polynomial estimator to have bounded first derivatives and to be convex, it should be noted from the proof that the proposition in fact applies to more general inequalities on the first two derivatives of the function,[2] not just the specific ones of interest in the context of estimating SPDs. The assumption that the kernel density function is log-concave is not much of a restriction since that class of kernel functions contains among others the Gaussian, uniform, Epanechnikov and Laplacian kernels, i.e., most of the kernels used in practice (see Mukerjee, 1988).

4. Monte-Carlo analysis

4.1. Comparison with unconstrained nonparametric estimators

We perform a Monte-Carlo analysis to determine the performance of the shape-constrained nonparametric SPD estimator and compare it to the standard unconstrained Nadaraya–Watson and locally linear nonparametric estimators. The natural terrain to apply these tools involve S&P 500 index options, so we calibrate our Monte-Carlo simulation experiments to match the basic features of this market. We assume that the true price function is the Black–Scholes/Merton model with an implied volatility smile curve. Naturally, the advantage of our nonparametric approach lies in its robustness. If the options were priced by another formula, the nonparametric approach should be able to approximate it as well since, by definition, it does not rely on any parametric specification for the underlying asset's price process. Therefore, similar Monte-Carlo simulation experiments can be performed for alternative option-pricing models. However, we choose to perform the simulation experiments under an implied volatility smile model designed to be realistic for a typical trading day in 1999. The smile curve used as

[2] If the inequalities are modified, then the constraints in the constrained least squares need, of course, to be modified accordingly.

the data generating process for the simulations was calibrated based on the smile observed on May 13, 1999 on options on the S&P 500 traded at the Chicago Board Options Exchange (CBOE) with expiration in July. The assumed smile is a linear function of the strike with volatility equal to 40% at the strike price 1000 and 20% at the strike price 1700. We set the spot price S_t at 1365. The short term interest rate and the dividend yield are set at $r_{t,\tau} = 4.5\%$ and $\delta_{t,\tau} = 2.5\%$, respectively. We consider both the 30 and 60 maturities and plot the results for the 30-day options. The 60-day results are qualitatively similar.

We assume that we observe $n = 25$ option prices with strike prices equally spaced between 1000 and 1700, as would be the case with actual data. To create simulated option prices, we add uniformly distributed noise to the theoretical option prices. There are two possible rationalizations for the amount of noise to introduce around the assumed "true" option prices in order to carry out simulations. First, the noise can model the bid-ask spread and the different liquidity of different options. Second, we can assume that there is a true set of option prices at one point in time and introduce noise to capture the time series variations of the option prices in a short window of time around that date, after accounting for the variation of the underlying asset price in the same window.

In the first approach, the assumed bid-ask spread, calibrated to the market data, is set to 5% of the option's ask price, with a floor at 50 cents and a cap at 2 dollars. The noise distribution around the theoretical price is then uniform between 0 and half of the bid-ask spread value. We also account for the different liquidity of options with different degrees of moneyness (most of the liquidity is near the money). Specifically, recall that the option's moneyness is $M_{t,\tau} \equiv X/F_{t,\tau}$ (strike divided by the forward value of the S&P 500). The noise distribution around the theoretical price is then uniform between 0 and half of the bid-ask spread value times a liquidity factor given by $1 + (2/0.2)|M_{t,\tau} - 1|$. This makes the liquidity factor 1 at the money ($M_{t,\tau} = 1$) and 2 at $M_{t,\tau} = 0.8$ or 1.2, and proxies for the observed differences in liquidity of these options.

In the second approach, we calibrate the noise to the typical intraday variation of S&P 500 option prices, using their range to calibrate the uniform distribution of the noise term. In percentage terms, the range of values reached stretches from 3% of the option value for deep in the money options to 18% for deep out of the money options. In terms of the performance of the estimators, both models for the noise term produce qualitatively similar results with the provision that the lower the amount of noise, the lower the RMSE performance advantage of the constrained estimator over the unconstrained locally linear estimator (since fewer simulated data samples violate the constraints). This being said, one may argue that *any* violation of arbitrage constraints (such as those produced by the unconstrained estimator) is potentially much more damaging than its mere RMSE effect (it could for instance induce trading on a false perceived arbitrage) and should be penalized accordingly when assessing an estimator's performance. Also, other things equal, more noise tends to increase the advantage of the constrained estimator. Nevertheless, the amount of noise we specified above is not unrealistically high. It is, in fact, if anything, too conservative: see the empirical evidence in Hentschel (2001).

For estimation, we use a Gaussian kernel. We select a range of bandwidths including those given in Section 3.5 and repeat the estimation steps for each bandwidth value. Then for each function to be estimated, we selected the optimal bandwidth on the basis of minimizing the *small sample* weighted mean integrated squared error given in (3.20). We discuss this further below. The Monte-Carlo averages and confidence intervals for each bandwidth, estimator and function to be estimated are based on 5000 simulations and we focus on simulations using the second specification of the noise term, the results being qualitatively similar to the first one.

Fig. 1 shows the average estimate, a 95% confidence interval, and the true functions for the unconstrained Nadaraya–Watson estimator. Panel A of Fig. 1 shows the call pricing function estimator $\hat{m}_{0,0}$, Panel B shows the first derivative $\hat{m}'_{0,0}$ of the pricing function with respect to the strike price, and Panel C shows the risk neutral density of the log-returns $\exp(r_{t,\tau}\tau)\hat{m}''_{0,0}/x$. As observed in Panel C of Fig. 1, standard unconstrained Nadaraya–Watson estimates are, on average, negative near the left boundary, where the true probabilities are low. Of course, kernel estimation near the boundaries is known to be problematic, see e.g., Wand and Jones (1995).

Fig. 2 shows the same results for the (unconstrained) locally linear estimator, $\hat{m}_{0,0}$ in Panel A, $\hat{m}_{0,1}$ in Panel B and $\exp(r_{t,\tau}\tau)\hat{m}'_{0,1}/x$ in Panel C. As observed in Panel C of Fig. 2, the locally linear estimator has much lower boundary bias than the Nadaraya–Watson estimator, but the SPD still can be negative in the left boundary where the true probabilities are low.

Figs. 3 and 4 report the results for the locally quadratic ($\hat{m}_{k,2}$ for $k = 0, 1, 2$) and locally cubic ($\hat{m}_{k,3}$ for $k = 0, 1, 2$) estimators, respectively. As expected, higher order locally polynomial estimators perform poorly in this context because they effectively correspond to more complex local models in the absence of large enough samples. The net result is that the estimator's biases can be entirely eliminated, but at the cost of a large variance penalty. At the optimal bandwidth choice (which is what is plotted in the figures), the trade-off between squared bias and variance results in relatively large biases and variances (see Panels C in Figs. 3 and 4). In addition, these estimators often violate the constraints near the boundaries (see Panels B and C). Comparing the results for the *unconstrained* locally polynomial estimators corresponding to $p = 0, 1, 2, 3$, it appears that locally linear estimators perform best in our context.

Fig. 5 reports the results for our estimator, $\hat{m}^{(0)}$, $\hat{m}^{(1)}$ and $\exp(r_{t,\tau}\tau)\hat{m}^{(2)}/x$. As observed in Panel C of Fig. 5, the constrained estimator does not share the drawbacks of the unconstrained estimators. First, the constrained SPD estimator does not have the same boundary bias as the locally constant Nadaraya–Watson—it behaves rather like the locally linear estimator that it is. Second, unlike the unconstrained locally linear estimator, the constrained estimator remains nonnegative even when the true probabilities are low. Intuitively, imposing the constraints has the effect of allowing lower bandwidths than would be optimal for a locally linear estimator in their absence. This lowers the bias of the estimator without increasing the variance correspondingly because the constraints prevent the large deviations (which would violate the constraints) from occurring. The net effect is a more accurate estimator on the basis of its mean squared error properties. Recall that we scale our density estimator and shift it, as discussed in Section 3.6. Even though these last two constraints are necessary to rule out arbitrage

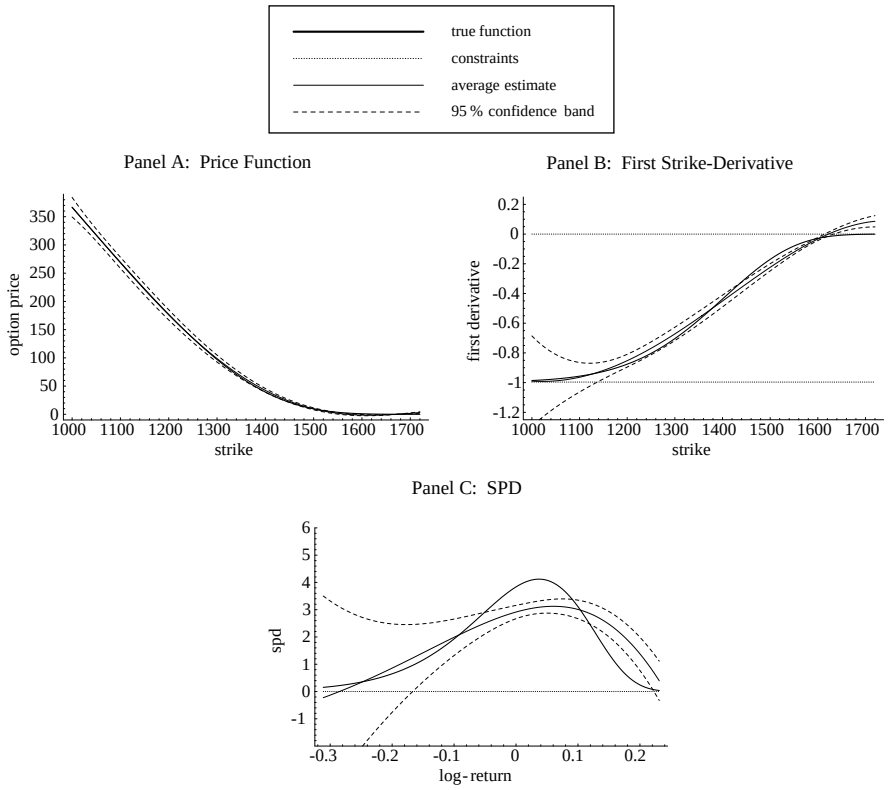

Fig. 4. Locally cubic estimator.

opportunities, our Monte-Carlo analysis reveals that, in practice, they make a small difference on the estimated function $\hat{m}^{(2)}(x)$. Indeed, for the optimal bandwidth case the average value of $\exp(r_{t,\tau}\tau)\int_0^{+\infty}\hat{m}^{(2)}(z)\,\mathrm{d}z$ is close to one (0.94) and the average shift z is 0.7% of the futures price.

We confirm that intuition by studying the mean squared error behavior of the various estimators, both pointwise and global, for the sample size under consideration. Fig. 6 reports the global root integrated mean squared error (RIMSE) of the five estimators of the pricing function, the first strike-derivative and the SPD, respectively. The RIMSE is the square root of the integral given in (3.20), unweighted. For each function to be estimated ($k=0,1,2$) and estimator ($p=0,1,2,3$, and shape-constrained estimator) we used the bandwidth resulting in the lowest RIMSE. The fact that smaller (resp. larger) bandwidths result in smaller (resp. larger) bias and larger (resp. smaller) variance produce these U-shaped RIMSE curves with the bottom of the U identifying for each function and estimator the globally optimal bandwidth used in Figs. 1 through 5, respectively. Comparing specifically our constrained estimator to the unconstrained locally linear estimator confirms the initial intuition: the shape-constrained estimator results in a lower RIMSE for lower bandwidths. For larger bandwidths, the two

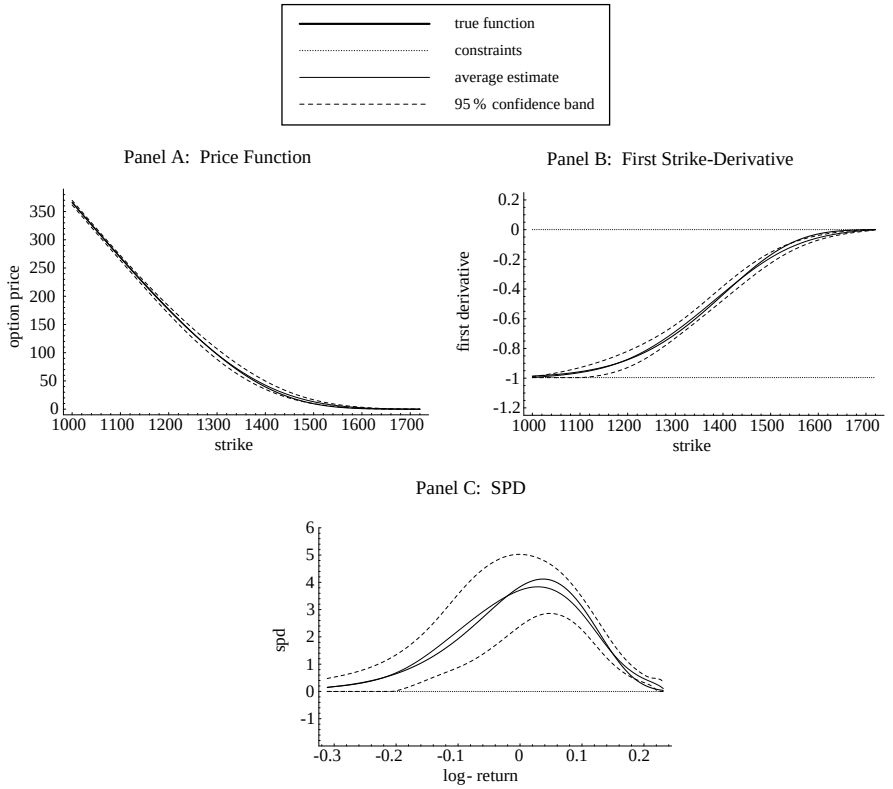

Fig. 5. Constrained estimator.

estimators converge because larger bandwidths result in flatter estimates, which consequently tend to satisfy the constraints. This explains why the RIMSE curves for these two estimators converge to one another to the right of their respective minima. However, the lowest RIMSE for the constrained estimator of the SPD is about 25% lower than that of the unconstrained estimator because lowering the bandwidth from the unconstrained optimum results in further decreases of the shape-constrained RIMSE. Fig. 7 shows the local, or pointwise, effect of oversmoothing (higher bias, lower variance) and undersmoothing (lower bias, higher variance) the constrained estimator relative to the optimal bandwidth.

Furthermore, we should note that, in all likelihood, MSE-based error measures alone underestimate the true cost of using an estimator that can violate the constraints. The mean-squared error does not attach any penalty to violations of the constraints by the unconstrained estimators. Economic measures of the cost of violating the constraints could be quite large. For example, hedges based on option deltas that violate the constraints could quickly become ineffective; pricing with an estimated SPD that is negative in the left tail leads to underestimation of out of the money put prices, trades could be put in place based on the false perception of arbitrage (locally negative SPD), etc.

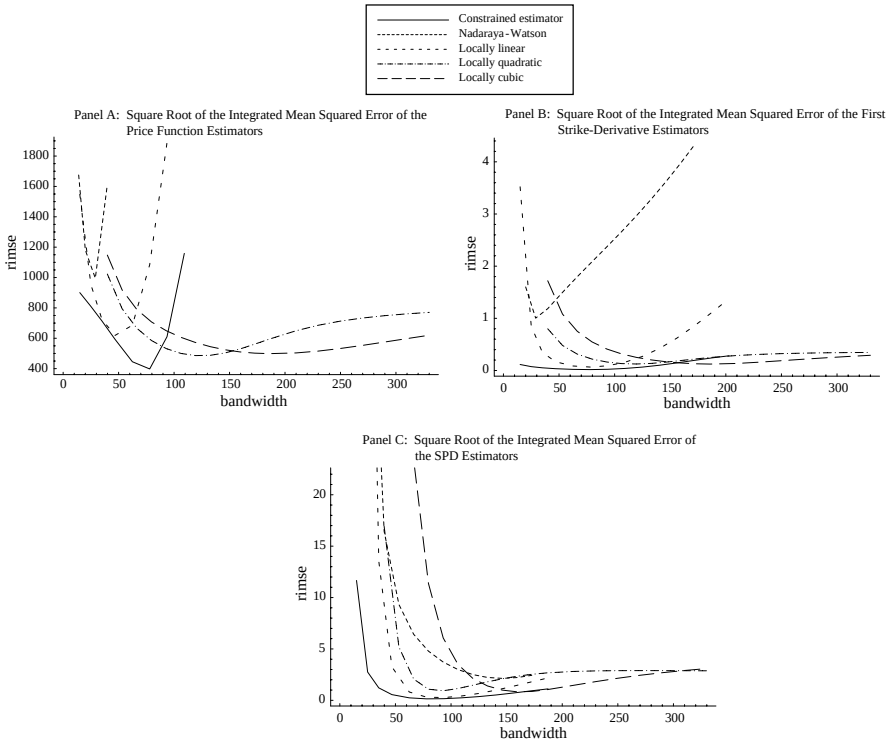

Fig. 6. Global root mean squared error and bandwidth selection.

Simulation results for $n = 50$ observations, and the first simulation design, are qualitatively similar. Overall, the results of the simulations suggest that for these types of sample sizes, imposing the shape constraints (2.5) results in a substantial improvement of the estimators.

4.2. Comparison with parametric alternatives

Finally, we also compare our estimator to two parametric alternatives. We consider the Jarrow and Rudd (1982) parametric extension of the Black–Scholes model where the lognormal density is replaced by a four-parameter expansion, namely

$$p(S_T|S_t) = \frac{\exp\{-z^2/2\}}{S_T\sqrt{2\pi\tau}\sigma}\left(1 + \frac{\mu_3}{6}\left(z^3 - 3z\right) + \frac{\mu_4}{24}(z^4 - 6z^2 + 3)\right), \tag{4.1}$$

where

$$z = z(S_T|S_t) = \frac{Ln(S_T/S_t) - (\mu_1 - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}$$

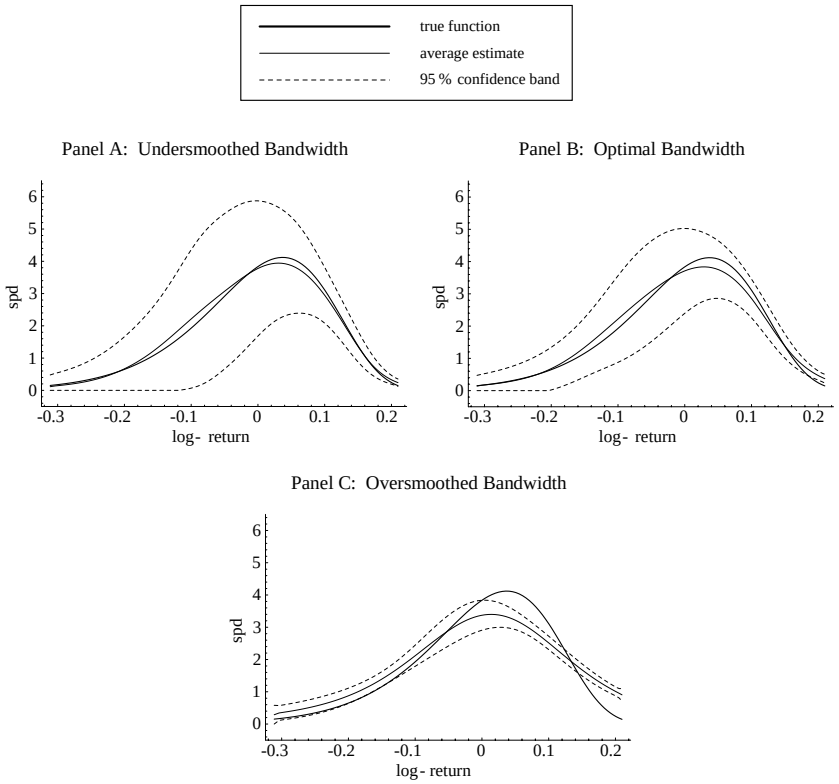

Fig. 7. Bias-variance trade-off for the constrained SPD estimator.

and the call price computed as

$$P_t = \mathrm{e}^{-r\tau} \int_K^{+\infty} (S_T - K)\, p(S_T|S_t)\, \mathrm{d}S_T. \tag{4.2}$$

The 4 parameters $\mu_1, \sigma, \mu_3, \mu_4$ are estimated by minimizing the squared deviations between market prices and parametric prices.[3] Since there is no bandwidth choice involved in this parametric formula, there is only one density per simulation. Given the sample sizes we consider, more flexible functional forms become essentially nonparametric in nature—if we have 25 observations and we are fitting a parametric model with, say, up to 10 parameters, then the choice of the number of parameters becomes akin to the choice of the bandwidth in nonparametrics.

The second parametric family we use in our comparisons is a five-parameter mixture of lognormal densities which has been used in this context by Bahra (1996). The

[3] See Christoffersen and Jacobs (2001) for a discussion of the influence of the choice of loss function in this context.

assumed model is

$$p(S_T|S_t) = \alpha p_{LN}(S_T|S_t;\mu_1,\sigma_1) + (1-\alpha)p_{LN}(S_T|S_t;\mu_2,\sigma_2), \tag{4.3}$$

where

$$p_{LN}(S_T|S_t;\mu,\sigma) = \frac{1}{S_T\sqrt{2\pi\tau}\sigma}\exp\left\{-\frac{1}{2\sigma^2\tau}(Ln(S_T/S_t)-(\mu_1-\sigma^2/2)\tau)^2\right\}$$

and the call price computed as in (4.2). The pricing formula corresponding to (4.3) is a linear combination of Black–Scholes formulae (α times the Black–Scholes formula corresponding to parameters (μ_1,σ_1) plus $1-\alpha$ times the Black–Scholes formula corresponding to parameters (μ_2,σ_2)).

The 5 parameters $\alpha,\mu_1,\sigma_1,\mu_2,\sigma_2$ are estimated by minimizing the squared percentage deviations between market prices and parametric prices. The reason for using squared price errors in one case and squared percentage errors in the other is that they produced the best results for the two methods, respectively. Attempting to minimize squared price errors with the mixture of lognormals often produces nonsensical results, where one of the two densities is tailor-made to fit in-the-money calls where pricing errors in dollars are costly, resulting in that density having a very low value of its σ parameter (in addition to a very negative value of its μ parameter).

Both parametric models provide a better contrast between the results of a true parametric procedure and those of nonparametric ones. Panels A and B of Fig. 8 report the results for the estimated SPD resulting from these two methods, in the same format as Panel C of Figs. 1–5. Because they are global in nature, as opposed to local, the two types of parametric estimators are unable to cope well with arbitrage violations in the data. This is not due to the inadequacy of the parametrizations: as we show in Panels C and D of Fig. 8, the two models can fit the true SPD assumed in the data generating process (with no noise) almost perfectly. The issues arise when we attempt to fit a set of price data that includes noise, i.e., sometimes *local* violations of convexity, as this produces a *global* distortion of the estimator—in other words, the error propagates from the local violation (which often occurs in one tail) throughout the estimated distribution (including near the peak and in the other tail).

This results in RIMSE measures that, for the same simulation designs as the other estimators we considered, are worse than what can be achieved by our proposed locally linear constrained estimator. After all, avoiding this local-to-global contamination due to outliers, bad data, etc., is often why one uses nonparametric estimators in the first place. Locally polynomial estimators are particularly apt at dealing with this issue.

5. Example: S&P 500 implied SPD under shape restrictions

Aït-Sahalia and Lo (1998, 2000) estimated the market call pricing function from a sample of 14,441 option prices on the S&P 500 index. They used the semiparametric approach described in (2.12). They found empirically, without imposing shape constraints, that their SPD estimator is convex but only because of the dimension

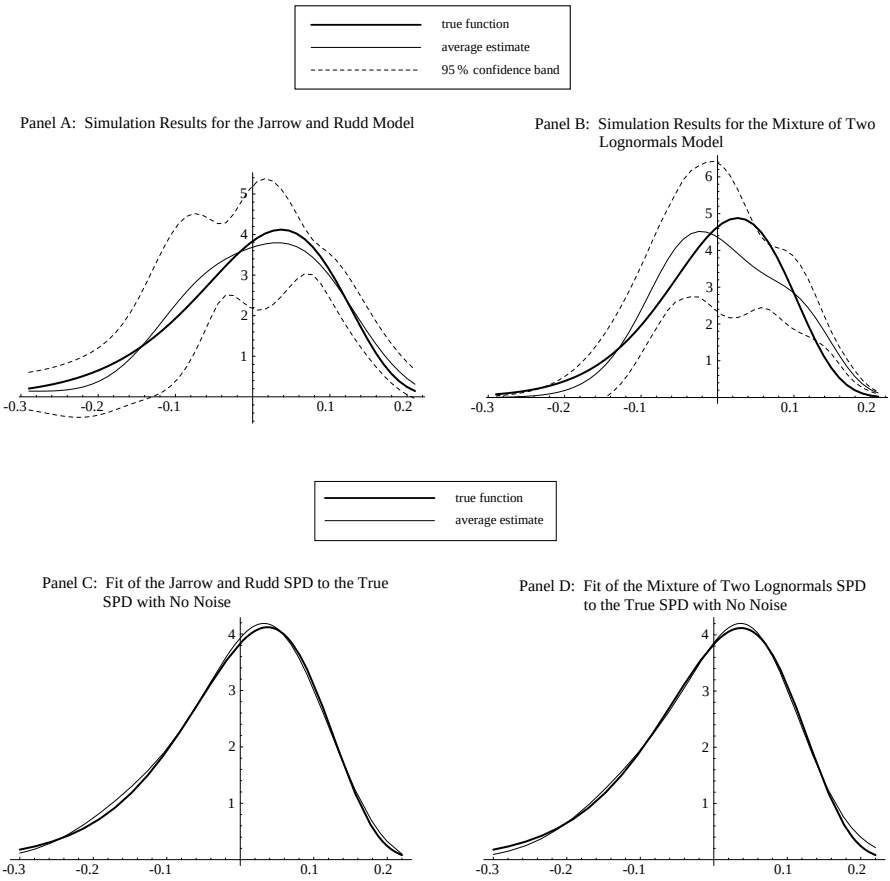

Fig. 8. Comparison with two parametric estimators.

reduction involved in the semiparametric specification, and because of the very large size of their sample. In practice, it would be desirable to have similar guarantees with substantially smaller samples. Indeed, as opposed to Aït-Sahalia and Lo (1998, 2000), we work with samples of tiny sizes (a typical cross-section at one point in time of 20 to 30 options versus a time-aggregated cross-section of 14,431 options).

The data consist of the closing prices on May 13, 1999 for call options on the S&P 500 traded at the CBOE for a maturity of 65 days corresponding to the July 1999 expiration (July 17). The closing spot price of the S&P 500 on that day was 1367.56, and the risk free interest rate for that maturity was 4.83%. The dividend yield is implied through put-call parity for the put-call pair at the money. The results from applying the five different estimators (unconstrained Nadaraya–Watson, unconstrained locally linear, quadratic and cubic, shape-constrained locally linear) are reported in Fig. 9. The bandwidths correspond to the optimum identified in the previous section. The three panels correspond to the three functions to be estimated. As is apparent

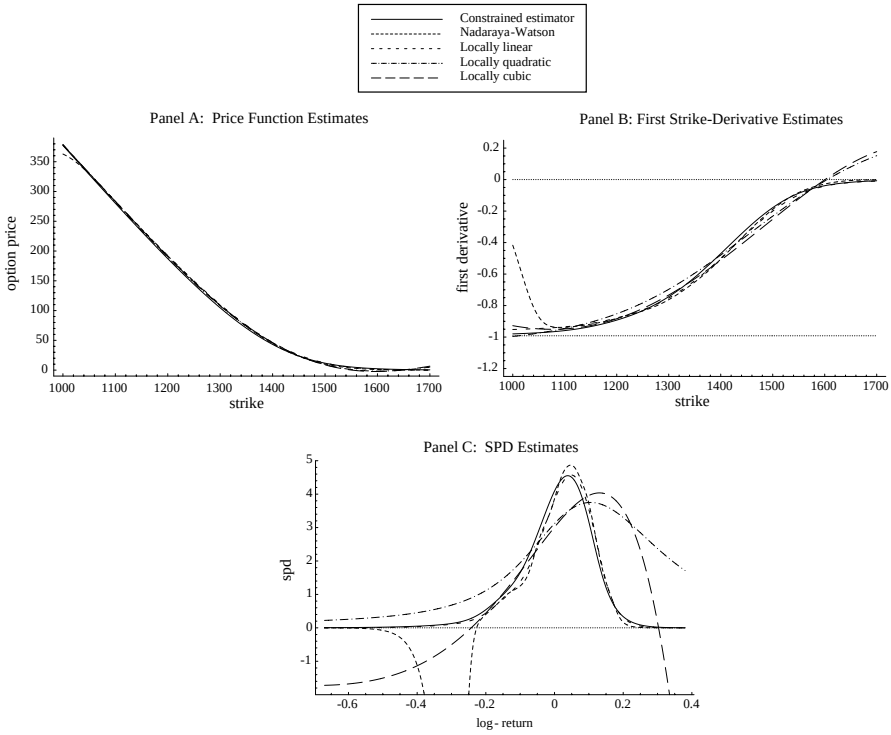

Fig. 9. S&P 500 Options, July expiration on May 13, 1999.

from Panel A, all estimators produce sensible looking (and visually indistinguishable) estimates for the pricing function as long as strikes remain relatively near the money (strikes between 1200 and 1500). However, for values of the strike price above 1600 the locally quadratic and locally cubic estimators display their high variability tendency which was clearly apparent in the simulations. And the Nadaraya–Watson estimator exhibits poor boundary behavior below 1100, clearly violating the convexity constraint on prices.

Naturally, differentiation tends to emphasize the differences between estimators. In Panel B, all remaining estimators except the two locally linear ones (constrained and unconstrained) violate the first derivative constraints somewhere. Regarding SPD estimates in Panel C, all the unconstrained estimators either violate the positivity constraint in the left tail of the density, or are too flat when evaluated at the globally optimal bandwidth. The unconstrained locally linear estimator tracks the constrained estimator relatively closely, except that the optimal bandwidth tends to produce an estimator that is slightly too flat, as was evidenced in the discussion of our simulation results. By contrast, the optimal amount of smoothing for the shape-constrained estimator is slightly lower which produces an estimator that is more sensitive to finer features of the data. Indeed, our shape-constrained estimator produces an estimate of the SPD which

Table 1
Occurrence of arbitrage restriction violations during 1999

Estimator	p	Violation frequency
Nadaraya–Watson	0	242/242
Locally linear	1	130/242
Locally quadratic	2	205/242
Locally cubic	3	241/242
Constrained locally linear		0/242

This table reports the percentage of trading days during year 1999 when the various estimators of the SPD violated the arbitrage constraints (i.e., positivity of the SPD). By construction, our constrained estimator will always satisfy the arbitrage restrictions. These results are for S&P 500 index options with 30 to 90 (calendar) days to expiration, and every day during year 1999 when such options are traded. Each day, we select the 25 most actively traded strikes, relying on put-call-parity as required to complete the range of traded in-the-money calls on the basis of out-of-the-money put prices (which are more actively traded). For each estimator, we used the bandwidths determined to be optimal in a sample of $n = 25$ strikes on the basis of our Monte-Carlo simulations reported earlier. The options data came from the CBOE.

looks quite plausible, displaying the expected level of negative skewness and excess kurtosis, while satisfying (by construction) the positivity constraint.

Finally, we report in Table 1 the results of repeating this analysis for every trading day during the year 1999. We repeated the analysis for different days (one set of quotes per day, each day treated separately) and report the frequency of arbitrage violations during that year. The unconstrained locally linear estimator violates the restrictions over 50% of the time, a percentage which rises to close to 100% as we move to the (unconstrained) locally quadratic and cubic estimators. By contrast, our estimator never violates the constraints (and still results in lower RIMSE). The violation of the arbitrage restrictions by the unconstrained estimators hold across a large spectrum of bandwidth values. Substantial oversmoothing is required to make the unconstrained estimator no longer violate the constraints. But this then results in a large bias.

6. Conclusions

This paper proposed a method to incorporate shape restrictions, such as monotonicity and convexity, into nonparametric locally linear estimators. The estimator is motivated by the practical problem of estimating state-price densities with option data, in a setting where no information other than monotonicity and convexity is available, yet the sample size is typically small. The simulations results indicate that nonparametric estimates can be quite feasible in sample sizes as small as twenty observations, provided that appropriate theory-motivated shape restrictions, such as monotonicity, and/or convexity, are imposed. As discussed in the Introduction, this is a frequent occurrence in other areas of economics as well.

In our specific context of SPD estimation, the shape-constrained SPD we estimated can have many uses. First, it provides us with an arbitrage-free method of pricing new, more complex, or less liquid securities, e.g., OTC derivatives or non-traded flexible

options, given a subset of observed and liquid "fundamental" prices, in this case basic call-option prices, that are used to estimate the SPD. We are able to achieve this in the context where very few fundamental securities are available, i.e., the observed cross-section is very sparse. Second, from a risk management perspective, our SPD estimates provide information that is crucial to understanding the nature of the fat tails of asset-return distributions implied by options data. Volatility cannot be used as a summary statistic for the entire distribution when typical return series display events that are three standard deviations from the mean approximately once a year. Our approach yields an estimate of the entire return distribution, from which single points, such as value-at-risk, can easily be derived. Third, our nonparametric estimator captures those features of the data that are most salient from an asset-pricing perspective and which ought to be incorporated into any successful parametric model. It also helps us understand what features are missed by tightly parametrized models, such as day-to-day or even intraday changes in the shape of the SPD, since we can now estimate such SPDs nonparametrically on the basis of very few observations. In fact, a nonparametric analysis can often be advocated as a prerequisite to the construction of any parsimonious parametric model, precisely because important features of the data are unlikely to be missed by nonparametric estimators.

Acknowledgements

We are grateful to seminar and conference participants, and in particular René Garcia, for their comments and suggestions. The comments of the Editors and three referees were very helpful. This research was conducted during the first author's tenure as an Alfred P. Sloan Research Fellow. Financial support from the NSF under grants SBR-9996023 and SES-0111140 (Aït-Sahalia) and from the Center for Research in Security Prices at the University of Chicago Graduate School of Business (Duarte) is gratefully acknowledged.

Appendix A. The constrained least square regression algorithm

A.1. Transforming the constrained least squares problem to one with conic constraints

We start by rewriting the constrained least squares problem in such a way as to reduce it to a convex cone problem which is then amenable to Dykstra's algorithm for constrained least squares under conic constraints. Goldman and Ruud (1995) contain ideas along those lines, although not a formal development. Write our constraints (3.3) in matrix form as $A.m - b \leqslant 0$, where A is n (the number of constraints) by n (the number of m_i's) and b is $n \times 1$. In its original form, our problem is therefore

$$\begin{aligned} \min_{m \in R^n} \quad & \|m - y\|^2 \\ \text{subject to} \quad & A.m - b \leqslant 0. \end{aligned} \tag{A.1}$$

Define

$$u = \begin{pmatrix} m - y \\ t \end{pmatrix} = \begin{pmatrix} z \\ t \end{pmatrix}, \quad v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad C = (A \,|\, A.y - b),$$

where t is 1×1, and the 0 block in the vector v is $n \times 1$. Then consider the problem

$$\begin{aligned} &\min_{u \in R^{n+1}} \quad \|u - v\|^2 = \|z\|^2 + |t - 1|^2 \\ &\text{subject to} \quad C.u \leqslant 0 \quad \text{and} \quad t = 1, \end{aligned} \tag{A.2}$$

where minimizing over u means minimizing over (z, t). The solution $u^{**} = (z^{**}, 1)$ to problem (A.2) gives the solution m^{**} of our original problem (A.1) as $m^{**} \equiv z^{**} + y$. Indeed, the solution u^{**} of (A.2) has set $t = 1$ and then minimized $\|z\|^2$ over z under the constraint that $C.u \leqslant 0$ and we have

$$C. \begin{pmatrix} z \\ 1 \end{pmatrix} \leqslant 0 \quad \Leftrightarrow \quad A.\, z + (A.y - b) \leqslant 0 \quad \Leftrightarrow \quad A.m - b \leqslant 0.$$

But problem (A.2) still does not have conic constraints (because of the constraint $t = 1$, which is again affine). So consider next the problem where we have relaxed the affine constraint $t = 1$ to the linear (or conic) constraint $t \geqslant 0$:

$$\begin{aligned} &\min_{u \in R^{n+1}} \quad \|u - v\|^2 = \|z\|^2 + |t - 1|^2 \\ &\text{subject to} \quad C.u \leqslant 0 \quad \text{and} \quad t \geqslant 0 \end{aligned} \tag{A.3}$$

Now this problem is in Dykstra's conic constraints form, and let its solution be denoted by $u^* = (z^*, t^*)$.

Let us see how the solutions to the two problems (A.2) and (A.3) are related. Note that because u^{**} satisfies the constraint $C.u^{**} \leqslant 0$, we have

$$A.\, z^{**} + (A.y - b) \leqslant 0.$$

Since $t^* \geqslant 0$, it follows that

$$A.\, z^{**} t^* + (A.y - b) t^* \leqslant 0.$$

Therefore $(z^{**} t^*, t^*)$ satisfies the constraints of problem (A.3). Since by definition the optimum of problem (A.3) is reached at $u^* = (z^*, t^*)$, it follows that

$$\|z^*\|^2 + |t^* - 1|^2 \leqslant \|z^{**} t^*\|^2 + |t^* - 1|^2$$

or

$$\|z^*\|^2 \leqslant \|z^{**}t^*\|^2. \tag{A.4}$$

Now, it is also the case that, since u^* satisfies the constraint $C.u^* \leqslant 0$, we have

$$A.\ z^* + (A.y - b)t^* \leqslant 0.$$

Since $t^* \geqslant 0$, it follows that

$$A.(z^*/t^*) + (A.y - b) \leqslant 0,$$

so that $((z^*/t^*), 1)$ satisfies the constraints of problem (A.2). But by definition the optimum of problem (A.2) is reached at $u^{**} = (z^{**}, 1)$, thus

$$\|z^{**}\|^2 \leqslant \|(z^*/t^*)\|^2. \tag{A.5}$$

Multiplying equation (A.5) by $(t^*)^2$ and combining with (A.4), it follows that $\|z^*\|^2 = \|z^{**}t^*\|^2$, so that the minimum of problem (A.2) is achieved at

$$z^{**} = z^*/t^*. \tag{A.6}$$

Therefore the solution $(z^{**}, 1)$ of problem (A.2) can be obtained from the solution (z^*, t^*) of problem (A.3). Recall that the solution m^{**} to our original problem (A.1) is obtained from the solution of problem (A.2) by $m^{**} \equiv z^{**} + y$. Hence solving problem (A.3) using Dykstra's algorithm to find (z^*, t^*) ultimately gives us the solution m^{**} to our original problem (A.1).

A.2. Algorithm for constrained least squares under conic constraints

We now briefly describe Dykstra (1983)'s algorithm to solve the constrained least square regression problem (A.3), which has conic constraints. Define the following cones in R^{n+1}. For $j = 1, \ldots, n-2$, let

$$C_j = \left\{ u \in R^{n+1} s.t. \frac{z_{j+2} - z_{j+1}}{x_{j+2} - x_{j+1}} - \frac{z_{j+1} - z_j}{x_{j+1} - x_j} + t \times \left(\frac{y_{j+2} - y_{j+1}}{x_{j+2} - x_{j+1}} - \frac{y_{j+1} - y_j}{x_{j+1} - x_j} \right) \leqslant 0 \right\} \quad j = \{1, \ldots, n-2\}$$

and

$$C_{n-1} = \{u \in R^{n+1} s.t.\ z_n - z_{n-1} + t \times (y_n - y_{n-1}) \leqslant 0\}$$

$$C_n = \{u \in R^{n+1} s.t. - z_2 + z_1 + t \times (-y_2 + y_1 - (x_2 - x_1) \times \mathrm{e}^{-r_{t,\tau}\tau}) \leqslant 0\}$$

$$C_{n+1} = \{u \in R^{n+1} s.t. - t \leqslant 0\}.$$

The minimization problem (A.3) can be written as

$$\min_{u \in \bigcap_{j=1}^{n+1} C_j} \sum_{i=1}^{n} (u_i - v_i)^2. \tag{A.7}$$

The algorithm consists in repeatedly projecting the vector u onto the cones C_j:

- Let $u_{1,1}$ denote the projection of u onto the cone C_1. Let $I_{1,1} = u_{1,1} - u$ denote the incremental change incurred by the projection, so that $u_{1,1} = u + I_{1,1}$.
- Let $u_{1,2}$ denote the projection of $u_{1,1}$ onto the cone C_2. Let $I_{1,2} = u_{1,2} - u_{1,1}$ denote the incremental change incurred by the projection, so that $u_{1,2} = u + I_{1,1} + I_{1,2}$.
- Let $u_{1,n+1}$ denote the projection of $u_{1,n}$ onto the cone C_{n+1}. Let $I_{1,n+1} = u_{1,n+1} - u_{1,n}$ denote the incremental change incurred by the projection, so that $u_{1,n+1} = u + I_{1,1} + I_{1,2} + I_{1,3} + \cdots + I_{1,n} + I_{1,n+1}$.
- After $u_{1,n+1}$ and $I_{1,n+1}$ are found. Let $u_{2,1}$ denote the projection of $u + I_{1,2} \cdots + I_{1,n+1}$ onto the cone C_1. Note that we have removed the increment $I_{1,1}$ before this projection. The new increment is $I_{2,1} = u_{2,1} - (u + I_{1,2} \cdots + I_{1,n+1})$.
- Continue, until $u_{\bullet,\bullet} \in \bigcap_{j=1}^{n+1} C_j$.

The projections of $u_{\bullet,\bullet}$ onto cones C_j are easily obtained. If we represent the cone C_j by $C_j = \{u \in R^{n+1} \ s.t. \ \sum_{i=1}^{n+1} a_{j,i} u_i \leqslant 0\}$, then the projection of u onto C_j is given by

$$P(u|C_j) = \begin{cases} u & \text{if } \sum_{i=1}^{n+1} a_{j,i} u_i \leqslant 0 \\ u' & \text{if } \sum_{i=1}^{n+1} a_{j,i} u_i > 0 \end{cases}$$

where

$$u_i' = u_i - \frac{(\sum_{l=1}^{n+1} a_{j,l} u_l) a_{j,i}}{\sum_{l=1}^{n+1} a_{j,l}^2}.$$

Appendix B. Proof of Proposition 1

Part 1: Proof that $\exp(-r_{t,\tau}\tau) \leqslant \hat{m}_{1,1}(x) \leqslant 0$.

The proof is based essentially on rearranging the terms in the numerators and the denominators of the locally linear estimators in such a way that they can be signed. With $k_i = K_h(x - x_i) = h^{-1}K(h^{-1}(x - x_i))$, the local linear estimator of the regression function is

$$\hat{m}_{0,1}(x) = \hat{\beta}_{0,1} = \frac{S_{n,2} T_{n,0} - S_{n,1} T_{n,1}}{S_{n,2} S_{n,0} - S_{n,1}^2}$$

$$= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x)^2 m_i k_i k_j - \sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x)(x_i - x) m_i k_i k_j}{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x)^2 k_i k_j - \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x)(x_j - x) k_i k_j}$$

$$= \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_j - x_i)((x_j - x)m_i - (x_i - x)m_j) k_i k_j}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i - x_j)^2 k_i k_j} \tag{B.1}$$

while the locally linear estimator of the first partial derivative of $m(x)$ with respect to x is given by

$$\hat{m}_{1,1}(x) = \hat{\beta}_{1,1} = \frac{S_{n,0}T_{n,1} - S_{n,1}T_{n,0}}{S_{n,2}S_{n,0} - S_{n,1}^2}$$

$$= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x) m_i k_i k_j - \sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x) m_i k_i k_j}{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x)^2 k_i k_j - \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x)(x_j - x) k_i k_j}$$

$$= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j) m_i k_i k_j}{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x)(x_i - x_j) k_i k_j}$$

$$= \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_j - x_i)(m_j - m_i) k_i k_j}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_j - x_i)^2 k_i k_j}. \tag{B.2}$$

Therefore if the bivariate sample $(x_1, m_1), \ldots, (x_n, m_n)$ satisfies the property that if $x_i < x_j$ then $(m_j - m_i)/(x_j - x_i) \geqslant \underline{c}$, for all i and $j > i$, where $\underline{c}$ is a constant then

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_j - x_i)(m_j - m_i) k_i k_j \geqslant \underline{c}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_j - x_i)^2 k_i k_j$$

and hence $\hat{m}_{1,1}(x) \geqslant \underline{c}$. If in addition the bivariate sample $(x_1, m_1), \ldots, (x_n, m_n)$ satisfies the property that if $x_i < x_j$ then $(m_j - m_i)/(x_j - x_i) \leqslant \bar{c}$, for all i and $j > i$, then

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_j - x_i)(m_j - m_i) k_i k_j \leqslant \bar{c}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_j - x_i)^2 k_i k_j$$

and hence $\hat{m}_{1,1}(x) \leqslant \bar{c}$. Applying this with $\underline{c} = \exp(-r_{t,\tau}\tau)$ and $\bar{c} = 0$ gives the result.

Part 2: Proof that $\hat{m}'_{1,1}(x) \geqslant 0$.

Let $M_{i,j} = (m_i - m_j)/(x_i - x_j) = (m_j - m_i)/(x_j - x_i)$ denote the local slope between x_i and x_j. Also define $k_{i,j} = (x_i - x_j)^2 k_i k_j$ and let $k'_{i,j}$ denote the partial derivative of $k_{i,j}$ with respect to x. Rewrite (B.2) as

$$\hat{m}_{1,1}(x) = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} M_{i,j} k_{i,j}}{\sum_{k=1}^{n-1}\sum_{l=k+1}^{n} k_{k,l}}$$

so that:

$$\hat{m}'_{1,1}(x) = \frac{\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} M_{i,j}k'_{i,j}\right)\left(\sum_{k=1}^{n-1}\sum_{l=k+1}^{n} k_{k,l}\right) - \left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} M_{i,j}k_{i,j}\right)\left(\sum_{k=1}^{n-1}\sum_{l=k+1}^{n} k'_{k,l}\right)}{\left(\sum_{k=1}^{n-1}\sum_{l=k+1}^{n} k_{k,l}\right)^2}. \tag{B.3}$$

Rearranging the terms in (B.3) yields

$$\left(\sum_{k=1}^{n-1}\sum_{l=k+1}^{n} k_{k,l}\right)^2 \hat{m}'_{1,1}(x) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=i+1}^{n-1}\sum_{l=k+1}^{n} (k'_{i,j}k_{k,l} - k_{i,j}k'_{k,l})(M_{i,j} - M_{k,l}) + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{l=j+1}^{n} (k'_{i,j}k_{i,l} - k_{i,j}k'_{i,l})(M_{i,j} - M_{i,l}). \tag{B.4}$$

We want to prove that $\hat{m}'_{1,1}(x) \geqslant 0$, i.e., that the right hand side of (B.4) is non-negative. Recall that we assumed that the kernel function $K(\cdot)$ was a log-concave probability density. That is, $\log(K)$ is concave, i.e., its first derivative is decreasing:

$$\frac{K'(a)}{K(a)} \geqslant \frac{K'(b)}{K(b)}$$

if $b \geqslant a$. Therefore if $k \geqslant i$ and $l \geqslant j$ we have

$$\frac{x - x_i}{h} \geqslant \frac{x - x_k}{h} \quad \text{and} \quad \frac{x - x_j}{h} \geqslant \frac{x - x_l}{h}$$

and hence

$$\frac{k'_i}{k_i} \leqslant \frac{k'_k}{k_k} \quad \text{and} \quad \frac{k'_j}{k_j} \leqslant \frac{k'_l}{k_l}$$

where $k_i = K_h(x - x_i)$ and $k'_i = h^{-1}K'_h(x - x_i)$. Therefore

$$\frac{k'_i}{k_i} - \frac{k'_k}{k_k} + \frac{k'_j}{k_j} - \frac{k'_l}{k_l} \leqslant 0$$

and

$$k'_{i,j}k_{k,l} - k_{i,j}k'_{k,l} = (x_i - x_j)^2(x_k - x_l)^2 k_i k_k k_j k_l \left(\frac{k'_i}{k_i} - \frac{k'_k}{k_k} + \frac{k'_j}{k_j} - \frac{k'_l}{k_l}\right) \leqslant 0 \tag{B.5}$$

if $k \geqslant i$ and $l \geqslant j$.

From now on, let

$$c_{i,j,k,l} \equiv (k'_{i,j}k_{k,l} - k_{i,j}k'_{k,l})(M_{i,j} - M_{k,l}) \tag{B.6}$$

denote the generic term in the sums (B.4). In addition to (B.5), it is also the case that $M_{i,j} \leqslant M_{k,l} \leqslant 0$, hence $M_{i,j} - M_{k,l} \leqslant 0$ for all (i,j,k,l) such that $k \geqslant i$ and $l \geqslant j$.

Therefore for such (i,j,k,l) we have $c_{i,j,k,l} \geqslant 0$. Throughout the first sum in (B.4), the indices satisfy $k > i$, and in the second sum $k = i$. Thus as long as $l \geqslant j$, the terms $c_{i,j,k,l}$ are nonnegative throughout the two sums in (B.4). That $l \geqslant j$ will be the case for all the terms in the second sum in (B.4), where $l \geqslant j+1$, but not necessarily in the first sum where there are quadruplets (i,j,k,l) such that $k \geqslant i$ but $l < j$. For these, we cannot be sure that $c_{i,j,k,l} \geqslant 0$.

Consider such a quadruplet (i,j,k,l) in the sum $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=i+1}^{n-1}\sum_{l=k+1}^{n} c_{i,j,k,l}$ for which nonnegativity of $c_{i,j,k,l}$ is not guaranteed. Such a quadruplet satisfies $i < k < l < j$. The key to the proof that these terms are not big enough to make the overall sum negative is to consider this problematic quadruplet (i,j,k,l) together with the two permutations (i,k,l,j) and (i,l,k,j). These two permutations are used up only with that particular quadruplet: any other problematic quadruplet would not need to re-use the same permutations. For these two permutations, we have $c_{i,k,l,j} \geqslant 0$ (since $l > i$ and $j > k$) and $c_{i,l,k,j} \geqslant 0$ (since $k > i$ and $j > l$) and it turns out that adding these two terms to the problematic term produces a nonnegative result, that is

$$c_{i,j,k,l} + c_{i,k,l,j} + c_{i,l,k,j} \geqslant 0. \tag{B.7}$$

To prove this, we now show that

$$\begin{aligned} c_{i,j,k,l} + c_{i,k,l,j} + c_{i,l,k,j} &= (k'_{i,j}k_{k,l} - k_{i,j}k'_{k,l})(M_{i,j} - M_{k,l}) + (k'_{i,k}k_{l,j} - k_{i,k}k'_{l,j}) \\ &\quad \times (M_{i,k} - M_{l,j}) + (k'_{i,l}k_{k,j} - k_{i,l}k'_{k,j})(M_{i,l} - M_{k,j}) \\ &= k_i k_j k_k k_l \left(\frac{k'_i}{k_i}\, t_i + \frac{k'_j}{k_j}\, t_j + \frac{k'_k}{k_k}\, t_k + \frac{k'_l}{k_l}\, t_l \right), \end{aligned} \tag{B.8}$$

where

$$\begin{aligned} t_i &\equiv (x_k - x_i)(x_l - x_i)(x_j - x_i)\{(M_{i,j} - M_{i,l})(2x_j - x_k - x_l) \\ &\quad + (M_{i,k} - M_{i,l})(2x_k - x_j - x_l)\}, \\ t_j &\equiv (x_j - x_l)(x_j - x_k)(x_j - x_i)\{(M_{i,j} - M_{k,j})(2x_i - x_k - x_l) \\ &\quad + (M_{l,j} - M_{k,j})(2x_l - x_i - x_k)\}, \\ t_k &\equiv (x_j - x_k)(x_l - x_k)(x_k - x_i)\{(M_{i,k} - M_{k,l})(2x_i - x_j - x_l) \\ &\quad + (M_{k,j} - M_{k,l})(2x_j - x_i - x_l)\}, \\ t_l &\equiv (x_j - x_l)(x_l - x_k)(x_l - x_i)\{(M_{i,l} - M_{k,l})(2x_i - x_j - x_k) \\ &\quad + (M_{l,j} - M_{k,l})(2x_j - x_i - x_k)\}. \end{aligned} \tag{B.9}$$

Note that

$$\begin{aligned} t_i + t_k &= 2(x_k - x_i)^2(x_j - x_l)^2(M_{i,k} - M_{l,j}), \\ t_j + t_l &= 2(x_k - x_i)^2(x_j - x_l)^2(M_{l,j} - M_{i,k}) \end{aligned} \tag{B.10}$$

therefore

$$t_i + t_k \leqslant 0,$$
$$t_j + t_l \geqslant 0,$$
$$t_i + t_k + t_j + t_l = 0. \tag{B.11}$$

Recall now that we are dealing with a quadruplet (i, j, k, l) such that $i < k < l < j$: therefore we have

$$M_{i,k} \leqslant M_{i,l} \leqslant M_{i,j} \leqslant M_{l,j},$$
$$M_{i,k} \leqslant M_{k,l} \leqslant M_{k,j} \leqslant M_{l,j},$$
$$M_{i,l} \leqslant M_{k,l}. \tag{B.12}$$

These inequalities follow from repeated application of the fact that for any triplet (i, k, l) such that $i < k < l$,

$$\frac{m_k - m_i}{x_k - x_i} \leqslant \frac{m_l - m_i}{x_l - x_i} \leqslant \frac{m_l - m_k}{x_l - x_k} \tag{B.13}$$

which itself follows from

$$\frac{m_l - m_i}{x_l - x_i} = \left(\frac{x_k - x_i}{x_l - x_i}\right) \frac{m_k - m_i}{x_k - x_i} + \left(1 - \frac{x_k - x_i}{x_l - x_i}\right) \frac{m_l - m_k}{x_l - x_k}$$

where $0 \leqslant (x_k - x_i)/(x_l - x_i) \leqslant 1$. Thus the middle slope $M_{i,l}$ is a weighted average of the extreme slopes $M_{k,l}$ and $M_{i,l}$.

As a consequence of (B.12), we have $t_k \geqslant 0$ and $t_j \geqslant 0$. Combined with (B.11), it follows that:

$$t_i \leqslant -t_k \leqslant 0,$$
$$-t_j \leqslant t_l \leqslant 0. \tag{B.14}$$

We can now return to (B.8). The sign of its right hand side is determined by the sign of

$$\left(\frac{k_i'}{k_i} t_i + \frac{k_j'}{k_j} t_j + \frac{k_k'}{k_k} t_k + \frac{k_l'}{k_l} t_l\right)$$

and since $i < k < l < j$, we have

$$\frac{k_i'}{k_i} \leqslant \frac{k_k'}{k_k} \leqslant \frac{k_l'}{k_l} \leqslant \frac{k_j'}{k_j}$$

by the log-concavity of the kernel function. Since $t_k \geqslant 0$,

$$\frac{k_i'}{k_i} t_k \leqslant \frac{k_k'}{k_k} t_k \quad \Rightarrow \quad \frac{k_i'}{k_i} t_i + \frac{k_k'}{k_k} t_k \geqslant \frac{k_i'}{k_i} (t_i + t_k)$$

and since $t_j \geqslant 0$,

$$\frac{k_j'}{k_j} t_j \geqslant \frac{k_l'}{k_l} t_j \quad \Rightarrow \quad \frac{k_j'}{k_j} t_j + \frac{k_l'}{k_l} t_l \geqslant \frac{k_l'}{k_l} (t_l + t_j).$$

Since now $t_l + t_j \geqslant 0$,

$$\frac{k_l'}{k_l} \geqslant \frac{k_i'}{k_i} \quad \Rightarrow \quad \frac{k_l'}{k_l} (t_l + t_j) \geqslant \frac{k_i'}{k_i} (t_l + t_j)$$

from which it follows that

$$\left(\frac{k_i'}{k_i} t_i + \frac{k_j'}{k_j} t_j + \frac{k_k'}{k_k} t_k + \frac{k_l'}{k_l} t_l \right) \geqslant \frac{k_i'}{k_i} (t_i + t_k + t_l + t_j) = 0 \tag{B.15}$$

hence the result (B.7).

Hence $\hat{m}'_{1,1}(x) \geqslant 0$, as desired. Setting $\hat{m}^{(1)}(x) = \hat{m}_{1,1}(x)$ and $\hat{m}^{(2)}(x) = \hat{m}'_{1,1}(x)$ we therefore have estimators of the slope and state-price density that will always satisfy the constraints in sample.

References

Abadir, K., Rockinger, M., 1998. Density-embedding functions. Working paper, HEC School of Management.

Afriat, S., 1967. The construction of a utility function from expenditure data. International Economic Review 8, 67–77.

Aït-Sahalia, Y., 1996a. Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527–560.

Aït-Sahalia, Y., 1996b. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9, 385–426.

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

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

Aït-Sahalia, Y., Wang, Y., Yared, F., 2001. Do option markets correctly price the probabilities of movement of the underlying asset? Journal of Econometrics 102, 67–110.

Bahra, B., 1996. Probability distributions of future asset prices implied by option prices. Bank of England Quarterly Bulletin 36, 299–311.

Banz, R., Miller, M., 1978. Prices for state-contingent claims: some estimates and applications. Journal of Business 51, 653–672.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D., 1972. Statistical Inference under Order Restrictions. Wiley, New York, NY.

Bates, D.S., 2000. Post-'87 crash fears in the S&P 500 futures option market. Journal of Econometrics 94, 181–238.

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

Bondarenko, O., 1997. Testing rationality of financial markets. Working paper, Caltech.

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

Brunk, H.D., 1970. Estimation of isotonic regression. In: Puri, M.L. (Ed.), Nonparametric Techniques in Statistical Inference, Cambridge University Press, Cambridge.

Christoffersen, P., Jacobs, K., 2001. The importance of the loss function in option pricing. Working paper, McGill University.

Cox, J.C., Ross, S.A., 1976. The valuation of options for alternative stochastic processes. Journal of Financial Economics 3, 145–166.

Diewert, W.E., 1973. Functional forms for profit and transformation functions. Journal of Economic Theory 6, 284–316.

Dole, D., 1999. Constrained scatterplot smoother for estimating convex, monotonic transformations. Journal of Business and Economic Statistics 17, 444–455.

Duffie, D., 1996. Dynamic Asset Pricing Theory, Second Edition. Princeton University Press, Princeton, NJ.

Dykstra, R.L., 1983. An algorithm for restricted least squares. Journal of the American Statistical Association 78, 837–842.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman & Hall, London.

Garcia, R., Gencay, R., 2000. Pricing and hedging derivative securities with neural networks and a homogeneity hint. Journal of Econometrics 94, 93–115.

Goldman, S.M., Ruud, P., 1995. Nonparametric multivariate regression subject to constraint. Working paper, UC Berkeley.

Haefke, C., White, H., Gottschling, A., 2000. Closed form integration of artificial neural networks with some applications in finance. Working paper, UC San Diego.

Hanson, D.L., Pledger, G., 1976. Consistency in concave regression. The Annals of Statistics 4, 1038–1050.

Hanson, D.L., Pledger, G., Wright, F.T., 1973. On consistency in monotonic regression. The Annals of Statistics 1, 401–421.

Harrison, M., Kreps, D., 1979. Martingales and arbitrage in multiperiod securities markets. Journal of Economic Theory 20, 381–408.

Hentschel, L., 2001. Errors in implied volatility estimation. Working paper, University of Rochester.

Hildreth, C., 1954. Point estimates of ordinates of concave functions. Journal of the American Statistical Association 49, 598–619.

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

Lucas, R.E., 1978. Asset prices in an exchange economy. Econometrica 46, 1429–1445.

Mammen, E., 1991. Estimating a smooth monotone regression function. The Annals of Statistics 19, 724–740.

Mammen, E., Thomas-Agnan, C., 1999. Smoothing splines and shape restrictions. Scandinavian Journal of Statistics 26, 239–252.

Matzkin, R.L., 1991. Semiparametric estimation of monotone and concave utility functions for polychotomous choice models. Econometrica 59, 1315–1327.

Matzkin, R.L., 1992. Nonparametric and distribution-free estimation of the binary choice and the threshold-crossing models of monotone and concave utility functions for polychotomous choice models. Econometrica 60, 239–270.

Matzkin, R.L., 1994. Restrictions of economic theory in nonparametric methods. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. 4, North Holland, Amsterdam.

Matzkin, R.L., Richter, M.K., 1991. Testing strictly concave rationality. Journal of Economic Theory 53, 287–303.

Mukerjee, H., 1988. Monotone nonparametric regression. The Annals of Statistics 16, 741–750.

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

Robertson, T., Wright, F.T., Dykstra, R.L., 1988. Order Restricted Statistical Inference. Wiley, New York.

Rubinstein, M., 1976. The valuation of uncertain income streams and the pricing of options. Bell Journal of Economics, 407–425.

Ruud, P., 1997. Restricted least squares subject to monotonicity and convexity constraints. In: Kreps, D.M., Wallis, K.F. (Eds.), Advances in Economics and Econometrics: Theory and Applications, Vol. III, Cambridge University Press, Cambridge.

Stone, C.J., 1983. Optimal uniform rate of convergence for nonparametric estimators of a density function or its derivatives. In: M.H. Rezvi, J.S. Rustagi and D. Siegmund (Eds.), Recent Advances in Statistics: Papers in Honor of Herman Chernoff. Academic Press, New York.

Varian, H.R., 1982. The nonparametric approach to demand analysis. Econometrica 50, 945–973.

Varian, H.R., 1983. Nonparametric tests of models of investor behavior. Journal of Financial and Quantitative Analysis 18, 269–278.

Varian, H.R., 1984. The nonparametric approach to production analysis. Econometrica 52, 579–597.

Von Neumann, J., 1950. Functional Operators, Volume II. Princeton University Press, Princeton, NJ.
Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman & Hall, London.
Wang, Y., 1993. The limiting distribution in concave regression. Working paper, University of Missouri-Columbia.
Wright, F.T., 1981. The asymptotic behavior of monotone regression estimators. The Annals of Statistics 9, 443–448.