Nonparametric Estimation of State-Price Densities Implicit In Financial Asset Prices

Yacine Aït-Sahalia and Andrew W. Lo*

First Draft: July 17, 1995
Latest Revision: June 3, 1997

Abstract

Implicit in the prices of traded financial assets are Arrow-Debreu prices or, with continuous states, the state-price density (SPD). We construct a nonparametric estimator for the SPD implicit in option prices and derive its asymptotic sampling theory. This estimator provides an arbitrage-free method of pricing new, complex, or illiquid securities while capturing those features of the data that are most relevant from an asset-pricing perspective, e.g., negative skewness and excess kurtosis for asset returns, volatility "smiles" for option prices. We perform Monte Carlo experiments and extract the SPD from actual S&P 500 option prices.

---

*Aït-Sahalia is from the University of Chicago and the NBER, and Lo is from MIT and the NBER. We received helpful comments from George Constantinides, Eric Ghysels, John Heaton, Jens Jackwerth, Mark Rubinstein, René Stulz (the editor), Jiang Wang, two referees, and especially Lars Hansen, as well as seminar participants at Duke University, the Fields Insitute, Harvard University, MIT, Tilburg University, The University of Chicago, ULB, Washington University, the Econometric Society Winter Meetings (1995), the NBER Asset Pricing Conference (1995), the NNCM Conference (1994) and the WFA Meetings (1996). A portion of this research was conducted during the first author's tenure as an IBM Corp. Faculty Research Fellow and the second author's tenure as an Alfred P. Sloan Research Fellow.

One of the most important theoretical advances in the economics of investment under uncertainty is the time-state preference model of Arrow (1964) and Debreu (1959) in which they introduce elementary securities each paying one dollar in one specific state of nature and nothing in any other state. Now known as *Arrow-Debreu* securities, they are the fundamental building blocks from which we derive much of our current understanding of economic equilibrium in an uncertain environment. In a continuum of states, the prices of Arrow-Debreu securities are defined by the *state-price density* or SPD, which gives for each state x the price of a security paying one dollar if the state falls between x and $x + dx$.

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 of Black and Scholes (1973) and Merton (1973). Both strands of the literature have adopted their own lexicon to denote closely related concepts.

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 (see, for example, Hansen and Jagannathan (1991) and Hansen and Richard (1987)).

Among the no-arbitrage models, the SPD is often called the *risk-neutral density*, based on the analysis of Ross (1976) and Cox and Ross (1976) who first observed that the Black-Scholes formula can be derived by assuming that all investors are risk neutral, implying that all assets in such a world—including options—must yield an expected return equal to the riskfree rate of interest. The SPD also uniquely characterizes the *equivalent martingale measure* under which all asset prices discounted at the riskfree rate of interest are martingales (see Harrison and Kreps (1979)).

Given the enormous informational content that Arrow-Debreu prices possess and the great simplification they provide for pricing complex state-contingent securities such as options and other derivatives, it is unfortunate that pure Arrow-Debreu securities are not yet traded on any organized exchange. However, they may be estimated or approximated from the prices of traded financial securities, as suggested by Banz and Miller (1978), Breeden and Litzenberger (1978) and Ross (1976), and three methods have been proposed to do just this.

In the first method, sufficiently strong assumptions are made on the underlying asset-price dynamics for the SPD to 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.[1] The second method requires specifying the SPD directly in some parametric form.[2] And the third method begins by specifying a "prior" parametric distribution as a candidate SPD, typically the Black-Scholes log-normal density. The SPD is then estimated by minimizing its distance to the prior parametric distribution under the constraints that it correctly prices a selected set of derivative securities.[3]

In this paper, we propose an alternative to these three methods in which the SPD is estimated *nonparametrically*, that is, with no parametric restrictions on either the underlying asset's price dynamics or on the family of distributions that the SPD belongs to, and no need for choosing any prior distribution for the SPD.[4] While parametric approaches are clearly preferable when the underlying asset's price process is known to satisfy particular parametric assumptions, e.g., geometric Brownian motion, nonparametric methods are preferred when such assumptions are violated. Since recent empirical evidence seems to cast some doubt on the more popular parametric specifications,[5] a nonparametric approach to estimating the SPD may be an important alternative to the more traditional methods.

In particular, our nonparametric SPD estimator can yield valuable insights in at least four contexts. 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, e.g., basic call-option prices, that are used to estimate the SPD.[6] From a risk management perspective, we 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.

Second, 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. For example, in our empirical application to S&P 500 index options in Section III, the nonparametric SPD estimator naturally captures the so-called *volatility smile* (see Figure 3) because this is a prominent feature of the data. But we also document changes in the shape of the volatility smile over different maturities which parametric models so far have not incorporated. The nonparametric SPD estimator also exhibits persistent negative skewness and excess kurtosis (see Figure 7) because these too are features of the data. Indeed, 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. In contrast, typical parametric stochastic-volatility models have difficulty eliminating biases in short-term option prices because they do not generate enough kurtosis, while typical jump models encounter difficulties with longer-term options because they revert too fast to the Black-Scholes prices as the maturity date grows.

Third, and perhaps most importantly, the nonparametric estimator highlights the empirical features of the data in a way that is robust to the classical "joint hypothesis" problem. Our estimator is free of the typical joint hypotheses on asset-price dynamics and risk premia that are typical of parametric arbitrage models, or on preferences in the equilibrium approach to derivative pricing. Of course, nonparametric techniques do require certain assumptions on the data-generating process itself, but these are typically weaker than those of parametric models and are less likely to be violated in practice.

Fourth, if we make the additional assumption that underlying asset prices follow diffusion processes, our estimator of the SPD can be used to estimate nonparametrically the instantaneous volatility function of the underlying asset-return process.[7] Thus if we restrict attention to diffusions, we obtain the continuous-state analog to the implied binomial trees proposed by Rubinstein (1994) (see also Derman and Kani (1994), Shimko (1993), and the implied volatility functions in Dupire (1994), and Dumas et al. (1995)).

In Section I we review briefly the relation between SPDs and the pricing of derivative securities, introduce our nonparametric SPD estimator, and compare it to alternative approaches. We present the results of extensive Monte Carlo simulation experiments in Section II in which we generate simulated price data under the Black-Scholes assumptions and show that our SPD estimator can successfully approximate the Black-Scholes SPD. In Section III, we apply our SPD estimator to the pricing and delta-hedging of S&P 500 index options. Options with different times-to-expiration yield a family of SPDs over different horizons. We document several empirical features of the SPD over time, including the term structures of mean returns, volatility, skewness and kurtosis that are implied by these distributions. Moreover, unlike many parametric option-pricing models, we show that the SPD-generated option-pricing formula is capable of capturing persistent volatility smiles and other empirical features of market prices. We then point to parametric modeling strategies that would incorporate these features. We conclude in Section IV, and give technical results and details in the Appendix, including the asymptotic distributions of our SPD estimators and corresponding specification test statistics.

I Nonparametric Estimation of SPDs

The intimate relation between SPDs, dynamic equilibrium models, and derivative securities is now well known, but for completeness and to develop notation we shall provide a brief summary here (see, for example, Huang and Litzenberger (1988, Chapters 5–8) for a more detailed discussion).

In a dynamic equilibrium model such as Lucas (1978) and Rubinstein (1976), the price of any financial security can be expressed as the expected net present value of its future payoffs, where the present value is calculated with respect to the riskless rate r and the expectation is taken with respect to the marginal-rate-of-substitution-weighted probability density function (PDF) of the payoffs. This PDF, which is distinct from the PDF of the payoffs, is the SPD or *risk-neutral* PDF (see Cox and Ross (1976)) or *equivalent martingale measure* (see Harrison and Kreps (1979)).

More formally, the date-t price P_t of a security with a single liquidating date-T payoff $Z(S_T)$ is given by:

$$P_t = e^{-r_{t,\tau}\tau} E_t^* \left[Z\left(S_T\right) \right] = e^{-r_{t,\tau}\tau} \int_0^\infty Z\left(S_T\right) f_t^*\left(S_T\right) dS_T \tag{1}$$

where S_T is a state variable, e.g., aggregate consumption in Lucas (1978), $r_{t,\tau}$ is the constant riskfree rate of interest between t and $T \equiv t + \tau$ and $f_t^*\left(S_T\right)$ is the date-t SPD for date-T payoffs.

The dynamic equilibrium approach illustrates the enormous information content that SPDs contain and the enormous information reduction that SPDs allow. For example, if parametric restrictions are imposed on the data-generating process of asset prices, e.g., geometric Brownian motion, the SPD estimator may be used to infer the preferences of the representative agent in an equilibrium model of asset prices (see, for example, He and Leland (1993)). Alternatively, if specific preferences are imposed, e.g., logarithmic utility, the SPD may be used to infer the data-generating process of asset prices. Indeed, any two of the following imply the third: (1) the representative agent's preferences; (2) asset-price dynamics; and (3) the SPD. From a pricing perspective, SPDs are "sufficient statistics" in an economic sense—they summarize all relevant information about preferences and business conditions for purposes of pricing financial securities.

A SPDs and Derivative Securities

In contrast to the PDF of the payoffs, the SPD cannot be easily estimated from the time series of payoffs because it is also influenced by preferences, i.e., the marginal rate of substitution. However, the SPD *can* be estimated from the time series of prices since prices represent the amalgamation of payoffs and preferences in an equilibrium context. In fact, building on Ross's (1976) insight that options can be used to create pure Arrow-Debreu state-contingent securities, Banz and Miller (1978) and Breeden and Litzenberger (1978) provide an elegant method for obtaining an explicit expression for the SPD from option prices: the SPD is the second derivative, normalized to have an integral of one, of a call option-pricing formula with respect to the strike price.

For example, recall that under the hypotheses of Black and Scholes (1973) and Merton (1973), the date-t price H of a call option maturing at date $T \equiv t + \tau$, with strike price

X, written on a stock with date-t price S_t, when the dividend yield is $\delta_{t,\tau}$ and the stock's volatility is σ, is given by:[8]

$$\begin{aligned} H_{\mathrm{BS}}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}; \sigma) &= e^{-r_{t,\tau}\tau} \int_0^{\infty} \max\left(S_T - X, 0\right) f^*_{BS,t}\left(S_T\right) dS_T \\ &= S_t e^{-\delta_{t,\tau}\tau} \Phi\left(d_1\right) - X e^{-r_{t,\tau}\tau} \Phi\left(d_2\right) \end{aligned} \tag{2}$$

where

$$d_1 \equiv \frac{\ln(S_t/X) + (r_{t,\tau} - \delta_{t,\tau} + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \qquad d_2 \equiv d_1 - \sigma\sqrt{\tau}. \tag{3}$$

In this case the corresponding SPD is a log-normal density with mean $((r_{t,\tau} - \delta_{t,\tau}) - \sigma^2/2)\tau$ and variance $\sigma^2\tau$:

$$\begin{aligned} f^*_{\mathrm{BS},t}(S_T) &= e^{r_{t,\tau}\tau} \frac{\partial^2 H}{\partial X^2}|_{X=S_T} \\ &= \frac{1}{S_T\sqrt{2\pi\sigma^2\tau}} \exp\left[-\frac{[\ln(S_T/S_t) - (r_{t,\tau} - \delta_{t,\tau} - \sigma^2/2)\tau]^2}{2\sigma^2\tau}\right] \end{aligned} \tag{4}$$

This expression shows that the SPD can depend on many quantities in general (although for simplicity we write it explicitly as a function of S_T only), and is distinct from but related to the PDF of the terminal stock price S_T.

Armed with (4), the price of any other derivative security can be calculated trivially, as long as the hypotheses that lead to the option-pricing formula (2) are satisfied.[9] For example, the date-t price of a *digital* call option—a security which gives the holder the right to receive a fixed cash payment of one dollar if S_T exceeds a prespecified level X—is given by:

$$\begin{aligned} D_t &= e^{-r_{t,\tau}\tau} E^*_t[I(S_T > X)] = e^{-r_{t,\tau}\tau} \int_0^{\infty} I(S_T > X) f^*_t(S_T) dS_T \\ &= e^{-r_{t,\tau}\tau} \left[1 - F^*_t(X)\right] \end{aligned} \tag{5}$$

where $I(S_T > X)$ is an indicator function that takes on the value one if $S_T > X$ and zero otherwise, and F^*_t is the CDF of the SPD. We shall return to this example in our empirical analysis of S&P 500 options (see Section III).

B Nonparametric Option Prices and SPDs

It is clear from (4) that the SPD is inextricably linked to the parametric assumptions underlying the Black-Scholes/Merton option-pricing model. If those parametric assumptions do not hold, e.g., if the dynamics of $\{S_t\}$ contain Poisson jumps or the volatility of the asset returns varies with the stock price, then the SPD (4) will yield incorrect prices, prices that are inconsistent with the dynamic equilibrium model or the hypothesized stochastic process driving $\{S_t\}$. A *nonparametric* estimator of the SPD—an estimator that does not rely on specific parametric assumptions such as geometric Brownian motion dynamics—can yield prices that are robust to such parametric specification errors.

Breeden and Litzenberger (1978) first observed that $f_t^*(S_T) = \exp(r_{t,\tau}\tau)\partial^2 H(\cdot)/\partial X^2$. We propose to estimate the SPD nonparametrically in the following way: use market prices to estimate an option-pricing formula $\hat{H}(\cdot)$ nonparametrically, then differentiate this estimator twice with respect to X to obtain $\partial^2 \hat{H}(\cdot)/\partial X^2$. Under suitable regularity conditions, the convergence (in probability) of $\hat{H}(\cdot)$ to the true option-pricing formula $H(\cdot)$ implies that $\partial^2 \hat{H}(\cdot)/\partial X^2$ will converge to $\partial^2 H(\cdot)/\partial X^2$ which is proportional to the SPD.

But how do we obtain $\hat{H}(\cdot)$ nonparametrically? Given a set of historical option prices $\{H_i\}$ and accompanying characteristics $\{\mathbf{Z}_i \equiv [\ S_{t_i}\ X_i\ \tau_i\ r_{t_i,\tau_i}\ \delta_{t_i,\tau_i}\ ]'\}$ we seek a *function* $H(\cdot)$ —not a set of parameters—that comes as close to $\{H_i\}$ as possible. In particular, using mean-squared-error as our measure of closeness, we wish to solve:

$$\min_{H(\cdot)\in\Gamma} \sum\nolimits_{i=1}^{n} [H_i - H(\mathbf{Z}_i)]^2 \tag{6}$$

where Γ is the space of twice-continuously-differentiable functions. It is well known that the solution is given by the conditional expectation of H given $\mathbf{Z}$.

To estimate this conditional expectation, we employ a statistical technique known as *nonparametric kernel regression*. Nonparametric kernel regression produces an estimator of the conditional expectation of H, conditioned on $\mathbf{Z}$, without requiring that the function $H(\cdot)$ be *parametrized* by a finite number of parameters (hence the term *nonparametric*). Kernel regression requires few assumptions other than smoothness of the function to be estimated and regularity of the data used to estimate it, and is robust to the potential misspecification of any given parametric call-pricing formula (see the Appendix). On the other hand, kernel regression tends to be data intensive. Financial applications are a natural outlet for kernel regression since typical parametric assumptions, e.g., normality or geometric Brownian motion, have been rejected by the data, yet large sample sizes of high quality data are not uncommon.

Another motivation for kernel regression that is particularly insightful is the local averaging or *smoothing* interpretation. Suppose we wish to estimate the relation between two variables $\mathbf{Z}_i$ and H_i that satisfy the following nonlinear relation:

$$H_i = H(\mathbf{Z}_i) + \epsilon_i\ , \qquad i = 1, \ldots, n \tag{7}$$

where $H(\cdot)$ is an unknown but fixed nonlinear function and $\{\epsilon_i\}$ is white noise. Consider estimating $H(\cdot)$ at a specific point $\mathbf{Z}_{i_0} = \mathbf{z}_0$ and suppose that for this one particular observation of $\mathbf{Z}_i$, we are able to obtain *repeated* observations of the variable H_{i_0}, say $H_{i_0}^{(1)}, \ldots, H_{i_0}^{(q)}$.

In this case, a natural estimator of the function $H(\cdot)$ at the point $\mathbf{z}_0$ is simply:

$$\hat{H}(\mathbf{z}_0) = \frac{1}{q}\sum_{j=1}^{q} H_{i_0}^{(j)} = H(\mathbf{z}_0) + \frac{1}{q}\sum_{j=1}^{q} \epsilon_{i_0}^{(j)}$$

and by the law of large numbers, $(1/q)\sum_{j=1}^{q}\epsilon_{i_0}^{(j)}$ becomes negligible for large q.

Of course, we do not have the luxury of repeated observations for a given $\mathbf{Z}_{i_0} = \mathbf{z}_0$. However, if we assume that the function $H(\cdot)$ is smooth, then for time series observations $\mathbf{Z}_i$ near the value $\mathbf{z}_0$, the corresponding values of H_i should be close to $H(\mathbf{z}_0)$. In other words, if $H(\cdot)$ is smooth, then in a small neighborhood around $\mathbf{z}_0$, $H(\mathbf{z}_0)$ will be nearly constant and may be estimated by taking an average of the H_i's that correspond to those $\mathbf{Z}_i$'s near $\mathbf{z}_0$. The closer the $\mathbf{Z}_i$'s are to the value $\mathbf{z}_0$, the closer an average of the corresponding H_i's will be to $H(\mathbf{z}_0)$. This argues for a *weighted* average of the H_i's, where the weights decline as the $\mathbf{Z}_i$'s get farther away from the point $\mathbf{z}_0$. Such a weighted average must be computed for *each* value of $\mathbf{z}$ in the domain of $H(\cdot)$ to estimate the entire function, hence computational considerations become important.

This weighted-average procedure of estimating $H(\mathbf{z})$ is the essence of smoothing. To implement such a procedure, we must define what we mean by "near" and "far". If we choose too large a neighborhood around $\mathbf{z}$ to compute the average, the weighted average will be too smooth and will not exhibit the genuine nonlinearities of $H(\cdot)$. If we choose too small a neighborhood around $\mathbf{z}$, the weighted average will be too variable, reflecting noise as well as nonlinearities in $H(\cdot)$. Therefore, the weights must be chosen carefully to balance these two considerations. The choice of weighting function—typically given by a probability density function (since such functions integrate to one), though the particular density function plays no probabilistic role here—determine the degree of local averaging (see Härdle (1990) and Wand and Jones (1995) for a more detailed discussion of nonparametric regression).

To specify a particular kernel regression model, we start with the natural assumption that the option-pricing formula H we seek to estimate is a function of a vector of option characteristics or "explanatory" variables, $\mathbf{Z} \equiv [\ S_t \ X \ \tau \ r_{t,\tau} \ \delta_{t,\tau}\ ]'$ so that each option price H_i, $i = 1, \ldots, n$, contained in our dataset is paired with the vector $\mathbf{Z}_i \equiv [\ S_{t_i} \ X_i \ \tau_i \ r_{t_i,\tau_i} \ \delta_{t_i,\tau_i}\ ]'$. [10] Since we have five explanatory variables, we select a five-dimensional weighting or *kernel* function $K(\mathbf{Z})$ which integrates to one.

The density $K(\mathbf{Z} - \mathbf{Z}_i)$, as a function of $\mathbf{Z}$, has a certain spread around the data point $\mathbf{Z}_i$. We can change the spread of the kernel K around $\mathbf{Z}$ using a *bandwidth* h, to form the new density function $(1/h)K((\mathbf{Z} - \mathbf{Z}_i)/h)$. An estimator of the conditional expectation of H conditioned on $\mathbf{Z}$ is then given by the following expression, called the *Nadaraya-Watson* kernel estimator, where h becomes closer to zero as the sample size n grows:

$$\hat{H}(\mathbf{Z}) = \hat{E}[H|\mathbf{Z}] = \frac{\sum_{i=1}^{n} K((\ \mathbf{Z} - \mathbf{Z}_i)/h)H_i}{\sum_{i=1}^{n} K((\mathbf{Z} - \mathbf{Z}_i)/h)}. \tag{8}$$

Intuitively, the estimate of the conditional expectation at a point $\mathbf{Z}$, i.e., the price of an option with characteristics $\mathbf{Z}$, is given by a weighted average of the observed prices H_i's with more weight given to the options whose characteristics $\mathbf{Z}_i$'s are closer to the characteristics $\mathbf{Z}$ of the option to be priced.

The closer h is to zero, the more peaked is this new density function around $\mathbf{Z}_i$, and hence more weight is given to realizations of the random variable $\mathbf{Z}_i$ that are close to $\mathbf{Z}$. To illustrate kernel regression in our context, consider a sample of options with different strike prices and the same time-to-expiration. Suppose that we are interested in learning how the implied volatilities of options of that maturity vary with the option's moneyness—the now familiar volatility smile. The choice of the kernel function typically has little influence on the end result, while the choice of the bandwidth h is the determining factor, since the spread of the density around $\mathbf{Z}_i$ is much more sensitive to the choice of h than that of K, provided of course that the variance of the class of potential kernel functions is restricted. Figure 1a demonstrates the effect of undersmoothing: h is 0.1 times the optimal value. The estimator is correctly centered around the true function (the dashed line), i.e., has low bias, but is highly variable. Figure 1b is optimally smoothed, according to the formula we derive in Appendix I. Figure 1c is oversmoothed: h is four times the optimal value. This estimator exhibits little variance, but is substantially biased. In Appendix I, we determine the bandwidth h that achieves the optimal trade-off between bias and variance.

C Practical Considerations: Dimension Reduction

We also show in Appendix I that obtaining accurate estimators of the regression function is more difficult when the number of regressors d is large (we start here with $d = 5$ as $\mathbf{Z}_i \equiv [\ S_{t_i}\ X_i\ \tau_i\ r_{t_i,\tau_i}\ \delta_{t_i,\tau_i}\ ]'$), and then derivatives of the function $H(\cdot)$ need to be computed: recall that to obtain the SPD we must differentiate the call-pricing function twice.[11]

To be fully nonparametric, all five regressors in $\mathbf{Z} \equiv [\ S_t\ X\ \tau\ r_{t,\tau}\ \delta_{t,\tau}\ ]'$must be included in the kernel regression (8) of call option prices H on $\mathbf{Z}$. To reduce the number of regressors, we examine the following possibilities.

First, we could assume that the option-pricing formula (and hence the SPD) is not a function of the asset price S_t, the riskfree rate $r_{t,\tau}$ and the dividend yield $\delta_{t,\tau}$ separately, but only depends on these three variables through the futures price $F_{t,\tau} = S_t e^{(r_{t,\tau} - \delta_{t,\tau})\tau}$ and the riskfree rate, that is: $H(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = H(F_{t,\tau}, X, \tau, r_{t,\tau})$. By no-arbitrage, the mean of the SPD depends only on—and, in fact, is equal to—$F_{t,\tau}$ and the assumption here would be that the entire distribution has this property. It is satisfied by the Black-Scholes SPD (4). Under this assumption, the number of regressors is reduced from $d = 5$ to $d = 4$. Note that the value of a European option written on a futures contract with the same maturity as the option is identical to the value of the corresponding option written on the asset. Further, the drift of the futures process is zero under the risk-neutral measure (hence independent of $\delta_{t,\tau}$), so it is quite reasonable to expect that the dividend yield does not enter the option-pricing formula other than through the futures value. Theoretically, it may still enter the function, for instance if the volatility of the asset's returns depends on $\delta_{t,\tau}$, but this would most likely be a fairly remote situation.

Second, we could assume that the option-pricing function is homogeneous of degree one in $F_{t,\tau}$ and X, as in the Black-Scholes formula. This assumption would also reduce the number of regressors from $d = 5$ to $d = 4$. Combining this assumption with the previous one, the dimension of the problem would be further reduced to $d = 3$. It can be shown however that the call-pricing function is homogeneous of degree one in the asset price and the strike price when the distribution of the returns is independent of the level of the asset price (see Merton (1973, Theorem 9) who also provides a counter example showing that the homogeneity property can fail if the distribution of the returns is not independent of S_t). An example of a pricing formula satisfying the homogeneity property would be the one generated by a stochastic volatility model where the drift and diffusion functions of the stochastic volatility process depend on the volatility itself but not on the asset price (see Renault (1995)). While this assumption may not be too restrictive in practice, this is nevertheless the type of assumption on the asset-price dynamics that we wish to avoid in the first place by using a nonparametric estimator.

Our third proposed approach to dimension reduction is *semiparametric*. Suppose that the call-pricing function is given by the parametric Black-Scholes formula (2) except that the implied volatility parameter for that option is a nonparametric function $\sigma(F_{t,\tau}, X, \tau)$:

$$H\left(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}\right) = H_{\text{BS}}\left(F_{t,\tau}, X, \tau, r_{t,\tau}; \sigma(F_{t,\tau}, X, \tau)\right) \tag{9}$$

In this semiparametric model, we would only need to estimate nonparametrically the regression of σ on a subset $\tilde{\mathbf{Z}}$ of the vector of explanatory variables $\mathbf{Z}$. The rest of the call-pricing function $H(\cdot)$ is parametric, thereby considerably reducing the sample size n required to achieve the same degree of accuracy as the full nonparametric estimator. We partition the vector of explanatory variables $\mathbf{Z} \equiv [\ \tilde{\mathbf{Z}}'\ F_{t,\tau}\ r_{t,\tau}\ ]'$ where $\tilde{\mathbf{Z}}$ contains $\tilde{d}$ nonparametric regressors. As a result, the effective number of nonparametric regressors d is given by $\tilde{d}$. In our empirical application, we will consider both $\tilde{\mathbf{Z}} \equiv [\ X\ \ F_{t,\tau}\ \ \tau\ ]'$ ($\tilde{d} = 3$) and $\tilde{\mathbf{Z}} \equiv [\ X/F_{t,\tau}\ \ \tau\ ]'$ ($\tilde{d} = 2$, by combining this dimension-reduction technique with the previous one).

Given the data $\{H_i, S_{t_i}, X_i, \tau_i, r_{t_i,\tau_i}, \delta_{t_i,\tau_i}\}$, the full-dimensional SPD estimator takes the following form. We construct the fully nonparametric call-pricing function as:

$$\hat{H}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = \hat{E}[H | S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}] \tag{10}$$

using a multivariate kernel K in (8), formed as a product of five univariate kernels where subscripts refer to regressors (so for example, $k_\tau(\cdot)$ is the kernel function used for time-to-expiration as a regressor, and h_τ is the corresponding bandwidth value):

$$\hat{H}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = \frac{\sum_{i=1}^n k_S(\frac{S_t - S_{t_i}}{h_S}) k_X(\frac{X - X_i}{h_X}) \quad k_\tau(\frac{\tau - \tau_i}{h_\tau}) k_r(\frac{r_{t,\tau} - r_{t_i,\tau_i}}{h_r}) k_\delta(\frac{\delta_{t,\tau} - \delta_{t_i,\tau_i}}{h_\delta}) H_i}{\sum_{i=1}^n k_S(\frac{S_t - S_{t_i}}{h_S}) k_X(\frac{X - X_i}{h_X}) k_\tau(\frac{\tau - \tau_i}{h_\tau}) k_r(\frac{r_{t,\tau} - r_{t_i,\tau_i}}{h_r}) k_\delta(\frac{\delta_{t,\tau} - \delta_{t_i,\tau_i}}{h_\delta})} \; . \tag{11}$$

For the reduced-dimension SPD estimators, we substitute the appropriate list of regressors for $(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})$ in equation (11). For example, in the semiparametric model with $\tilde{\mathbf{Z}} \equiv [\; X \;\; F_{t,\tau} \;\; \tau \;]'$, we form the three-dimensional kernel estimator of $E[\sigma | F_{t,\tau}, X, \tau]$ as:

$$\hat{\sigma}(F_{t,\tau}, X, \tau) = \frac{\sum_{i=1}^n k_F(\frac{F_{t,\tau} - F_{t_i,\tau_i}}{h_S}) k_X(\frac{X - X_i}{h_X}) k_\tau(\frac{\tau - \tau_i}{h_\tau}) \sigma_i}{\sum_{i=1}^n k_F(\frac{F_{t,\tau} - F_{t_i,\tau_i}}{h_S}) k_X(\frac{X - X_i}{h_X}) k_\tau(\frac{\tau - \tau_i}{h_\tau})} \tag{12}$$

where σ_i is the volatility implied by the price H_i. We then estimate the call-pricing function as:

$$\hat{H}\left(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}\right) = H_{\mathrm{BS}}\left(F_{t,\tau}, X, \tau, r_{t,\tau}, \delta_{t,\tau}; \hat{\sigma}(F_{t,\tau}, X, \tau)\right). \tag{13}$$

In what follows, we will refer to any one of the above estimators $\hat{H}\left(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}\right)$ as nonparametric estimators and specify in each case the appropriate vector $\tilde{\mathbf{Z}}$. The option's delta and the SPD estimators follow by taking the appropriate partial derivatives of $\hat{H}(\cdot)$:

$$\hat{\Delta}_t = \frac{\partial \hat{H}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial S_t} \tag{14}$$

$$\hat{f}_t^*(S_T) = e^{r_{t,\tau}\tau} \left[\frac{\partial^2 \hat{H}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial X^2} \right]_{|X = S_T} . \tag{15}$$

D Comparisons with Other Approaches

Several other approaches to fit derivative prices have been proposed in the recent literature, hence a comparison of their strengths and weaknesses to those of the nonparametric kernel estimator is appropriate before turning to the Monte Carlo analysis and empirical application of Sections II and III.

D.1 Learning Networks

Hutchinson, Lo, and Poggio (1994) apply several nonparametric techniques to estimate option-pricing models that they describe collectively as *learning networks*: artificial neural networks, radial basis functions, and projection pursuit. Although they show that these techniques can recover option-pricing models such as the Black-Scholes/Merton model, they do not consider extracting the SPD from their nonparametric estimators.

Of course, SPDs can be extracted from many nonparametric option-pricing estimators—including learning-network models—simply by taking the second derivative of the nonparametric estimator with respect to the strike price (see Section II.A). However, the second derivative of a nonparametric estimator need not be a good estimator for the second derivative of the function to be estimated. This concern is particularly relevant for estimating nonlinear functions that are not smooth, i.e., infinitely differentiable, or where the degree of smoothness is unknown. Since the focus of our paper is the SPD and not the option-pricing function, we have constructed our estimator to reflect this focus. For example, the Appendix outlines our choices for the kernel function, the order of the kernel, and the bandwidth parameter, each choice determined to some degree by our interest in the second derivative of the estimator. Therefore, the properties of our estimator are likely to be quite different from and superior to the second derivatives of other nonparametric option-pricing estimators.

Also, Hutchinson, Lo, and Poggio (1994) do not provide any formal statistical inference to gauge the accuracy of their estimators. This is a problem endemic to learning networks and other recursive estimators and is extremely difficult to resolve as White (1992) has shown. In contrast, nonparametric regression and other non-recursive smoothing methods are better suited to hypothesis testing and other standard types of statistical inference. If such inferences are important in empirical applications, our estimator is preferable. In particular, we derive both pointwise and global confidence intervals for our estimator in Appendix II and III respectively, as well as stability tests in Appendix IV.

D.2 Implied Binomial Trees

Perhaps the closest alternative to our approach is Rubinstein's (1994) *implied binomial tree*, in which the risk-neutral probabilities $\{\pi_i^*\}$ associated with the binomial terminal stock price S_T are estimated by minimizing the sum of squared deviations between $\{\pi_i^*\}$ and a set of *prior* risk-neutral probabilities $\{\tilde{\pi}_i^*\}$, subject to the restrictions that $\{\pi_i^*\}$ correctly price an existing set of options and the underlying stock, in the sense that the optimal risk-neutral probabilities yield prices that lie within the bid-ask spreads of the options and the stock (see also Jackwerth and Rubinstein (1996) for smoothness criteria). This approach is similar in spirit to Jarrow and Rudd (1982) and Longstaff's (1995) method of fitting risk-neutral density functions using a four-parameter Edgeworth expansion (however, Rubinstein (1994) points out several important limitations of Longstaff's method when extended to a binomial model, including the possibility of negative probabilities. See also Derman and Kani (1994) and Shimko (1993)).

Although Rubinstein (1994) derives his risk-neutral probabilities within a binomial tree model—a parametric family—his approach can be interpreted as nonparametric in the sense that virtually any SPD can be approximated to any degree of accuracy by a binomial tree. But there are two main differences between his approach and ours: (1) he requires a prior $\{\tilde{\pi}_i^*\}$ for the risk-neutral probabilities; and (2) he fits one set of risk-neutral probabilities $\{\pi_i^*\}$ for *each* cross-section of options, whereas kernel SPDs aggregate option data over time to get a single SPD.

The first difference is not substantial since Rubinstein (1994) shows that the prior has progressively less influence on the implied binomial tree as the number of options used to constrain the probabilities increases. However, the second difference is significant. In particular, an n-step implied binomial tree calculated at date t_1 will, in general, be different from an n-step implied binomial tree calculated at date t_2. In contrast, since the nonparametric kernel estimator is based on both cross-sectional and time-series option prices, the resulting n-step SPD is the same for t_1 and t_2.

This difference is the result of a difference in the modeling strategies of the two approaches. The implied binomial tree is an attempt to obtain the risk-neutral probabilities that come closest to correctly pricing the existing options at a single point in time. As such, it can and will change over time. The nonparametric kernel estimator is an attempt to estimate the risk-neutral probabilities as a fixed *function* of certain economic variables, e.g., current stock price, riskless rate, etc. If it is successful, the functional form of the estimated SPD should be relatively stable over time, though the risk-neutral probabilities for any particular event, say $S_{t+\tau} > X$, can change over time (in a specific way) if the economic variables on which the SPD is based change over time (we test this hypothesis below in Section III.F).

Both approaches have advantages and disadvantages. The implied binomial tree is completely consistent with all option prices at each date, but is not necessarily consistent across time. The nonparametric kernel SPD estimator is consistent across time, but there may be some dates for which the SPD fits the cross section of option prices poorly and other dates for which the SPD performs very well. Whether or not consistency over time is a useful property depends on how well the economic variables used in constructing the kernel SPD can account for time variation in risk-neutral probabilities. In addition, the kernel SPDs take advantage of the data temporally surrounding a given date. Tomorrow's and yesterday's SPDs contain information about today's SPD—this information is ignored by the implied binomial trees but not by kernel-estimated SPDs.

Implied binomial trees are less data-intensive—by construction, they require only one cross section of option prices whereas the kernel SPD requires many cross sections. However, kernel SPDs are smooth functions by construction and easily (and optimally) interpolate between strike prices, maturities, and other kinds of discreteness. As we show, it is also possible to conduct statistical inference with kernel-estimated SPDs, something not easily handled by implied binomial trees.

In summary, while the two approaches have similar objectives, they also have important differences in their implementation. We examine below in Section III.G how these methods perform on actual data, notably by comparing both their in-sample and out-of-sample forecast errors.

II Monte Carlo Analysis

To examine the practical performance of the nonparametric SPD estimator, we perform several Monte Carlo simulation experiments under the assumption that call-option prices are truly determined by the Black-Scholes formula. Our nonparametric approach should be able to approximate Black-Scholes prices, from which the SPD may be extracted according to Breeden and Litzenberger (1978). The nonparametric pricing formula and SPD may then be compared to the Black-Scholes formula and theoretical SPD, respectively, to gauge the accuracy of the nonparametric approach.

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 the Black-Scholes assumptions since this is the leading case from which most applications and extensions are derived.

A Calibrating the Simulations

Our empirical application involves S&P 500 index options, so we perform Monte Carlo simulation experiments to match the basic features of our dataset (see Section III and Table I for further details). We start by simulating one year, i.e., 252 days, of daily index prices generated by a geometric Brownian motion with constant drift and diffusion parameters that match the moments of the data. Specifically, the values of the initial index level, the interest rate, and the index return mean and standard deviation are fixed at 455, 3 percent, 7.95 percent, and 10.28 percent, respectively. Throughout this study we follow the common convention of reporting returns and their means and standard deviations at an annual frequency. For the purpose of calculating option prices, annual parameter values are converted to a daily frequency by assuming that a year consists of 252 trading days.

At the start of this one-year sample of simulated daily prices, we create call options with strike prices and times-to-maturity that follow the Chicago Board Options Exchange

(CBOE) conventions for introducing options to the market. As the index price changes from one simulated day to the next, existing options may expire and new options may be introduced, again according to CBOE conventions, with strike prices that bracket the index in five-point increments. Therefore, on any given simulated day, the number of options is an endogenous function of the prior sample path of the index price—there are approximately 80 different call options on any one simulation day, including both existing and new ones.

We then price these options by the Black-Scholes formula using the actual index volatility, and add a small white-noise term to those prices. By doing so, we seek to reflect the existence in real data of a bid-ask spread and other possible sources of error in the recorded prices.[12] Since by construction the option prices in the simulation satisfy the Black-Scholes formula on average, when we apply our nonparametric pricing function to the simulated data, we should be able to "recover" the Black-Scholes formula. By this, we mean that $\hat{H}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})$ should approximate the Black-Scholes formula (2) numerically, not necessarily algebraically—in practice, the functional form of $\hat{H}(\cdot)$ may be quite different from the Black-Scholes formula (2), but both expressions will produce similar prices over the range of input values in the data. The two objectives of our simulation experiments is to determine how close $\hat{H}$ is to (2), and how close the corresponding nonparametric SPD is to the theoretical SPD (4).

Specifically, we take the Black-Scholes prices in the simulated dataset and the option characteristics as the inputs $\{H_i, S_{t_i}, X_i, \tau_i, r_{t_i,\tau_i}, \delta_{t_i,\tau_i}\}$ of our procedure, and then compute the smooth nonparametric call-pricing function of Section I.B given by (13). We then construct the option delta estimator $\hat{\Delta}$ and SPD estimator $\hat{f}^*$ according to equations (14) and (15), respectively. This entire procedure is repeated 5,000 times, to provide an indication of the accuracy of our estimators.

B Accuracy of Prices, Deltas, and SPDs

To assess the performance of nonparametric option-pricing formula and its corresponding delta and SPD, we consider the percentage differences between the nonparametric option-pricing formula, delta, SPD, and their theoretical Black-Scholes counterparts, respectively. In Figure 2, the theoretical values for prices, deltas, and SPDs are plotted on the left, and the average differences between the theoretical values and the nonparametric ones, averaged over the 5,000 replications, are plotted on the right. The figures show that the nonparametric quantities are within one percent of their theoretical counterparts—the estimators are virtually free of any bias. The dispersion of the estimates across the simulation runs yields the sampling distribution of the estimator and allowed us to confirm the accuracy of the asymptotic distribution derived in Appendix II for the sample size relevant for our empirical study in Section III. These simulations also illustrate empirically the fact that higher-order derivatives—SPD relative to deltas, deltas relative to prices—are estimated at a slower rate of convergence, which we demonstrate theoretically in Appendix I.

III Estimating SPDs From S&P 500 Options Data

To assess the empirical relevance of our nonparametric option-pricing formula and the corresponding SPD estimator, we present an application to the pricing and hedging of S&P 500 index options using data obtained from the CBOE for the sample period from January 4, 1993 to December 31, 1993.

A The Data

Table I describes the main features of our data set. This sample contains a total of 16,923 pairs of call- and put-option prices—we take averages of bid- and ask-prices as our raw data. Observations with time-to-maturity less than one day, implied volatility greater than 70 percent, and price less than 1/8 are dropped, which yields a final sample of 14,431 observations and this is the starting point for our empirical analysis.

During 1993, the mean and standard deviation of continuously-compounded daily returns of the S&P 500 index are 7.95 percent and 10.28 percent, respectively. During this period short-term interest rates exhibit little variation: they range from 2.85 percent to 3.21 percent. The options in our sample vary considerably in price and terms—for example, the time-to-maturity varies from one day to 350 days, with a median of 66 days. Given the volatility and movement in the index during this period and CBOE rules for introducing new options to the market, our sample contains a fairly broad cross section of options.

S&P 500 Index Options (symbol: SPX) are among the most actively traded financial derivatives in the world. Average total daily volume during the sample period was 65,476 contracts. The minimum tick for series trading below 3 is 1/16 and for all other series 1/8. Strike price intervals are five points, and 25 for far months. The expiration months are the three near-term months followed by three additional months from the March quarterly cycle (March, June, September, December). The options are European, and the underlying asset is an index, the most likely case for which a lognormal assumption (with continuous dividend stream) can be justified. By the simple effect of diversification, jumps are less likely to occur in indices than in individual equities. In other words, this market is as close as one can get to satisfying the assumptions of the Black-Scholes model. This is, therefore, a particularly promising context to test our approach: how different is our estimated SPD from the Black-Scholes SPD?

Even though the options are European and do not have a wildcard feature, the raw data present three challenges that must be addressed. First, because in-the-money options are very infrequently traded relative to at- and out-of-the-money options, in-the-money option prices are notoriously unreliable. For example, the average daily volume for puts that are 20 points out-of-the-money is 2,767 contracts; in contrast, the volume for puts that are 20 points in-the-money is 14 contracts. This reflects the strong demand by portfolio managers for protective puts (a phenomenon which started in late 1987 for obvious reasons).

Second, it is difficult to observe the underlying index price at the exact times that the option prices are recorded. In particular there is no guarantee that the closing index value reported is recorded at the same time as the closing transaction for each option. S&P 500 index futures are traded on the Chicago Mercantile Exchange (CME), not the CBOE, and time-stamped reported quotes may not necessarily be perfectly synchronized across the two markets. Even slight mismatches can lead to economically significant but spurious pricing anomalies.

Third, the index typically pays a dividend and the future rate of dividend payment is difficult, if not impossible, to determine. Standard and Poor's does provide daily dividend payments on the S&P 500, but by nature these data are backward-looking, and there is no reason to assume that the actual dividends recorded ex-post correctly reflect the expected future dividends at the time the option is priced.

We propose to address these three problems by the following procedure. Since all option prices are recorded at the same time on each day, we require only one temporally-matched index price per day. To circumvent the unobservability of the dividend rate $\delta_{t,\tau}$, we infer the futures price $F_{t,\tau}$ for each maturity τ. By the spot-futures parity, $F_{t,\tau}$ and S_t are linked through:[13]

$$F_{t,\tau} = S_t e^{(r_{t,\tau} - \delta_{t,\tau})\tau} \tag{16}$$

To derive the implied futures, we use the put-call parity relation which must hold if arbitrage opportunities are to be avoided, independently of any parametric option-pricing model:[14]

$$H(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) + Xe^{-r_{t,\tau}\tau} = G(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) + F_{t,\tau}e^{-r_{t,\tau}\tau}. \tag{17}$$

where G denotes the put price. To infer the futures price $F_{t,\tau}$ from this expression, we require reliable call and put prices—prices of actively traded options—at the *same* strike price X and time-to-expiration τ. To obtain such reliable pairs, we must use calls and puts that are closest to at-the-money (recall that in-the-money options are illiquid relative to out-the-money ones, hence any matched pair that is not at-the-money would have one potentially unreliable price). The average daily volume for at-the-money calls and puts is 4,360 contracts and we are therefore very confident in both prices. On every day t, we do this for all available maturities τ to obtain for each maturity the implied futures price from put-call parity.

Given the derived futures price $F_{t,\tau}$, we then replace the prices of all illiquid options, i.e., in-the-money options, with the price implied by put-call parity at the relevant strike prices. Specifically, we replace the price of each in-the-money call option with $G(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) + F_{t,\tau}e^{-r_{t,\tau}\tau} - Xe^{-r_{t,\tau}\tau}$ where, by construction, the put with price $G(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})$ is out-of-the-money and therefore liquid. After this procedure, all the information contained in liquid put prices has been extracted and resides in corresponding call prices via put-call parity, therefore put prices may now be discarded without any loss of reliable information.

B A Nonparametric S&P 500 Index Option Pricing Formula

We use price data on every option traded during 1993, for a total of n = 14,431 options after applying the filters described in the previous section. We estimated the SPD using each of the dimension reduction techniques discussed in Section I.C. In the interest of saving space, we only report below the results for the estimator given by equation (13). Note from Table I that the recorded interest rate data exhibited little significant variation during 1993 and thus we could reasonably have excluded $r_{t,\tau}$ from the regressor list—that is, treating it as constant—given our sample. This would have further reduced the dimensionality of the regression. In the semiparametric approach, this turns out to be unnecessary since $r_{t,\tau}$ does not enter the semiparametric regression. Naturally a different time period where interest rates are more volatile would require that $r_{t,\tau}$ be kept in $\mathbf{Z}$ for the full nonparametric model. The kernel and bandwidth values are given in Table II.

The estimator $\hat{\sigma}(F_{t,\tau}, X, \tau)$ necessary in (13) generates a strong volatility smile with respect to moneyness (see Figure 3). We follow this market's convention of quoting (and hedging) the options in terms of the futures rather than the cash index and therefore define moneyness as the ratio of strike X to futures prices $F_{t,\tau}$. The implied volatility at a fixed maturity is a decreasing nonlinear function of moneyness. Note in particular that our estimated smile is strongly asymmetric, pointing to the anecdotal evidence that out-of-money *put* prices have been consistently bid up since the crash of 1987 by investment managers looking for protection against future downward index movements. In contrast, stochastic volatility models—the class of models most commonly used to generate smile effects—typically produce symmetric smiles (see Bates (1995) and Renault (1995)).

A further result of our multidimensional approach is the changing shape of the smile as time-to-maturity increases, i.e., the variation of $\hat{\sigma}(F_{t,\tau}, X, \tau)$ with τ. This is best seen in Figure 4. The one-month smile is the steepest. We find that the implied volatility curves are generally flatter for longer times-to-maturity, but we document a persistence in the smile over longer maturities that is not captured by existing stochastic volatility models. In a typical such model, mean reversion in stochastic volatility induces a rapid disappearance of the smile as the time horizon increases—unlike long-memory effects. Our results therefore suggest that modeling long-term memory in stochastic volatility, for instance along the lines of Harvey (1995), could be a promising empirical approach.

Note also that the curves for all maturities intersect at approximately the same level of moneyness (0.975). In other words, options with moneyness of 0.975 are priced at about the same volatility for all maturities. At-the-money options (moneyness equals one) have an implied volatility which increases slightly with maturity. The implied volatility of out-of-the-money puts (calls) decreases (increases) with maturity. This suggests that it may be misleading to focus, as is often the case in practice, on the term structure of *at-the-money* volatilities as a way of fitting the Black-Scholes model to the data for different maturities. Our nonparametric approach documents that the small differences in at-the-money implied volatilities across maturities, where all the curves are close together, understate the overall variation of implied volatilities over the full range of traded strikes.

We report in Table III the nonparametric prices and deltas (with respect to the futures) for a sample of calls and puts for maturities of four months, priced for a current futures price of 455. To compute deltas with respect to the stock, note that $\partial H/\partial S = (\partial H/\partial F)(\partial F/\partial S) = (\partial H/\partial F)e^{(r-\delta)\tau}$). We give the price of every option with a delta greater or equal to 0.05 in absolute value. Not surprisingly, compared to Black-Scholes prices, our prices are consistent with the features of the actual market prices. We also report the prices of the standardized butterflies over five-point strike spreads which, in light of the result of Breeden and Litzenberger (1978), yield a discrete approximation to the value of the SPD at that strike level.

C S&P 500 Index Option SPDs

In Figure 5 the nonparametric SPDs are overlaid with the corresponding Black-Scholes SPDs at the same maturities (the Black-Scholes log-normal SPDs are evaluated at the at-the-money implied volatility for that maturity). Figure 5 also reports the 95 percent confidence intervals around each estimated SPD. The confidence interval is constructed from the asymptotic distribution theory derived in Appendix II.

Figure 6 shows the estimated nonparametric density of the continuously compounded τ-period return, $u_\tau \equiv \ln(S_T/S_t)$, that is compatible with our nonparametric SPD estimate. We compute the density of u_τ by noting that:

$$\Pr\left(\ln\left(\frac{S_T}{S_t}\right) \leq u\right) = \Pr\left(S_T \leq S_t e^u\right) = \int_0^{S_t e^u} f_t^*(S_T) dS_T. \tag{18}$$

The density of continuously compounded τ-period returns equivalent to the SPD for prices is then:

$$\frac{\partial}{\partial u} \Pr\left(\ln\left(\frac{S_T}{S_t}\right) \leq u\right) = S_t e^u f_t^*(S_t e^u). \tag{19}$$

We compare this density to the Gaussian Black-Scholes density $\mathrm{N}((r_{t,\tau} - \delta_{t,\tau} - \sigma^2/2)\tau,\ \sigma^2\tau)$ for each maturity.

Not surprisingly, the differences in Figure 6 between the nonparametric and Black-Scholes continuously compounded return distributions are quite similar to those of the estimated SPDs for prices in Figure 5. However, computing the densities for returns allows us to illustrate the magnitude of the differences by plotting in Figure 7 the term structures of implied mean, standard deviation, skewness, and kurtosis of the SPD-generated return distributions along with their Black-Scholes counterparts. We define kurtosis as the value in excess of the standard Gaussian distribution. All the moments of the returns in Figure 7 are annualized. Denote by $U_\tau \equiv u_\tau/\tau$ the annualized continuously compounded τ-period return and observe that

$$\begin{aligned}
\mathrm{E}\left[U_\tau\right] &= (1/\tau)\ \mathrm{E}\left[u_\tau\right] \\
\mathrm{Var}\left[U_\tau\right] &= (1/\tau)\ \mathrm{Var}\left[u_\tau\right] \\
\mathrm{Skew}\left[U_\tau\right] &= \sqrt{\tau}\,\mathrm{Skew}\left[u_\tau\right] \\
\mathrm{Kurt}\left[U_\tau\right] &= \tau\,\mathrm{Kurt}\left[u_\tau\right].
\end{aligned}$$

Figure 7 highlights the differences between the nonparametric and Black-Scholes return distributions. Although the nonparametric return densities in Figure 6 have comparable mean and standard deviations to those obtained from the Black-Scholes formula (we estimate the Black-Scholes SPD at the actual at-the-money implied volatility) they exhibit considerably different skewness and kurtosis. Specifically, for all four maturities the nonparametric SPDs are negatively skewed, have fatter tails and the amount of skewness and kurtosis both increase with maturity. Table III quantifies the differences in the nonparametric and Black-Scholes SPDs for the four maturities that are the focus of our empirical application. The skewness and kurtosis of the 21-day nonparametric density is -0.1976 and 0.0748, respectively, becoming progressively more severe as the maturity date increases, reaching -0.5165 and 0.2907, respectively, at the 126-day horizon.

D An Application to Digital Options

These results point to important differences between the nonparametric SPD and the Black-Scholes SPD, which implies correspondingly important differences in the pricing implications of the two. To illustrate these differences, we compare the prices of digital call options under the nonparametric and Black-Scholes SPDs in the last two columns of Table IV.

Recall from Section I.A that a digital call option is a security which gives the holder the right to receive a fixed cash payment of one dollar if S_T exceeds a prespecified level X. Its price is directly related to the CDF F_t^* of the SPD in the simple manner given in (5). This represents a typical application of our estimator: we infer the SPD from the prices of highly liquid calls and puts, and then price an over-the-counter contract consistently with the liquid prices.

The nonparametric SPD yields a higher price for the digital calls at strikes of 415 and 420: 0.06 versus 0.04 and 0.08 versus 0.06 respectively. The prices are then similar prices at a strike of 425: 0.10. But as the strike price increases, the Black-Scholes digital price increases more rapidly than the nonparametric—the at-the-money nonparametric digital is priced at 0.39, whereas the corresponding Black-Scholes price is 0.46, an 18 percent difference. However, these differences decline steadily until the 480 strike when the two prices are 0.79 and 0.80 respectively.

The behavior of the digital prices is symptomatic of the differences between the nonparametric and Black-Scholes log-returns SPDs: the Black-Scholes SPD is Gaussian and obviously cannot accommodate fat tails or skewness, whereas the nonparametric SPD can and does. In other words, the differences in Table IV between the nonparametric and Black-Scholes prices are the translation at the level of prices (i.e., integrals of the SPDs) of the differences in SPDs, or their skewness and kurtosis, that were identified in Table III.

E Specification Tests: Parametric vs. Nonparametric SPDs

Figures 5 and 6 suggest that the SPDs we estimate are substantially different from the benchmark Black-Scholes case. The 95 percent confidence interval reported in Figure 5 is pointwise, and therefore does not provide a global answer to the question: could the nonparametric family of SPDs have been generated by the Black-Scholes model? To answer this question, we derive in Appendix III a test based on the integrated squared distance between the two densities, nonparametric and Black-Scholes, and derive the asymptotic distribution of the test statistic. Empirical results are in Table VI. We strongly reject the null hypothesis that the densities are equal. We show in Appendix III that, since the alternative model is nonparametric, it is not equivalent to test the Black-Scholes model at the level of prices, deltas or SPDs. The curse of differentiation that we discuss in Appendix I plays a role here as well, just as in Figure 2. We test each one of these three null hypotheses, and find that in each case the Black-Scholes model is rejected with p-values equal to 0.00.

F Is the SPD Stable over Time?

Our approach assumes that the SPD is a constant function of a vector of state variables $\mathbf{Z}$ over the time period that is used to estimate it. A natural question that arises is whether the data actually validates this hypothesis. We derive in Appendix IV a diagnostic test based on the integrated squared difference between the two SPDs estimated over two different time periods $f_l^*(\mathbf{Z})$, $l = 1, 2$. The intuition behind our test is simple: under the null, the expected value of $(f_1^*(\mathbf{Z}) - f_2^*(\mathbf{Z}))^2$ should be small, whereas it will be large if the SPD is not the same over the two subperiods. We compute the asymptotic distribution of the test statistic and apply the test to pairwise comparisons of the SPD estimated separately over each of the four quarters of the year 1993. We report the results in Table VII. The tests do not reject the null hypothesis for any pair of quarters: the p-value for the pairwise quarterly comparisons are 0.41, 0.42 and 0.27 (recall in interpreting the values of the standardized $\hat{S}\left(\hat{f}_1^*, \hat{f}_2^*\right)$ that the test is one-sided: we only reject when the test statistic is large and positive, i.e., when $\hat{f}_1^*$ is sufficiently far away from $\hat{f}_2^*$). These results therefore provide strong evidence in favor of the stability of the SPD function $f^*(\mathbf{Z})$ over subperiods of our sample.

G In-Sample and Out-of-Sample Forecasts

As we discussed in Section I.D, one of the potential advantages of our approach to estimating SPDs lies in its ability to use a set of option prices with adjacent option characteristics, i.e., close values of $\mathbf{Z}$, to estimate the SPD, resulting in smooth and potentially more stable estimates. Relative to methods that rely exclusively on the current cross-section of option prices to infer the SPD, we would therefore expect our estimator to result in lower out-of-sample forecasting errors. This would typically be achieved at the expense of deteriorating the in-sample fit. We will verify this intuition by comparing the in-sample and out-of-sample fits of four models: (1) Jackwerth and Rubinstein's (1996) (JR) method of extracting densities by minimizing a criterion function which penalizes for unsmoothness; [15] (2) an extension of Hutchinson, Lo and Poggio's (1994) (HLP) approach in which we estimate a single-hidden-layer feedforward neural network with a logistic activation function and differentiate it twice numerically; (3) the Black and Scholes (1973) (BS) model; and (4) our SPD estimator (AL).

We present in Table VIII two types of evidence that confirm empirically the arguments made in Section I.D. In Panel A, we examine the ability of the SPDs—estimated on day t for the traded maturity closest to six months—to predict the SPD for the same six-month maturity on day $t + \tau$, for $\tau = 0$ (in-sample), and $\tau = 1, 5, 10, 15,$ and 20 trading days (out-of-sample) corresponding to forecast dates that are getting progressively farther away. We use the options data from the first three quarters of the year to construct the AL and HLP estimators on day t; the corresponding information on day t to construct the JR density;

and the implied volatility from the at-the-money option on day t to estimate the BS density. When calculating the densities, we annualize the returns to adjust for the differences in the actual numbers of days-to-maturity of the six-month options on day t and day $t + \tau$.

We then examine the out-of-sample forecasting properties of these estimators by repeating the computations for each day t in the fourth quarter of the year, and rolling the estimation period to retain the most recent nine months. We calculate the square root of the mean-squared difference between the density predicted by each method for day $t+\tau$, based on the market information available on day t, and the density that is realized on day $t + \tau$. The realized density on day $t + \tau$ is represented by either the JR or the AL estimators (see the two parts of Panel A). For scaling purposes, we report these differences as a percentage of the comparison density value at its mode. In Panel B, we run the same comparison, but focusing this time on differences between the option prices predicted on day t by each model, obtained by integrating the call payoff function against the model's density, and the realized market prices on day $t + \tau$ (again adjusting for differences in days-to-maturity). We also examine how the simplest possible procedure—using the market price at date t to predict the market price at date $t + \tau$, i.e., treating the implied volatility smile as a martingale—would have performed during the same period (this is the last column of Panel B, labeled MKT). We forecast the prices of the 400, 425, 450 and 475 strike options, and report the forecast error as a percentage of the market price over the comparison interval.

Both panels lead to similar conclusions. Comparing densities, the JR densities produce a better fit in-sample and over short horizons (for τ up to 15 days), but only when the forecasted density is estimated by the JR method (the left part of Panel A). However the AL estimates are uniformly better over every horizon at predicting the future density, when it is measured by the AL method (the right part of Panel A), but also if it is measured by the JR density for $\tau = 20$ days (the left part of Panel A). The HLP estimator, designed to fit prices rather than their derivatives, produce comparatively larger errors, and how to optimize it for that purpose would be unclear in the absence of some clear asymptotic guidance. Comparing market prices, the JR density estimates lead to a better fit in-sample or over less than a week ($\tau = 0, 1, 5$ days) than the AL densities; simply using the market prices would have done better however (see the last column in Panel B). Over longer forecasting horizons, using the AL densities leads to better forecasts of the market prices on day $t+\tau$ than using either the JR densities ($\tau = 10, 15, 20$ days) or the market prices on day t (for $\tau = 15$ and 20 days).

In each scenario, the BS densities lead to large forecasting errors relative to the JR and AL methods. This suggests that the features of the implied volatility smile that these methods identify are sufficiently persistent to be of some value as we forecast the future SPDs and market prices. The speed of mean reversion of the implied volatility smile to its average pattern is sufficiently slow that over short horizons the most recent data remains the best predictor of the future data (a martingale approach). However, there is enough mean reversion for a method based on estimating a stable function over time (AL) to forecast better over a longer horizon than one that bases the forecast exclusively on the most recent day's data (JR).

IV Conclusion

In this paper, we propose a nonparametric technique for estimating state-price densities based on the relation between state-price densities and option prices. We also derive a number of statistical properties of the estimator, including pointwise asymptotic distributions, specification test statistics and a test for stability across subsamples. Although this approach is data intensive, generally requiring several thousand datapoints for a reasonable level of accuracy, it offers a promising alternative to standard parametric pricing models when parametric restrictions are violated.

This trade-off between parametric restrictions and data requirements lies at the heart of the nonparametric approach. While parametric formulae are surely preferable to nonparametric ones when the underlying asset's price dynamics are well-understood, this is rarely the case in practice. Since they do not rely on restrictive parametric assumptions such as lognormality or sample-path continuity, nonparametric alternatives are robust to the specification errors that plague parametric models. A nonparametric approach is particularly valuable in option-pricing applications since the typical parametric restrictions have been shown to fail, sometimes dramatically. For example, Figure 7 shows that the primary failure of the Black-Scholes model in pricing S&P 500 index options is its inability to account for the skewness and kurtosis apparent in the nonparametric estimates of the returns distribution, and even more so for longer-maturity options. Parametric extensions of the Black-Scholes model, whether they capture stochastic volatility, jumps, or are multi-parameter extensions of the basic formula, should focus on capturing these empirical facts. The amount of persistence in the smile is such that parametric models incorporating long-term memory in stochastic volatility may be the most promising.

Also, by their very nature nonparametric methods are adaptive, responding to structural shifts in the data-generating processes in ways that parametric models do not. And finally, they are flexible enough to encompass a wide range of derivative securities and fundamental asset-price dynamics, yet relatively simple and computationally efficient to implement.

Appendix

In this Appendix we describe in more detail our procedure for selecting the kernel functions and the bandwidths in our option-pricing context. We also discuss the rate of convergence of our kernel estimator of the call-pricing function to the true function, its partial derivative with respect to the asset price to the true option delta and its second derivative with respect to the strike price to the true SPD.

I Kernel and Bandwidth Selection

We have to choose both the univariate kernel function k and the bandwidth parameters in h for each regressor. We need to differentiate twice the right-hand-side of (8) with respect to the strike price X to obtain an estimate of $\partial^2 H/\partial X^2$, and once with respect to S_t to estimate the option delta. We therefore require that k_X (k_S) be at least twice (once) continuously-differentiable.

Four elements determine the choice of the kernel and bandwidth: the sample size n, i.e., the number of options used to construct the estimator $\hat{H}(\cdot)$, the total number d of regressors included in the nonparametric regression, the number p_j of existing continuous partial derivatives of the true option-pricing function $H(\cdot)$ with respect to the j-th regressor $\mathbf{Z}_j$ and finally the order m_j of the partial derivative with respect to the j-th regressor that we wish to estimate (with the convention that $m_j = 0$ when no partial differentiation is required with respect to $\mathbf{Z}_j$).

We naturally assume that the call-pricing function $H(\cdot)$ to be estimated is sufficiently smooth, i.e., $p_j \geq m_j$ for all $j = 1, \ldots, d$. In our problem, the highest derivative that we shall estimate is the second partial derivative of the call price with respect to the strike price. We assume that the function $H(\cdot)$ admits four continuous derivatives with respect to each of its regressors.[16]

Define the order s of a kernel function k as the even integer that satisfies the following relation: $\int z^l k(z)du = 0$ for $l = 1, \ldots, s-1$, and $\int |z|^s k(z)du < \infty$. For example, the popular Gaussian kernel

$$k_{(2)}(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{A1}$$

is of order $s = 2$, while the following kernel is of order $s = 4$:

$$k_4(z) = \frac{3}{\sqrt{8\pi}} \left(1 - \frac{z^2}{3}\right) e^{-z^2/2}. \tag{A2}$$

Using higher-order kernels has the effect of accelerating the speed of convergence of the estimator to the true function as the sample size increases, in a mean-squared sense. We set the order of the kernel for each regressor j to be $s_j = p_j - m_j$. When estimating the m_j-th partial derivative of $H(\cdot)$ with respect to the j-th regressor, we obtain a bias term of order $h_j^{p-m_j}$ and a variance term of order $n^{-1}h_j^{-(2m_j+d)}$. We are minimizing the mean-squared error of the estimator, which consists of the sum of the squared bias and variance terms. Since the bandwidth h_j goes to zero as the sample size n grows, the larger the order of the kernel, the lower the bias term (the curve fit improves) but the larger the variance term (the curve estimator becomes noisier). The choice of the bandwidth given in (A3) optimally balances these two effects. However higher-order kernels can be cumbersome to use in practice—they are no longer uniformly positive for example. Experience suggests that it is desirable to limit the choice to kernels of orders no larger than $s = 4$.

Therefore, to estimate the call-pricing function where $m_j = 0$ for each regressor we set in the full nonparametric model $k_X = k_S = k_\tau = k_r = k_\delta = k_{(4)}$, and in the semiparametric model (9), $k_X = k_F = k_\tau = k_{(4)}$. To estimate the option delta where $m_j = 0$ for each regressor except S_t for which $m_j = 1$: in the full nonparametric model, we set $k_X = k_\tau = k_r = k_\delta = k_{(4)}$ and $k_S = k_{(2)}$, and in the semiparametric model, $k_X = k_\tau = k_{(4)}$ and $k_F = k_{(2)}$. Finally, to estimate the SPD where $m_j = 0$ for each regressor except X for which $m_j = 2$: in the full nonparametric model, we set $k_S = k_\tau = k_r = k_\delta = k_{(4)}$ and $k_X = k_{(2)}$, and in the semiparametric model, $k_F = k_\tau = k_{(4)}$ and $k_X = k_{(2)}$.

Monte Carlo evidence in Section II shows that for the choices of kernel functions above, $\hat{H}(\cdot)$ is very accurate for the typically large sample size n considered in our empirical application. For each of the d regressors in $\mathbf{Z}$, we set the corresponding bandwidth parameter h_j according to the relation:

$$h_j = c_j s(\mathbf{Z}_j) n^{-1/(d+2p)} \tag{A3}$$

where $\sigma_j \equiv s(\mathbf{Z}_j)$ is the unconditional standard deviation of the regressor $\mathbf{Z}_j$, $j = 1, \ldots, d$. This bandwidth choice is such that our estimator $\hat{H}(\cdot)$ achieves the optimal rate of convergence in the mean-squared sense among all possible nonparametric estimators of $H(\cdot)$:

$$n^{(p-m)/(d+2p)} \tag{A4}$$

where $m \equiv \sum_{j=1}^{d} m_j$.

The parameter c_j depends on the choice of the kernel and the function to be estimated; it is typically of the order of one and small deviations from the exact value have no large effects. It can be selected by *cross-validation* (see Härdle (1990) for example), a technique which ensures that we minimize the mean-squared error of our estimator $\hat{H}(\cdot)$. Any choice of bandwidth going to zero at a strictly faster rate than given by (A3) will center the asymptotic distribution at zero (see Appendix II). In practice, we select $c_j = c_{j0}/\ln(n)$, with the constant c_{j0} selected by cross-validation. This results in a rate of convergence for the estimator that is slightly slower than the rate given in (A4), but in exchange we eliminate asymptotically the bias term.

Note from the rate (A4) that the lower the dimensionality d of the regression function, i.e., the smaller the number of regressors in $\mathbf{Z}$, the faster the convergence of the estimator $\hat{H}(\mathbf{Z}) = \hat{E}[H|\mathbf{Z}]$ and its derivatives to the true function. This is the technical motivation for our proposed approaches to dimension direction in Section I.C.

Note also from the rate (A4) that higher-order derivatives converge at a slower speeds, hence the SPD estimator (for which $m = 2$) converges slower than the delta estimator (for which $m = 1$), which in turn converges slower than the price estimator (for which $m = 0$). One additional order of differentiation slows down the rate of convergence by as much as two additional regressors, i.e., the "curse of differentiation" (the decrease in rate of convergence as m increases) is twice as damning as the "curse of dimensionality" (the decrease in rate of convergence as d increases). As a theoretical matter, we can still get arbitrarily close to the parametric rate of convergence $n^{1/2}$ if the call-pricing function has enough continuous derivatives, and we use a kernel of sufficiently high order $p - m$: as p increases, the rate of convergence $n^{(p-m)/(2p+d)}$ converges to the parametric rate $n^{1/2}$.

Suppose now that we have estimated the option-pricing function and extracted the SPD. One of the main practical reason for doing so is the ability to price other derivative securities consistently. If the payoff function of the derivative is smooth, i.e., $\psi(S_T)$ is twice continuously differentiable in S_T everywhere on the range of integration, and its derivatives are bounded at the boundary of that range, then the derivative price can be rewritten as:

$$\begin{aligned} e^{-r_{t,\tau}\tau}\int_0^\infty \psi(S_T)\hat{f}_t^*(S_T)dS_T &= \int_0^\infty \psi(S_T)\left[\frac{\partial^2 \hat{H}_t}{\partial X^2}\right]|_{X=S_T}\ dS_T \\ &= \int_0^\infty \frac{\partial^2\psi(S_T)}{\partial S_T^2}\hat{H}_t(S_T)dS_T \end{aligned}$$

where $\hat{H}_t(S_T) \equiv \hat{H}(\mathbf{Z})$. Integrating a smooth function against $\hat{H}(\cdot) = \hat{E}[H|\cdot]$ speeds up the convergence rate of the estimator relative to the pointwise rate for $\hat{H}(\mathbf{Z})$: $\hat{H}(\mathbf{Z})$ converges at speed $n^{p/(d+2p)}$, but its integral against a smooth function of S_T converges at the faster rate $n^{p/((d-1)+2p)}$. We would therefore gain a factor $n^{-p/\{(d+2p)(d-1+2p)\}}$ in terms of rate of convergence when integrating the second derivative of the payoff function against $\hat{H}$, i.e., when computing the price of another derivative security with smooth payoff function consistently with the pricing function $H(\cdot)$.

II Asymptotic Distributions

We collect option prices in the form of panel data, and assume stationarity of the data. Methods based on local averaging can be sensitive to departures from stationarity in the form of a unit root. In such a case, the same region of the support of the distribution fails to revisited by the sample path often enough. One way to impose stationarity is to derive the asymptotics with respect to the cross-sectional dimension, holding the time series dimension as fixed. We obtain the following from the general results in Aït-Sahalia (1995), where precise regularity conditions are stated:

For the full nonparametric model in which $\mathbf{Z} \equiv [\ X\ S_t\ \tau\ r_{t,\tau}\ \delta_{t,\tau}\ ]'$, we have:

$$n^{1/2} h_1^{1/2} \prod_{j=2}^{d} h_j^{1/2} \left[\hat{H}(\mathbf{Z}) - H(\mathbf{Z})\right] \xrightarrow{d} \mathcal{N}(0, \sigma_H^2) \tag{A5}$$

$$n^{1/2} h_1^{3/2} \prod_{j=2}^{d} h_j^{1/2} \left[\hat{\Delta}(\mathbf{Z}) - \Delta(\mathbf{Z})\right] \xrightarrow{d} \mathcal{N}(0, \sigma_\Delta^2) \tag{A6}$$

$$n^{1/2} h_1^{5/2} \prod_{j=2}^{d} h_j^{1/2} \left[\hat{f}_t^*(\ \mathbf{Z}) - f_t^*(\mathbf{Z})\right] \xrightarrow{d} \mathcal{N}(0, \sigma_{f^*}^2) \tag{A7}$$

where

$$\begin{aligned}
\sigma_H^2 &\equiv \left(\int_{-\infty}^{+\infty} k_{(4)}^2(\omega) d\omega\right)^d s^2(\mathbf{Z})\ /\pi(\mathbf{Z}_i) \\
\sigma_\Delta^2 &\equiv \left(\int_{-\infty}^{+\infty} {k'}_{(2)}^2(\omega) d\omega\right) \left(\int_{-\infty}^{+\infty} k_{(4)}^2(\omega) d\omega\right)^{d-1} s^2(\mathbf{Z})\ /\pi(\mathbf{Z}) \\
\sigma_{f^*}^2 &\equiv \left(\int_{-\infty}^{+\infty} {k''}_{(2)}^2(\omega) d\omega\right) \left(\int_{-\infty}^{+\infty} k_{(4)}^2(\omega) d\omega\right) s^2(\mathbf{Z})\ /\pi(\mathbf{Z}),
\end{aligned}$$

$\pi(\mathbf{Z})$ is the joint density function of the vector $\mathbf{Z}$, i.e., the marginal distribution of option characteristics, and $s^2(\mathbf{Z})$ the conditional variance of the nonparametric regression. The bandwidths are chosen as discussed in Appendix I.

In the semiparametric model, we partition the vector of explanatory variables $\mathbf{Z} \equiv [\ \tilde{\mathbf{Z}}'\ F_{t,\tau}\ r_{t,\tau}\ ]'$ where $\tilde{\mathbf{Z}}$ contains $\tilde{d}$ nonparametric regressors. In our specific case, we consider both $\tilde{\mathbf{Z}} \equiv [\ X\ F_{t,\tau}\ \tau\ ]'$ ($\tilde{d} = 3$) and $\tilde{\mathbf{Z}} \equiv [\ X/F_{t,\tau}\ \tau\ ]'$ ($\tilde{d} = 2$), and we estimate the implied volatility function $\sigma(\tilde{\mathbf{Z}})$ nonparametrically. This yields the following asymptotic distributions:

$$\begin{aligned}
&n^{1/2} h_1^{1/2} \prod_{j=2}^{\tilde{d}} h_j^{1/2} \left[\hat{\sigma}(\tilde{\mathbf{Z}}) - \sigma(\tilde{\mathbf{Z}})\right] \xrightarrow{d} \mathcal{N}(0, \sigma_\sigma^2) \\
&n^{1/2} h_1^{3/2} \prod_{j=2}^{\tilde{d}} h_j^{1/2} \left[\frac{\partial \hat{\sigma}}{\partial X}(\tilde{\mathbf{Z}}) - \frac{\partial \sigma}{\partial X}(\tilde{\mathbf{Z}})\right] \xrightarrow{d} \mathcal{N}(0, \sigma_{d\sigma}^2) \\
&n^{1/2} h_1^{5/2} \prod_{j=2}^{\tilde{d}} h_j^{1/2} \left[\frac{\partial^2 \hat{\sigma}}{\partial X^2}(\tilde{\mathbf{Z}}) - \frac{\partial^2 \sigma}{\partial X^2}(\tilde{\mathbf{Z}})\right] \xrightarrow{d} \mathcal{N}(0, \sigma_{d^2\sigma}^2)
\end{aligned} \tag{A8}$$

where

$$\sigma_\sigma^2 \equiv \left(\int_{-\infty}^{+\infty} k_{(4)}^2(\omega) d\omega\right)^{\tilde{d}} s^2(\tilde{\mathbf{Z}})\ /\pi(\tilde{\mathbf{Z}})$$

$$
\begin{aligned}
\sigma^2_{d\sigma} &\equiv E[\tilde{\epsilon}^2|\tilde{\mathbf{Z}}] \left(\int_{-\infty}^{+\infty} k'^2_{(2)}(\omega)d\omega\right) \left(\int_{-\infty}^{+\infty} k^2_{(4)}(\omega)d\omega\right)^{\tilde{d}-1} s^2(\tilde{\mathbf{Z}})\Big/\pi(\tilde{\mathbf{Z}}) \\
\sigma^2_{d^2\sigma} &\equiv E[\tilde{\epsilon}^2|\tilde{\mathbf{Z}}] \left(\int_{-\infty}^{+\infty} k''^2_{(2)}(\omega)d\omega\right) \left(\int_{-\infty}^{+\infty} k^2_{(4)}(\omega)d\omega\right)^{\tilde{d}-1} s^2(\tilde{\mathbf{Z}})\Big/\pi(\tilde{\mathbf{Z}}),
\end{aligned}
$$

$\pi(\tilde{\mathbf{Z}})$ is the joint density function of the vector $\tilde{\mathbf{Z}}$, and $s^2(\tilde{\mathbf{Z}})$ the conditional variance of the nonparametric regression $\sigma(\tilde{\mathbf{Z}})$. In Table V we provide the values of the kernel integrals that appear in the expressions above. The asymptotic distributions of the semiparametric estimators for prices, deltas, and SPDs then follow from the delta method, i.e.,

$$
\begin{aligned}
\hat{H}(\mathbf{Z}) &= H_{\mathrm{BS}}(\hat{\sigma}(\tilde{\mathbf{Z}}), \mathbf{Z}) \\
&= H_{\mathrm{BS}}(\sigma(\tilde{\mathbf{Z}}), \mathbf{Z}) + \frac{\partial H_{\mathrm{BS}}}{\partial \sigma}(\sigma(\tilde{\mathbf{Z}}), \mathbf{Z}) \left[\hat{\sigma}(\tilde{\mathbf{Z}}) - \sigma(\tilde{\mathbf{Z}})\right] \\
&\quad + \mathbf{O}_p\left(\|\hat{\sigma}(\tilde{\mathbf{Z}}) - \sigma(\tilde{\mathbf{Z}})\|^2_{L^2}\right).
\end{aligned}
$$

so $H_{\mathrm{BS}}(\hat{\sigma}(\tilde{\mathbf{Z}}), \mathbf{Z}) - H_{\mathrm{BS}}(\sigma(\tilde{\mathbf{Z}}), \mathbf{Z})$ behaves asymptotically as $\partial H_{\mathrm{BS}}/\partial\sigma \ \left(\hat{\sigma}(\tilde{\mathbf{Z}}) - \sigma(\tilde{\mathbf{Z}})\right)$; and because derivatives of $\hat{\sigma}(\tilde{\mathbf{Z}})$ converge at a slower rate (see the rates of convergence (A8)) the asymptotic distribution of $\partial H_{\mathrm{BS}}/\partial F(\hat{\sigma}(\tilde{\mathbf{Z}}), \mathbf{Z}) - \partial H_{\mathrm{BS}}/\partial F(\sigma(\tilde{\mathbf{Z}}), \mathbf{Z})$ is that of $\partial H_{\mathrm{BS}}/\partial\sigma \ (\partial\hat{\sigma}/\partial F - \partial\sigma/\partial F)$, and the asymptotic distribution of $\partial^2 H_{\mathrm{BS}}/\partial X^2(\hat{\sigma}(\tilde{\mathbf{Z}}), \mathbf{Z}) - \partial^2 H_{\mathrm{BS}}/\partial X^2(\sigma(\tilde{\mathbf{Z}}), \mathbf{Z})$ is that of $\partial H_{\mathrm{BS}}/\partial\sigma \ (\partial^2\hat{\sigma}/\partial K^2 - \partial^2\sigma/\partial K^2)$.

III Specification Tests

We propose to test the null hypothesis H_0

$$
\Pr\left(H(X, S, \tau, r, \delta) = H_{\mathrm{BS}}(X, S, \tau, r, \delta)\right) = 1
$$

against the alternative hypothesis H_A

$$
\Pr\left(H(X, S, \tau, r, \delta) = H_{\mathrm{BS}}(X, S, \tau, r, \delta)\right) < 1.
$$

A natural test statistic is the sum of squared deviations between the two option-pricing formulas:

$$
D(H, H_{\mathrm{BS}}) \equiv E\left[(H(\mathbf{Z}) - H_{\mathrm{BS}}(\mathbf{Z}))^2 \omega_H(\mathbf{Z})\right] \tag{A9}
$$

where $\omega_H(\mathbf{Z})$ is a weighting function. The intuition behind our proposed test is simple: if the null Black-Scholes model is correctly specified, then its call-pricing formula should be close to the formula estimated nonparametrically. An estimator for $D(H, H_{\mathrm{BS}})$ is either the sample analog

$$
\hat{D}(\hat{H}, \hat{H}_{\mathrm{BS}}) \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\hat{H}(\mathbf{Z}_i) - \hat{H}_{\mathrm{BS}}(\mathbf{Z}_i)\right)^2 \omega_H(\mathbf{Z}_i), \tag{A10}
$$

or any other evaluation of the integral on the right-hand-side of (A9). In practice, we evaluate numerically the integral on a rectangle of values of $\mathbf{Z}$ representing the support of the density π, and use the binning method to evaluate the kernels (see e.g., Wand and Jones (1995) for a description of the binning method).

To test the hypotheses that $\Delta(\mathbf{Z}) = \Delta_{\text{BS}}(\mathbf{Z})$ and $f^*(\mathbf{Z}) = f^*_{\text{BS}}(\mathbf{Z})$, we define $D(\Delta, \Delta_{\text{BS}})$ and $D(f^*, f^*_{\text{BS}})$ in a similar fashion. To form a test, we derive the asymptotic distributions of these test statistics. Under the null hypothesis of the Black-Scholes model H_0, we can show that:

$$nh_1^{1/2} \prod_{j=2}^{d} h_j^{1/2} \hat{D}(\hat{H}, \hat{H}_{\text{BS}}) - \prod_{j=1}^{d} h_j^{-1/2} B_H \quad \xrightarrow{d} \mathcal{N}(0, \mathbf{\Sigma}_H^2) \tag{A11}$$

$$nh_1^{5/2} \prod_{j=2}^{d} h_j^{1/2} \hat{D}(\hat{\Delta}, \hat{\Delta}_{\text{BS}}) - \prod_{j=1}^{d} h_j^{-1/2} B_\Delta \quad \xrightarrow{d} \mathcal{N}(0, \mathbf{\Sigma}_\Delta^2) \tag{A12}$$

$$nh_1^{9/2} \prod_{j=2}^{d} h_j^{1/2} \hat{D}(\hat{f}^*, \hat{f}^*_{\text{BS}}) - \prod_{j=1}^{d} h_j^{-1/2} B_{f^*} \quad \xrightarrow{d} \mathcal{N}(0, \mathbf{\Sigma}_{f^*}^2) \tag{A13}$$

where d is the number of included nonparametric regressors (the dimension of $\mathbf{Z}$). We order the regressors in $\mathbf{Z}$ so that the first one is the variable with respect to which the call-pricing function is to be differentiated m times. For simplicity, we use a common bandwidth for the variables $j = 2, \ldots, d$ (if these regressors have different scales, they should first be standardized). The bandwidths h_j, $j = 1, \ldots, d$ are given in each case $m = 0, 1, 2$ corresponding respectively to prices, deltas and SPDs, by:

$$h_1 = \mathbf{O}(n^{-1/\delta_1}), \qquad h_j = \mathbf{O}(n^{-1/\delta_2}), \quad j = 2, \ldots, d \tag{A14}$$

with $\delta_1 = \delta_2 + 2m$ and δ_2 satisfying:

$$\left(\frac{1+m}{\delta_2 + 2m}\right) + \left(\frac{d-1}{\delta_2}\right) < \frac{1}{2} < \left(\frac{1/2 + 2m + r}{\delta_2 + 2m}\right) + \left(\frac{(d-1)/2}{\delta_2}\right) \tag{A15}$$

and

$$\left(\frac{3/2}{\delta_2 + 2m}\right) > \left(\frac{(d-1)/2}{\delta_2}\right) \tag{A16}$$

where we use a common kernel k of order r on every variable. Equation (A16) ensures that derivatives of lower order do not affect the distribution of the highest-order derivative term. Note that these bandwidth selection inequalities result in bandwidth values that under-smooth relative to the values in Appendix I. The bias and variance terms are given by:

$$B_H = \left(\int_w k^2(w)dw\right) \int_{\mathbf{Z}} s^2(\mathbf{Z}) \tilde{\omega}_H(\mathbf{Z}) d\mathbf{Z}$$

$$
\begin{aligned}
B_{\Delta} &= \left(\int_w k'^2(w)dw\right)\left(\int_w k^2(w)dw\right)^{d-1}\int_{\mathbf{Z}} s^2(\mathbf{Z})\tilde{\omega}_{\Delta}(\mathbf{Z})d\mathbf{Z} \\
B_{f^*} &= \left(\int_w k''^2(w)dw\right)\left(\int_w k^2(w)dw\right)^{d-1}\int_{\mathbf{Z}} s^2(\mathbf{Z})\tilde{\omega}_{f^*}(\mathbf{Z})d\mathbf{Z}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{\Sigma}_H^2 &= 2\left[\int_v\left(\int_w k(w)k(w+v)dw\right)^2 dv\right]^d \int_{\mathbf{Z}} s^4(\mathbf{Z})\tilde{\omega}_H^2(\mathbf{Z})d\mathbf{Z} \\
\mathbf{\Sigma}_{\Delta}^2 &= 2\left[\int_v\left(\int_w k'(w)k'(w+v)dw\right)^2 dv\right]\left[\int_v\left(\int_w k(w)k(w+v)dw\right)^2 dv\right]^{d-1}\times \\
&\qquad \int_{\mathbf{Z}} s^4(\mathbf{Z})\tilde{\omega}_{\Delta}^2(\mathbf{Z})d\mathbf{Z} \\
\mathbf{\Sigma}_{f^*}^2 &= 2\left[\int_v\left(\int_w k''(w)k''(w+v)dw\right)^2 dv\right]\left[\int_v\left(\int_w k(w)k(w+v)dw\right)^2 dv\right]^{d-1}\times \\
&\qquad \int_{\mathbf{Z}} s^4(\mathbf{Z})\tilde{\omega}_{f^*}^2(\mathbf{Z})d\mathbf{Z}.
\end{aligned}
$$

where $\tilde{\omega} = \omega$ for the full nonparametric model. As for the pointwise asymptotic distributions, we report in Table V the values of the kernel integrals that appear in the expressions above. To estimate consistently the conditional variance of the regression, $s^2(\mathbf{Z}) = E\left[(y - M(\mathbf{Z}))^2 |\mathbf{Z}\right] = E[y^2|\mathbf{Z}] - E[y|\mathbf{Z}]^2$, where y is the dependent variable and $M(\mathbf{Z}) \equiv E[y|\mathbf{Z}]$, we calculate the difference between the kernel estimator of the regression of the squared dependent variable y^2 on $\mathbf{Z}$ and the squared of the regression $M(\mathbf{Z})$. The two regressions appearing in $s^2(\mathbf{Z})$ are estimated with bandwidth $h_{cv} = \mathbf{O}(n^{-1/\delta_{cv}}/\ln(n))$ where $\delta_{cv} = 2r + d$ (recall that we only need a consistent estimator of $s^2(\mathbf{Z})$, so there is no need to under-smooth as in (A15)-(A16)).

The test statistics are formed by standardizing the asymptotically normal distance measures $\hat{D}$: we estimate consistently the asymptotic mean B and variance $\mathbf{\Sigma}^2$, then subtract the mean and divide by the standard deviation. The test statistic then has an asymptotic $N(0,1)$ distribution. Since the test is one-sided (we only reject when $\hat{D}$ is too large, hence when the test statistic is large and positive), the 10 percent critical value is 1.28, while the 5 percent value is 1.64.

For the semiparametric case, we partition $\mathbf{Z}$ into the nonparametric regressors $\tilde{\mathbf{Z}}$ and the remaining variables, hence we seek to test $H(\mathbf{Z}) = H_{\text{BS}}(\sigma(\tilde{\mathbf{Z}}), \mathbf{Z})$. The test statistic for prices is given by:

$$
D(H, H_{\text{BS}}) \equiv E\left[(H_{\text{BS}}(\sigma(\tilde{\mathbf{Z}}), \mathbf{Z}) - H_{\text{BS}}(\sigma_{\text{BS}}, \mathbf{Z}))^2 \omega_H(\mathbf{Z})\right]
$$

which behaves asymptotically like

$$
E\left[\left(\frac{\partial H_{\text{BS}}(\sigma_{\text{BS}}, \mathbf{Z})}{\partial \sigma}\right)^2 \left(\sigma(\tilde{\mathbf{Z}}) - \sigma_{\text{BS}}\right)^2\right].
$$

For deltas and SPDs, the corresponding test statistics test the specification of the first and second partial derivatives, respectively, of $\sigma(\tilde{\mathbf{Z}})$ with respect to X, which is contained in the first argument of $\tilde{\mathbf{Z}}$. They also behave asymptotically like their leading terms, which are

$$E\left[\left(\frac{\partial H_{\mathrm{BS}}(\sigma_{\mathrm{BS}},\mathbf{Z})}{\partial\sigma}\right)^2\left(\frac{\partial\sigma(\tilde{\mathbf{Z}})}{\partial\tilde{Z}_1}\frac{\partial\tilde{Z}_1}{\partial F}\right)^2\right],\quad E\left[\left(e^{r_{t,\tau}\tau}\frac{\partial H_{\mathrm{BS}}(\sigma_{\mathrm{BS}},\mathbf{Z})}{\partial\sigma}\right)^2\left(\frac{\partial^2\sigma(\tilde{\mathbf{Z}})}{\partial\tilde{Z}_1^2}\left(\frac{\partial\tilde{Z}_1}{\partial X}\right)^2\right)^2\right]$$

where $\tilde{Z}_1$ is the first component of the vector $\tilde{\mathbf{Z}}$. Therefore, in the formulas above, we set:

$$\tilde{\omega}_H(\mathbf{Z})\equiv\left(\frac{\partial H_{\mathrm{BS}}(\sigma_{\mathrm{BS}},\mathbf{Z})}{\partial\sigma}\right)^2\omega_H(\mathbf{Z}) \tag{A17}$$

for prices and

$$\tilde{\omega}_\Delta(\mathbf{Z})\equiv\left(\frac{\partial H_{\mathrm{BS}}(\sigma_{\mathrm{BS}},\mathbf{Z})}{\partial\sigma}\frac{\partial\tilde{Z}_1}{\partial F}\right)^2\omega_\Delta(\mathbf{Z}),\quad \tilde{\omega}_{f^*}(\mathbf{Z})\equiv\left(e^{r_{t,\tau}\tau}\frac{\partial H_{\mathrm{BS}}(\sigma_{\mathrm{BS}},\mathbf{Z})}{\partial\sigma}\left(\frac{\partial\tilde{Z}_1}{\partial X}\right)^2\right)^2\omega_{f^*}(\mathbf{Z}) \tag{A18}$$

for deltas and SPDs. To construct the test, we fix the variables in $\mathbf{Z}$ that are excluded from $\tilde{\mathbf{Z}}$ at their sample means and let $\tilde{\mathbf{Z}}\equiv[X/F_{t,\tau}\ \tau]'$. Thus $M(\mathbf{Z})$ above is $\sigma(X/F_{t,\tau},\tau)$, $\partial\tilde{Z}_1\,/\,\partial F = -X/F^2$ and $\partial\tilde{Z}_1\,/\,\partial X = 1/F$. Then we can apply the results above with $\tilde{d}=2$ instead of the full value of d, $p=4$, and $m=0,1,2$ for prices, deltas, and SPDs, respectively. Values of δ_j, $j=1,2$, corresponding bandwidth parameters and test results can be found in Table VI.

IV Testing the Stability of the SPD Across Time Periods

We now construct a test of the null hypothesis that the pricing function, delta, and SPD in one subsample are the same in another subsample, hence the null hypothesis is:

$$\mathrm{H}_0:\Pr\left(H_1(\mathbf{Z})=H_2(\mathbf{Z})\right)=1$$

and the alternative hypothesis is:

$$\mathrm{H}_A:\Pr\left(H_1(\mathbf{Z})=H_2(\mathbf{Z})\right)<1$$

under the maintained assumption that the marginal distribution of option characteristics, $\pi(\mathbf{Z})$, is identical over the two subsamples.

Let n be the sample size of the two subsamples (assumed to be the same across subsamples for simplicity). In what follows, the subscript $l=1$ or 2 denotes the subsample that was used to estimate the function, so H_1 denotes the call-pricing function estimated on subsample 1, etc. Relying on the same intuition as in Appendix III, we form the (suitably normalized) sum of squared deviations between the two option-pricing formulas estimated on the two subsamples:

$$S(H_1,H_2)\equiv E\left[(H_1(\mathbf{Z})-H_2(\mathbf{Z}))^2\omega_H(\mathbf{Z})\right]. \tag{A19}$$

The intuition behind our proposed test is straightforward: if the null Black-Scholes model is correctly specified, then its call-pricing formula should be close to the formula estimated nonparametrically.

To test the hypotheses that $\Delta_1(\mathbf{Z}) = \Delta_2(\mathbf{Z})$ and $f_1^*(\mathbf{Z}) = f_2^*(\mathbf{Z})$, we define $S(\Delta_1, \Delta_2)$ and $S(f_1^*, f_2^*)$ in a similar fashion. Under the null hypothesis of equality of the pricing function, delta, and SPD over the two subsamples, we can show that:

$$nh_1^{1/2} \prod_{j=2}^{d} h_j^{1/2} \hat{S}(\hat{H}_1, \hat{H}_2) - \prod_{j=1}^{d} h_j^{-1/2} C_H \quad \xrightarrow{d} \mathcal{N}(0, \mathbf{\Omega}_H^2) \tag{A20}$$

$$nh_1^{5/2} \prod_{j=2}^{d} h_j^{1/2} \hat{S}(\hat{\Delta}_1, \hat{\Delta}_2) - \prod_{j=1}^{d} h_j^{-1/2} C_\Delta \quad \xrightarrow{d} \mathcal{N}(0, \mathbf{\Omega}_\Delta^2) \tag{A21}$$

$$nh_1^{9/2} \prod_{j=2}^{d} h_j^{1/2} \hat{S}(\hat{f}_1^*, \hat{f}_2^*) - \prod_{j=1}^{d} h_j^{-1/2} C_{f^*} \quad \xrightarrow{d} \mathcal{N}(0, \mathbf{\Omega}_{f^*}^2) \tag{A22}$$

where the bandwidths h_j, $j = 1, \ldots, d$ (common to both subsamples) are given in each case $m = 0, 1, 2$—corresponding to prices, deltas, and SPDs respectively—by $h_1 = \mathbf{O}(n^{-1/\delta_1})$, $h_j = \mathbf{O}(n^{-1/\delta_2})$, $j = 2, \ldots, d$ with $\delta_1 = \delta_2 + 2m$ and δ_2 satisfying:

$$\left(\frac{1+m}{\delta_2 + 2m}\right) + \left(\frac{d-1}{\delta_2}\right) < \frac{1}{2} < \left(\frac{1/2 + 2m + 3r}{2\,(\delta_2 + 2m)}\right) + \left(\frac{(d-1)/2}{2\delta_2}\right) \tag{A23}$$

(note that the right-hand-side of this inequality is different from that of (A15)), and

$$\left(\frac{3/2}{\delta_2 + 2m}\right) > \left(\frac{(d-1)/2}{\delta_2}\right). \tag{A24}$$

We again use a common kernel k of order r for every variable. The bias and variance terms are given by:

$$\begin{aligned}
C_H &= \left(\int_w k^2(w)dw\right) \int_{\mathbf{Z}} \left\{s_1^2(\mathbf{Z}) + s_2^2(\mathbf{Z})\right\} \tilde{\omega}_H(\mathbf{Z})d\mathbf{Z} \\
C_\Delta &= \left(\int_w k'^2(w)dw\right) \left(\int_w k^2(w)dw\right)^{d-1} \int_{\mathbf{Z}} \left\{s_1^2(\mathbf{Z}) + s_2^2(\mathbf{Z})\right\} \tilde{\omega}_\Delta(\mathbf{Z})d\mathbf{Z} \\
C_{f^*} &= \left(\int_w k''^2(w)dw\right) \left(\int_w k^2(w)dw\right)^{d-1} \int_{\mathbf{Z}} \left\{s_1^2(\mathbf{Z}) + s_2^2(\mathbf{Z})\right\} \tilde{\omega}_{f^*}(\mathbf{Z})d\mathbf{Z}
\end{aligned}$$

$$\begin{aligned}
\mathbf{\Omega}_H^2 &= 2\left[\int_v \left(\int_w k(w)k(w+v)dw\right)^2 dv\right]^d \int_{\mathbf{Z}} \left\{s_1^2(\mathbf{Z}) + s_2^2(\mathbf{Z})\right\}^2 \tilde{\omega}_H^2(\mathbf{Z})d\mathbf{Z} \\
\mathbf{\Omega}_\Delta^2 &= 2\left[\int_v \left(\int_w k'(w)k'(w+v)dw\right)^2 dv\right] \left[\int_v \left(\int_w k(w)k(w+v)dw\right)^2 dv\right]^{d-1} \times \\
&\qquad \int_{\mathbf{Z}} \left\{s_1^2(\mathbf{Z}) + s_2^2(\mathbf{Z})\right\}^2 \tilde{\omega}_\Delta^2(\mathbf{Z})d\mathbf{Z}
\end{aligned}$$

$$
\begin{aligned}
\mathbf{\Omega}^2_{f^*} &= 2\left[\int_v \left(\int_w k''(w)k''(w+v)dw\right)^2 dv\right]\left[\int_v \left(\int_w k(w)k(w+v)dw\right)^2 dv\right]^{d-1} \times \\
&\qquad \int_{\mathbf{Z}} \left\{s_1^2(\mathbf{Z}) + s_2^2(\mathbf{Z})\right\}^2 \tilde{\omega}^2_{f^*}(\mathbf{Z})d\mathbf{Z}.
\end{aligned}
$$

where $\tilde{\omega} = \omega$ in the full nonparametric case. For the semiparametric case, we proceed as in Appendix III and the weighting functions are identical to those defined in (A17)–(A18). Note that the distributions depend on the two conditional variances in the two subsamples, $s_l^2(\mathbf{Z})$, $l = 1, 2$, which need not be equal. Bandwidth values and test results for pairwise quarterly comparisons in 1993 can be found in Table VII.

Footnotes

1. See for example Goldenberg (1991) for formulas in integral form for the case of diffusions other than geometric Brownian motion, Heston (1993) for an implicit characterization of the SPD in a stochastic volatility model, and Bates (1995) in a model with stochastic volatility and jumps.

2. See Jarrow and Rudd (1982), Shimko (1993), Longstaff (1995), and Madan and Milne (1994).

3. See Rubinstein (1994) and Jackwerth and Rubinstein (1996) who experiment with different distance criterions.

4. In several other contexts the potential benefits of nonparametric methods for asset-pricing applications have recently begun to be explored. Aït-Sahalia (1996a) constructs nonparametric estimators of the diffusion process followed by the underlying asset return, as a basis to price interest rate derivatives nonparametrically. Aït-Sahalia (1996b) tests parametric specifications of the spot interest rate process against a nonparametric alternative. The estimators all use discrete data, and require no discrete approximation even though the estimated model is in continuous-time. Hutchinson, Lo, and Poggio (1994) show how a neural network can approximate the Black-Scholes formula and other derivative-pricing models. Boudoukh et al. (1995) price mortgages while Stutzer (1996) estimates the empirical distribution of stock returns.

5. See, for example, the discussion in Campbell, Lo, and MacKinlay (1997, Chapter 2).

6. Of course, markets must be dynamically complete for such prices to be meaningful—see, for example, Constantinides (1982), and Duffie and Huang (1985). This assumption is almost always adopted, either explicitly or implicitly, in parametric derivative-pricing models, and we shall adopt it here as well.

7. This complements the approach in Aït-Sahalia (1996a) where a nonparametric estimator of the same volatility function was obtained from the time series of the underlying asset returns. We can also infer the volatility function of the underlying price process from derivative prices. Girsanov's Theorem implies that they should be the same, which is a testable implication of the no-arbitrage pricing paradigm.

8. Let $F_{t,\tau}$ denotes the value at t of a futures contract written on the asset, with the same maturity τ as the option. At the maturity of the futures, the futures price equals the asset's spot price. Thus a European call option on the asset has the same value as a European

call option on the futures contract with the same maturity. As a result, we will often rewrite the Black-Scholes formula as $H_{BS}(F_{t,\tau}, X, \tau, r_{t,\tau}; \sigma) = e^{-r_{t,\tau}\tau}(F_{t,\tau}\Phi(d_1) - X\Phi(d_2))$, with $d_1 \equiv (\ln(F_{t,\tau}/X) + (\sigma^2/2)\tau)/(\sigma\sqrt{\tau})$ and $d_2 \equiv d_1 - \sigma\sqrt{\tau}$.

9. That is: frictionless markets, unlimited riskless borrowing and lending opportunities at the same instantaneous constant rate r, geometric Brownian motion dynamics for S_t with a known and constant diffusion coefficient. See Merton (1973) for further discussion.

10. Our approach assumes that the vector of state variables Z is correctly specified and is then designed to be very flexible in the way it accomodates the influence of included state variables. Plausible candidates as omitted variables include stochastic market volatility, option and market trading volumes, relative supply and demand for particular options due to portfolio considerations, etc. These variables will be partially reflected in the estimated SPD to the extent that some of the variation in these variables can be accounted for by the included variables, e.g., if, as the empirical evidence suggest, the market volatility covaries negatively with the market price level. Note however that, unlike the clasical regression case, our estimator is not made inconsistent by the omission of relevant variables. Suppose that the true vector $\mathbf{Z}$ contains two sets of variables, $\mathbf{Z}_1$ and $\mathbf{Z}_2$, but that we only include $\mathbf{Z}_1$ to construct our estimator. In that case, we would be estimating $E[H|Z_1]$ instead of $E[H|Z] = E[H|Z_1, Z_2]$; but by the law of iterated expectations $E[H|Z_1] = E[\ E[H|Z]\,|Z_1]$, and so our estimator would still be consistent for the reduced-form call-pricing function as a function of $\mathbf{Z}_1$. In a classical linear regression, the reduced estimator (on $\mathbf{Z}_1$ only) would be inconsistent (unless $E[Z_2|Z_1] = 0$).

11. As the sample size increases, the estimator (8), as well as its derivatives with respect to $\mathbf{Z}$, converge to the true function $H(\cdot)$ and its derivatives at every point. Therefore, when the true function satisfies certain shape restrictions, e.g., monotonicity and convexity with respect to certain variables, the estimator $\hat{H}(\cdot)$ will also have these properties. However, since the asymptotic convergence of the higher-order derivatives may in practice be slow, it can be useful in small samples to modify the estimator to force it to satisfy these restrictions. One simple way to enforce monotonicity is to run an isotonic regression after the kernel regression; this guarantees that the resulting estimator $\hat{H}(\cdot)$ is monotonic. To enforce convexity in small samples, we can extend this procedure one step further: differentiate the isotonic kernel estimator to obtain $\hat{H}'(\cdot)$, and then run an isotonic regression on $\hat{H}'(\cdot)$. We are then guaranteed that $\hat{H}'(\cdot)$ also is monotonic, and hence $\hat{H}(\cdot)$ is convex.

12. The white-noise term is Gaussian with standard deviation equal to either one or two price ticks, depending upon whether the option characteristics made it a high- or low-volume option.

13. This relationship requires that interest rates do not covary with the futures price. During 1993, short term rates did not exhibit much variation at all [see Table 1].

14. Since any violation of put-call parity would give rise to a pure arbitrage opportunity, it can be expected to hold with some degree of confidence. CBOE floor traders of S&P 500 options have confirmed that put-call parity is almost never violated in practice. See also Black and Scholes (1972), Harvey and Whaley (1992), Kamara and Miller (1995) and Rubinstein (1985).

15. We thank Jens Jackwerth and Mark Rubinstein for graciously providing us with the empirical densities from their estimation method.

16. It is possible to make primitive assumptions on the data-generating process which imply the necessary smoothness of the true call-pricing function. In particular, if we assumed that the underlying asset price followed a stochastic differential equation with diffusion function σ^2 which admitted at least p continuous derivatives, $p \geq 2$, then the call-pricing function would also admit at least p continuous derivatives. This follows by writing the call-pricing function as the solution of the generalized Black-Scholes partial differential equation derived from using the standard dynamic replicating strategy (see Merton (1973)). It is a parabolic partial differential equation satisfying all the hypotheses in Friedman (1964, Chapter I) whose solution is then known to have at least p derivatives.

References

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

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

Aït-Sahalia, Yacine, 1995, The delta method for nonparametric kernel functionals, Working paper, Graduate School of Business, University of Chicago.

Arrow, Kenneth, 1964, The role of securities in the optimal allocation of risk bearing, *Review of Economic Studies* 31, 91–96.

Banz, Rolf, and Merton Miller, 1978, Prices for state- contingent claims: Some estimates and applications, *Journal of Business* 51, 653–672.

Bates, David S., 1995, Post-'87 crash fears in S&P 500 futures options, Working paper, The Wharton School, University of Pennsylvania.

Black, Fisher, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–659.

Black, Fisher, and Myron Scholes, 1972, The valuation of option contracts and a test of market efficiency, *Journal of Finance* 27, 399–418.

Boudoukh, Jacob, Matthew Richardson, Richard Stanton, and Robert Whitelaw, 1995, Pricing mortgage-backed securities in a multifactor interest rate environment: A multivariate density estimation approach, Working paper, Stern School of Business, New York University.

Breeden, Douglas, and Robert H. Litzenberger, 1978, Prices of state-contingent claims implicit in option prices, *Journal of Business* 51, 621–651.

Campbell, John, Andrew W. Lo and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).

Constantinides, George M., 1982, Intertemporal asset pricing with heterogeneous consumers and without demand aggregation, *Journal of Business* 55, 253–268.

Cox, John, and Stephen Ross, 1976, The valuation of options for alternative stochastic processes, *Journal of Financial Economics* 3, 145–166.

Debreu, Gerard, 1959, *Theory of Value* (John Wiley and Sons, New York).

Derman, Emanuel, and Iraj Kani, 1994, Riding on the smile, *RISK* 7, 32–39.

Duffie, Darrell, 1996, *Dynamic Asset Pricing Theory*, Second Edition (Princeton University Press, Princeton, NJ).

Duffie, Darrell, and Chi-fu Huang, 1985, Implementing Arrow-Debreu equilibria by continuous trading of few long-lived securities, *Econometrica* 53, 1337–1356.

Dumas, Bruno, Jeffrey Flemming, and Robert E. Whaley, 1995, Implied volatility functions: Empirical tests, Working paper, Fuqua School of Business, Duke University.

Dupire, Bruno, 1994, Pricing with a smile, *RISK* 7, 18–20.

Friedman, Avner, 1964, *Partial Differential Equations of Parabolic Type* (Prentice-Hall, Englewood Cliffs, NJ).

Galai, Dan, 1978, Empirical tests of boundary conditions for CBOE options, *Journal of Financial Economics* 6, 187–211.

Garman, Mark, 1978, The pricing of supershares, *Journal of Financial Economics* 6, 3–10.

Goldenberg, David H., 1991, A unified method for pricing options on diffusion processes, *Journal of Financial Economics* 29, 3-34.

Hansen, Lars Peter, and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587-613.

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Hakansson, Nils H., 1976, Purchasing power funds: A new kind of financial intermediary, *Financial Analysts Journal* 32, 49–59.

Härdle, Wolfgang, 1990, *Applied Nonparametric Regression* (Cambridge University Press, Cambridge, UK).

Harrison, J. Michael and David Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381–408.

Harvey, Andrew, 1995, Long memory in stochastic volatility, Working paper, London School of Economics.

Harvey, Campbell R., and Robert E. Whaley, 1991, S&P 100 index option volatility, *Journal of Finance* 46, 1251–1261.

Harvey, Campbell R., and Robert E. Whaley, 1992, Market volatility prediction and the efficiency of the S&P 100 index option market, *Journal of Financial Economics* 31, 43–73.

He, Hua, and Hayne Leland, 1993, On equilibrium asset price processes, *Review of Financial Studies* 6, 593–617.

Heston, Steven L., 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327–343.

Huang, Chi-fu, and Robert H. Litzenberger, 1988, *Foundations For Financial Economics* (Elsevier Publishing Company, New York).

Hutchinson, James M., Andrew W. Lo and Tomaso Poggio, 1994, A nonparametric approach to the pricing and hedging of derivative securities via learning networks, *Journal of Finance* 49, 851–889.

Jackwerth, Jens C., and Mark Rubinstein, 1996, Recovering probability distributions from contemporary security prices, *Journal of Finance* 51, 1611-1631.

Jarrow, Robert, and Andrew Rudd, 1982, Approximate option valuation for arbitrary stochastic processes, *Journal of Financial Economics* 10, 347-369.

Kamara, Avraham, and Thomas W. Miller Jr., 1995, Daily and intradaily tests of European put-call parity, *Journal of Financial and Quantitative Analysis* 30, 519-539.

Longstaff, Francis, 1995, Option pricing and the martingale restriction, *Review of Financial Studies* 8, 1091–1124.

Lucas, Robert E., 1978, Asset prices in an exchange economy, *Econometrica* 46, 1429–1446.

Madan, Dilip B., and Frank Milne, 1994, Contingent claims valued and hedged by pricing and investing in a basis, *Mathematical Finance* 4, 223-245.

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

Renault, Eric, 1995, Econometric models of option pricing errors, working paper, Université de Toulouse; to appear in the collected invited papers at the 7th World Congress of the Econometric Society (Cambridge University Press, Cambridge, UK).

Ross, Stephen, 1976, Options and efficiency, *Quarterly Journal of Economics* 90, 75–89.

Rubinstein, Mark, 1976, The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics* 7, 407-425.

Rubinstein, Mark, 1985, Nonparametric tests of alternative option-pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978, *Journal of Finance* 40, 455–480.

Rubinstein, Mark, 1994, Implied binomial trees, *Journal of Finance* 49, 771–818.

Shimko, David, 1993, Bounds of probability, *RISK* 6, 33–37.

Stutzer, Michael, 1996, A simple nonparametric approach to derivative security valuation, *Journal of Finance* 51, 1633-1652.

White, Halbert, 1992, *Artificial Neural Networks* (Blackwell Publishers, Oxford, UK).

Figure Titles and Captions

Figure 1: Kernel Regression and Bandwidth Selection

In each panel, the solid line is the kernel-regression estimator of the theoretical relation given by the dashed line. This figure exemplifies the practical importance of bandwidth selection. In Panel (a) the bandwidth is too small by a factor of ten, in Panel (b) the bandwidth is optimal (see Appendix I), and in Panel (c) the bandwidth is too large by a factor of four. The sample size equals 1,000.

Figure 2: Pricing and Hedging Errors of Nonparametric Estimator of the Black-Scholes Formula

Theoretical values for prices, deltas, and SPDs are plotted in Panels (a), (c), and (e), and the average differences (in percent) between the theoretical values and the nonparametric estimators, averaged over the 5,000 replications, are plotted in Panels (b), (d), and (f).

Figure 3: Implied Volatility Surface of The Nonparametric Estimator

Plot of the nonparametric estimator of the implied volatility as a function of time-to-maturity and moneyness. This corresponds to the semiparametric model where we partition the vector of explanatory variables $\mathbf{Z}$ as $[\ \tilde{\mathbf{Z}}'\ F_{t,\tau}\ r_{t,\tau}]'$ with $\tilde{\mathbf{Z}} \equiv [\ X/F_{t,\tau}\ \tau\ ]'$. The nonparametric estimator generates a strong volatility smile, defined as the relationship between the option's moneyness $X/F_{t,\tau}$ and its implied volatility $\sigma(\tilde{\mathbf{Z}})$.

Figure 4: Implied Volatility Curves For The Nonparametric Estimator

Implied volatility curves for the nonparametric option estimator for various times-to-maturity as a function of the option's moneyness, corresponding to the function $X/F_{t,\tau} \longmapsto \sigma(\tilde{\mathbf{Z}})$ for the four expirations τ, where $\tilde{\mathbf{Z}} \equiv [\ X/F_{t,\tau}\ \tau\ ]'$.

Figure 5: Nonparametric SPD Estimates

Nonparametric SPD estimates (solid lines) are overlaid with the corresponding Black-Scholes SPDs (dashed lines) at the same maturities. The Black-Scholes log-normal SPDs are evaluated at the at-the-money implied volatility for that maturity. Dotted lines around each SPD estimate are 95 percent confidence intervals constructed from the asymptotic distribution theory derived in Appendix II. These estimates of the SPD are derived from the semiparametric model where $\tilde{\mathbf{Z}} \equiv [\ X\ F_{t,\tau}\ \tau\ ]'$. The second derivative of the call-pricing function is evaluated at $X = S_T$, and we plot the functions $S_T \mapsto \hat{f}_t^*(\mathbf{Z})$ and $S_T \mapsto \hat{f}_{\mathrm{BS},t}^*(\mathbf{Z})$ for values of the other elements in $\mathbf{Z}$ fixed at their sample means. The volatility for the Black-Scholes SPD is the average at-the-money implied volatility from the market prices of options with the corresponding maturity.

Figure 6: Nonparametric Estimate of SPD-Generated Densities for Continuously-Compounded Returns

Estimated nonparametric distribution of the continuously-compounded τ -period returns, $u_\tau \equiv \ln(S_T/S_t)$, that is compatible with our nonparametric SPD estimator, for the same four maturities as in Figures 4 and 5. The corresponding Black-Scholes distribution, evaluated at the average at-the-money implied volatility for each maturity, are overlaid.

Figure 7: Term Structure of Implied Moments of SPD-Generated Return Densities

The term structures of implied mean, standard deviation, skewness, and kurtosis of the SPD-generated continuously-compounded return distributions along with their Black-Scholes counterparts. All the moments of the returns are annualized. This figure documents the increase in the levels of skewness and kurtosis that are implied by the nonparametric SPD-generated return distributions as a function of the options' maturities.

Table I: Summary Statistics For S&P 500 Index Options Data

Summary statistics for the sample of all traded CBOE daily call and put option prices on the S&P 500 index from January 4, 1993 to December 31, 1993 (14,431 observations). "Implied σ" denotes the implied volatility of the option, and "Implied ATM σ" denotes the implied volatility of at-the-money options. τ denotes the options time-to-expiration, r the riskless rate, X the options strike price and F the S&P 500 futures value implied from the at-the-money call and put prices. "S.D." denotes the sample standard deviation of the variable. During this period, the sample daily mean and standard deviation of continuously-compounded returns of the S&P 500 index were 7.95 and 10.28 percent (annualized with a 252-day year), respectively. The average daily value of the S&P 500 index was 451.5.

Variable	Mean	S.D.	Min	Percentiles					Max
				5%	10%	50%	90%	95%	
Call Price H ($)	24.16	25.29	0.13	0.32	0.76	16.62	59.83	74.21	121.06
Put Price G ($)	9.73	12.53	0.13	0.30	0.54	4.73	25.41	33.37	101.89
Implied σ (%)	11.38	3.29	5.07	7.45	7.84	10.72	15.69	17.41	36.83
Implied ATM σ (%)	9.40	0.87	6.10	7.97	8.29	9.39	10.42	10.68	16.47
τ (Days)	86.64	72.32	1.00	11.00	21.00	66.00	196.00	259.00	350.00
X (Index Points)	440.80	33.02	350.00	390.00	400.00	440.00	480.00	490.00	550.00
F (Index Points)	455.43	10.28	429.19	436.13	441.64	457.82	467.49	469.16	474.22
r (%)	3.07	0.08	2.85	2.96	2.97	3.08	3.18	3.19	3.21

Table II: Bandwidth Values for the SPD Estimator

Bandwidth selection for the SPD estimator with kernels k_F, k_X, and k_τ according to the relation $h_j = c_j \sigma_j n^{-1/[2p+d]}$ for regressor j, where n is the sample size, p is the number of continuous derivatives of the function to be estimated, d is the number of regressors, σ_j is the unconditional standard deviation of the regressor, and $c_j = c_{j0}/\ln(n)$, depends on the particular regressor and the kernel function, with c_{j0} determined by cross-validation. m_j is the order of the partial derivative with respect to the regressor that we wish to estimate, and s_j is the order of the corresponding kernel (see Appendix I for further details). The nonparametric regression corresponds to the semi-parametric model with $\tilde{d}$ equals three and $\tilde{\mathbf{Z}} \equiv [\ X\ F_{t,\tau}\ \tau\ ]'$. The sample size n equals 14,431. The coefficient of determination $R^2 = 0.86$ measures the goodness-of-fit of the nonparametric kernel regression $\hat{\sigma}(F_{t,\tau}, X, \tau)$.

Kernel	s_j	p	m_j	d	c_j	σ_j	h_j
$k_F = k_{(4)}$	4	4	0	3	2.040	10.275	8.776
$k_X = k_{(2)}$	2	4	2	3	1.260	33.018	17.418
$k_\tau = k_{(4)}$	4	4	0	3	0.1014	72.324	3.071

Table III: Moments of Nonparametric Return Densities

Annualized moments of nonparametric and Black-Scholes densities of the continuously-compounded τ-period return, $u_\tau \equiv \ln(S_T/S_t)$, that are compatible with the SPD estimates for prices. The estimates are based on a sample of 14,431 CBOE daily call and put option prices on the S&P 500 index from January 4, 1993 to December 31, 1993. The nonparametric regression for the semiparametric model has $\tilde{d}$ equals three and $\tilde{\mathbf{Z}} \equiv [\ X\ F_{t,\tau}\ \tau\ ]'$. This table quantifies the differences between the nonparametric and Black-Scholes return densities in terms of their first four moments, for maturities of one, two, four and six months. Note that the mean of the price SPD is given by the futures price $F_{t,\tau}$, and is the same for the nonparametric and Black-Scholes price SPDs. However, as a result of Itô's Lemma, the means of the return densities are not equal since the estimated volatility of the nonparametric diffusion for the S&P 500 index is not constant.

SPD Estimator	Mean	Standard Deviation	Skewness	Kurtosis
Time-to-Maturity: 21 Days				
Nonparametric	0.0136	0.0991	−0.1976	0.0748
Black-Scholes	0.0136	0.0990	0.0000	0.0000
Time-to-Maturity: 42 Days				
Nonparametric	0.0147	0.0980	−0.3431	0.2175
Black-Scholes	0.0149	0.0954	0.0000	0.0000
Time-to-Maturity: 84 Days				
Nonparametric	0.0151	0.0989	−0.4406	0.2281
Black-Scholes	0.0152	0.0974	0.0000	0.0000
Time-to-Maturity: 126 Days				
Nonparametric	0.0155	0.1001	−0.5165	0.2907
Black-Scholes	0.0153	0.1019	0.0000	0.0000

Table IV: Nonparametric Estimates of S&P 500 Index Option Prices

Estimated nonparametric call option, put option, normalized butterfly-spread and digital option prices on the S&P 500 index for a four-month maturity (84 days) and strike prices with deltas greater than or equal to 0.05 in absolute value, priced for a current index value of 455.00. The nonparametric kernel estimator is based on a sample of 14,431 CBOE daily call and put option prices on the S&P 500 index from January 4, 1993 to December 31, 1993. The nonparametric regression corresponds to the semiparametric model with $\tilde{d}$ equals three and $\tilde{\mathbf{Z}} \equiv [\ X\ \ F_{t,\tau}\ \ \tau\ ]'$. The Black-Scholes prices are computed at the actual at-the-money implied volatility. Option deltas are computed with respect to the futures price, not the spot price. Butterfly prices are standardized by the squared of the strike price increment (five points), to provide a discrete approximation to the SPD value. "SPD Value" denotes the exact nonparametric estimate of the SPD. Both values are multiplied by 100. "Digital Price" and "BS Digital Price" denote the nonparametric and Black-Scholes prices of the digital option, respectively. Their differences illustrate the differences between the nonparametric and Black-Scholes estimates of the cumulative distribution functions corresponding to the two SPDs.

Strike Price	Call Price	Call Delta	Put Price	Put Delta	Implied Vol.	Butterfly Price	SPD Value	Digital Price	BS Digital Price
Futures: \$458.04, Interest Rate: 3.05%, Time-to-Maturity: 84 Days									
415	44.04	0.93	1.44	−0.06	13.03	0.31	0.32	0.06	0.04
420	39.47	0.91	1.82	−0.08	12.64	0.39	0.40	0.08	0.06
425	35.00	0.88	2.30	−0.11	12.25	0.49	0.49	0.10	0.10
430	30.65	0.85	2.90	−0.14	11.86	0.60	0.61	0.13	0.14
435	26.45	0.81	3.65	−0.18	11.47	0.74	0.75	0.17	0.18
440	22.44	0.77	4.58	−0.22	11.09	0.91	0.91	0.21	0.24
445	18.66	0.71	5.75	−0.27	10.71	1.10	1.11	0.26	0.31
450	15.15	0.65	7.19	−0.34	10.34	1.31	1.33	0.32	0.38
455	11.96	0.58	8.95	−0.41	9.98	1.51	1.53	0.39	0.46
460	9.16	0.50	11.10	−0.49	9.65	1.67	1.69	0.47	0.54
465	6.77	0.41	13.66	−0.57	9.35	1.74	1.77	0.55	0.61
470	4.82	0.33	16.66	−0.66	9.08	1.70	1.73	0.64	0.68
475	3.29	0.25	20.08	−0.74	8.84	1.56	1.58	0.72	0.74
480	2.16	0.18	23.89	−0.81	8.63	1.33	1.34	0.79	0.80
485	1.35	0.12	28.04	−0.87	8.45	1.06	1.07	0.85	0.84
490	0.81	0.08	32.44	−0.91	8.29	0.78	0.79	0.90	0.88

Table V: Kernel Constants For Asymptotic Distribution of SPD Estimator, Specification and Stability Tests

Kernel constants that characterize the asymptotic variances of the nonparametric and semiparametric estimators for option prices, deltas, and SPDs, which are given in Appendix II, as well as the specification and stability tests in Appendix III and IV respectively. The kernel functions $k_{(2)}$ and $k_{(4)}$ are defined in Appendix I. k' and k'' denote the first and second derivative of the kernel function, respectively.

Functional	$k = k_{(2)}$	$k = k'_{(2)}$	$k = k''_{(2)}$	$k = k_{(4)}$	$k = k'_{(4)}$	$k = k''_{(4)}$
$\int_\omega k^2(\omega)d\omega$	$\frac{1}{2\sqrt{\pi}}$	$\frac{1}{4\sqrt{\pi}}$	$\frac{3}{8\sqrt{\pi}}$	$\frac{27}{32\sqrt{\pi}}$	$\frac{175}{64\sqrt{\pi}}$	$\frac{273}{128\sqrt{\pi}}$
$\int_\omega \left[\int_{\tilde{\omega}} k(\tilde{\omega})k(\omega+\tilde{\omega})d\tilde{\omega}\right]^2 d\omega$	$\frac{1}{2\sqrt{2\pi}}$	$\frac{3}{32\sqrt{2\pi}}$	$\frac{105\sqrt{2}}{1024\sqrt{\pi}}$	$\frac{7881\sqrt{2}}{16384\sqrt{\pi}}$	—	—

Table VI: Nonparametric Specification Test of the Black-Scholes Model

Nonparametric specification tests of the Black-Scholes option-pricing model based on the option-pricing formula H_{BS} and its corresponding delta Δ_{BS} and SPD f^*_{BS}, based on the sample of 14,431 CBOE daily call and put option prices on the S&P 500 index from January 4, 1993 to December 31, 1993. The nonparametric regression for the semiparametric model has $\tilde{d}$ equals two and $\tilde{\mathbf{Z}} \equiv [\; X/F_{t,\tau} \quad \tau \;]'$. The bandwidths are chosen to be $h_1 = h_{01}\sigma_1 n^{-1/\delta_1}$ with $\delta_1 = \delta_2 + 2m$ and $h_2 = h_{02}\sigma_2 n^{-1/\delta_2}$, where σ_j, $j = 1, 2$ is the unconditional standard deviation of the j-th regressor. The bandwidths to estimate the conditional variance of the nonparametric regression are $h_{1,cv} = 0.014957$ and $h_{2,cv} = 14.658$. The Black-Scholes volatility value, estimated as the mean of the nonparametric regression of implied volatility on $\tilde{\mathbf{Z}}$, is $\sigma_{BS} = 12.097$ percent. The weighting functions ω_H, ω_Δ and ω_{f^*} are trimming indices, i.e., only observations with estimated density above a certain level, and away from the boundaries of the integration space, are retained. The two numbers in the column "Trim" refer respectively to the trimming level (as a percentage of the mean estimated density value), and the percentage trimmed at the boundary of the integration space when calculating the test statistics. For instance, if the latter is 5 percent, the trimming index retains the values between 1.05 times the minimum evaluation value and 0.95 times the maximum value. "Integral" refers to the percentage of the estimated density mass on the integration space that is kept by the trimming index. "Test Statistic" refers to the standardized distance measure between the nonparametric and Black-Scholes estimates (remove the bias term, divide by the standard deviation). The distance measure for each of the three null hypotheses considered is $\hat{D}(\hat{H}, \hat{H}_{BS})$, $\hat{D}(\hat{\Delta}, \hat{\Delta}_{BS})$ and $\hat{D}(\hat{f}^*, \hat{f}^*_{BS})$ respectively. The integrals $\hat{D}$ are calculated over the integration space given by the rectangle $[0.85, 1.10] \times [10, 136]$ in the moneyness $\times$ days-to-expiration space. The kernel weights are constructed using the binning method with 30 bins in the moneyness dimension and 20 in the days-to-expiration dimension. The test is described in Appendix III.

Null	d	m	s	k	δ_1	δ_2	h_{01}	h_{02}	h_1	h_2	Trim	Integral	Test Statistic	p-Value
$H = H_{\text{BS}}$	2	0	2	$k_{(2)}$	4.75	4.75	0.48	0.48	0.0047168	4.62255	50/ 5	55.4	6045.8	0.0000
$\Delta = \Delta_{\text{BS}}$	2	1	2	$k_{(2)}$	6.75	4.75	0.48	0.48	0.0085724	4.62255	50/ 5	54.9	480.8	0.0000
$f^* = f^*_{\text{BS}}$	2	2	2	$k_{(2)}$	8.75	4.75	0.48	0.48	0.0118562	4.62255	50/ 5	54.5	61.9	0.0000

Table VII: Stability Tests for the SPD Estimator

Tests of stability of the SPD estimator across subperiods corresponding to each quarter of 1993. Each row reports a test of the null hypothesis that the SPD's in quarters Q_i and Q_j are identical, where Q_i and Q_j are non-adjacent quarters of 1993 to reduce the effects of dependence. Each quarter contains n equals 3,607 observations and the bandwidths are chosen to be $h_1 = h_{01}\sigma_1 n^{-1/\delta_1}$ with $\delta_1 = \delta_2 + 2m$ and $h_2 = h_{02}\sigma_1 n^{-1/\delta_2}$, where σ_j, $j = 1, 2$ is the unconditional standard deviation of the j-th regressor. The bandwidths to estimate the conditional variance of the nonparametric regression are $h_{1,cv} = 0.018845$ and $h_{2,cv} = 18.468$. The weighting function ω_{f^*} is a trimming index. The two numbers in the column "Trim" refer respectively to the trimming level (as a percentage of the mean estimated density value), and the percentage trimmed at the boundary of the integration space when calculating the test statistics. For instance, if the latter is 5 percent, the trimming index retains the values between 1.05 times the minimum evaluation value and 0.95 times the maximum value. "Integral" refers to the percentage of the estimated density mass on the integration space that is kept by the trimming index. The estimates $\hat{f}_l^*$, $l = 1, 2$, of the SPD are calculated for the semiparametric model where the nonparametric regression has $\tilde{d}$ equals two and $\tilde{\mathbf{Z}} \equiv [\ X/F_{t,\tau} \quad \tau\ ]'$. The integrals $\hat{S}$ are calculated over the integration space given by the rectangle $[0.85, 1.10] \times [10, 94]$ in the moneyness $\times$ days-to-expiration space. This rectangle has the highest concentration of options observed in each of the four quarters. The kernel weights are constructed using the binning method with 30 bins in the moneyness dimension and 20 in the days-to-expiration dimension. The test is described in Appendix IV.

Null	d	m	s	k	δ_1	δ_2	h_{01}	h_{02}	h_1	h_2	Trim	Integral	Test Statistic	p-Value
$Q_1 = Q_3$	2	2	2	$k_{(2)}$	8.75	4.75	0.48	0.48	0.0063157	6.1894	50/ 5	52.5	0.23	0.41
$Q_1 = Q_4$	2	2	2	$k_{(2)}$	8.75	4.75	0.48	0.48	0.010527	6.1894	50/ 5	52.1	0.21	0.42
$Q_2 = Q_4$	2	2	2	$k_{(2)}$	8.75	4.75	0.48	0.48	0.013892	6.1894	50/ 5	50.3	0.63	0.27

Table VIII: In and Out-of-Sample Forecasts

Comparison of average forecast errors of the densities produced by: (JR) Jackwerth and Rubinstein's (1996) method of extracting densities by minimizing a criterion function which penalizes for unsmoothness; (HLP) an extension of Hutchinson, Lo and Poggio's (1994) where we use their learning network (with five variables and four hidden units) and differentiate it twice numerically; (BS) the Black-Scholes density with the volatility taken as the at-the-money implied volatility of the corresponding options; and finally our SPD estimator (AL) where the nonparametric regression corresponds to the semiparametric model with $\tilde{d}$ equals three and $\tilde{\mathbf{Z}} \equiv [\ X/F_{t,\tau}\ \ \tau\ ]'$. In Panel A, the density forecast error is expressed as a percentage of the mode value of the comparison density. In Panel B, the price forecast error is reported as a percentage of the options market price over the comparison interval. When forecasting prices, we consider the 400, 425, 450 and 475 strike options. (MKT) refers to the actual CBOE market price. Option-pricing models and their corresponding SPD's are estimated with daily data from January 4 to September 30, 1993 and out-of-sample forecasts are generated for various forecast horizons τ on a daily rolling basis from October 1 to December 31, 1993.

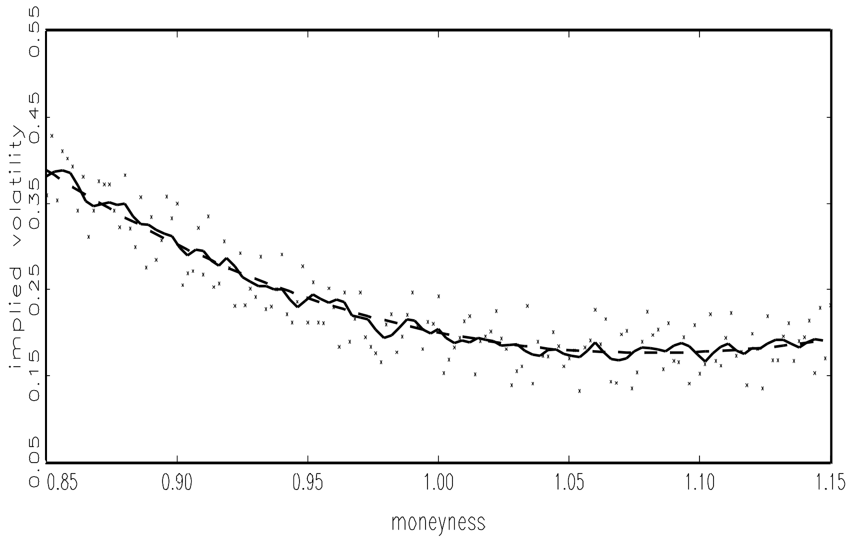
Panel A: Density Forecast Error

τ	Forecast of JR-Density τ Trading Days Ahead				Forecast of AL-Density τ Trading Days Ahead			
	JR	HLP	AL	BS	JR	HLP	AL	BS
0	0.00	8.71	5.14	11.57	6.06	6.36	0.00	9.14
1	3.82	8.85	5.41	11.67	6.55	6.74	2.48	9.42
5	5.80	9.41	6.62	12.40	7.47	7.52	4.63	10.49
10	7.07	9.43	7.06	12.65	8.03	7.55	4.91	10.83
15	8.42	9.36	7.50	12.74	8.82	7.65	5.42	11.09
20	9.34	8.89	7.73	12.47	9.85	7.69	6.22	11.20

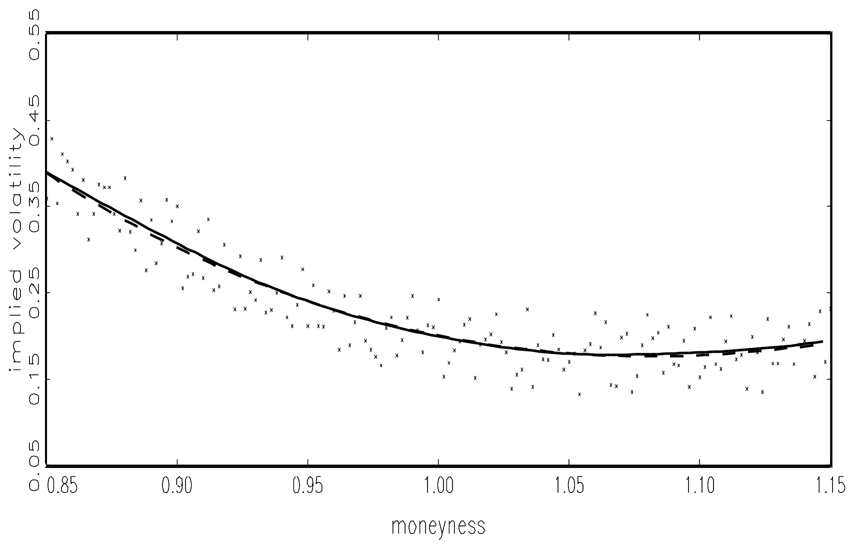
Panel B: Market-Price Forecast Error

τ	JR	HLP	AL	BS	MKT
0	2.81	5.90	3.31	7.27	0.00
1	2.89	5.90	3.36	7.27	0.98
5	3.11	5.92	3.24	7.28	1.57
10	3.80	5.75	2.95	7.24	2.46
15	4.39	5.56	2.68	7.28	3.10
20	4.97	5.49	2.70	7.40	3.54

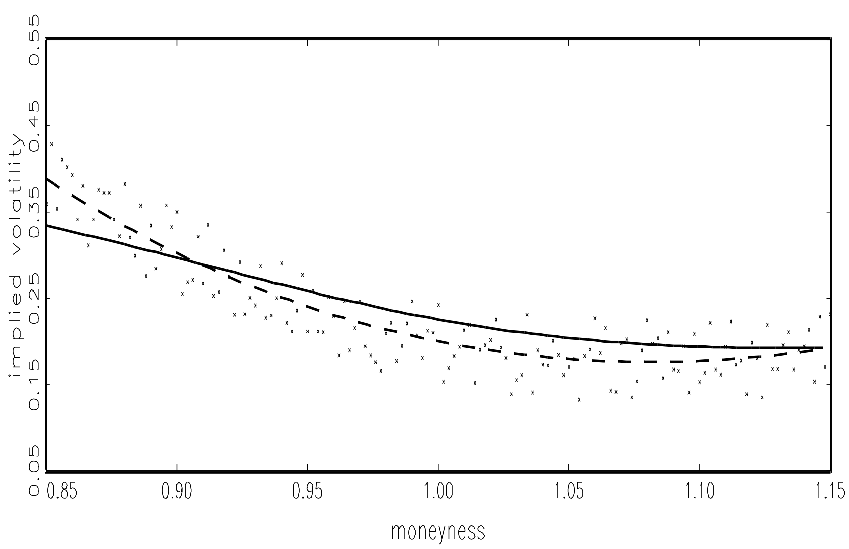
a: Kernel Regression with Undersmoothed Bandwidth

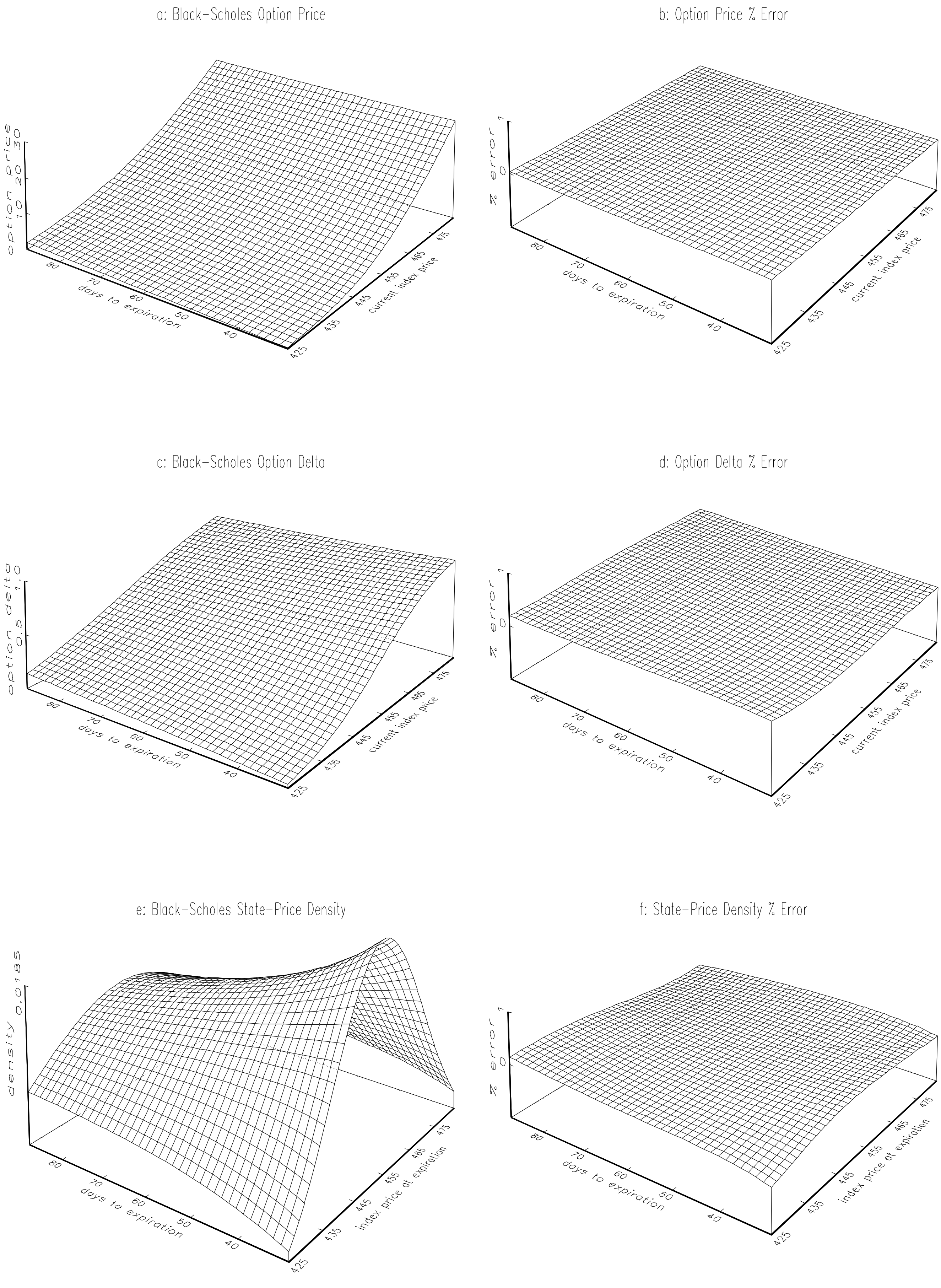


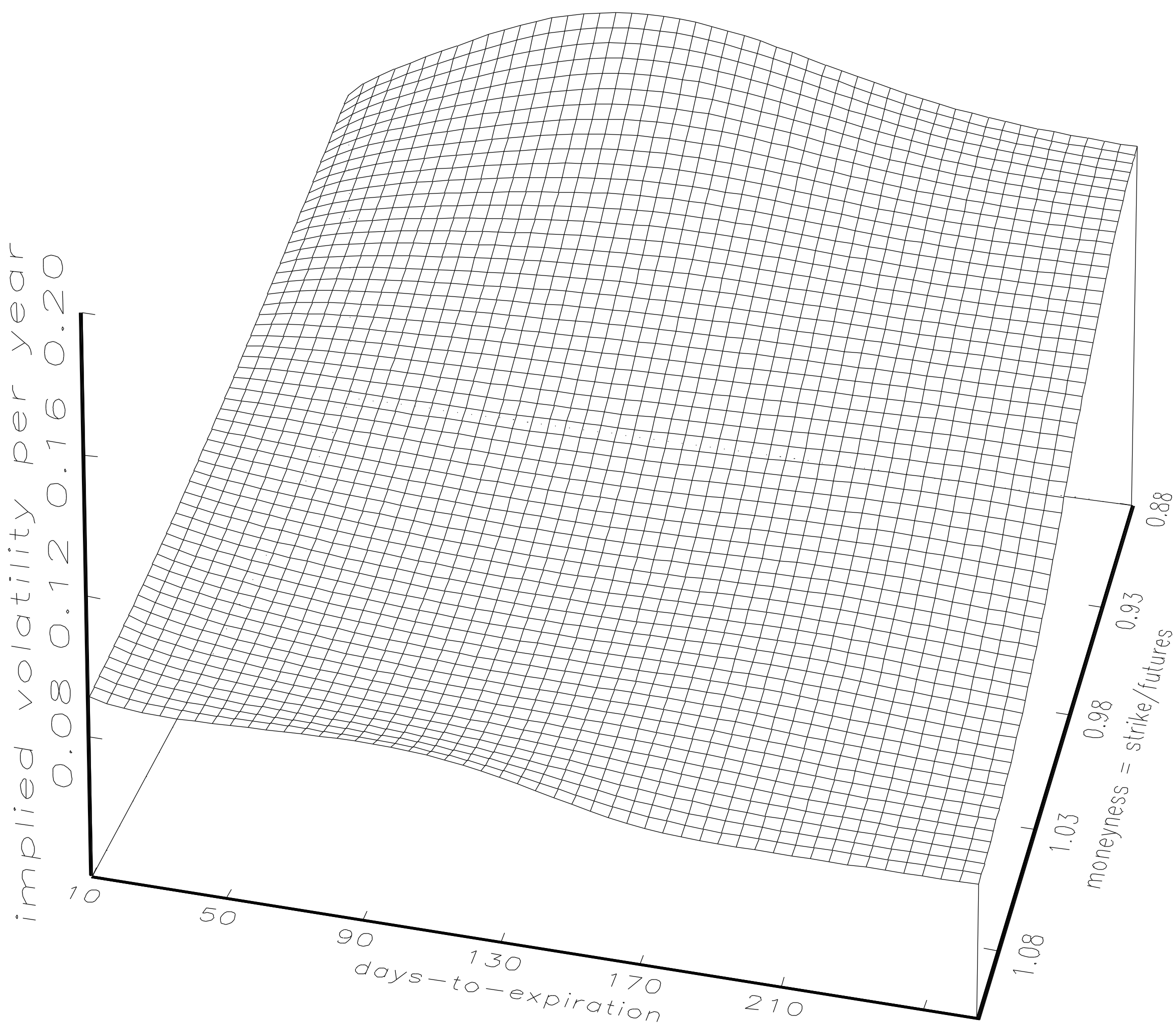
b: Kernel Regression with Optimal Bandwidth



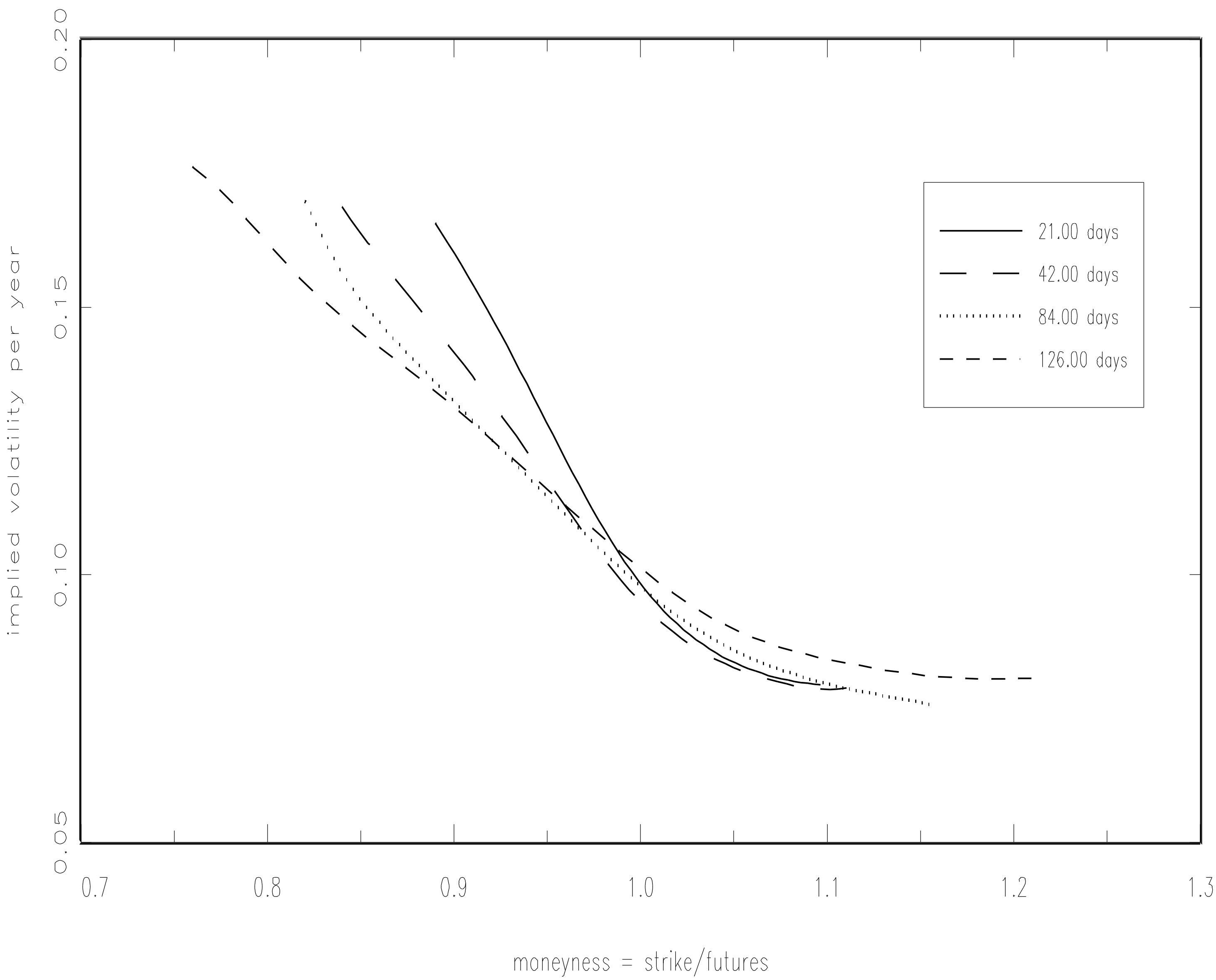
c: Kernel Regression with Oversmoothed Bandwidth



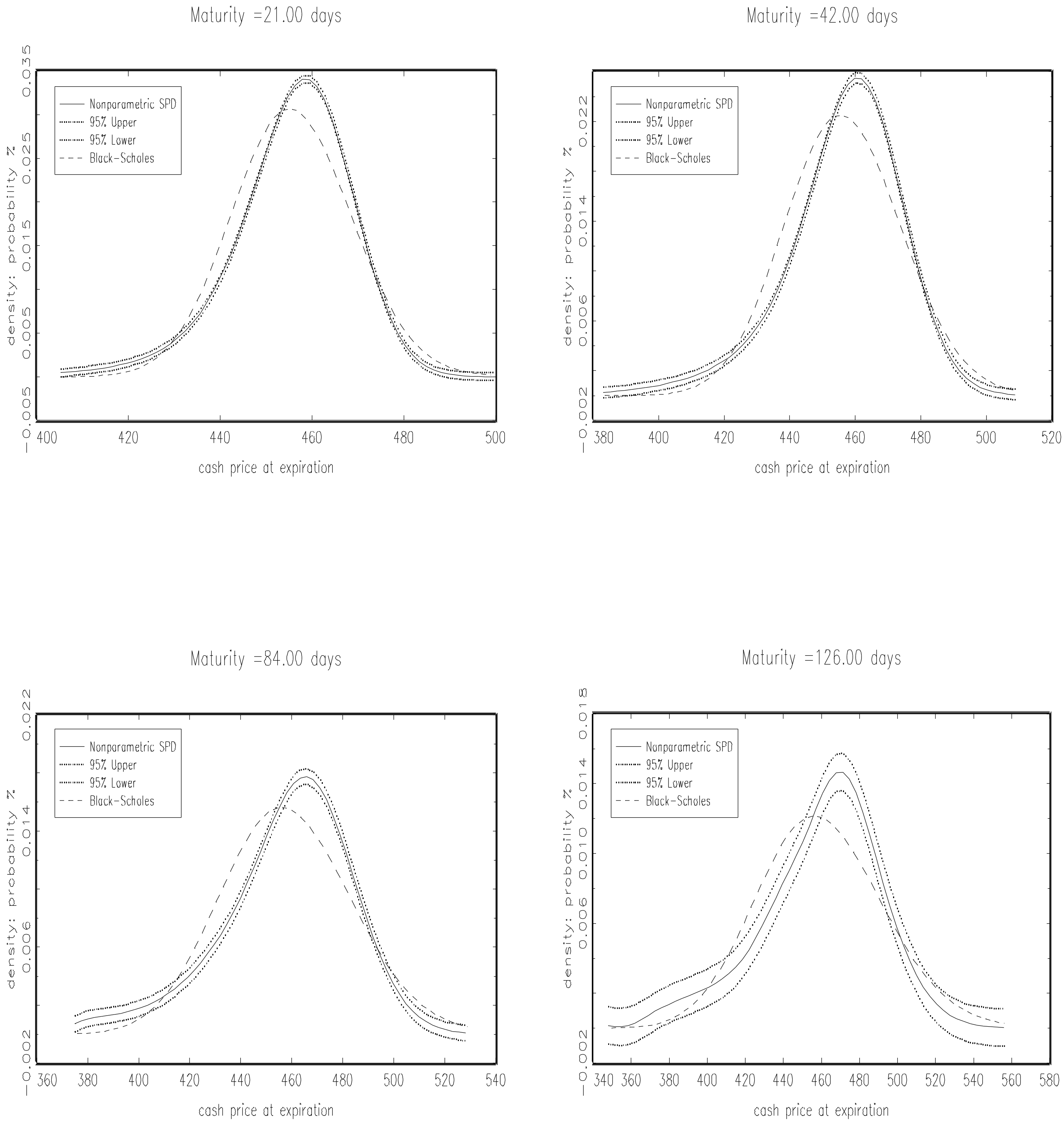

a: Black-Scholes Option Price
option price
0
10
20
30
days to expiration
80
70
60
50
40
current index price
425
435
445
455
465
475
b: Option Price % Error
% error
0
1
days to expiration
80
70
60
50
40
current index price
425
435
445
455
465
475
c: Black-Scholes Option Delta
option delta
0.5
1.0
days to expiration
80
70
60
50
40
current index price
425
435
445
455
465
475
d: Option Delta % Error
% error
0
1
days to expiration
80
70
60
50
40
current index price
425
435
445
455
465
475
e: Black-Scholes State-Price Density
density
0.0185
days to expiration
80
70
60
50
40
index price at expiration
425
435
445
455
465
475
f: State-Price Density % Error
% error
0
1
days to expiration
80
70
60
50
40
index price at expiration
425
435
445
455
465
475



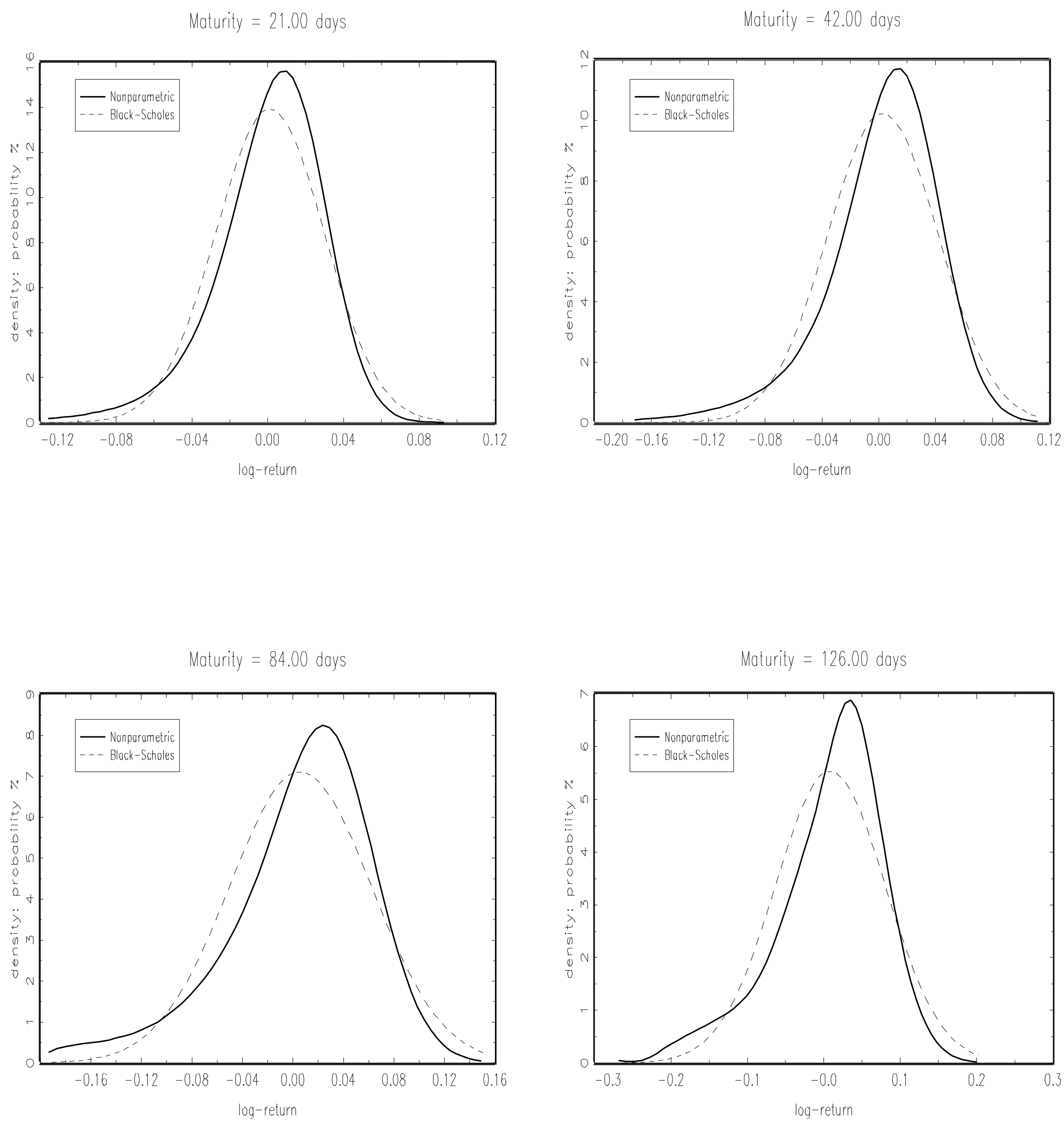
implied volatility per year
0.08
0.12
0.16
0.20
10
50
90
130
170
210
days-to-expiration
moneyness = strike/futures
0.88
0.93
0.98
1.03
1.08



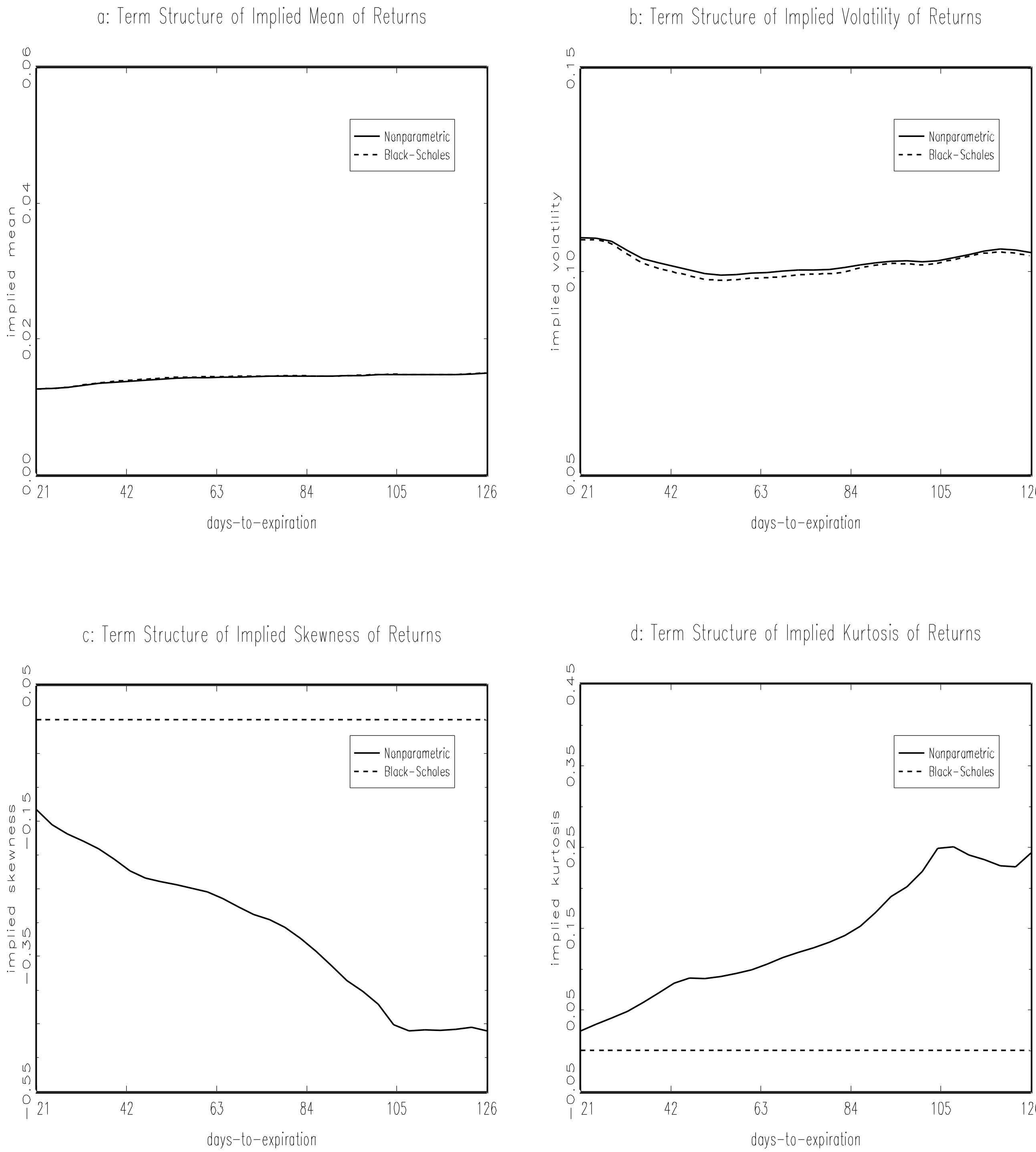
0.20
0.15
0.10
0.05
implied volatility per year
0.7
0.8
0.9
1.0
1.1
1.2
1.3
moneyness = strike/futures
21.00 days
42.00 days
84.00 days
126.00 days



Maturity =21.00 days
Nonparametric SPD
95% Upper
95% Lower
Black-Scholes
density: probability %
cash price at expiration
Maturity =42.00 days
Nonparametric SPD
95% Upper
95% Lower
Black-Scholes
density: probability %
cash price at expiration
Maturity =84.00 days
Nonparametric SPD
95% Upper
95% Lower
Black-Scholes
density: probability %
cash price at expiration
Maturity =126.00 days
Nonparametric SPD
95% Upper
95% Lower
Black-Scholes
density: probability %
cash price at expiration



Maturity = 21.00 days
Nonparametric
Black-Scholes
density: probability %
log-return
Maturity = 42.00 days
Nonparametric
Black-Scholes
density: probability %
log-return
Maturity = 84.00 days
Nonparametric
Black-Scholes
density: probability %
log-return
Maturity = 126.00 days
Nonparametric
Black-Scholes
density: probability %
log-return


a: Term Structure of Implied Mean of Returns
Nonparametric
Black-Scholes
implied mean
0.0
0.02
0.04
0.06
21
42
63
84
105
126
days-to-expiration
b: Term Structure of Implied Volatility of Returns
Nonparametric
Black-Scholes
implied volatility
0.05
0.10
0.15
21
42
63
84
105
126
days-to-expiration
c: Term Structure of Implied Skewness of Returns
Nonparametric
Black-Scholes
implied skewness
-0.55
-0.35
-0.15
0.05
21
42
63
84
105
126
days-to-expiration
d: Term Structure of Implied Kurtosis of Returns
Nonparametric
Black-Scholes
implied kurtosis
-0.05
0.05
0.15
0.25
0.35
0.45
21
42
63
84
105
126
days-to-expiration