Nonparametric Estimation of Daily State-Price Densities Using Transaction-Level Data on S&P500 Index Options

Andrew Maurer

September, 2012

An essay submitted to the Department of Economics

in partial fulfillment of the requirements for

the degree of Master of Arts

Queen's University

Kingston, Ontario, Canada

Abstract

A good estimate for the state-price density has important applications both practically, to derivatives valuation, and in terms of theoretical economics, to our understanding of preferences and risk tendencies. I estimate the state-price density using transaction level data on S&P500 call and put options from 1993 using a nonparametric estimation technique. By using transaction-level data, I am able to obtain estimates for the state-price density for individual days. The estimates are compared to the Black-Scholes lognormal state-price density and are seen to be negatively skewed, have higher kurtosis, and a higher center, particularly for high times to maturity.

Contents

1 Introduction

At the core of financial theory is the problem of how to optimally allocate capital over time. Since what will happen in the future is not known for sure, agents must invest in the presence of uncertainty. The time-state preference model was developed by Arrow (1964) and Debreu (1959) in order to formalize the problem of agents investing under uncertainty. In this model, there is a set of possible states of nature that could unfold in the future and an associated price that a consumer today would pay to be guaranteed \$1 of consumption in each particular state. This is termed the state-price, or the price of an elementary security. Consumption in a state that has a high probability of occurring would have a higher value today than consumption in a state with a low probability of occurring, all else equal. A portfolio consisting of all elementary securities guarantees \$1 of consumption at time T, no matter what state is realized. Then the price of such a portfolio must equal the price of a bond that pays \$1 at time T, or $B(0,T) = 1/(1+r)$.

A well-known way of interpreting state-prices is as a measure of probability. That is, the future value of the state prices sum to one, like the traditional measure of probability and the "probability" of any individual event must be greater than zero and less than unity. This alternate measure of probability is known as the risk-neutral probability (RNP). The name comes from the interpretation that, for a risk-neutral representative agent, the RNP would be equivalent to the objective probability. The objective probability, traditional probability, and physical probability are all synonymous. The degree of risk aversion of the market plays an important role in the formation of state prices. Investors who seek to hedge downside risk may more highly value states associated with unfavourable outcomes. The associated risk neutral probability, or state price, of such events would tend be higher than the objective probability.

Extending the discrete time model of Arrow and Debreu is the continuous time counterpart, the state-price density (SPD). [1] The interpretation of the state price density is that the area under the curve between x and $x + dx$ is the cost today of \$1 of consumption in all states between x and $x + dx$. Since the concept of "state" is somewhat abstract, we may think of states as the possible prices a stock could take at some future time, T, conditional on today's information set. Then the SPD signifies the value that the market places on consumption conditional on each possible

[1] The state-price density is also referred to as the stochastic discount factor, the pricing kernel, and the risk neutral density. These are all equivalent.

realization of the stock price at time T, S_T.

Being able to model the state price density is useful both for theoretical and practical purposes. In terms of economic theory, the state price density characterizes the preferences of the market. Under certain restrictions, namely complete markets and homothetic preferences, there is a representative agent whose preferences are the sum of all individual investor's preferences. Given a model of the state price density, that is, the market implied or risk neutral density, in addition to a model of the objective probabilities, the degree of risk aversion of the representative agent may be surmised. The market would be considered risk averse if the market-implied probabilities of extreme events were lower than the actual probabilities. On a more practical level, a good model of the SPD would be invaluable in pricing derivatives. Derivatives are just portfolios of elementary securities. Since the SPD prices the continuum of elementary securities, it may also be used to price any derivative, since derivatives are just portfolios of elementary securities.

In Section 1.1 I introduce the state price density and show how it may be derived from a call option pricing function, estimated from an observed set of option prices. I derive the state price density in the Black-Scholes case to provide a reference for my nonparametric estimates. In Section 1.3 I discuss how the Black-Scholes implied volatility may be calculated. I explore filtering techniques in the context of implied volatility to improve upon the Newton-Raphson algorithm for calculating the minimum of a function. In Section 1.4 I give a mathematical background of nonparametric estimation techniques, particularly on the Nadaraya Watson estimator. Given the tools to estimate the state price density, I simulate a set of option prices in Section 2 according to the Black-Scholes formula and show how the nonparametric estimator may be used to recover the SPD. In Section 3.1 I turn to estimating the SPD using transaction level options data for puts and calls traded on the S&P500 index in 1993. In particular, I estimate the SPD for individual trading days, avoiding the common assumption that the SPD does not change over the year. I argue why transaction level data provides a more realistic measure of value than the bid-ask midpoint. I conclude in Section 4 and derive the semiparametric estimator I use for estimation in the Appendix.

1.1 The State-Price Density

Modeling the state-price density is simply the continuous-time equivalent of modeling the prices of all individual Arrow-Debreu (A-D) securities. Such a model can be used to infer the prices of A-D securities implied by the prices of options observed in the marketplace. Pricing securities based on taking combinations of other securities, a cornerstone of financial theory, is known as no arbitrage pricing. For example, consider a world in which there are only two possible future states of the world: up and down. Suppose there are three securities in the market place, with respective payoffs of [up,down] for the three securities of [1,1], [1,2], and [2,3]. In order for there to not be an arbitrage opportunity and because the third security has a payoff equivalent to the sum of the first and second, the price of the third must equal the sum of the prices of the first and second. Moreover, since there are two securities with linearly independent payoffs and only two states of the world, the no-arbitrage prices of A-D securities may be inferred through simple linear algebra. Consider a portfolio long in security two and short in security one. Then the payoff is $[2,1]-[1,1]=[1,0]$. This is the Arrow-Debreu security corresponding to the up state and must be priced the same as the corresponding portfolio in order for there to be no arbitrage opportunity. Finally, the price of the A-D security corresponding to the down state is just the price of the first security less the price of the AD1. That is, [1,1] - [1,0] = [0,1]. Thus, by observing the market prices of a set of securities with payoffs that span all states of nature, the complete set of AD prices may also be priced.

However, the real world is more complicated than the two-state example just discussed. When considering the random path of the price of a stock, there is a continuum of possible prices that the stock could take for some future date[2]. However, there are only a finite number of traded securities. Not only is the number of traded securities finite, but it is small relative to the infinity of states of nature. That is, the market is not complete. For example, options traded on the Chicago Board Options Exchange have strike prices quoted in intervals of \$5 near the money, with far fewer options traded deep out of the money. Thus, a model is needed in order to "complete the market" by interpolating the theoretical prices of all possible options. Only after such an option pricing function is modeled can the associated state-price density be recovered.

The idea of recovering the SPD from a set of option prices was first introduced by Ross (1976) in which he presents the idea of building up derivatives from a set of "primitives", or Arrow Debreu securities. Banz and Miller (1978) expand on this by discussing how the primitive securities implied by a market portfolio may be used to value projects. However, the SPD was not formally characterized until 1978, when

[2]In practice, there is not actually a continuum of possible prices that the stock could take, since there is the real world constraint that options are quoted in 1/16ths of a dollar and may thus only take values in multiples of 1/16 of a dollar (CBOE conventions in 1993).

Breeden and Litzenberger derived the analytical solution for the SPD, given a call option pricing function. First, the price of a call option is expressed in terms of the risk neutral measure Q, or "equivalent martingale measure":

$$\begin{aligned} C(S_t, K, T, r_{t,T}, \delta_{t,T}) &= e^{-r_{t,T}(T-t)} E_Q[\max(S_T - K, 0)] \\ &= e^{-r_{t,T}(T-t)} \int_K^{+\infty} (S_T - K) f^*(S_T) dS_T \end{aligned}$$

The result from Breeden and Litzenberger is that the second derivative of the call option pricing function (C) with respect to the strike price (K) is the state price density.

$$\begin{aligned} \frac{\partial^2 C}{\partial K^2} &= e^{-r_{t,T}(T-t)} f^*(K) \\ &= \pi_{S_T=K} \end{aligned} \quad (1)$$

That is, $\partial^2 C/\partial K^2$ is the state price of the event that the terminal stock price exactly equals the strike price ($\pi_{S_T=K}$). Notably, the statement makes no assumptions on the functional form of the call option pricing function.[3] In particular, the call is not assumed to be priced according to the Black-Scholes formula.

1.2 The Black-Scholes Case

In the most basic case, where the stock price follows a geometric Brownian motion with constant volatility and interest rates held constant, the SPD may be derived algebraically. The parameters may then be estimated from data. First, the call option pricing function must be derived. When the Black-Scholes assumptions are satisfied, the call option price is described by Equation (3), Black and Scholes (1973). The equation is derived through a no-arbitrage argument. The payoff of a call option is replicated through a portfolio of the stock and risk free bond, implying the price of the call option must equal the price of the portfolio. The assumptions are that the stock price follows a geometric Brownian Motion with constant drift and volatility, there are no market frictions, markets are complete, and there are no arbitrage opportunities. The Black-Scholes formula for the price of a call option is:

[3] A trivial assumption is that the call price is a function of the strike price - but this is obvious.

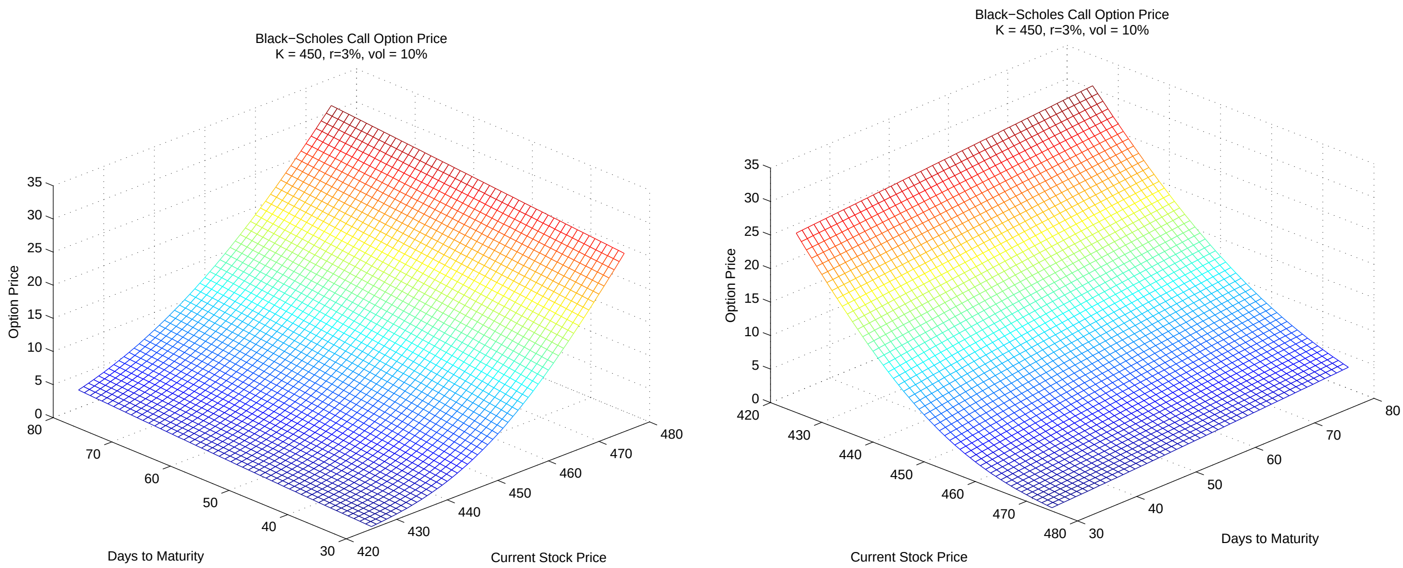
$$C_{BS}(S_t, K, T, r_{t,T}, \delta_{t,T}; \sigma) = S_t e^{-\delta_{t,T}T}\Phi(d_1) - Ke^{-r_{t,T}T}\Phi(d_2) \tag{2}$$

where

$$d_1 = \frac{\ln(S_t/K) + (r_{t,T} - \delta_{t,T} + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \; d_2 = d1 - \sigma\sqrt{T}$$

or all together, with $\delta = 0$

$$C_{BS} = S_t\Phi(\frac{\ln(S_t/K) + (r_{t,T} + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}) - Ke^{-r_{t,T}T}\Phi(\frac{\ln(S_t/K) + (r_{t,T} - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}) \tag{3}$$



(a) Varying S_t

(b) Varying K

Figure 1: Notice that in (a), we see call option price increasing in S_0 and in (b) we see call option price decreasing in K. Both of these make sense. A call option is the right to buy the underlying asset at price K. Clearly this is more valuable for either higher S_0 (given fixed K) or lower K (given fixed S_0). These figures represent neither real data nor simulated data, but are simply the geometric representations of Equation (3)

Now the state price density is derived, here denoted f^*, by plugging C_{BS} from Equation (3) into the general form of the SPD derived by Breeden and Litzenberger, Equation (1).

$$f^*_{BS,t}(S_T) = e^{r_{t,T}T}\frac{\partial^2 C}{\partial K^2}|_{S_T=K} \qquad (4)$$
$$= \frac{1}{S_T\sqrt{2\pi\sigma^2 T}}\exp[-\frac{[\ln(S_T/S_t) - (r_{t,T} - \delta_{t,T} - \sigma^2/2)T]^2}{2\sigma^2 T}]$$

Figure 2: Black-Scholes State-Price Density

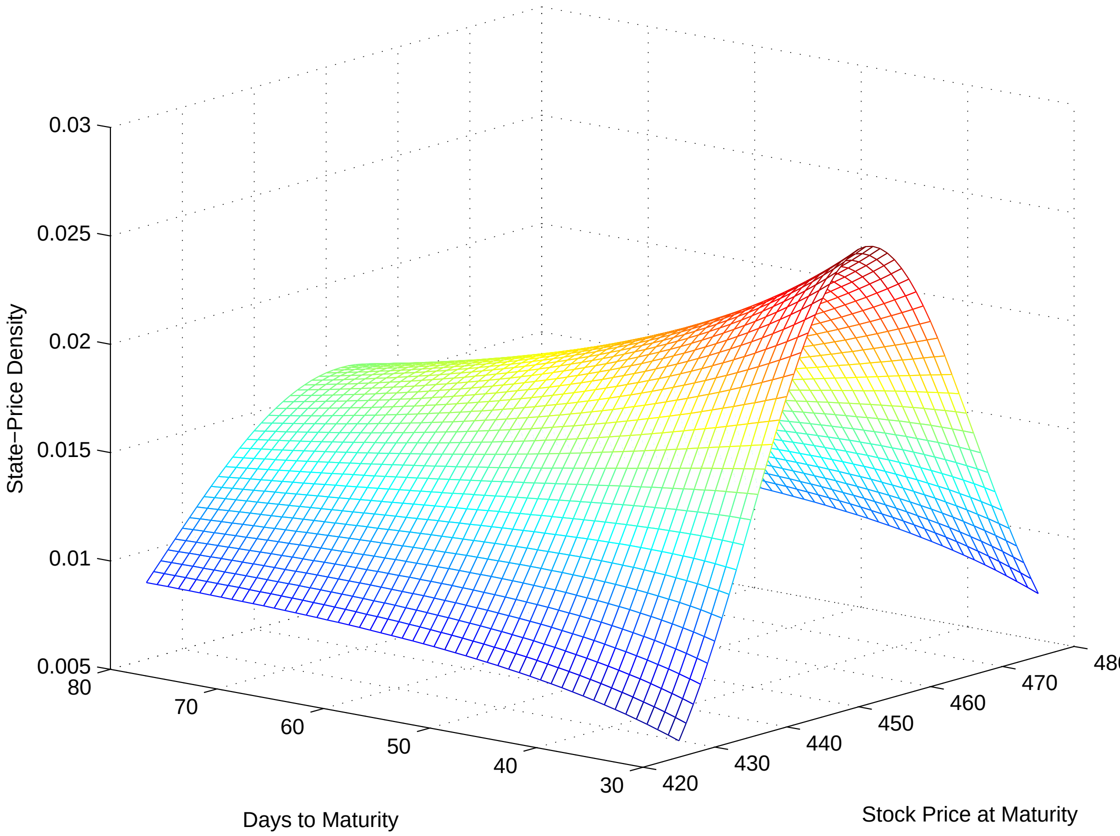


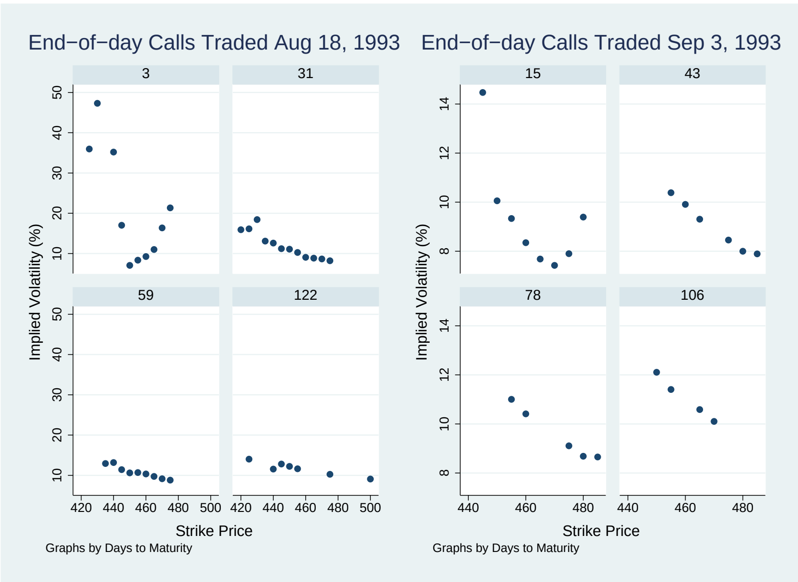
Figure 2 is the geometric representation of Equation (4), where $S_t = 450, r = 3\%,$ and $\sigma = 10\%$. This may be thought of loosely as the market-implied likelihood of the index price falling in a particular range for a contract with a particular number of days to maturity.

However, it is well known that the assumption that stock prices follow a geometric Brownian motion with constant variance is misspecified. This may be observed as follows. First, suppose that the stock price does follow a geometric Brownian motion.

Then, by no arbitrage, and assuming constant volatility, the price of a call option must be as described.

Then from observations of option price data in the market, we observe each variable in Equation (3) apart from σ. The implied volatility is simply the unique value of σ such that Equation (3) holds. Since we observe data on a cross section of options, the implied volatility may be calculated from each individual transaction. Given that Equation (3) is true, the implied volatility should be the same across time to maturity and strike price. However, in practice, we observe all sorts of shapes when graphing the implied volatility against strike price: volatility "smiles", "smirks", and "sneers".

Figure 3: Volatility "smile" for different days to maturity and strike price



Notice that for options traded on both Aug 18 and Sep 3, there is an apparent smile for the options with the closest maturity date, which flattens and becomes more of a "sneer" for longer times to maturity. These two dates were chosen randomly as illustrative examples.

1.3 Computation of Implied Volatility

A well known way to numerically solve an equation in one unknown is by using the Newton-Raphson Method. In general, an initial guess is first made for the value of the variable being solved for, x_0. The algorithm proceeds by making a subsequent guess x_1, which is the intersection between the tangent line from a first guess and the x axis. That is:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \tag{5}$$

The algorithm proceeds through subsequent guesses of x until the difference between subsequent guesses is sufficiently small, or $abs(x_{n+1} - x_n) < a$, for some threshold value a.

Before initiating the computation of implied volatility, the data must be filtered of observations for which the implied volatility cannot be calculated or for which the calculation may be wrong. First, the actual option price must be less than the Black-Scholes price of an option with the same strike price, time to maturity, and interest rate, with a volatility that approaches zero. That is:

$$C_{BS}(S_t, K, T, r_{t,T}; \sigma_{0^+}) \geq C(S_t, K, T, r_{t,T}; \sigma) \tag{6}$$

This is clear from the fact that volatility must always be greater than or equal to zero and the realization that an increase in volatility decreases the Black-Scholes value of the option, C_BS. The filter uses a positive number close to zero rather than zero, since the Black-Scholes formula for the price of a call option includes the d1 and d2 terms, which both require dividing by the volatility (Equation (3)).

To assess how well the method for calculating implied volatility performs, I run a simulation. A sample dataset of 20,160 options is simulated, corresponding to 80 options per day, with various strike prices and times to maturity, over a 252 day year (see section 2.2 for details on the data creation process). I then price the simulated options according to the Black-Scholes formula, all with $\sigma = 0.10$. Now that there is a dataset of options with strike price, time to maturity, stock price, option price, and a constant interest rate of 3%, the implied volatility may be calculated. Since the option prices were generated according the Black-Scholes assumptions, the implied volatility should just be 0.10, given that the method for calculating implied volatility is valid. It is useful to test the method first using simualted data, rather than actual data, since the true value of σ is known with simulated data. For the actual SP500 options data, the true value of σ is unkown (given that the stock process follows a Brownian Motion with constant volatility) or does not exist (given that the stock process does not follow a Brownian Motion with constant volatility).

With any gradient method of optimization, such as Newton-Raphson, one must always start with an initial guess for x_0. There is no obvious value to start at, although a guess between 0% and 70% is reasonable (as discussed, all options with implied volatility of greater than 70% are filtered out due to the noisiness of such data). For my computations, I repeat the optimization over initial values σ_0 of 0.1%, 1%, 2%, 3%, and 5%. If the algorithm fails to converge for a given initial guess, I move to the next one.

Implied volatility is then calculated with a threshold set to 10^{-6} (Table 1). At the start, 131 observations are filtered which violate Equation (6). After calculating the implied volatility, of 20,029 remaining options, 19,526 were calculated correctly (within 1% of the true value), while 634 had errors of greater than 1%. Of the 634 bad options, 131 had been filtered out already, leaving 503 remaining bad estimates. All 131 options that were filtered out, would have given an incorrect volatility estimate had they been left in. The mean error of the unfiltered options with errors greater than 1% is 2.6%. After repeating the experiment for a number of different threshholds, the optimal threshhold is found to be 10^{-12} (Table 2). Only 107 options are estimated incorrectly, or 0.5% of the filtered population, with a mean error of 1.6%.

Table 1: Accuracy of Implied Volatility Calculations, Threshold= 10^{-6}

Filtered?	$\sigma \sim= 0.10$	$\|0.10 - \sigma\| > .001$	Total
No	19526	503	20029
Yes	0	131	131
Total	19526	634	20160

Table 2: Accuracy of Implied Volatility Calculations, Threshold= 10^{-12}

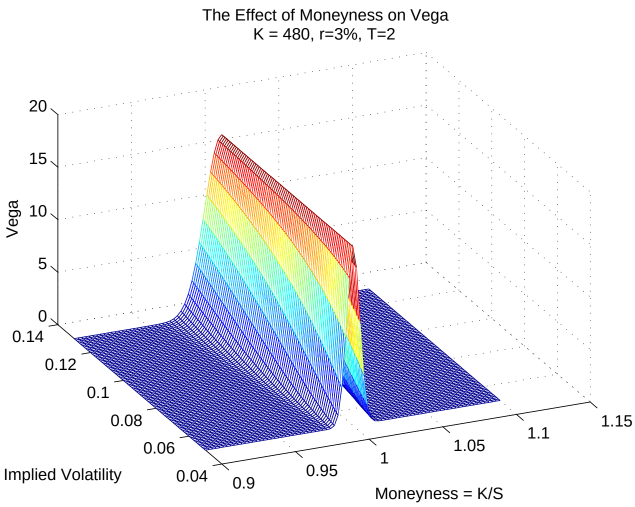
Filtered?	$\sigma \sim= 0.10$	$\|0.10 - \sigma\| > .001$	Total
No	19922	107	20029
Yes	0	131	131
Total	19922	238	20160

Looking to explain the implied volatilities estimated with error, I examine the relationship between moneyness and probability of having an incorrect estimate for implied volatility. All poor estimates have moneyness, K/S, such that $K/S < 0.94$

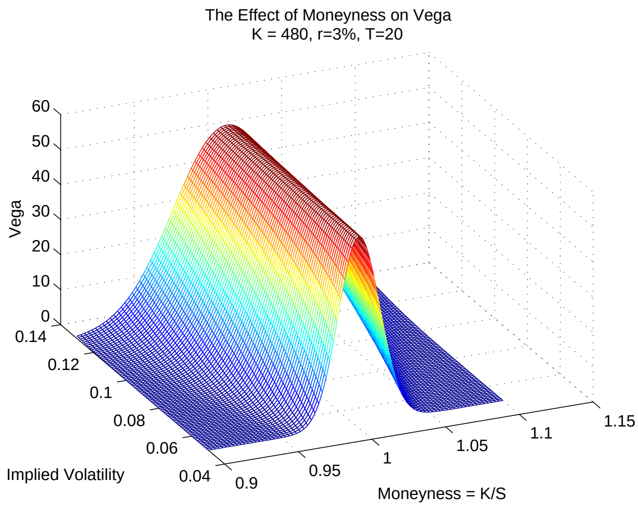
or $K/S > 1.06$. The filter from Equation (6) eliminates 131 of these poor estimates, and those that it eliminates are all deep in the money options. This relationship between distance from the money and probability of having implied volatility estimated incorrectly may be explained by an understanding of how implied volatility is calculated. In a dataset of option characteristics and prices, implied volatility is calculated by taking the characteristics of an option, excluding the price, and plugging in these characteristics along with guesses for σ into the Black-Scholes formula for the price of a call option. For at the money options, a small change in volatility may make a big difference to the probability that an option will be exercised and the expected profit given that it is exercised. However, for far from the money options, along with low volatility of the underlying stock, the option will almost certainly be called in the case of in the money options (ie.; the option is like cash), and the option will almost certainly never be called in the case of out of the money options. When volatility is low, small changes in the volatility will have only minor impact on the Black-Scholes implied price of these options. For example, the impact of a one percentage point change in volatility may only have a \$$10^{-100}$ impact on the Black Scholes implied price of the option. Then the limit of the largest generally available numeric storage type, "double", begins to bind. That is, the precision of double is $1.4x10^{-16}$, with a closest to 0 without being 0 value of 10^{-323}.

The derivative of the Black Scholes option price with respect to σ, known in finance as vega, is:

$$\text{vega} = \frac{\partial C}{\partial \sigma} = S_t N'(\frac{\ln(S_t/K) + (r_{t,T} - \delta_{t,T} + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}})\sqrt{T-t} \tag{7}$$



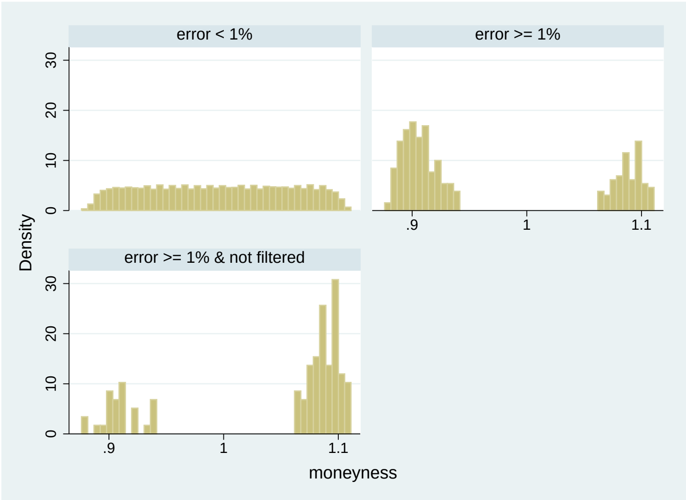
(a) T = 2 Days



(b) T = 20 Days

Vega is highest for at the money options and declines sharply with distance from the money. For options with moneyness outside [0.95,1.05] when T=2, vega is very tiny, which becomes an issue for the Newton-Raphson algorithm of calculating implied volatility. Recall that the algorithm updates with: $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$. In the case of the Black-Scholes price of a call option $f'(x_0)$ is defined as vega. The reciprocal of a very tiny number is very large and at some point, becomes too large for modern computer programs to store. Vega increases with time to maturity, so the problem arises with options that are both far from the money and have a short time to maturity.

Figure 4: Error in Implied Volatility Estimates, by Moneyness



Of the 20,160 implied volatility estimates, 238 were greater than 1% away from the true value. All of these poor estimates correspond to deep in the money or deep out of the money options, with moneyness, K/S, such that $K/S < 0.94$ or $K/S > 1.06$. The filter from Equation (6) elimates 131 of these poor estimates, and those that it eliminates are all deep in the money

1.4 Nonparametric Methods

1.4.1 Kernel Regression

One method to estimate the state-price density is to estimate it non-parametrically, without making any parametric restrictions on the underlying stock-price dynamics. This method was first proposed by Aït-Sahalia and Lo (1998). In contrast to parametric estimation, which makes certain assumptions on the functional form of the option pricing function, and then estimates those parameters, nonparametric estimation essentially involves taking local weighted averages across the observed data. Parametric models are relatively rigid, in that they specify the same functional form across all data. In contrast, a nonparametric model allows greater flexibility, by using only data in the neighborhood of the point being estimated to perform the estimate. Thus, the shape of the estimated function in one neighborhood may be vastly different from the shape of the estimated function in another neighborhood.

The primary method of nonparametric estimation is kernel regression. This estimation technique estimates the value of a dependent variable Y given a set of regressors, X. More precisely, if Y and X are jointly distributed, then the desired object is the conditional expectation, $\mu(x) = E(Y|x)$ given a set of data. This may be generally modeled as:

$$Y = g(X) + \epsilon$$

Then the conditional expectation of Y is [4]:

$$\begin{aligned} E(Y|X = x) &= \int y f_{Y|X}(y|X = x) dy \\ &= \int y \frac{f_{Y,X}(y, x)}{f_X(x)} dy \\ &= \frac{\int y f_{Y,X}(y, x) dy}{f_X(x)} \end{aligned}$$

where $f_{Y,X}$ is the joint density of Y and X and $f_X(x)$ is the marginal density of X. Since only a finite sample of x is observed, the integral is estimated by summation

[4] Bierens, Herman J. 1987, Kernel Estimators of Regression Functions, *Advances in Econometrics: Fifth World Congress* 1, 99-144.

over the observed x. This is the Nadaraya-Watson nonparametric estimator, which simplifies to:

$$E(Y|X=x) = \frac{\sum_{i=1}^{n} y_i k_i}{\sum_{i=1}^{n} k_i}, \; k_i = k(\frac{x - x_i}{h}) \tag{8}$$

where h is the bandwidth parameter and k is the kernel function.

Like in ordinary least squares estimation (OLS) and maximum likelihood estimation (MLE), the estimator is characterized by an objective function. The estimator brings the estimated function as close to the true data as possible, in the sense of mean-squared-error. That is, it solves the estimating equation: [5]

$$\sum_{i=1}^{n} k_i(y_i - \hat{\mu}_h(x)) = 0, \; \mu(x) = E(Y|x)$$

An alternative, more intuitive, interpretation of Equation (8) is as follows. Consider some *value*, x, of the explanatory variable, X. If this were linear regression, for example $E(Y|X=x) = 2x$, then for a value x of 5, the predicted value of Y would be $2 * 5 = 10$. With the Nadaraya-Watson estimator, the predicted value of Y is calculated by taking a weighted average of the y_i's, with weights greater for the associated x_i's that are close to $x = 5$. The kernel function $k(.)$ is applied to each datapoint in the sample, with the resulting weight for the i'th (x_i, y_i) pair being k_i.

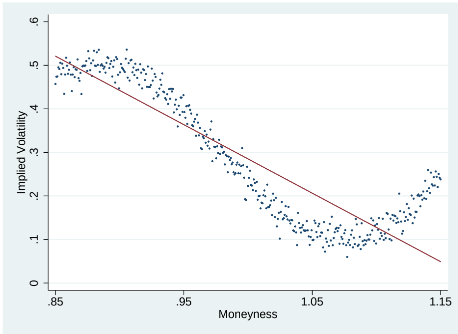
It is by no means obvious how to choose a particular kernel function. A kernel function must integrate to one and must be symmetric about $x = 0$. They generally give more weight to the center than to points far away, although this is not required. The normal probability density function is often used, since it has nice properties (it is smooth, continuous, differentiable, etc). Alternatives are the uniform kernel, triangular kernel, and Epanechnikov kernel, as seen in Figure 6. The uniform kernel equally weights observations near x_i and gives no weight to observations far from x_i (with closeness determined by the bandwidth parameter), which is undesirable. The triangular kernel is not differentiable at 0, which is undesirable. While the Epanechnikov kernel has nice mathematical properties strictly within the domain, it is not differentiable at the endpoints. The Gaussian kernel is chosen due to its nice properties, it is differentiable everywhere, and also because it performs better

[5] Davidson, Russell, and James G. MacKinnon, 2003. *Econometric Theory and Methods* (Oxford University Press, USA).

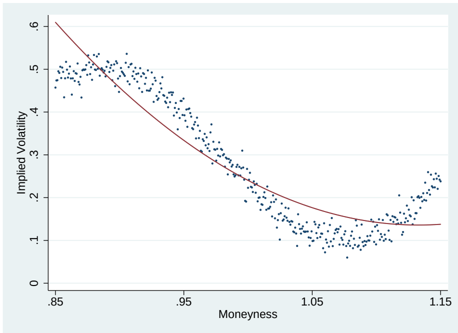
in situations where there are relatively large gaps between datapoints. That is, an estimate may still be made for a value of x that is between two datapoints that are far away, while the Epanechnikov kernel would not be able to compute an estimate for a value of x that is not within the bandwidth of an actual datapoint.

Figure 5: Comparison of Parametric Estimation and Nonparametric Estimation

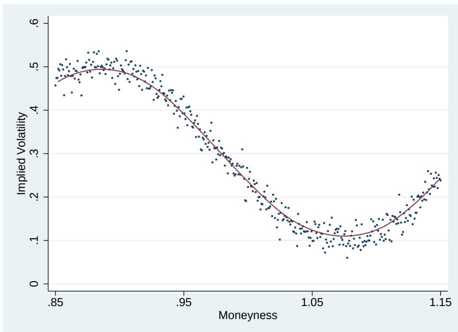
(a) Linear Regression



(b) Quadratic Regression

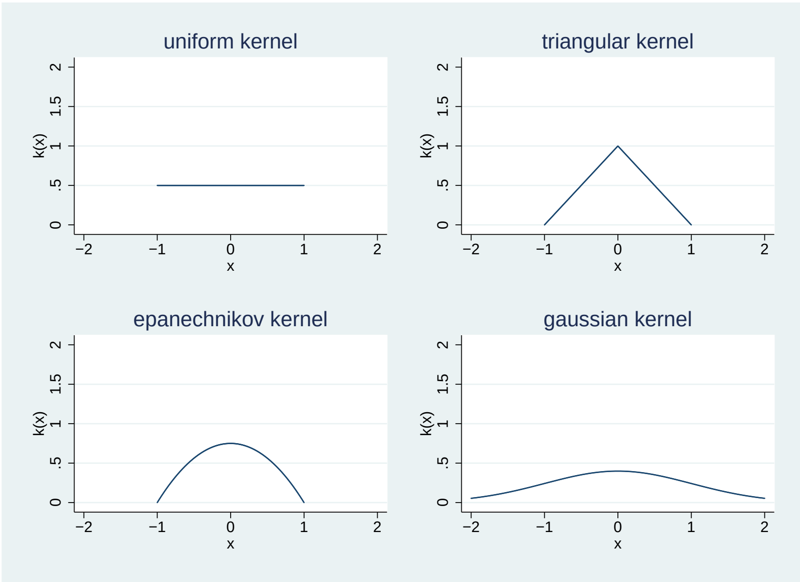


(c) Nonparametric Kernel Regression



Data is simulated in order to illustrate the effect of model mis-specification and how nonparametric estimation overcomes this. The linear regression model of panel (a) is clearly mis-specified. In panel (b), a quadratic specification may fit the data well for options with moneyness between 0.95 and 1.15, however it cannot accommodate the change in behavior for options with moneyness between 0.85 and 0.95. In panel (c), the fit is very good and accommodates the change in behaviour.

Figure 6: Different Kernel Functions



These are all examples of kernel functions. They integrate to one and define how to weight observations of x. The respective equations are as follows, with I() being the indicator function, equal to one if the statement is true and zero otherwise. The definitions of the kernels are:
uniform kernel: $k(x) = \frac{1}{2}I(|x| \leq 1)$
triangular kernel: $k(x) = (1 - |x|)I(|x| \leq 1)$
Epanechnikov kernel: $k(x) = \frac{3}{4}(1 - x^2)I(|x| \leq 1)$
Gaussian kernel: $k(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$

In order to be useful for nonparametrically estimating the state-price density, the estimator must be multidimensional, since the SPD is a function of five variables: $S_t, K, R_{t,T}, \delta_{t,T}$, and T. The framework is the same as in the one dimensional case, but the objective is now to estimate the conditional expectation of Y, conditioned on a *set* of regressors, rather than just one. Then the multivariate kernel considers the joint probability distribution of Y along with the vector of regressors, X. The result is the multidimensional Nadaraya-Watson estimator:

$$E(Y|X = x) = \frac{\sum_{i=1}^{n} y_i K_i}{\sum_{i=1}^{n} K_i}, \;\; K_i = K(\frac{\mathbf{x}\text{-}\mathbf{x}_i}{\mathbf{h}}) \tag{9}$$

where the multivariate kernel is generally chosen to be the product of the univariate kernels.[6]

$$K(x) = k(x_1) \cdot k(x_1) \cdots k(x_q)$$

1.4.2 Nonparametric Estimation of the State Price Density

Now armed with the estimator of Equation (9), it may be applied to the call option pricing function. Once the call option pricing function has been modeled, the SPD may be recovered. The fully nonparametric call option pricing function outputs the price of a call option given the set of inputs $\{S_t, K, T, r_{t,T}, \delta_{t,T}\}$ and returns the price of the associated option. Plugging the five explanatory variables into Equation (9) yields:

$$\begin{aligned} E(C|S_t, K, T, r_{t,T}, \delta_{t,T}) &= \hat{C}(S_t, K, T, r_{t,T}, \delta_{t,T}) \qquad (10) \\ &= \frac{\sum_{i=1}^{n} C_i \cdot k_S(\frac{S_t - S_{t_i}}{h_S}) k_K(\frac{K - K_i}{h_K}) k_T(\frac{T - T_i}{h_T}) k_r(\frac{r_{t,T} - r_{t_i,T_i}}{h_r}) k_\delta(\frac{\delta_{t,T} - \delta_{t_i,T_i}}{h_\delta})}{\sum_{i=1}^{n} k_S(\frac{S_t - S_{t_i}}{h_S}) k_K(\frac{K - K_i}{h_K}) k_T(\frac{T - T_i}{h_T}) k_r(\frac{r_{t,T} - r_{t_i,T_i}}{h_r}) k_\delta(\frac{\delta_{t,T} - \delta_{t_i,T_i}}{h_\delta})} \end{aligned}$$

While it is theoretically possible to run this estimation, it is computationally highly expensive. It is also unnecessary. The computational expense is due to the mechanics of kernel regression. In contrast to linear regression, for example, in which the whole function is estimated during the estimation process; the Nadaraya-Watson estimator is a pointwise estimator, in which each point is estimated separately. Consider the example of data along the line $y = x$. Linear regression involves estimating $\hat{\beta}$ from $\hat{y} = \hat{\beta}x$. The result is a *function.* In contrast, kernel regression involves estimating a *point*, $\hat{y}_i$, given the set of data. In the linear regression case, the line $\hat{y} = \hat{\beta}x$ is recovered. In the case of kernel regression, a set of estimated values is recovered, only as fine as the mesh over which the kernel regression was run over.

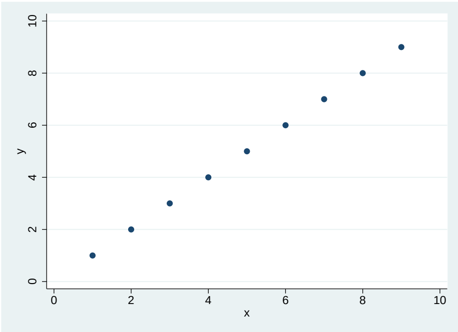
Figure 7 illustrates two distinctions. First, comparing the linear estimation from panel b with the nonparametric estimation from the lower panels, the difference between a functional estimator and pointwise estimator is seen. A functional estimator estimates the entire function, while a pointwise estimator only estimates individual points. Second, the grid of points to be estimated may be any set within the domain of the data. It may be coarse, it may have holes, it may be regular or irregular and

[6] Note that here the subscript on x denotes the particular regressor x, while in Equation (9), the i subscript denotes the i'th *observation*

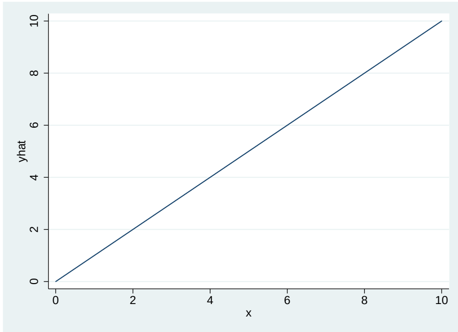
it may or may not include actual data points. For example, there is an estimate of y for x=0.3, even though there is no actual data point that takes this value.

Figure 7: Functional vs Pointwise Estimation

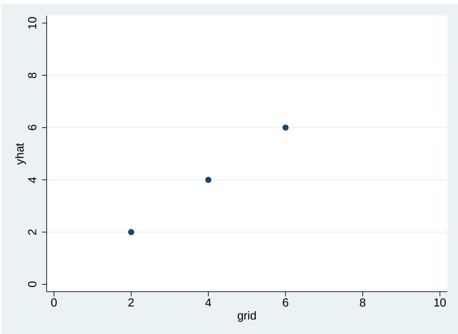
(a) Raw Data



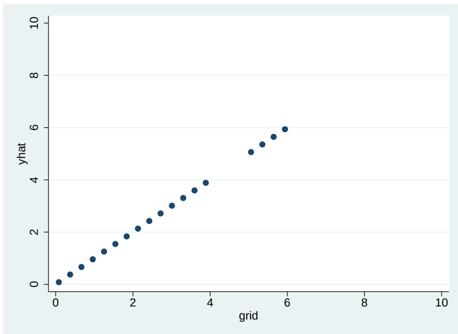
(b) Estimated Function (Linear Regression)



(c) Nonparametrically Estimated Points



(d) Nonparametrically Estimated Points



It is also unnecessary to estimate Equation (10). Instead of estimating the fully nonparametric estimator, methods to reduce the dimensionality may be employed. Aït-Sahalia and Lo (1998) show how the results from a fully nonparametric estimator are virtually indistinguishable from the results of a *semiparametric* estimator, in which the Black-Scholes option pricing function is used, and is a function of a nonparametrically estimated volatility surface. Local differences in shape are then captured through the estimated implied volatility function.

This semiparametric estimator is much faster, since only three dimensions need be estimated. The implied volatility surface is estimated nonparametrically as a function of the current stock price, strike price, and time to maturity. This results in a significant reduction in estimation time. To understand this, recall that the Kadaraya-Watson estimator is a pointwise estimator. That is, values are estimated over a grid. With one explanatory dimension (ie: explanatory variable x and estimated y), a grid with ten intervals would require ten estimations. For example, if the range of x is [1,10] the estimated points may be $\{1, 2, ..., 10\}$. In two dimensions (ie: y, x_1, and x_2), a grid with ten intervals would require 100 estimations. Eg: $\{x_1 = 1, x_2 = 1\}, \{x_1 = 1, x_2 = 2\}, ..., \{x_1 = 1, x_2 = 9\}, \{x_1 = 1, x_2 = 10\}, \{x_1 = 2, x_2 = 1\}, \{x_1 = 2, x_2 = 2\}, ...\{x_1 = 10, x_2 = 9\}, \{x_1 = 10, x_2 = 10\}$. Thus, the difference in speed between a three-dimensional semiparametric estimator and five-dimensional fully nonparametric estimator is 10^3 vs 10^5 for a 10x10x10... grid. In practice, this may be far too coarse. If instead a grid of 100 intervals is estimated, this is the difference between 100^3 vs 100^5. That is, the fully nonparametric would be 1000 times slower.

An important component of the Nadaraya-Watson estimator is the choice of bandwidth. That is, the estimation technique is never truly nonparametric, since the bandwidth parameter must be chosen. There have been many propositions as to the optimal choice of bandwidth, but there is no singular best method. For example, the method of Hall, Sheather, Jones, and Marron (1991) has been shown to be optimal, but only in large samples. Others have better finite sample properties, but may be biased. Since the focus of this paper is not on the intricacies of bandwidth selection, I use an accepted optimal bandwidth selector of $sd_x * N^{-1/5}$, where sd_x is the standard deviation of variable x.[7] This is derived by minimizing the asymptotic mean squared error of the Nadaraya-Watson estimator with respect to the choice of bandwidth.

[7]Li, Qi, and Jeffrey S. Racine, 2006 *Nonparametric Econometrics: Theory and Practice* (Princeton University Press)

2 Simulation

Before applying the estimation methodology to observed market data, I first use simulated data, following the outline of A-S&L (1998). While actual market data contains noise and irregularities that must be dealt with through filtering, simulated data is generated exactly according to the specifications of the data generating process (DGP) developed for it. By starting with simulated data, the nonparametrically estimated option pricing function may be compared to the *true* option pricing function, since the data has been generated according to a *known* function. Given that the estimation technique satisfactorily recovers the option pricing function when it is known, it may then be tested against actual market data, where the existence of a true function is less obvious.

2.1 Calibration of the Stock Price Process

The stock price is simulated according to the Black-Scholes assumptions, with moments matching those of the observed market data. Given the set of historical data, moments are estimated as follows. Begin with the assumption that the stock price follows a Brownian Motion with constant drift μ and volatility σ. Then:

$$\ln(S_t) = \ln(S_{t-1}) + (\mu - \frac{1}{2}\sigma^2)dt + \sigma\epsilon\sqrt{dt},\ \epsilon \sim N(0,1) \tag{11}$$

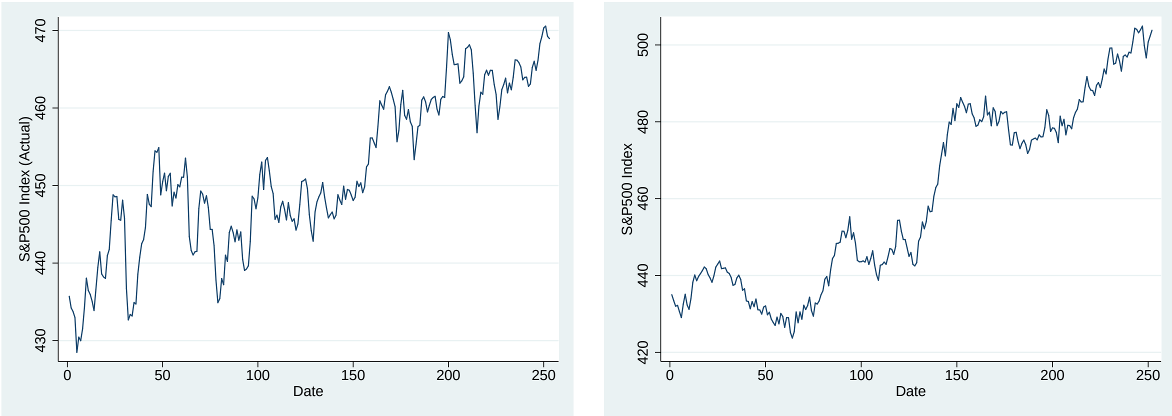
The volatility, σ, and drift, μ, of the Brownian Motion are then calculated from:

1. Generate $u_t = \ln(\frac{S_t}{S_{t-1}})$ as the continuously compounded daily return between day t and $t-1$
2. Mean return is then calculated as $\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i = 0.00030$, or $0.00030 * 252 =$ 7.5% annualized with a 252 day year
3. The standard deviation of the daily return is then $\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(u_i - \bar{u})^2} =$ 0.0050, or $0.0050 * \sqrt{252} = 7.9\%$ annualized
4. Then from Equation (11), the drift term is $\mu = \bar{u} + \frac{1}{2}\sigma^2 = 0.00031$

Given $\mu = 0.00031$ and $\sigma = 0.0050$ as calculated above, the stock price process is simulated. The initial index level is set to 435, and the interest rate is held constant at 3%.

Figure 8: S&P500 Historical and Simulated Index Levels (1993)

(a) S&P500 Historical Index Level (1993) (b) S&P500 Simulated Index Level (1993)



2.2 Options Data

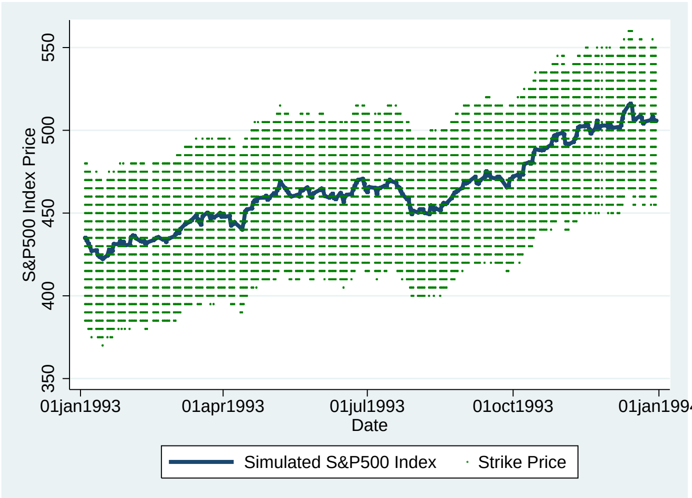
Call option data is simulated according to the CBOE conventions for introducing options to the market. In the actual data, there is an average of 80 unique contracts per day, which includes puts and calls with the same terms. This yields 58 unique call prices including both actual call prices and call prices implied by put prices. While the actual data has holes, ie: there may have been bid and ask prices for a particular contract, but no completed transactions, the simulation process is simplified by assigning a price to each contract available to be traded. There are 80 contracts available per day, with new contracts becoming available and old contracts that stop being traded depending on how the stock price changes.

1. Options expire on the third Saturday of the month of expiration
2. For every strike price, options are available for the next 4 maturity dates (these must be at least two days from the current date)
3. On every day, for each maturity date, there are 20 contracts available with strike prices in multiples of 5, nearest the current strike price. Of these, 10 are out of the money and 10 are in the money.

This yields a dataset of 20,160 contracts. The set of available contracts is seen in Figure 9.

Given the assigned characteristics of each contract, namely: the strike price, time to maturity, interest rate, stock price, and the true volatility of the underlying index price; the options are priced according to the Black Scholes formula. Indeed, since the data is constructed according to a Black-Scholes world, the BS assumptions are satisfied and the option prices are valid.

Figure 9: Simulated Set of Option Contracts Available over Time



20,160 contract prices are simulated according to the specification described in Section 2.2. There are 10 in-the-money and 10 out-of-the-money call options available each business day of 1993. A business calendar was used to exclude weekends and holidays to ensure that the characteristics of the simulated data match those of the actual data.

2.3 Estimation of Simulated Option Prices

Once the simulated dataset is generated, the semiparametric estimation technique of Section 1.4.2 may be employed. First, implied volatility is calculated according to the Newton-Raphson method and filtering technique discussed in Section 1.3. This brings the dataset from 20,160 to 20,029 contracts. Second, the volatility surface is estimated nonparametrically, as seen in Figure 11. Third, the estimated volatility surface is plugged back in to the Black Scholes option pricing function, completing the estimation of the simulated option prices.

While the true volatility surface should be flat at 10%, volatility is over-estimated for a small subset of options that are far out of the money with a short time to maturity (and to a lesser extent, for options far in the money with short time to maturity). This is a result of the calculation of implied volatility and not of the nonparametric estimation technique. Clearly the estimation technique performs very well, recovering 99.5% of the volatility parameters within 1% accuracy. Moreover, the parameters with error correspond to options whose prices are very insensitive to changes in volatility. Thus, although the volatility is estimated with error, the option price is still estimated correctly for all options. For example the estimated price of a deep-in-the-money option may be the same (to an accuracy of 10^{-100}) whether the volatility is 10% or 13%. Thus, although the true implied volatility is 10%, the price of the option is still correctly estimated using the wrong implied volatility of 13%.

Table 3: Summary Statistics for Simulated S&P 500 Index Options Data

Variable	Mean	Std. Dev.	Min	5%	10%	50%	90%	95%	Max
Call Price	17.99	17.10	0.00	0.04	0.38	12.17	45.63	50.61	62.52
Implied σ (%)	10.01	0.16	10.00	10.00	10.00	10.00	10.00	10.00	16.53
Days to Maturity	63.10	36.35	2.00	8.00	15.00	64.00	110.00	117.00	183.00
Simulated SPX Index	463.95	23.16	422.35	430.87	433.33	461.92	501.89	505.82	516.12
Strike Price	461.80	36.84	370.00	405.00	415.00	460.00	510.00	525.00	560.00
3 Month T-Bill Rate	3.00	0.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
Observations	20029								

Figure 10: Nonparametric Estimation of Simulated Black-Scholes Implied Volatility Surface

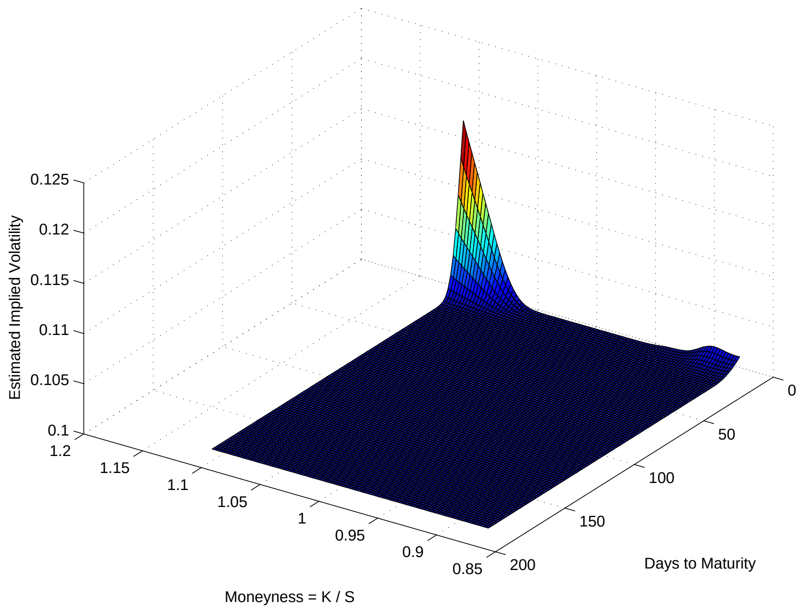
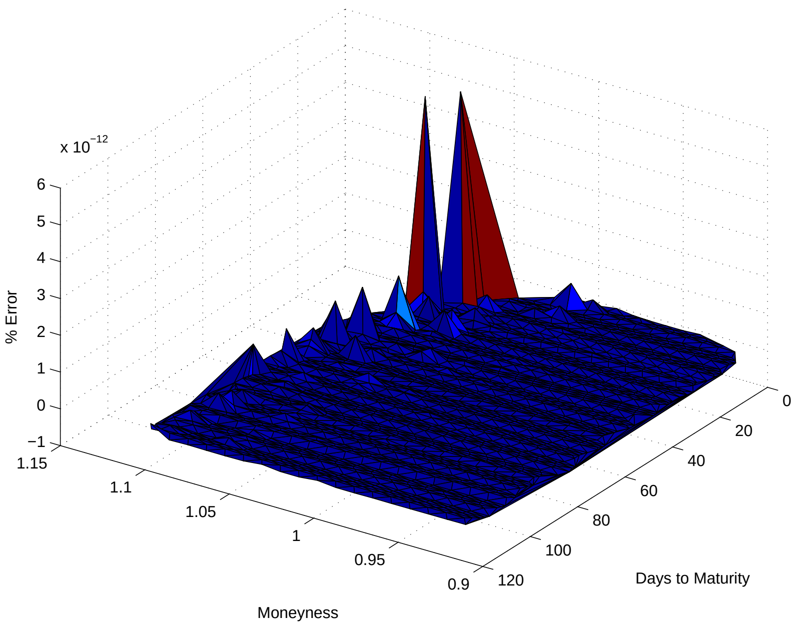


Figure 11: Error in Estimation of Simulated Option Prices



% Error is defined as $(\hat{C} - C)/C \cdot 100$. The maximum error is on the order of 10^{-12}, ie: essentially zero.

3 S&P 500 Data and Estimation

3.1 Data and Summary Statistics

Table 4: Summary Statistics for S&P 500 Index Options Data

Variable	Mean	Std. Dev.	Min	5%	10%	50%	90%	95%	Max
Call Price	5.62	8.54	0.06	0.19	0.38	3.25	12.38	18.25	121.25
Put Price	4.39	4.54	0.06	0.25	0.50	3.25	9.25	12.13	109.50
Implied σ (%)	12.31	3.99	0.85	8.56	9.13	11.46	16.16	18.63	69.82
Days to Maturity	39.18	46.93	2.00	3.00	4.00	26.00	82.00	124.00	362.00
SPX Index	451.71	10.02	426.88	434.33	438.04	450.19	464.80	466.46	471.28
Strike Price	447.00	19.12	330.00	415.00	425.00	450.00	465.00	475.00	550.00
3 Month T-Bill Rate	3.05	0.08	2.84	2.92	2.95	3.05	3.17	3.18	3.20
Observations	215993								

When Breeden and Litzenberger first derived the formula for the state-price density, Equation (1), in 1978, markets were much less complete than they are now. With the automation of the stock market, the greater range of options traded on the same stock, and the sheer volume of trades that occur every day, there is now a huge amount of data that may be exploited to recover information about investor preferences.

One year of transaction level data has been obtained from the Berkeley Options Database, including all completed trades and quotes for S&P 500 index call and put options between January 4, 1993 and December 31, 1993. The data includes all variables needed to estimate the option pricing function, and thus, the state-price density. Summary statistics are shown in Table 4

As is expected with a dataset of this size, it must be cleaned and filtered to remove errors. This is an intensive task in itself. Due to the sheer amount of data, it would be infeasible to inspect all data manually and filter out observations manually. Instead, a set of rules is applied on what constitutes a valid and reliable entry and all observations not meeting these rules are dropped.[8]

[8]Dacorogna, Michel, Ramazan Gencay, Ulrich A. Müller, Richard B. Olsen and Olivier V. Pictet, 2001. *An Introduction to High-Frequency Finance* (Elsevier Science)

1. Observations with implied volatility of greater than 70% are deemed to be outliers and are dropped.
2. Observations with price less than 2 ticks (\$1/8) are dropped, since these correspond to far out of the money options, for which prices are very noisy.
3. Prices of far in the money options are replaced with the prices implied by put-call parity, since the market for far in the money options is also illiquid. Actors in an illiquid market may be more sensitive to liquidity pressures, so the influence of whether a trade is buyer or seller initiated would tend to have a greater impact on price than in a liquid market.
4. Any remaining far in the money or far out of the money options are dropped: those with moneyness, K/S, such that $K/S < 0.94$ or $K/S > 1.06$.

A feature that distinguishes the actual data from the simulation is dividend payments from individual stocks within the index. The Black Scholes formula requires that if dividend payments are made, they must be known in advance. That is, either a constant dividend yield or a series of lump sum dividends are acceptable so long as there is no uncertainty as to the timing and value of the payments. To correct for this, data on cumulative dividends paid on the S&P500 Index to date is obtained. A dividend payout has the effect of decreasing the value of a call option. This is due to the expectation that the underlying stock decreases by the value of the dividend payout. To compensate for this, I adjust the stock price corresponding to each option by the total amount of dividends that are paid out over the option's lifetime. That is:

$$S_{t,\text{effective}} = S_t - \sum_{t=t_0}^{T} \delta_t \tag{12}$$

where δ_t is the total dividend payout on date t. While investors do not know the future dividend stream with certainty, the ex-post stream serves as a strong proxy for ex-ante beliefs. That is, investors are assumed to get it right on average. An additional practical benefit of correcting for dividends in this manner is that it reduces the fully nonparametric estimation method from five to four dimensions.

3.2 Estimation of the Implied Volatility Surface

Now with all data necessary to estimate the state price density cleaned, the SPD may be estimated. As with the simulated data, the semiparametric estimation technique of Section 1.4.2 is used to recover the call option pricing function. The semiparametric option pricing function is derived in Appendix A as:

$$\hat{C}(S_t, K, T, r_{t,T}) = \tag{13}$$

$$S_t\Phi(\frac{\ln(S_t/K) + (r_{t,T} + \frac{1}{2}(\frac{\sum_{i=1}^{n}\sigma_i A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)}{\sum_{i=1}^{n} A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)})^2)T}{(\frac{\sum_{i=1}^{n}\sigma_i A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)}{\sum_{i=1}^{n} A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)})\sqrt{T}})$$

$$-Ke^{-r_{t,T}T}\Phi(\frac{\ln(S_t/K) + (r_{t,T} - \frac{1}{2}(\frac{\sum_{i=1}^{n}\sigma_i A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)}{\sum_{i=1}^{n} A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)})^2)T}{(\frac{\sum_{i=1}^{n}\sigma_i A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)}{\sum_{i=1}^{n} A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)})\sqrt{T}})$$

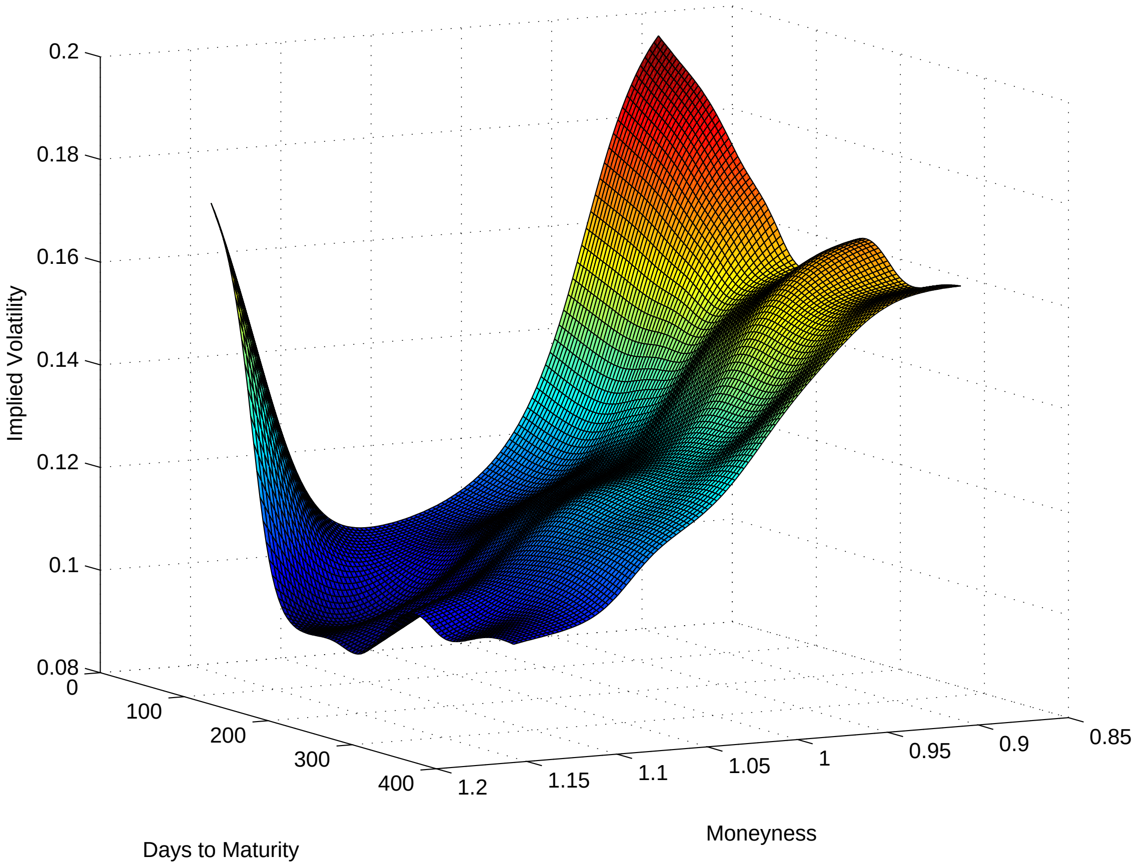
Once $\hat{C}$ is estimated, the state price density is simply $\partial^2\hat{C}/\partial K^2$. The estimation of the volatility over the full year of 1993 yields a surface, Figure 12, that is as expected. That is, there is the characteristic volatility "smile" for short maturity dates, and the smile flattens as time to maturity increases.

Figures 13 to 16 are shown to highlight the power of intraday data. It is much more common to estimate a volatility surface like Figure 12, where the surface is assumed to be non-stochastic and a time series of data is used to construct one surface. Using intraday data, there are enough options traded that an entire surface may be estimated limiting oneself to only trades that occurred that day. While the volatility surface may still change throughout the day, the magnitude of variation is expected to be much lower than the variation of the surface over an entire year.

The next four figures, referring to maturity dates of March 20th and December 18th, 1993 do make the assumption of a non-stochastic volatility surface, but only during the 40 days preceding each respective maturity date. This is still a more reasonable assumption than assuming that the volatility surface does not change throughout the entire year. There is more detail available in these samples for two reasons. First, there is regularity in the time to maturity variable. That is, the sample includes options with $T = 2, 3, 4, ..., 40$, rather than the limited set of times to maturity available from the November 5th data, for example, of $T = 15, 43, 78, ...$ Second, there is clearly a much higher number of trades that occur over any 40 days relative to one single day. This leads to a much finer estimated surface.

Figure 12: Estimated Implied Volatility Surface

1993 S&P500 Options Data



The surface is estimated based upon all out of the money calls and the implied call prices of all out of the money puts for all options traded in 1993. Data is filtered according to the conditions described in Section 3.1.

Figure 13: Implied Volatility from Data
Options Traded Nov 5, 1993 and $T < 100$

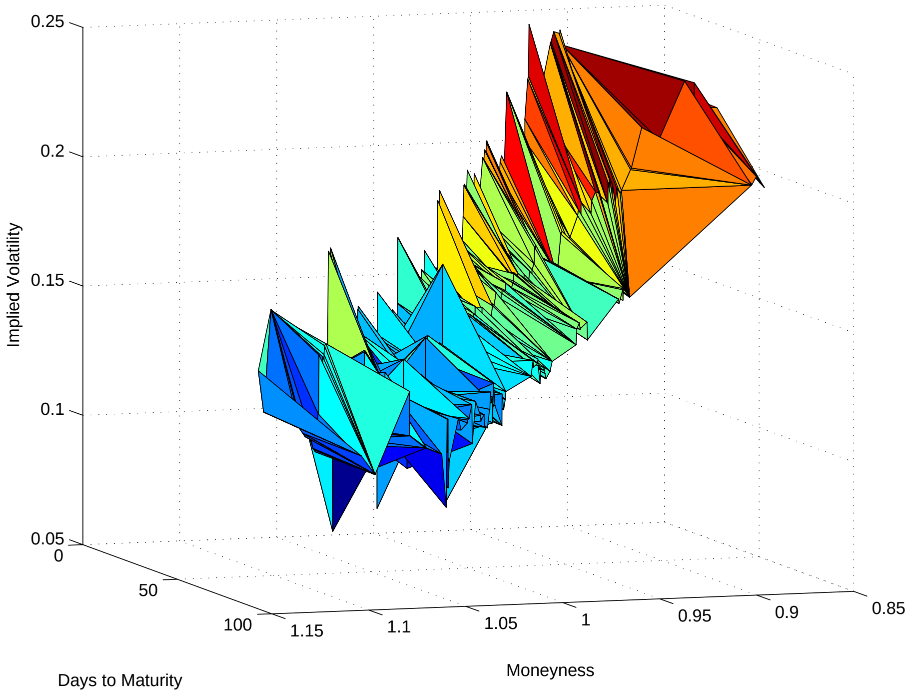
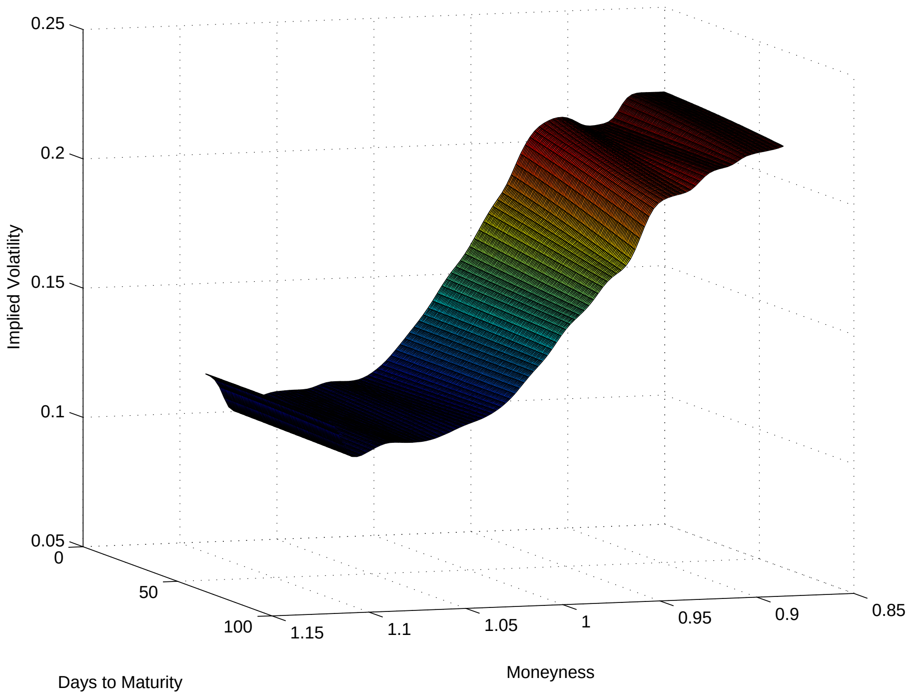


Figure 14: Estimated Implied Volatility Surface



The surface from November 5th is shown since it is the day with the most observations after filtering in a single day, with 722 observations (relative to a mean of 253 observations per day). It is difficult to estimate the volatility surface using one day of data due to the lack of regularity of data. That is, since options are only available with maturities on the third Saturday of the month, there is not very much information in the time-to-maturity dimension.

Figure 15: Implied Volatility from Data
Options Traded Apr 2, 1993 and $T < 100$

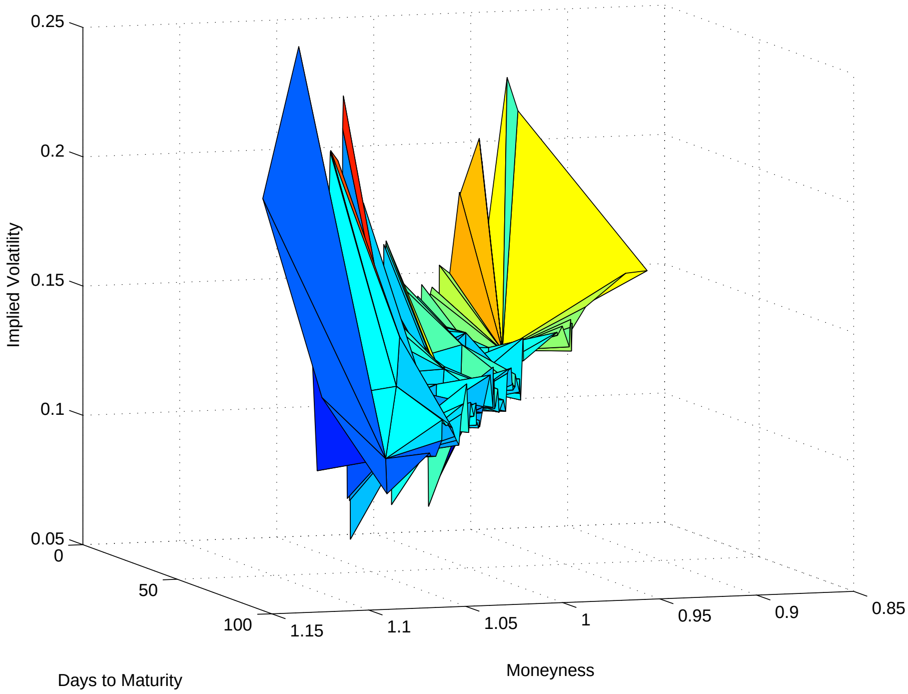
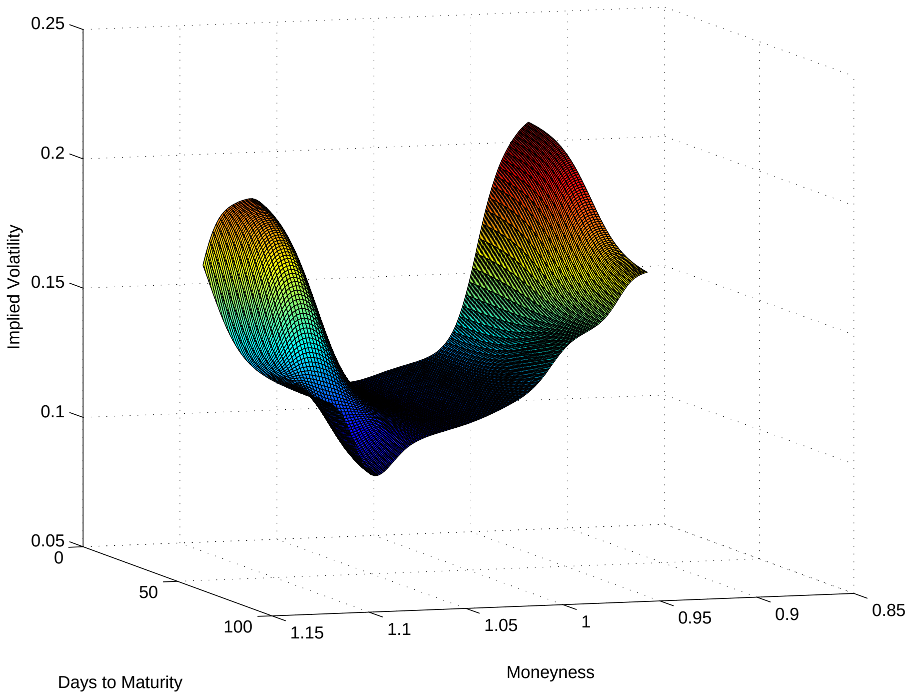


Figure 16: Estimated Implied Volatility Surface



April 2nd was the second highest trading day of 1993 (of filtered options), with a total number of observations of 703. The surface exhibits the same general characteristics as the surface much later in the year on November 5 over the area in common. (There were few far out of the money trades on April 2 relative to Nov 5, but more in the money trades on April 2 relative to November 5. This gives the surfaces different domains, but does not mean that the surfaces are different over the domain in common.) This may support the hypothesis that the state price density is stable over time, though a formal test would be be more conclusive.

Figure 17: Implied Volatility from Data

Options with Maturity March 20, 1993 and $T < 40$

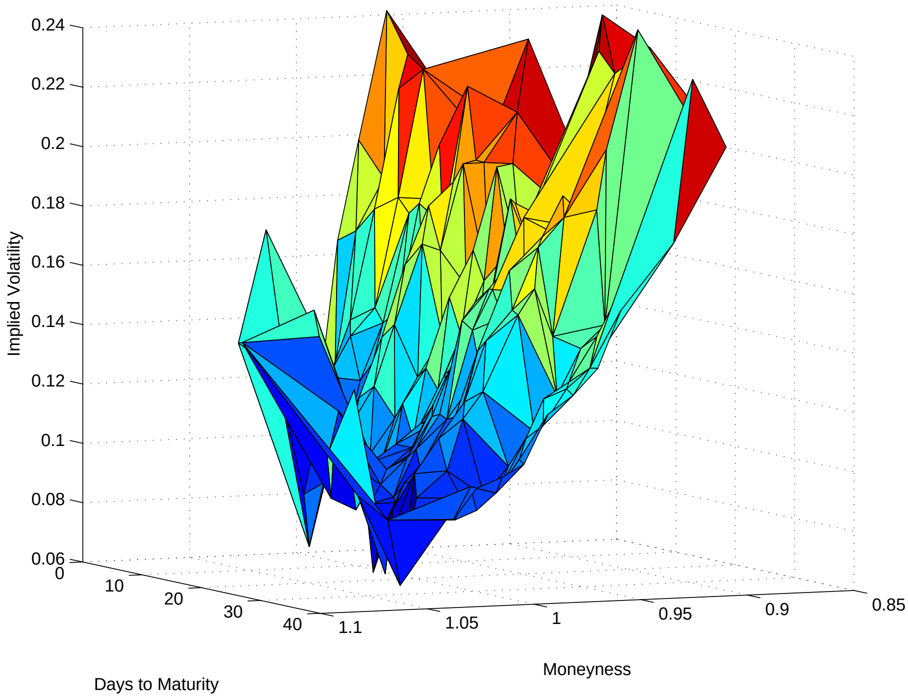
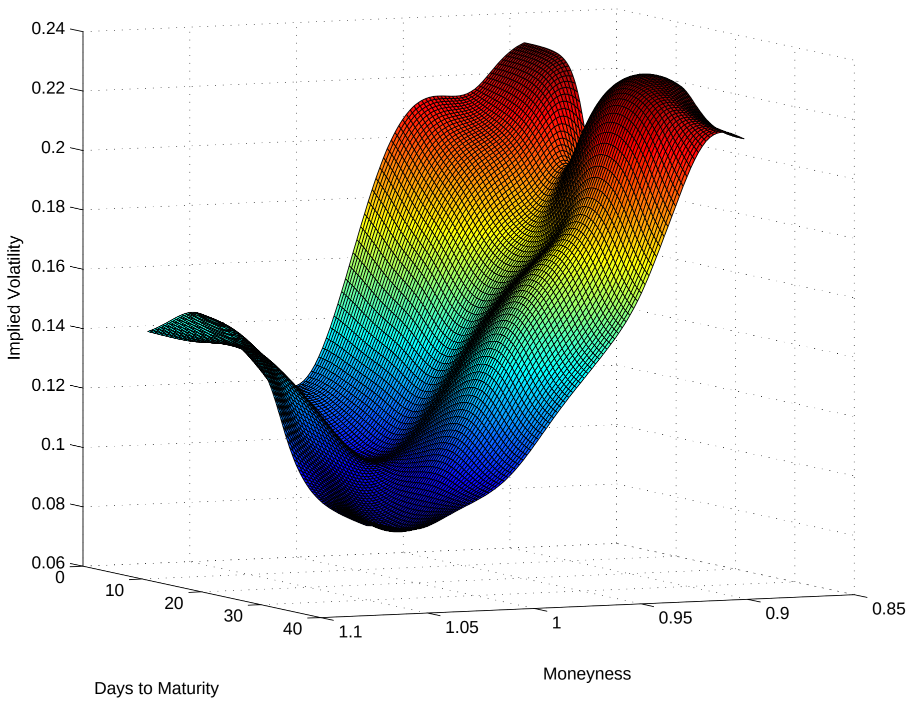


Figure 18: Estimated Implied Volatility Surface



The surface with maturity of March 20th is shown since it is the maturity date with the most observations after filtering, with 5,157 observations (relative to a mean of 3,537 observations per maturity date). With great variation over the time to maturity dimension, notice how the volatility "smile" fades to a "smirk" as T increases.

Figure 19: Implied Volatility from Data

Options with Maturity December 18, 1993 and $T < 40$

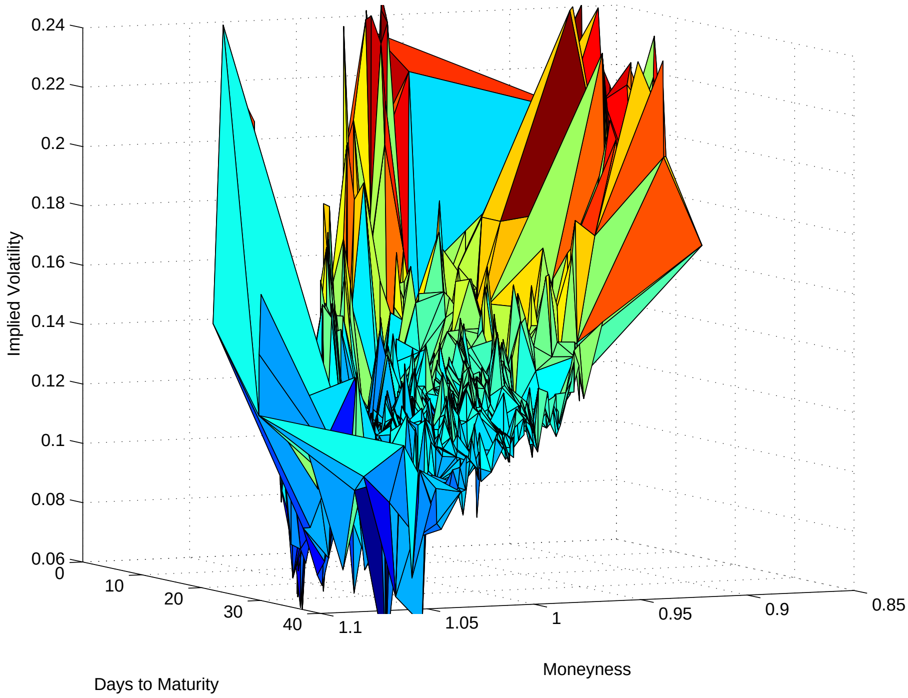
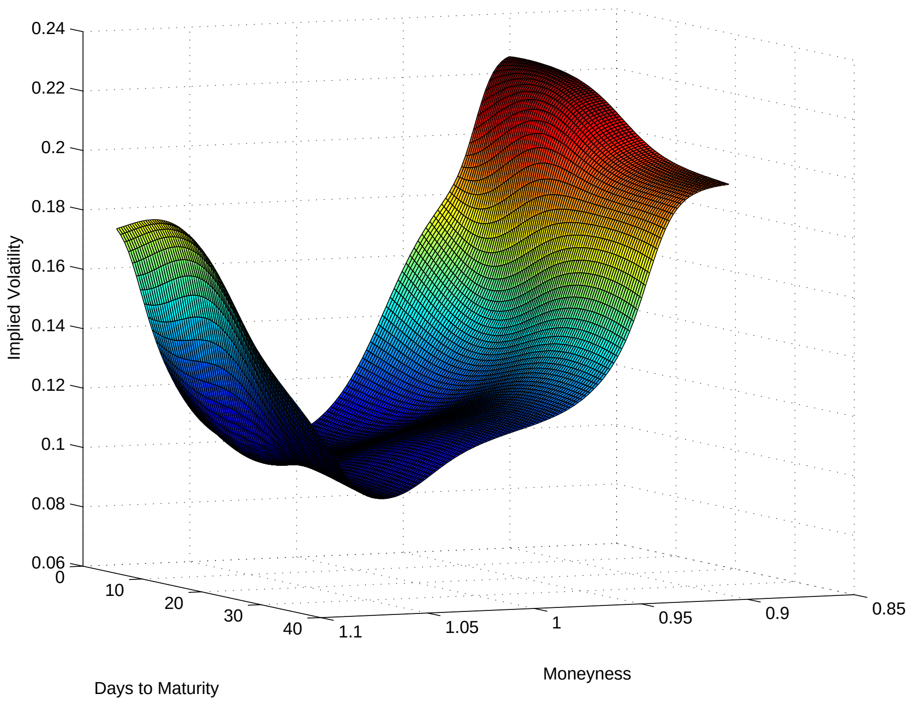


Figure 20: Estimated Implied Volatility Surface



There were a total of 4,186 observations included with a December 18th maturity date.

3.3 Estimation of the State Price Density

To illustrate how the state price density may be estimated in practice, Equation (13) is estimated, holding fixed a subset of the dependent variables. Rather than analytically solve for the second derivative, a simplified approach is to numerically differentiate $\hat{C}$ in order to recover the SPD. By definition, the derivative of a function $f(x)$, with respect to variable x is:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

This is then applied twice to the call option pricing, $\hat{C}(S_t, K, T, r_{t,T}; \sigma)$, in order to get the second derivative. First, the variables T and $r_{t,T}$ are fixed at constant values. $\sigma(S_t, K)$ is then estimated nonparametrically. Finally, $\hat{C}$ is estimated over a grid as a function of K, holding S_t fixed. The grid must be fine enough so as to avoid "the curse of differentiation", in which any noise or wrinkles in the original function are exacerbated when the derivative is taken.

Figure 21 illustrates the consequences associated with taking the derivative of an insufficiently smoothed, nonparametrically estimated function. The second derivative associated with the under-smoothed function does not have meaning.[9] Similarly, an over-smoothed function would miss potentially important local characteristics of the function. Figure 22 illustrates an optimally smoothed function, in which the bandwidth is set to $sd \cdot N^{-1/5}$ (where sd is the standard deviation of the variable to which the bandwidth corresponds.) It is difficult to distinguish between the under-smoothed and optimally-smoothed call pricing functions, however, the differences become apparent in the first derivative. By the second derivative, most of the resemblance between the two functions is lost.

Bandwidth selection is thus a very important aspect of nonparametric estimation and should not be overlooked. A small bandwidth parameter picks up small, local characteristics of the function being estimated, but is also more prone to being noisy. A relatively high bandwidth parameter does not suffer as much from outliers and noise in the data and it provides smoother estimates, but can miss important features of the function by smoothing them out. The motivation for using nonparametric estimation in the first place is to identify these local features. At the extreme, for

[9] This mathematical phenomenon was compared rather succinctly to a phenomenon of the animal kingdom by Professor Ted Neave. "In any event, a flea crawling over a warty elephant would find its skin much rougher than would, say, the mahout stroking the elephant."

the nonparametrically estimated implied volatility surface, a very high bandwidth parameter would lead to a flat estimated surface, completely smooth. This is just the same as the Black Scholes case, which assumes a constant implied volatility. Thus, the optimal bandwidth parameter is characterized by balancing the competing objectives of minimizing the impact of noise while also capturing important local characteristics of the observed data.

For SPDs derived using only data from transactions that occurred on a particular day, the SPD is often incomplete. That is, the SPD is estimated only over the range of terminal stock prices for which contracts with that strike price are available. So far example, on November 23, 1993 and for contracts with 25 days to maturity, contracts were only traded with strike prices between \$410 and \$490. As seen in Figure 22, the upper tail of the SPD is not estimated. This enforces the point that nonparametric estimation may not be used to estimate a model beyond the range of observed data. A potential solution would be to use bid-ask midpoints to estimate the density beyond the range of actual transactions. However, the midpoint between a bid and ask price may be a bad proxy for the value of a contract, particularly when the bid-ask spread is high. Moreover, it is for those contracts that are not traded that the bid-ask spread tends to be higher.

An alternative to using bid-ask midpoints to estimate the SPD beyond the range of traded contracts for a particular day is to make the assumption that the SPD is non-stochastic. This is the case with much of the literature, including Aït-Sahalia and Lo (1998) in which data over a whole year is pooled as a cross section in order to estimate SPDs. This assumption has been made to generate the estimates of Figure 23, in which the nonparametric SPDs are compared to the Black Scholes SPDs for different times to maturity. The cross sections are then stacked side-by-side in Figure 24 to generate the full SPD estimate across both the index price at expiration and time to maturity. The nonparametrically estimated SPD is quite similar to the Black Scholes SPD, as derived in Figure 4. This is expected and supports the idea that the Black Scholes SPD is a reasonable approximation, at least in the case of S&P500 index options traded in 1993.

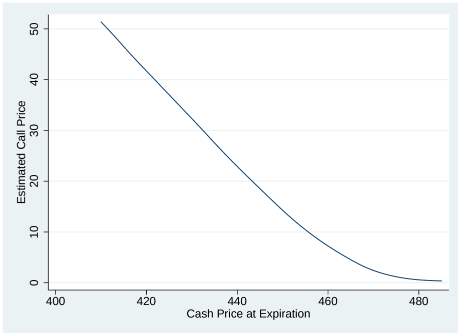
However, looking at individual cross sections of the state price density compared to the Black-Scholes SPD, the differences are more apparent. In Figure 23, the nonparametric and Black-Scholes SPD estimates are overlaid for four different times to maturity. The differences are minor for the shorter times to maturity, $T = 2$ and $T = 10$. The nonparametric density is slightly more peaked in both and is slightly shifted to the the right for $T = 10$. However, for the longer times to maturity, the nonparametric density exhibits important differences from the Black-Scholes.

First, the center of the distribution is shifted to the right, implying that the Black Scholes density slightly understates the expected return on the S&P500. Second, the nonparametric distribution is slightly skewed left, implying that slight downside events occur with less (risk-neutral) probability, but deep downside events occur with a greater probability than predicted by the Black Scholes SPD. Third, the risk neutral SPD appears to have higher kurtosis, meaning it is more "peaked", and assigns relatively higher weight in the tails.

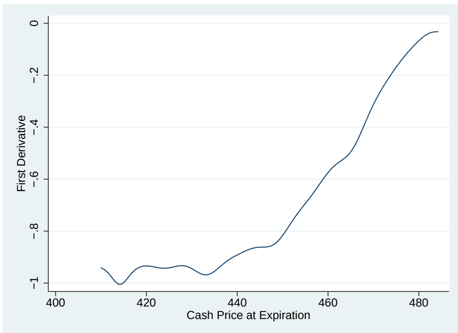
It should be kept in mind that these differences between the Black Scholes and nonparametrically estimated SPDs are expected, but are of relatively low magnitude. This would be supported by arguing that 1993 was a relatively stable year, in which there was not significant market turmoil or disruption. At least in terms of interest rates, 1993 was stable, with interest rates being very close to 3% without significant variation over the year. The Black Scholes and nonparametrically estimated SPDs would be expected to be farther apart during turbulent times, when investors are more risk averse or there is more downside risk.

Figure 21: Suboptimal Bandwidth: Estimation of the State Price Density
S&P500 on Nov 23, 1993 for T = 25 days

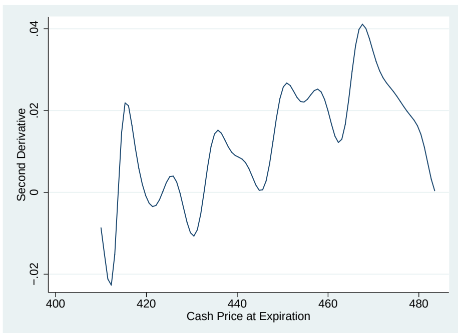
(a) Estimated Call Pricing Function: $\hat{C}$



(b) First Derivative: $\partial\hat{C}/\partial K$



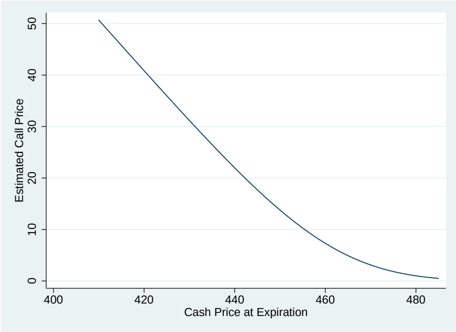
(c) Second Derivative (SPD):
$\partial^2\hat{C}/\partial K^2$



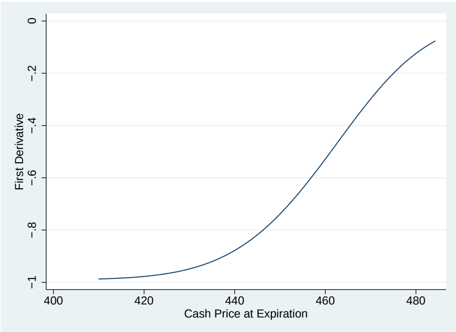
Bandwidth $= sd \cdot N^{-1/5}$. $S_t = 459.27$. These figures illustrate the high sensitivity of derivatives to the smoothness of the original function. In this case, the bandwidth is too small and the call option pricing function is under-smoothed. Minor noise in the original function gets picked up in the first derivative and is even more pronounced in the second derivative. In this particular case, the problem is pronounced due to the approximate linearity of $\hat{C}$ at low strike prices.

Figure 22: Optimal Bandwidth: Estimation of the State Price Density
S&P500 on Nov 23, 1993 for T = 25 days

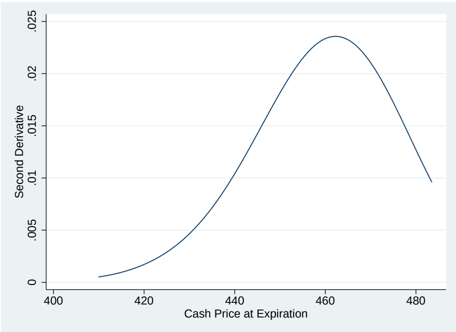
(a) Estimated Call Pricing Function: $\hat{C}$



(b) First Derivative: $\partial\hat{C}/\partial K$

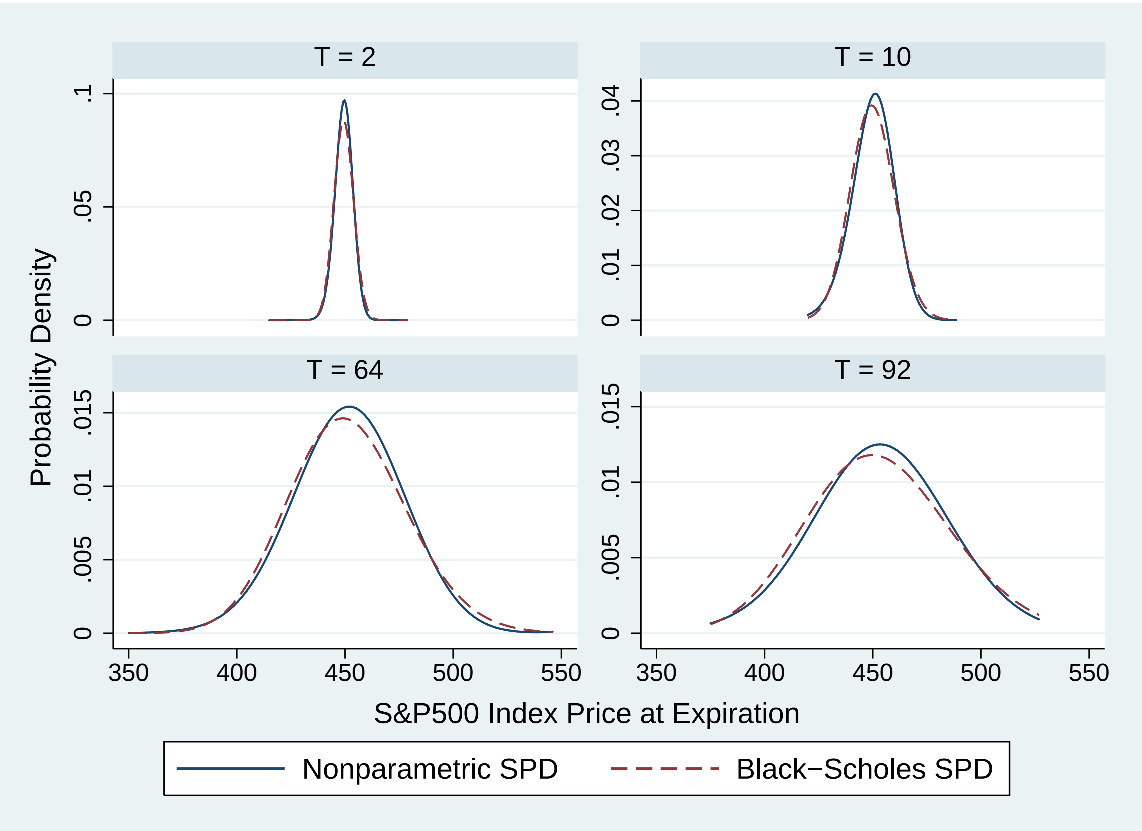


(c) Second Derivative (SPD):
$\partial^2\hat{C}/\partial K^2$



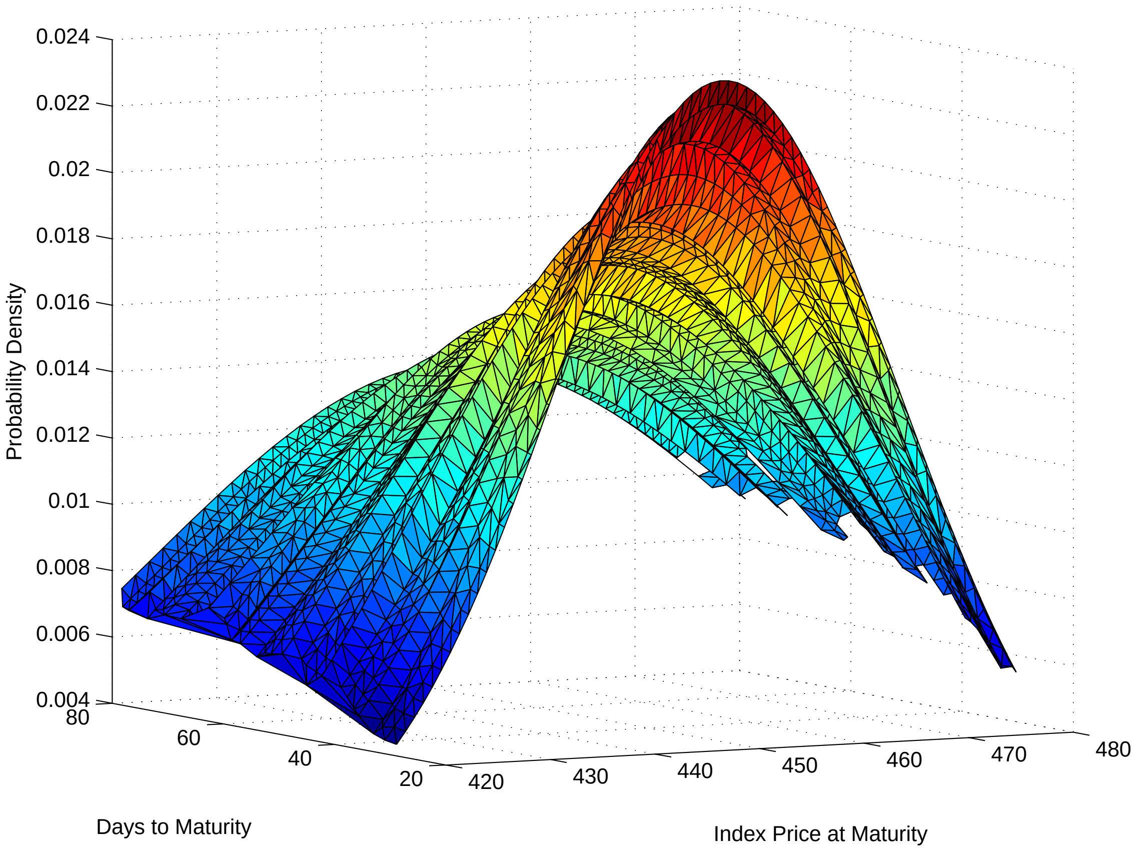
Bandwidth $= 5 \cdot sd \cdot N^{-1/5}$. $S_t = 459.27$. By increasing the smoothness of the call option pricing function, through an increase in the bandwidth parameter, the SPD estimate improves.

Figure 23: Nonparametric SPD Estimates: Cross Sections Compared to Black-Scholes SPD
All transactions in 1993 for each time to maturity



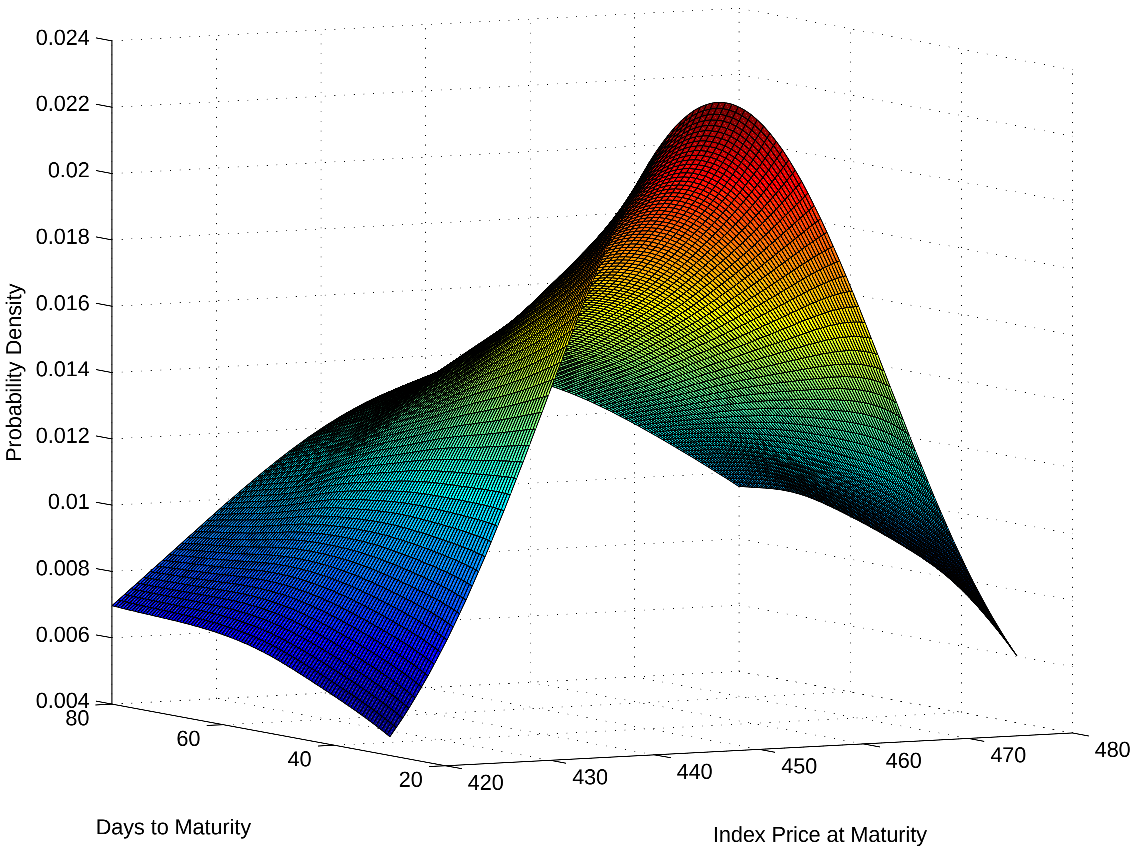
The State Price Density is estimated using all contracts traded in 1993 with maturity specified in each panel. S_t is fixed at 450 for each time to maturity so that the estimates are comparable. The Black-Scholes SPD is evaluated with implied volatility held at the average implied volatility of the options included in the estimation.

Figure 24: SPD Estimates over Maturity



The SPD corresponding to each time to maturity between T = 2 and T = 130 are estimated individually. These "cross sections" are then placed side-by-side to generate the full SPD estimate across both the index price at expiration and time to maturity.

Figure 25: SPD Estimate over Maturity (smoothed)



The SPD estimate from Figure 24 is smoothed (through nonparametric regression of course) to generate the figure seen here.

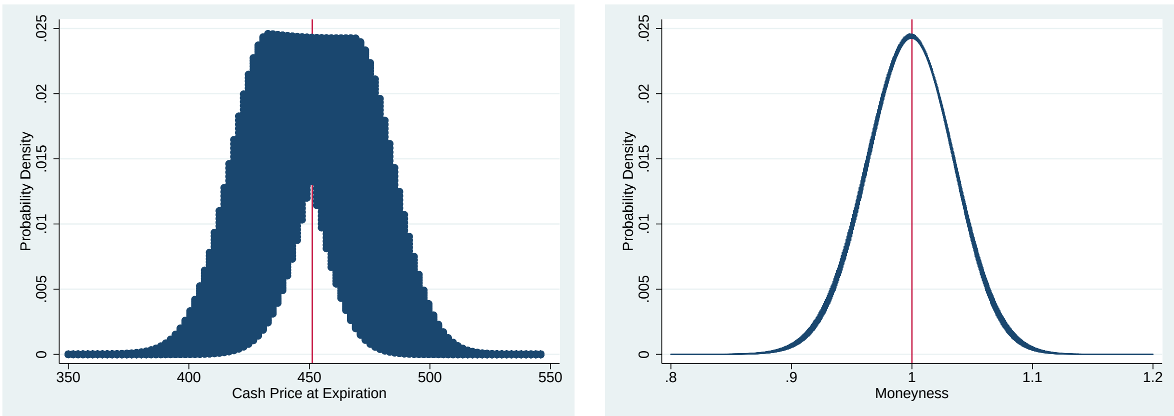
In the previous figures of Section 3.3, the state price density estimates are displayed with strike price on the x-axis. This is shown so that the figures may be compared to the estimates of other authors. However, a more meaningful alternative is to display the SPD with moneyness on the x-axis. This ensures that SPD estimates for different time periods are comparable, even when the stock price has changed over the period. There is no change in shape, since by defining moneyness at expiration as S_T/S_t, the axis is just a linear transformation of S_T.

In Figure 26 the results of this transformation are seen. Since the data includes all options traded over the year with maturity = 64 days, the SPD is "thick", implying that the stock price moved around over the period. However, when the SPD is plotted against moneyness, the shape is completely intuitive, having corrected for changes in stock price. The interpretation of this is that the risk neutral probability of the stock price being over 475 after 64 days, for example, has changed within the data. However, the risk neutral probability of the stock increasing in price by over 5% after 64 days has remained constant.

Figure 26: Comparison of SPD by S_T vs SPD by Moneyness
Options Traded in 1993 with Maturity = 64 Days

(a) SPD vs Cash Price at Expiration

(b) SPD vs Moneyness



While the SPD with respect to stock price has changed within the data, the SPD with respect to moneyness, for maturity = 64 days, has remained relatively constant.

4 Conclusion

In this paper, I nonparametrically estimate the state price density implied by the prices of S&P500 call and put options traded in 1993. Nonparametric estimation is chosen in order to avoid making restrictive assumptions on the functional form of the SPD, as is the case with parametric alternatives. In order to make the procedure computationally feasible, the dimensionality of the problem is reduced by using the semiparametric approach proposed by Aït-Sahalia and Lo (1998). In comparison to a fully nonparametric call option pricing function, $C(S_t, K, T, r_{t,T})$, which is estimated nonparametrically over five variables, the semiparametric estimator, $C(S_t, K, T, r_{t,T}; \sigma(S_t, K, T))$, requires nonparametric estimation over only three dimensions. With this approach, the implied volatility surface is estimated nonparametrically and is then plugged in to the Black-Scholes formula for the price of a call option. In this way, any local shape changes observed in call-price data are captured through the estimated implied volatility surface. Aït-Sahalia and Lo demonstrate how this semiparametric method is not significantly different from a fully nonparametric estimation.

While other authors assume that the SPD is constant over the whole period of data being used (typically one year), I avoid making this assumption by using high frequency transaction level data. With this large amount of data I am able to estimate the state price density for individual days. The advantage of less frequent, daily data, is that the data is regular, or evenly spaced over time. This facilitates nonparametric regression. However, I argue that this benefit of regularity is outweighed by the cost of being required to assume that the SPD is nonstochastic for an entire year.

Additionally, the SPD is typically estimated using bid-ask midpoints, rather than actual prices of completed transactions. A bid price represents a willingness to pay and an ask price represents a willingness to sell. However, using the bid-ask midpoint assumes that the average of the bid and ask prices is the "price" of the option for that day. That is, by nature of using bid-ask pairs instead of completed transactions, an exact value of the option is unknown. All that is known for sure is that the true value lies within the interval formed by the bid-ask pair. This may be the only route possible if transaction level data is not available.

However, with the availability of the transaction level Berkeley Options Data, analysis is now possible using the prices of trades that actually occurred. Thus, instead of using bid-ask pairs, I use completed transactions in my analysis. Moreover, the set of all transactions occurring within the year is much larger than just the set of closing prices. Therefore, the SPD may be estimated taking into account all intra-day variation, and is not subject to the limitation that a single closing bid-ask pair may be randomly high or low due to fluctuations in the market.

The SPD estimates seen in Figure 23 show important differences in comparison to the Black-Scholes SPD. The differences are small for short times to maturity, but increase as time to maturity increases. In particular, the nonparametric estimates exhibit higher kurtosis (implying a higher probability of extreme events) and are skewed downwards, which supports the idea that investors are risk averse. This may be interpreted in that investors are willing to pay more to avoid adverse outcomes. Alternatively it may represent that adverse outcomes occur with higher objective probability than suggested by the Black-Scholes lognormal SPD.

A Deriving the Semiparametric SPD Estimator

The semiparametric SPD involves first nonparametrically estimating the implied volatility and then plugging the volatility estimates into the Black Scholes call option pricing function. Second, the second derivative of the estimated call option pricing function is taken with respect to the strike price, K in order to recover the state price density (Breeden and Litzenberger).

The estimated Black Scholes formula for the price of a call option is:

$$\begin{aligned} E(C|S_t,K,T,r_{t,T},\delta_{t,T}) &= \hat{C}(S_t,K,T,r_{t,T}) \\ &= S_t\Phi(\frac{\ln(S_t/K)+(r_{t,T}+\frac{1}{2}\hat{\sigma}^2)T}{\hat{\sigma}\sqrt{T}}) - Ke^{-r_{t,T}T}\Phi(\frac{\ln(S_t/K)+(r_{t,T}-\frac{1}{2}\hat{\sigma}^2)T}{\hat{\sigma}\sqrt{T}}) \end{aligned} \tag{14}$$

where $\hat{\sigma}$ is estimated nonparametrically. The multivariate kernel used is the product of univariate Guassian kernels, with σ estimated as a function of S_t, K, and T. Importantly, the Gaussian kernel is twice differentiable.

$$\hat{\sigma}(S_t,K,T) = \frac{\sum_{i=1}^n \sigma_i \cdot k_S(\frac{S_t-S_{t_i}}{h_S})k_T(\frac{T-T_i}{h_T})k_K(\frac{K-K_i}{h_K})}{\sum_{i=1}^n k_S(\frac{S_t-S_{t_i}}{h_S})k_T(\frac{T-T_i}{h_T})k_K(\frac{K-K_i}{h_K})},\ k(x) \equiv \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

$$\begin{aligned} &= \frac{\sum_{i=1}^n \sigma_i \cdot k_S(\frac{S_t-S_{t_i}}{h_S})k_T(\frac{T-T_i}{h_T})\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{K-K_i}{h_K})^2)}{\sum_{i=1}^n k_S(\frac{S_t-S_{t_i}}{h_S})k_T(\frac{T-T_i}{h_T})\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{K-K_i}{h_K})^2)} \\ &= \frac{\sum_{i=1}^n \sigma_i A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h_K})^2)}{\sum_{i=1}^n A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h_K})^2)},\ A_i \equiv k_S(\frac{S_t-S_{t_i}}{h_S})k_T(\frac{T-T_i}{h_T}) \end{aligned}$$

Plugging in $\hat{\sigma}$ into (14) yields:

$\hat{C}(S_t, K, T, r_{t,T}) =$

$$S_t \Phi\left(\frac{\ln(S_t/K) + (r_{t,T} + \frac{1}{2}(\frac{\sum_{i=1}^{n} \sigma_i A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)}{\sum_{i=1}^{n} A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)})^2)T}{(\frac{\sum_{i=1}^{n} \sigma_i A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)}{\sum_{i=1}^{n} A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)})\sqrt{T}}\right)$$

$$-Ke^{-r_{t,T}T} \Phi\left(\frac{\ln(S_t/K) + (r_{t,T} - \frac{1}{2}(\frac{\sum_{i=1}^{n} \sigma_i A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)}{\sum_{i=1}^{n} A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)})^2)T}{(\frac{\sum_{i=1}^{n} \sigma_i A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)}{\sum_{i=1}^{n} A_i \exp(-\frac{1}{2}(\frac{K-K_i}{h})^2)})\sqrt{T}}\right)$$

Now the state-price density is simply the second derivative of $\hat{C}$ with respect to K, or $\partial^2 \hat{C}/\partial K^2$

References

[1] Aït-Sahalia, Yacine, and Andrew W. Lo, 1998, Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices, *Journal of Finance* 53, 499-547

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

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

[4] Bierens, Herman J. 1987, Kernel Estimators of Regression Functions, *Advances in Econometrics: Fifth World Congress* 1, 99-144.

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

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

[7] Campbell, John, Andrew W. Lo, and A. Craig MacKinlay, 1997. *The Econometrics of Financial Markets*(Princeton University Press, Princeton, N.J.).

[8] Cochrane, John, 2001. *Asset Pricing*(Princeton University Press, Princeton, N.J.).

[9] Dacorogna, Michel, Ramazan Gencay, Ulrich A. Müller, Richard B. Olsen and Olivier V. Pictet, 2001. *An Introduction to High-Frequency Finance* (Elsevier Science).

[10] Davidson, Russell, and James G. MacKinnon, 2003. *Econometric Theory and Methods* (Oxford University Press, USA).

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

[12] Härdle, Wolfgang, 1990. *Applied Nonparametric Regression* (Cambridge University Press, Cam- bridge, U.K.).

[13] Jackwerth, Jens C., and Mark Rubinstein, 1996, Recovering probability distributions from con- temporary security prices, *Journal of Finance* 51, 1611-1631.

[14] Jackwerth, Jens C., 1999, Option Implied Risk-Neutral Distributions and Implied Binomial Trees: A Literature Review, *Journal of Derivatives* 7, 66-82.

[15] Jackwerth, Jens C., 1999, Recovering risk aversion from option prices and realized returns, *Review of Financial Studies* 13, 433-451.

[16] Li, Qi, and Jeffrey S. Racine, 2006 *Nonparametric Econometrics: Theory and Practice* (Princeton University Press)

[17] Ross, Stephen, 1976, Options and efficiency, *Quarterly Journal of Economics* 90, 75-89.

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