

A Linearization-Based Solution to the Ill-Posed Local Volatility Estimation Problem

James N. Bodurtha, Jr.
Georgetown University

June 1997

Last Revised: December 2000

Abstract

Given a set of option quotes at a point in time, we estimate a local volatility surface. The heterogeneous time- and state-dependent parameters in this linearized model specification are identified by the Tikhonov regularization method. Specifically, the optimal parameter estimates satisfy a least squares error criterion and are minimum sum of squared deviations from the Black-Scholes constant variance specification. Applying Wahba's (1983) generalized degrees of freedom (GDF) measure, we specify a test of time-dependent, state-dependent and maturity time- and state-dependent restrictions of the proposed time- and state-dependent local volatility specification. As an application, we evaluate these restrictions for the November 25, 1991 set of Philadelphia Stock Exchange European Deutschemark option trades. The restricted specifications either fail to meet the least squares error criterion or require more generalized degrees-of-freedom than the time- and state-dependent local volatility specification which we propose.

This work is an extension of an earlier draft, "Regular Smiles." with Martin Jermakyan, and I thank him for his input. For their comments, I appreciate the comments of Bardia Kamrad, Peter Lenk, Vadim Linetsky, Akhtar Siddique and Martin Young, as well as participants in the following seminars: University of Chicago Applied Mathematics, Georgetown University Mathematics, University of Maryland School of Business, IAFE-Stanford Computational Finance Conference, IEEE/IAFE/INFORMS Conference on Computational Intelligence for Financial Engineering (CIFEr) - New York, Goldman-Sachs Research and Georgetown School of Business.

A Linearization-Based Solution to the Ill-Posed Local Volatility Estimation Problem

Abstract

Given a set of option quotes at a point in time, we estimate a local volatility surface. The heterogeneous time- and state-dependent parameters in this model specification are identified by the Tikhonov regularization method. Specifically, the optimal parameter estimates satisfy a least squares error criterion and are minimum sum of squared deviations from the Black-Scholes constant variance specification. Applying Wahba's (1983) generalized degrees of freedom (GDF) measure, we specify a test of time-dependent, state-dependent and maturity time- and state-dependent restrictions of the proposed time- and state-dependent local volatility specification. As an application, we evaluate these restrictions for the November 25, 1991 set of Philadelphia Stock Exchange European Deutschemark option trades. The restricted specifications either fail to meet the least squares error criterion or require more generalized degrees-of-freedom than the time- and state-dependent local volatility specification which we propose.

We develop a local volatility surface estimator that is implied in a set of contemporaneous option prices. Even in its discretized variant, this fundamental nonlinear local volatility coefficient estimation problem is ill-posed because the number of volatility parameters far exceeds the number of available option prices. In resolving this problem, we regularize and linearize a trinomial lattice-based estimator of the time- and state-dependent local volatility surface.

Following Tikhonov's approach, our regularization stabilizes and identifies the ill-posed local volatility parameter estimates.¹ The proposed estimator overcomes the undesirable interpolation feature of the Derman-Kani (1994), Dupire (1994) and Andersen and Brotherton-Ratcliffe (1998) local volatility surface estimators, and extends Rubinstein's (1994) optimization-based procedure for estimating term transition probabilities to the more general local case. Previously, Avellandè et. al. (1996), Bodurtha-Jermakyan (1999), Brown-Toft

(1996), Derman-Kani-Zou (1996), Jackwerth-Rubinstein (1996) and Lagnado-Oscher (1996) have implemented the regularization approach.² These authors identify quote or bid-ask spread error as the source of model error or discrepancy.

Our linearization method substitutes directly for linearizations that are inherent in non-linear optimization algorithms, and it leads to an appealing and tractable ridge regression specification. The linearization introduces additional model error that is termed "operator error" in the Tikhonov paradigm. This second source of error is analogous to regression errors-in-variables. Therefore, our discrepancy error convergence criterion includes both bid-ask price error and the added operator error that must arise in any discrete diffusion-based model approximation to derivative values.

Our modeling approach is developed through an extensive example and implemented for November 25, 1991 Philadelphia Stock Exchange Deutschemark European options. Based on these option trade prices, we compare the time- and state-dependent local volatility specification against three alternatives: time-dependent volatility, state-dependent volatility, and option maturity time- and state-dependent volatility.

These specifications are motivated by the seminal contributions of Merton (1973), Cox-Ross (1976) and Cox-Rubinstein (1985), and Rubinstein (1994), respectively. The Wahba (1983) generalized degrees of freedom (GDF) measure is used to differentiate between the specifications. This measure indicates that the time- and state-dependent local volatility specification is preferable to the three other specifications.

¹ See Tikhonov and Arsenin (1977).

² Among these works, the Avellaneda et. al. and Bodurtha-Jermakyan (1999) regularizers are analogous to the one that we use in estimation. The other works either explicitly or implicitly estimate volatility curves for particular maturities and then interpolate local volatilities from these estimates. As these works are, effectively, curve fitting implementations, the regularizers used are defined by second derivatives (the Laplacian) alone. For surface estimation, such regularizers are not sufficient to control poles (spikes) in the parameter estimate surface. For some related discussion, also see Silverman(1985).

The first section of this work defines the example market environment, develops our time- and state-dependent variance trinomial lattice model and the regularized nonlinear estimation example. Section two contains our linearized model development and illustrates model estimation under restrictions. Section three describes the data that is used for estimation. Section four presents both further discussion of the estimation method and our estimation results. Section five concludes the main body of our work, and the appendix outlines the singular value decomposition procedure underlying our calculations.

1. A Local Volatility Estimation Example

Our example begins with the Table 1 market scenario.

Table 1: Time Zero Market Scenario

	Maturity	Market Price
Spot price (S_0)		100.0
Constant interest rate (r)	1.0	6.0%
Constant yield (y)	1.0	6.0%
European calls by strike price		
81.87	1.0	18.739
100.0	1.0	5.844
122.14	1.0	0.291

Conditional only on this current market data, we will extract estimates of the local time- and state-dependent spot price volatilities. Separate estimates must be identified for all spot price states and at all times that the spot price evolves up to the option maturity.

1.1 A Binomial Model Preface

As the first step in generating local volatility estimates, we calculate a single implied volatility by fitting a four time-step Cox-Ross-Rubinstein or CRR (1979) binomial model to the three example option prices. Given the Table 1 spot price, interest rate and yield, a 20% volatility estimate minimizes the squared deviations of the three Table 1 option prices from the

corresponding four time-step CRR model prices.

To illustrate our method, we construct an example tree with Figure 1. The volatility is specified to be 20 percent, $\sigma_0=20\%$. The time step length is one quarter of a year, $h_0 = 1/4$, and time is aggregated by the number of time steps, j_0 , that have occurred, $t=j_0h_0$. The binomial tree step size is set in the standard way, $u_0 = e^{\sigma_0\sqrt{h_0}} = 1.1052$.³

At each time step, spot price states are indexed from the initial zero state level. The initial spot price is defined $S(0,0)$. In each subsequent time period, a positive spot price movement is indicated by $k_0=1$, and the binomial spot state index, i_0 , is increased by k_0 . A negative spot price movement is indicated by $k_0 = -1$, and one is subtracted from the state index, i_0 . Spot prices evolve $S_0(i_0+k_0, j_0+1) = S_0(i_0, j_0)u_0^k$ for $k_0=[1,-1]$, and $S_0(0,0) = S(0,0)$. For example, $S_0(1,1) = S_0(0,0)u_0 = 110.52$, and $S_0(-1,1) = S_0(0,0)u_0^{-1} = 90.48$.

The no-arbitrage condition requires the spot price to equal to discounted risk-neutral expected value:

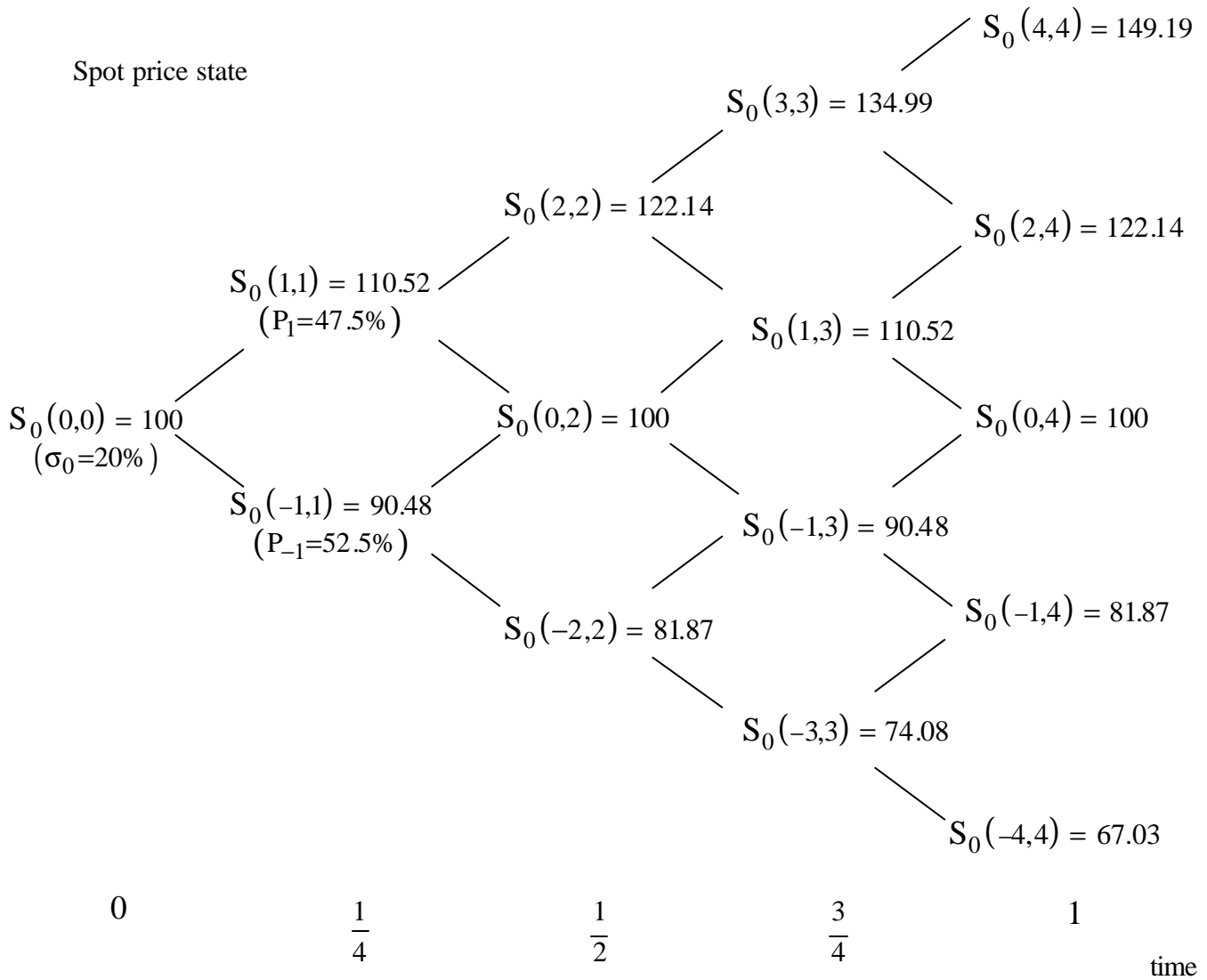
$$S_0(0,0) = S_0(0,0) \left[P_1 u_0 + (1-P_{-1}) u_0^{-1} \right] e^{(y-r)h_0} \quad 1)$$

The no arbitrage condition 1) implies the following risk-neutral probabilities:

$$P_1 = \frac{\left[e^{(r-y)h_0} - u_0^{-1} \right]}{u_0 - u_0^{-1}} = 0.475, \quad P_{-1} = 1 - P_1 = 0.525 \quad 2)$$

³ See Karmrad and Ritchken's (1991) specification and analysis of a closely related binomial and trinomial constant volatility model. We favor their notation and identify the down step size, d , as u^{-1}

Figure 1: Example Binomial Tree



With $u_0 = e^{\sigma_0 \sqrt{h_0}} = 1.1052$, $S_0(i_0+k_0, j_0+1) = S_0(i_0, j_0) u_0^k$ for $k_0=[1, -1]$.

The CRR binomial option model prices are calculated as the discounted expected value taken over the risk neutral probabilities, P_1 and P_{-1} .

Table 2 reports the results of our example valuation exercise:

Table 2: Example European Option Contracts and Pricing

Strike Price	Maturity	Market Price	CRR Model Price with 20% volatility	Option Specific CRR Implied Volatility
81.87	1.0	18.739	18.134	20.74%
100.0	1.0	5.844	7.052	16.57%
122.14	1.0	0.291	1.297	4.83%

Note, CRR stands for Cox-Ross-Rubinstein

From Table 2, we see that the market option prices and the CRR model prices are not equal.

The last column of Table 2 translates the market prices into a separate and misspecified CRR model implied volatility for each option. Though commonly used, we emphasize that these estimates are misspecified price functions aggregated over a true local volatility surface. The in-the-money (81.87 strike price) call has a 20.74% CRR implied volatility, the at-the-money (100.0 strike price) call has a 16.57% CRR implied volatility, and the out-of-the-money (122.14 strike price) call has only a 4.83% CRR implied volatility. Such a cross-option implied volatility pattern has been called the volatility "skew" or "smirk." Were this curve symmetric, the implied volatility pattern is called the volatility "smile."

To address CRR model pricing errors, and yet remain in the limiting diffusion model class, we permit time- and state-dependent volatilities. Specifically, we introduce a trinomial lattice that is equivalent to the binomial tree if volatility is constant (i.e., the CRR model volatility assumption is true).

1.2 A Time- and State-Dependent Local Volatility Trinomial Lattice Option Model

Recombining binomial trees are not sufficiently flexible to invert option prices into time- and state-dependent volatilities. To gain this flexibility, we use the constant volatility binomial tree time and state spot price to define the trinomial lattice spot price outcomes over time and

state.

1.2.1 A Trinomial Lattice

We define the trinomial time index, j , to increment by one with passage of the time interval $h=2h_0$, which is twice the binomial model time interval. We also define a spot price change indicator, k , that incorporates an additional no change state. With $k = [-1, 0, 1]$, define the trinomial up step as $u=u_0^2$. The state indicator i defines the net number of up price movements. From any time jh and state i , the trinomial lattice spot price evolves as $S(i+k, j+1) = S(i, j) u^k$ for $k=[-1, 0, 1]$.

Under these definitions and a constant volatility, M period trinomial lattice-based valuation is consistent with $2M$ period binomial tree-based valuation. Figure 2 illustrates the correspondence between our example trinomial lattice and the Figure 1 binomial tree.

1.2.1 Trinomial Lattice-based Valuation

At any lattice node, risk-neutral pricing implies that the spot price equals its discounted expected value⁴:

$$S(i, j+1) = S(i, j) \left[P_1(i, j)u_0 + P_0(i, j) + P_{-1}(i, j)u_0^{-1} \right] e^{(y-r)h} \quad 3)$$

where, $P_0(i, j) = 1 - P_1(i, j) - P_{-1}(i, j)$.

Equation 3) implies

$$P_1(i, j) = \frac{\left[e^{(r-y)h} - 1 \right] - P_{-1}(i, j)(u^{-1} - 1)}{u - 1} \quad 4)$$

⁴ Schwartz (1977), Parkinson (1977) and Mason (1979) pioneered the trinomial method.

Substituting from equation 4) into equation 5), we obtain

$$P_{-1}(i,j) = \sigma^2(i,j)D - R, \quad D = \frac{u-1}{2\sigma_0^2(u-u^{-1})}, \quad R = \frac{e^{(r-y)h} - 1}{u-u^{-1}} \quad 6)$$

Finally, solving for the up and no-move probabilities, we find

$$P_1(i,j) = \sigma^2(i,j)U + R, \quad U = \frac{1-u^{-1}}{2\sigma_0^2(u-u^{-1})}, \quad P_0(i,j) = 1 - \sigma^2(i,j)(U+D) \quad 7)$$

With two time periods, the associated trinomial option-pricing model is the following:

$$\begin{aligned} e^{r2h}\hat{C}(0,0;K_n,2h) = & P_1(0,0)P_1(1,1)\text{Max}[S(2,2)-K_n,0] \\ & + [P_1(0,0)P_0(1,1)+P_0(0,0)P_1(0,1)]\text{Max}[S(1,2)-K_n,0] \\ & + [P_1(0,0)P_{-1}(0,1)+P_0(0,0)P_0(0,1)+P_{-1}(0,0)P_1(-1,1)]\text{Max}[S(0,2)-K_n,0] \\ & + [P_0(0,0)P_{-1}(0,1)+P_{-1}(0,0)P_0(-1,1)]\text{Max}[S(-1,2)-K_n,0] \\ & + P_{-1}(0,0)P_{-1}(-1,1)\text{Max}[S(-2,2)-K_n,0], \quad n = 1,2,3 \end{aligned} \quad 8)$$

This trinomial model generates the Table 1 example call option prices.⁵ In Table 3, The local volatility values that generate these example option prices are listed.

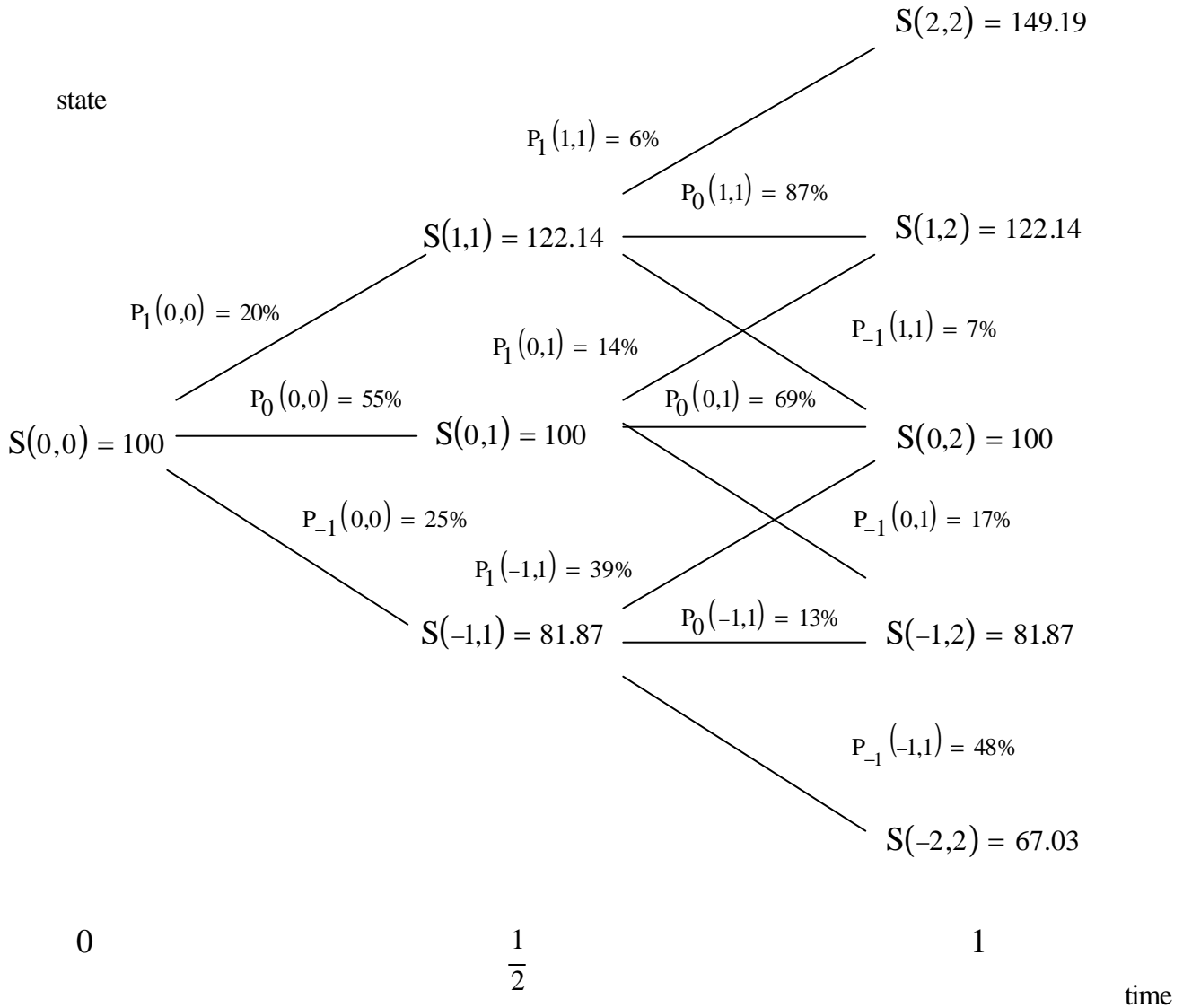
Table 3: Example Benchmark Local Volatility Surface

state and time - $S(i,j)$	$S(0,0)=100.0$	$S(1,1)=122.14$	$S(0,1)=100.0$	$S(-1,1)=81.87$
$\sigma(i,j)$	19.00%	10.00%	15.81%	26.46%

⁵ Note, no ½ year maturity options are traded and the market is incomplete in time and state. The one period pricing model is the following:

$$e^{rh}\hat{C}(0,0;K_n,h) = P_1(0,0)\text{Max}[S(1,1)-K_n,0] + P_0(0,0)\text{Max}[S(0,1)-K_n,0] + P_{-1}(0,0)\text{Max}[S(-1,1)-K_n,0] \quad 9)$$

Figure 3: Example Trinomial Lattice Manifesting Local Volatility Tree
 $[\sigma(0,0) = 19\%, \sigma(1,1) = 10\%, \sigma(0,1) = 15,81\%, \sigma(-1,1) = 26.46\%]$



At the 100 initial spot price, the time zero local volatility, $\sigma(0,0)$, is 19%. From Table 3, we see that volatility falls significantly to 10% with a price increase ($\sigma(1,1)$ at 122.14), and also falls to 15.81%, $\sigma(0,1)$, if the initial spot price remains unchanged. Conversely, volatility rises to 26.46% with a price decline ($\sigma(-1,1)$ at 81.87). From equation 8), we see that high local volatility comes from high local probability of up and down outcomes, relative to the local

probability of no move.

We see that high volatility (26.46%) is manifest in the down (-1,1) state with relatively high probabilities of the up and down outcomes (39% and 48%, respectively) as well as a relatively low probability of remaining in the down state (13%). Analogously, low volatility (10%) is manifest in the up (1,1) state with relatively low probabilities of the up and down outcomes (6% and 7%, respectively) as well as a relatively low probability of remaining in the down state (87%). (Figure 3 also reports these time- and state-dependent local probabilities.)

The Table 3 example benchmark exhibits a *local* volatility smile. However, this local volatility smile is not as severe as the mis-specified term volatility smile that is implied by ad-hoc use of the CRR model in a changing volatility scenario. (See the Table 2 Option Specific CRR Implied Volatility column.)

In order to restate the equation 8) trinomial model in terms of variance, we define state- and time-dependent variances as follows:

$$\sigma^2(i,j) = \sigma_0^2 + a(i,j) \quad 10)$$

To value the two-period maturity options, we work through two backward induction steps. At time h, the up, no change, and down state-contingent option values are functions of three quantities: constant volatility probabilities, P_k , the local volatility probability components, $a(i,j)$, and the one step ahead option values, $C_0(i+k,j+1;K_n,2h)$:

$$e^{rh} \hat{C}(i,j;K_n,2h) = \quad 11)$$

$$P_1 C_0(i+1,j+1;K_n,2h) + P_0 C_0(i,j+1;K_n,2h) + P_{-1} C_0(i-1,j+1;K_n,2h)$$

$$+ a(i,j) \left[UC_0(i+1,j+1;K_n,2h) - (U+D)C_0(i,j+1;K_n,2h) + DC_0(i-1,j+1;K_n,2h) \right]$$

$$= e^{rh} C_0(i, j, K_n, 2h) + a(i, j) \left[UC_0(i+1, j+1; K_n, 2h) - (U+D)C_0(i, j+1; K_n, 2h) + DC_0(i-1, j+1; K_n, 2h) \right]$$

At time 0, the option values are

$$\begin{aligned} e^{r2h} \hat{C}(0, 0; K_n, 2h) = & \left[P_1 C_0(1, 1; K_n, 2h) + P_0 C_0(0, 1; K_n, 2h) + P_{-1} C_0(-1, 1; K_n, 2h) \right] e^{rh} \quad 12) \\ & + a(0, 0) \left[UC_0(1, 1; K_n, 2h) - (U+D)C_0(0, 1; K_n, 2h) + DC_0(-1, 1; K_n, 2h) \right] e^{rh} \\ & + a(1, 1) P_1 \left[UC_0(2, 2; K_n, 2h) - (U+D)C_0(1, 2; K_n, 2h) + DC_0(0, 2; K_n, 2h) \right] \\ & + a(0, 1) P_0 \left[UC_0(1, 2; K_n, 2h) - (U+D)C_0(0, 2; K_n, 2h) + DC_0(-1, 2; K_n, 2h) \right] \\ & + a(-1, 1) P_{-1} \left[UC_0(0, 2; K_n, 2h) - (U+D)C_0(-1, 2; K_n, 2h) + DC_0(-2, 2; K_n, 2h) \right] \\ & + a(0, 0) a(1, 1) U \left[UC_0(2, 2; K_n, 2h) - (U+D)C_0(1, 2; K_n, 2h) + DC_0(0, 2; K_n, 2h) \right] \\ & - a(0, 0) a(0, 1) (U+D) \left[UC_0(1, 2; K_n, 2h) - (U+D)C_0(0, 2; K_n, 2h) + DC_0(-1, 2; K_n, 2h) \right] \\ & + a(0, 0) a(-1, 1) D \left[UC_0(0, 2; K_n, 2h) - (U+D)C_0(-1, 2; K_n, 2h) + DC_0(-2, 2; K_n, 2h) \right] \end{aligned}$$

The interaction of the time and state-specific local volatility parameters illustrates the inherent nonlinearity of the implied local volatility estimation problem. As we more accurately value options by introducing more time steps, n , the nonlinear aspects of the inference problem are manifest in recursive polynomials of order n .

For example, the likelihood of up moves in each of 30 time steps is the following:

$$\begin{aligned} P(u^{30}) = & P_1(0, 0) P_1(1, 1) P_1(2, 2) \dots P_1(29, 29) \\ = & \left(\left[\sigma_0^2 + a(0, 0) \right] U + R \right) \left(\left[\sigma_0^2 + a(1, 1) \right] U + R \right) \dots \left(\left[\sigma_0^2 + a(29, 29) \right] U + R \right) \end{aligned}$$

Clearly, this probability and all others are highly nonlinear in the variance parameters, $a(i, j)$'s.

In general, this optimization problem faces multiple local optima and instability. The estimation problem is, fundamentally, ill-posed.

Another source of ill-posed estimation arises because there are always fewer available option quotes than there are unknown time- and state-dependent local volatility parameters. In

the context of our example, we have a set of four local volatility parameters to estimate:

$\psi(a) = \{a(0,0), a(1,1), a(0,1), a(-1,1)\}$. However, we have only three option quotes from which to estimate these four parameters. Our inference exercise is over-parameterized, and, again, ill-posed.⁶ Since options trade only at exercise prices and on maturity dates that represent a small fraction of possible prices and dates, local volatility estimation is always under-identified.

To address both the unstable characteristics of our non-linear estimation problem and under-identification of the local volatility parameters (too few option quotes), we employ the Tikhonov regularization method.

1.3 A Nonlinear Regularized Volatility Surface Estimator

Resolution our under-identified and unstable local volatility estimation problem requires explicit (or implicit) use of prior information.⁷ Specifically, we advocate a method for identifying volatility parameter estimates around reasonable prior estimates. Our specification penalizes volatility parameter adjustments from the BSM-CRR constant volatility case.

⁶ The overidentification case does not occur at any given instant. In real-time application, we would need many more grid points and associated local volatilities than currently available quotes or trades. To overidentify the system, some time-series assumptions should be made regarding volatility surface evolution. For an example, see Ait-Sahalia and Lo (1998).

⁷ Implicit regularization was imposed in early volatility curve and surface estimation. For volatility and probability curves, Shimko (1993) fitted a three parameter (quadratic) function to the more than three option implied volatility smile for each maturity, and Rubinstein (1994) limited the parameter space by restricting certain path probabilities to be equal. For volatility surfaces, Dupire (1994) and Derman-Kani (1994) interpolated traded option-specific implied term volatilities into a volatility surface and used this generated surface to provide as many option price pseudo-observations as necessary to identify the estimated parameters. In an alternative attempt to resolve Black-Scholes option model pricing errors, Bakshi, Cao and Chen (1997), Bates (1996) and Scott (1997) estimate time-homogeneous parameter stochastic volatility specifications. These specifications are sufficiently parsimonious to be identified in the sense that their are fewer estimated parameters than fitted option prices. Nevertheless, non-regularized estimation of these highly nonlinear models remains ill-posed because solution stability and uniqueness are not guaranteed.

We define a set of perturbation parameters to the BSM-CRR volatility as follows:

$$\sigma^2(i,j) = \sigma_0^2 + a(i,j) = [\sigma_0 + a'(i,j)]^2, \quad a(i,j) = 2\sigma_0 a'(i,j) + a'(i,j)^2 \quad (13)$$

$$a(i,j) \in [a(0,0), a(1,1), a(0,1), a(-1,1)], \quad a'(i,j) \in [a'(0,0), a'(1,1), a'(0,1), a'(-1,1)]$$

The $a(i,j)$ are variance perturbations, and the $a'(i,j)$ are the associated volatility perturbations.

Table 4 reports the volatility perturbations:

Table 4: Benchmark Perturbations $[a'(i,j)]$ from 20% Constant Volatility

state-time	0	u	m	d
$a'(k)$	-1.00%	-10.00%	-4.91%	+6.46%

Our two-period example has four unknown variance parameters. To estimate these parameter values, a natural alternative is to minimize the squared model errors.

In the context of our Table 1 example option prices, $C(0,0,K_n,T)$ for $n = 1, 2, 3$, and one maturity of $m = 2$ periods, the least squares estimation problem is the following:

$$\text{Min}_{\psi(a)} \sum_{m=2}^{M=2} \sum_{n=1}^{N=3} [C(0,0;K_n,mh) - \hat{C}(0,0;K_n,mh)]^2 \quad (14)$$

$$\text{where } \psi(a) = \{a(0,0), a(1,1), a(0,1), a(-1,1)\}$$

Of course, there are only three option trade quotes and four local variance parameters to estimate. Therefore, the minimand in equation 14) is under-identified. In this case, an infinite number of four local variance parameter estimates will minimize the optimand. Therefore, we implement a regularized estimator. Our regularization chooses volatility parameter adjustments that provide both minimum deviations from the CRR constant volatility, σ_0 , and satisfactory model “goodness-of-fit” to the quoted prices.

The proposed regularizer has two components: a penalty weight or multiplier for the prior constraint, α , and the sum of the squared local variance parameters. The penalty weight is

chosen to provide sufficient least squares fit. Dependent on this fit, the estimated local variance set is minimum mean square distance from the constant BSM-CRR variance estimate.

The regularized estimation problem is the following¹:

$$\begin{aligned} \text{Min}_{\psi(a)} \sum_{m=2}^{M=2} \sum_{n=1}^{N=3} \left[C(0,0;K_n, mh) - \hat{C}(0,0;K_n, mh) \right]^2 + \alpha \sum_{i=-1}^1 \sum_{j=0}^1 a(i,j)^2 \quad 15) \\ \text{where } \psi(a) = \{a(0,0), a(1,1), a(0,1), a(-1,1)\} \end{aligned}$$

Minimand 15) has two components: the leading least squares goodness-of-fit term and the second regularizer or smoother. Adding the regularizer to the under-identified least squares estimation problem ensures that unique and stable local variance estimates are found.

To resolve the regularizer weight choice, α , we follow Tikhonov's approach. That is, we specify an accuracy tolerance for the data, δ , and solve the least squares problem to this level of error.² For this discrete example, the only error source is rounding. We estimate the rounding error to be $\delta^2 = 1E-07$. For fixed a δ , an iterative three-step procedure is followed to ensure convergence. Starting the initial iteration, we set $p=1$:

Step 1 Given α_p , minimize the nonlinear smoothed minimand 15).

Step 2 Obtain the discrepancy between the least squares model fit and the error tolerance level, δ^2 . This function is known as the discrepancy function:

$$\rho(\alpha_p) = \sum_{m=2}^{M=2} \sum_{n=1}^{N=3} \left[C(0,0;K_n, mh) - \hat{C}(0,0;K_n, mh) \right]^2 - \delta^2 \quad 16)$$

¹ Following usual convention, iterations over index j that decrease the index are set to zero.

² The outlined method is a special case of the general partial differential equation-based method of Bodurtha-Jermakyan (1999). In that work, additional smoothness constraints are placed on the estimated volatility function. Given the fundamental option pricing assumption that underlying prices follow a univariate diffusion process, these additional constraints guarantee convergence of estimated option values, deltas, gammas, etc. to their true values. Bodurtha-Jermakyan (1999) also incorporates the forward price adjustment suggested by Barle-Cakici (1995).

- Step 3** If $\rho(\alpha_p) > 0$, set $\alpha_{p+1} = \frac{1}{2}\alpha_p$ and increment p by 1. Go to step 1.
- If $\rho(\alpha_p) < 0$, set $\alpha_{p+1} = \alpha_p + \frac{1}{3}\alpha_p$ and increment p by 1. Go to step 1.
- If $\rho(\alpha_p)=0$, terminate procedure.

The outlined solution procedure uses the bisection method to solve for the root of the discrepancy function. Other nonlinear search methods (e.g. secant method) will yield convergence in fewer steps.

Intuitively, $\rho(\bullet) > 0$ implies too little fit and too smooth a local volatility surface. To adjust, the smoothness (or roughness penalty) parameter, α_p , should be decreased.

Conversely, $\rho(\bullet) < 0$ implies too tight a fit. Given δ , a smoother set of estimates would be appropriate. In this latter case, the smoothness parameter should be increased. $\rho(\bullet)=0$ indicates the requisite balance between smoothness and fit.

For our two period example, we start our search procedure with $\alpha_p=1$. We seek a goodness-of-fit, $\tilde{\delta}^2$, that equals $\delta^2 = 1E-07$, and Table 5 illustrates convergence steps:

Table 4: Steps to the Nonlinear Volatility Estimates

Step	α_p	$\tilde{\delta}^2$	$\sum_{\psi(a)} a^2(k)$	Minimand 14)	$\sigma(0,0)$	$\sigma(1,1)$	$\sigma(0,1)$	$\sigma(-1,1)$
1	1	2.2E-06	0.1907%	0.001909	19.94%	9.54%	14.13%	25.20%
2	0.5	4.6E-07	0.1909%	0.000955	19.97%	9.52%	14.08%	25.17%
3	0.25	1.1E-07	0.1910%	0.000478	19.97%	9.52%	14.08%	25.17%
...	...							
*	0.2337	1E-07	0.1910%	0.000447	19.97%	9.52%	14.08%	25.17%
...	...							
4	0.125	2.9E-08	0.1910%	0.000239	19.97%	9.52%	14.08%	25.17%
actual					19.00%	10.00%	15.81%	26.46%

The optimal estimates fall within two percent of the true volatilities.

In cases with few time steps, nonlinear estimation of the local volatility parameters can be solved with standard optimizers. A fairly common alternative to true local volatility estimation is estimation of maturity- and state-specific volatility functions.³ In the sequel, we treat this alternative as a restriction of our an iterative-linearized local volatility estimator.

2. A Linearized Local Volatility Trinomial Model

We now impose a linearized approximation to the nonlinear trinomial lattice model. In equation 12), we set all $a(\bullet)$ cross-product terms to zero:

$$\begin{aligned}
 C_1(0,0;K_n,2h) &= C_0(0,0;K_n,2h) \\
 &+ a(0,0) \left[UC_0(1,1;K_n,2h) - (U+D)C_0(0,1;K_n,2h) + DC_0(-1,1;K_n,2h) \right] e^{-r h} \\
 &+ a(1,1) P_1 \left[UC_0(2,2;K_n,2h) - (U+D)C_0(1,2;K_n,2h) + DC_0(0,2;K_n,2h) \right] e^{-r 2h} \\
 &+ a(0,1) P_0 \left[UC_0(1,2;K_n,2h) - (U+D)C_0(0,2;K_n,2h) + DC_0(-1,2;K_n,2h) \right] e^{-r 2h} \\
 &+ a(-1,1) P_{-1} \left[UC_0(0,2;K_n,2h) - (U+D)C_0(-1,2;K_n,2h) + DC_0(-2,2;K_n,2h) \right] e^{-r 2h}
 \end{aligned} \tag{17}$$

This approximate option pricing model, $C_1(0,0;K_n,2h)$, is linear in the local variance adjustments $[a(i,j)]$. All other terms in equation 17) are specified by the constant volatility CRR model: all $C_0(i,j;K_n,2h)$'s and P_k 's. The accuracy of this linear approximation depends on the CRR model being a reasonably good model, so that local variance adjustments to the CRR constant volatility are not too large.

³ For example, see Bouchoev (1997).

2.1 Estimating the Linearized Local Volatility Trinomial Model

The linearized estimation problem is the following⁴:

$$\text{Min}_{\psi(a)} \sum_{m=2}^{M=2} \sum_{n=1}^{N=3} \varepsilon_1(0,0;K_n,mh)^2 + \alpha \sum_{i=-1}^1 \sum_{j=0}^i a_1(i,j)^2 \quad 18)$$

where $C(0,0;K_n,2h) = C_1(0,0;K_n,2h) + \varepsilon_1(0,0;K_n,2h)$, and

$$\psi(a_1) = \{a_1(0,0), a_1(1,1), a_1(0,1), a_1(-1,1)\}$$

We restate model equation 17):

$$C_1(0,0;K_n,2h) = C_0(0,0;K_n,2h) + X_0(1,1;K_n,2h) a_1(1,1) \quad 19)$$

$$+ X_0(0,1;K_n,2h) a_1(0,1) + X_0(-1,1;K_n,2h) a_1(-1,1) + X_0(0,0;K_n,2h) a_1(0,0)$$

$$X_0(i,1;K_n,2h) = P_1 \left[U \text{Max}(S(i+1,2) - K_n, 0) - (U+D) \text{Max}(S(i,2) - K_n, 0) \right. \\ \left. + D \text{Max}(S(i-1,2) - K_n, 0) \right] e^{-r2h}$$

$$X_0(0,0;K_n,2h) = X_0(1,1;K_n,2h) + X_0(0,1;K_n,2h) + X_0(-1,1;K_n,2h)$$

Clearly, the model is linear in a_1 , the parameter vector. In matrix notation,

$$\text{Min}_{\psi(a_1)} (Y - X_0 a_1)' (Y - X_0 a_1) + \alpha_1 a_1' a_1, \quad \psi(a_1) = [a_1(0,0), a_1(1,1), a_1(0,1), a_1(-1,1)]' \quad 20)$$

$$X_0 = \begin{bmatrix} X_0(0,0;K_1,2h) & X_0(1,1;K_1,2h) & X_0(0,1;K_1,2h) & X_0(-1,1;K_1,2h) \\ X_0(0,0;K_2,2h) & X_0(1,1;K_2,2h) & X_0(0,1;K_2,2h) & X_0(-1,1;K_2,2h) \\ X_0(0,0;K_3,2h) & X_0(1,1;K_3,2h) & X_0(0,1;K_3,2h) & X_0(-1,1;K_3,2h) \end{bmatrix}$$

$$Y = \begin{bmatrix} C(0,0;K_1,2h) - C_0(0,0;K_1,2h) \\ C(0,0;K_2,2h) - C_0(0,0;K_2,2h) \\ C(0,0;K_3,2h) - C_0(0,0;K_3,2h) \end{bmatrix}$$

⁴ Following usual convention, iterations over index j that decrease the index are set to zero.

Given α_1 , we solve the first-order condition equation:

$$a_1 = \left(X_0' X_0 + \alpha_1 I \right)^{-1} X_0' Y \quad (21)$$

Given a positive regularizer weight, the inverse in this equation always exists. With regard to choosing the regularizer weight, we follow the Tikhonov approach as defined with equation 16). Given a regularizer weight, parameter estimation is linear, and any standard search algorithm (e.g. bisection or secant method) solves the discrepancy function for the "optimal" weight quickly. Table 6 presents the linearized estimates for our example:

Table 6: Local Volatility Parameters

	α	δ^2	$\sigma(0,0)$	$\sigma(1,1)$	$\sigma(0,1)$	$\sigma(-1,1)$
Linearized estimates	0.00088	1E-07	18.15%	12.68%	16.33%	26.50%
Nonlinear estimates	0.2337	1E-07	19.97%	9.52%	14.08%	25.17%
<i>actual</i>			19.00%	10.00%	15.81%	26.46%

From Table 6, we see that the estimated linearized local volatility estimates roughly match the actual volatilities. The linearized estimates also compare favorably to the nonlinear estimates of Table 5, which were reported.

Given the estimated volatility surface, non-traded European options may be marked to market. For a K^* strike price, T^* maturity and X_0^* regressor vector option, the estimated linear approximation value is

$$C_1(0,0,K^*,T^*) = C_0(0,0,K^*,T^*) + X_0^* a_1 \quad (22)$$

Alternatively, we may use our estimated variance parameters to solve the true (nonlinear) option model, equation 12). Table 7 reports the results of this calculation.

Table 7: Linear and Nonlinear Model Values
based on Linearized Variance Parameter Estimates

Model / Options	Out of the money	At the money	In the money
Linear (\cong actual price)	0.291	5.844	18.739
Nonlinear (based on linear estimates)	0.428	5.707	18.597
CRR Model	1.297	7.052	18.134

Though based on a linearized estimator, the nonlinear model values are quite close to the actual prices (and much closer to the actual values than the CRR model values.) Table 7 illustrates that the solution to the direct nonlinear option value problem with a linearized and regularized set of variance parameter estimates is reasonable.

2.2 Two Restricted Estimates: Time-dependent and State-dependent Volatility

In our example's context, the finance literature motivates two restrictions of the full time- and state-dependent local volatility specification. The local volatilities may be restricted to be equal across spot price states or across time periods. Restricting the local volatilities to be equal across spot price states results in a time-dependent set of estimates – e.g. Merton (1973). Restricting the local volatilities to be equal across time at spot price state results in a state-dependent set of estimates – e.g. Cox-Ross (1976) and Cox-Rubinstein (1985). These restrictions reduce our surface estimation problem to the curve estimation sub-case.

These restricted specifications may be nested in the local volatility ridge regression specification. The parameter set is reduced by the equality constraints. The regressor set is similarly reduced by summation of all time- and state-dependent specification regressors that are associated with a particular restricted coefficient.

In the context of our two-period example, the time-dependent local volatility parameters are

the minimizers of the following optimand.

$$\begin{aligned} \text{Min}_{\psi(a_t)} & \left(Y - X_{0,t} a_t \right)' \left(Y - X_{0,t} a_t \right) + \alpha_t \left[a_t(0,0)^2 + 3a_t(0,1)^2 \right], \quad \psi(a_t) = [a_t(0,0), a_t(0,1),] \\ X_{0,t} & = \left[\begin{array}{c} \left(X_0(0,0;K_1,2h) \right) \left(X_0(1,1;K_1,2h) + X_0(0,1;K_1,2h) + X_0(-1,1;K_1,2h) \right) \\ \left(X_0(0,0;K_2,2h) \right) \left(X_0(1,1;K_2,2h) + X_0(0,1;K_2,2h) + X_0(-1,1;K_2,2h) \right) \\ \left(X_0(0,0;K_3,2h) \right) \left(X_0(1,1;K_3,2h) + X_0(0,1;K_3,2h) + X_0(-1,1;K_3,2h) \right) \end{array} \right] \end{aligned} \quad (23)$$

For the restricted minimand, the regularizer must account for the number of times that the parameters enter the trinomial model lattice. In the time-dependent volatility specification, the parameter $a(0,0)$ occurs once, but the parameter $a(0,1)$ occurs three times in substituting for all of the second period coefficients, $a(1,1)$, $a(0,1)$ and $a(-1,1)$. The regressor associated with the $a(0,0)$ parameter is the same as in the unrestricted case. The regressor associated with the restricted $a(0,1)$ parameter is equal to the sum of the remaining volatility regressors.

The state-dependent volatility specification restricts all volatility parameters that are associated with a particular spot price outcome to be equal across time. In our example, the $a(0,0)$ and $a(0,1)$ parameters are both associated with the 100.0 spot price.

For this example, imposing this restriction reduces both the parameter set and the regressor set by one. The minimand becomes

$$\begin{aligned} \text{Min}_{\psi(a_s)} & \left(Y - X_{0,s} a_s \right)' \left(Y - X_{0,s} a_s \right) + \alpha \left[2a_s(0,1) + a_s(1,1) + a_s(-1,1) \right], \\ \psi(a_s) & = [a_s(0,1), a_s(1,1), a_s(-1,1)] \end{aligned} \quad (24)$$

$$X_{0,s} = \left[\begin{array}{c} \left(X_0(0,0;K_1,2h) + X_0(0,1;K_1,2h) \right) \left(X_0(1,1;K_1,2h) \right) \left(X_0(-1,1;K_1,2h) \right) \\ \left(X_0(0,0;K_2,2h) + X_0(0,1;K_2,2h) \right) \left(X_0(1,1;K_2,2h) \right) \left(X_0(-1,1;K_2,2h) \right) \\ \left(X_0(0,0;K_3,2h) + X_0(0,1;K_3,2h) \right) \left(X_0(1,1;K_3,2h) \right) \left(X_0(-1,1;K_3,2h) \right) \end{array} \right]$$

Table 8 presents the restricted time-dependent and state-dependent volatility parameter estimates.

Table 8: Restricted Local Volatility Surface Estimates

	α_1	δ^2	$\sigma(0,0)$	$\sigma(1,1)$	$\sigma(0,1)$	$\sigma(-1,1)$
<i>Time-dependent</i>	9098.0	0.90	18.64%	19.55%	19.55%	19.55%
<i>State-dependent</i>	0.0037	1E-07	17.27%	13.86%	17.27%	27.08%
<i>Time- and state-dependent</i>	0.0009	1E-07	18.15%	12.68%	16.33%	26.50%
actual			19.00%	10.00%	15.81%	26.46%

Of the two restricted models, the state-dependent specification is sufficient to match the desired discrepancy level. As the minimizer for the time-dependent restriction does not meet the desired discrepancy level, this specification is overly restrictive.

2.3 Some Intuition and Example Extension to Practice

In the linearized model, the X ridge regressor variables have an appealing interpretation. The (i,j)th regressor is proportional to the expected gamma taken over all paths passing through the ith state and jth time step volatility point, $\sigma(i,j)$. Importantly, the expectation is taken with respect to the CRR risk-neutral probabilities, which are fixed, for all j.

A particular strike price, K_n , and maturity, T_m , expected gamma may be calculated from two lattices: one for the probability of reaching a node and the other for the expected gamma beyond the node. To illustrate the general structure of these lattices, we represent a four time-step lattice in matrix table form.

The probability of reaching a node with i net spot price up moves in the jth period is $\bar{P}(i,j)$. The sum i is aggregated over all previous spot price state moves. Following our previous definition of the indicator $k = [1,0,-1]$, an up move adds one to i, no move leaves i as is,

and a down move subtracts one from i . For instance, $\bar{P}(1,3)$ is the probability of one-net spot price up move after 3h time is passed. Therefore, $\bar{P}(1,3)$ equals the cumulative probability of reaching state $S(1,3)$. For the K_n strike price and T_m maturity option contract, the (i, j) node expected gamma value is $G(i, j, K_n, T_m)$.

The associated regressor variable is defined as the discounted product of the probability of reaching a lattice node and the expected gamma at that node:

$$X_0(i, j, K_n, T_m) = \bar{P}(i, j)G(i, j, K_n, T_m)e^{-rT_m} \quad 25)$$

Table 9 illustrates the recursions that underlie the $\bar{P}(i, j)$ and $G(i, j, K_n, T_m)$ values. Given $M=4$ time steps, value calculations don't require the $\bar{P}(i,4)$ column. Nevertheless, we report these values to further demonstrate the probability recursion.

To program the $\bar{P}(\bullet)$, $G(\bullet)$ and $X(\bullet)$ values with more than four time steps, we simply propagate the probability lattice *forward* to step $M-1$, while the expectation component of the gamma recursion propagates *back* from step $M-2$. The number of X regressor variables for any one option is dependent on the number of time periods, M . The regressor variable count and the number of separate variance parameters in the lattice both equal $M^2 \left(= \sum_{i=0}^{M-1} 2i+1 \right)$.

A 60 time-step implementation involves solution of a 3600 equation system. We solve this equation system by the singular value decomposition (SVD) method. This method, which is stable and relatively fast, is outlined in the appendix.⁵

⁵ We use G.W. Stewart, www.netlib.no/netlib/linpack/dsvdc.f, version dated 08/14/78 with correction made to shift 2/84, University of Maryland, Argonne National Lab.Routine. See Golub and VanLoan (1989) for discussion.

Table 9: Time- and State-Dependent Probabilities and Expected Gammas - Components of Regressor Variable, $G(i,j,K_i,4h)$

$\bar{P}(i,j)$	1	2	3	4
4				$\bar{P}(3,3) P_u$
3			$\bar{P}(2,2) P_u$	$\bar{P}(3,3) P_m + \bar{P}(2,3) P_u$
2		$\bar{P}(1,1) P_u$	$\bar{P}(2,2) P_m + \bar{P}(1,2) P_u$	$\bar{P}(3,3) P_d + \bar{P}(2,3) P_m + \bar{P}(1,3) P_u$
1	P_u	$\bar{P}(1,1) P_m + \bar{P}(0,1) P_u$	$\bar{P}(2,2) P_d + \bar{P}(1,2) P_m + \bar{P}(0,2) P_u$	$\bar{P}(2,3) P_d + \bar{P}(1,3) P_m + \bar{P}(0,3) P_u$
0	P_m	$\bar{P}(1,1) P_d + \bar{P}(0,1) P_m + \bar{P}(-1,1) P_u$	$\bar{P}(1,2) P_d + \bar{P}(0,2) P_m + \bar{P}(-1,2) P_u$	$\bar{P}(1,3) P_d + \bar{P}(0,3) P_m + \bar{P}(-1,3) P_u$
-1	P_d	$\bar{P}(-1,1) P_m + \bar{P}(0,1) P_d$	$\bar{P}(0,2) P_d + \bar{P}(-1,2) P_m + \bar{P}(-2,2) P_u$	$\bar{P}(0,3) P_d + \bar{P}(-1,3) P_m + \bar{P}(-2,3) P_u$
-2		$\bar{P}(-1,1) P_d$	$\bar{P}(-2,2) P_m + \bar{P}(-1,2) P_d$	$\bar{P}(-1,3) P_d + \bar{P}(-2,3) P_m + \bar{P}(-3,3) P_u$
-3			$\bar{P}(-2,2) P_d$	$\bar{P}(-3,3) P_m + \bar{P}(-2,3) P_d$
-4				$\bar{P}(-3,3) P_d$

$G(i,j;K_n,4h)$ = $G(i,j)$	1	2	3	4
4				$\text{Max}(S(4,4)-K_i,0)$
3			$UG(4,4)-(U+D)G(3,4)+DG(2,4)$	$\text{Max}(S(3,4)-K_i,0)$
2		$P_u * G(3,3) + P_m * G(2,3) + P_d * G(1,3)$	$UG(3,4)-(U+D)G(2,4)+DG(1,4)$	$\text{Max}(S(2,4)-K_i,0)$
1	$P_u * G(2,2) + P_m * G(1,2) + P_d * G(0,2)$	$P_u * G(2,3) + P_m * G(1,3) + P_d * G(0,3)$	$UG(2,4)-(U+D)G(1,4)+DG(0,4)$	$\text{Max}(S(1,4),K_i,0)$
0	$P_u * G(1,2) + P_m * G(0,2) + P_d * G(-1,2)$	$P_u * G(1,3) + P_m * G(0,3) + P_d * G(-1,3)$	$UG(1,4)-(U+D)G(0,4)+DG(-1,4)$	$\text{Max}(S(0,4)-K_i,0)$
-1	$P_u * G(0,2) + P_m * G(-1,2) + P_d * G(-2,2)$	$P_u * G(0,3) + P_m * G(-1,3) + P_d * G(-2,3)$	$UG(0,4)-(U+D)G(-1,4)+DG(-2,4)$	$\text{Max}(S(-1,4)-K_i,0)$
-2		$P_u * G(-1,3) + P_m * G(-2,3) + P_d * G(-3,3)$	$UG(-1,4)-(U+D)G(-2,4)+DG(-3,4)$	$\text{Max}(S(-2,4)-K_i,0)$
-3			$UG(-2,4)-(U+D)G(-3,4)+DG(-4,4)$	$\text{Max}(S(-3,4)-K_i,0)$
-4				$\text{Max}(S(-4,4)-K_i,0)$
$G(0,0)=$	$P_u * G(1,1) + P_m * G(0,1) + P_d * G(-1,1)$			

The time index runs across the table, and the state index runs down the table.

3. PHLX Options Data

We extracted one day of trades for the Philadelphia Stock Exchange (PHLX) Deutschemark European options.⁶ The trades and the associated strike prices, maturities, simultaneous spot prices and times of trade are all listed in Table 10.

Table 10: 11/25/95 PHLX DM European Option Trades

Strike Price	Days to maturity	DM spot	Trade time	Strike Price	Days to maturity	DM spot	Trade time
62.0	18	63.19	757	63.0	18	63.13	1532
62.0	18	63.19	757	63.0	18	63.13	1532
62.0	18	63.23	814	63.5	18	63.12	728
62.0	18	63.23	814	63.5	18	63.12	728
62.0	18	63.12	931	64.5	18	63.11	729
62.0	18	63.12	931	64.5	18	63.11	729
62.5	18	63.20	756	66.0	53	63.11	943
62.5	18	63.20	756	66.0	53	63.11	943
62.5	18	63.21	823	58.0	109	63.04	710
62.5	18	63.21	823	58.0	109	63.04	710
62.5	18	63.21	824	62.0	109	63.06	712
62.5	18	63.21	824	62.0	109	63.06	712
62.5	18	63.20	825	65.0	109	63.06	712
62.5	18	63.20	825	65.0	109	63.06	712

Among the 28 available trades, in-, at- and out-of-the-money options are traded at three maturities. Most trades were in the shortest (18-day) maturity and were concentrated from 7:00 - 8:30 am. . Necessary Eurodollar and EuroDeutschemark interest rates were obtained from Datastream. Figure 4 plots the different implied volatilities associated with each option trade. Across all 28 trades, the least squares minimizing implied volatility estimate is 12.88%.⁷ The 12.88% implied volatility estimate is our (constant) volatility surface prior, σ_0 .

⁶ Note that the PHLX European options are much less frequently traded than the companion American options. Nevertheless, the prices are representative due to the close substitutability of the options.

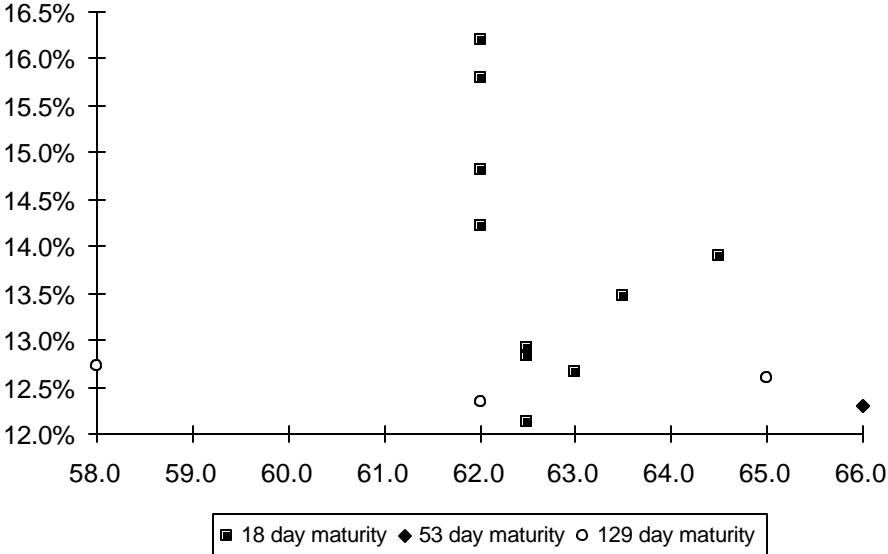
⁷ We use the Whaley (1982) procedure.

We must also define a discrepancy level for this particular set of option trades. Unlike the simple example discrepancy of sections 1.2 and 2.1, which depended only on rounding error, fitted option values are subject to both input bid-ask spread error and model error.

To provide rudimentary estimates of these errors, we have sampled the bid-ask spreads for PHLX Deutschemark options from February 9, 1995 through April 14, 1997. During this time period, the average PHLX bid-ask volatility spread was 0.64% with a 0.52% standard error.

Figure 4

PHLX DM Implied Volatilities - November 25, 1991



4. Estimation and Results

Bid-ask spread error and model error give rise to errors in equation and errors in variable problems of the standard regression application, respectively. Since spread and model errors swamp rounding and discretization errors, we define the discrepancy as a function of estimated spread and model errors. The errors are estimated as functions of half of the average bid-ask volatility spread, $\bar{b} = 0.32\%$, or defined as functions of half of the average bid-ask volatility spread plus one standard deviation, $\hat{b} = 0.58\%$. Spread-induced error creates a feasible price

bound around a traded price. We estimate this error as a sum of squared differences of CRR model option values.

The two values used to estimate the spread-induced price differences are determined by the CRR implied volatility plus and minus the spread estimate, b :

$$\text{spread error: } \delta(b)^2 = \sum_{n=1}^{\#\text{options}} [C_0(0, K_n, T_n, \sigma_0 + b) - C_0(0, K_n, T_n, \sigma_0 - b)]^2 \quad 26)$$

Incremental model error arises in our fitted values because we do not know the true (gamma) regressor matrix, $X(\bullet)$. We model this error by calculating a perturbed set of regressors. To generate the perturbation, we subtract the spread from the common implied volatility estimate and calculate the associated regressor matrix. Then, we subtract this matrix from the original regressor matrix. Following the Tikhonov theory, we estimate the model regressor (or operator) error as the maximum singular value of the perturbation matrix:

$$\text{model error - } h = \|H_0(\sigma_0, b)\| = \|X_0(\sigma_0) - X_0(\sigma_0 - b)\| = \sqrt{\max(\text{singularvalue})} \quad 27)$$

Our discrepancy regularizer combines the bid-ask spread pricing error estimate and the model error estimate. The resulting discrepancy function is the following:⁸

$$\rho(\alpha) = (Y - X_0 a_1)' (Y - X_0 a_1) - \alpha_1 \left[\delta + h \sqrt{\sum a_1^2} \right]^2 = 0 \quad 28)$$

Given a smoothing parameter α_1 , our estimation problem is linear. Conditional on a suitable regularizer function, Tikhonov has shown that the optimal smoothing parameter search, which solves discrepancy function 28), is monotonic and non-increasing. For our application,

⁸ The model error term, $h \sqrt{\sum a_1^2}$, is an upper bound estimate of the textbook errors in variable contribution to total squared error, $E \left[H_0(\sigma_0, b) a_1 a_1' H_0(\sigma_0, b)' \right]$.

the sum of squared parameter estimates is a suitable regularizer, and a grid, bi-section or secant method search yields the optimal smoothing parameter efficiently.

We fit the 28 option prices listed in Table 10. A 60 time-step lattice is used, and this lattice has 3,600 local variance parameters. To handle three maturities, variable time steps are used over the three maturity intervals. The parameters minimize the following optimand:

$$\text{Min}_{\{a_1\}} \sum_{n=1}^{N=28} \varepsilon_1(0,0,K_n,T_n)^2 + \alpha_1 a_1' a_1 \quad 29)$$

where $\Psi(a_1) = \{a_1(0,0), a_1(-1,1), a_1(0,1), a_1(1,1), a_1(-2,2), \dots, a_1(-59,59), \dots, a_1(59,59)\}$

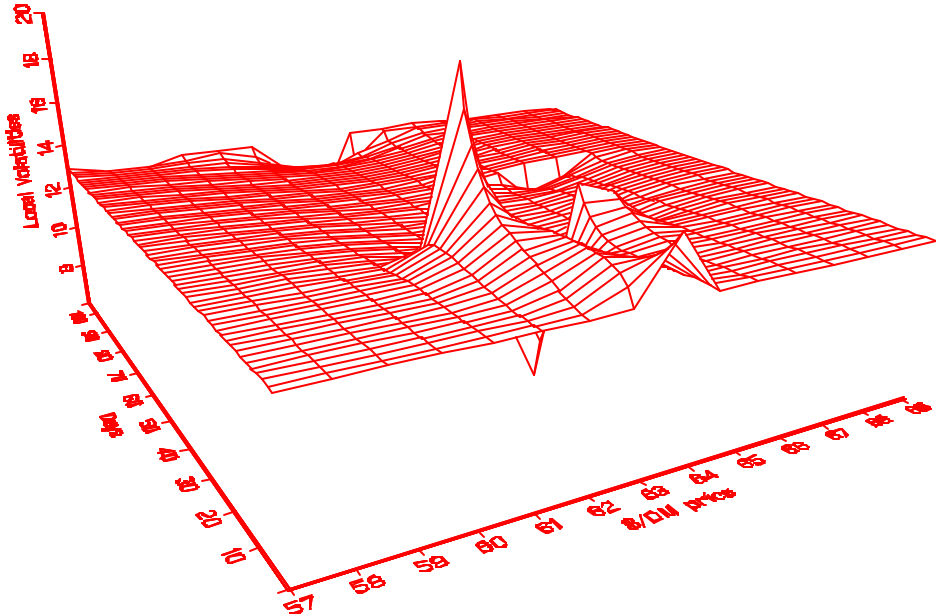
Figure 5 plots our local volatility surface estimates. Figure 5 contains two panels, and we differentiate panel plots by the discrepancy level used in estimation. In panel A, the average bid-ask volatility spread (b=0.64%) is used to define the discrepancy function, and the function root is found with a 0.00111 least squares fit and a regularizing coefficient (α) of 0.14. The associated local volatility plot is quite peaked. The maximum estimate is 19.8% and the minimum estimate is 6.3%.

Interestingly, the maximum and minimum estimates are associated with the same time, 18 days. Furthermore, the respective spot outcomes, 61.58 and 63.19, are separated by only one state point, 62.38. At the optimal $\alpha_1 = 0.14$ level, these estimates indicate that the least squares criterion is relatively flat with regard to offsetting changes in two parameters. For the equation 29) optimand, small improvements in least squares are largely driven by linked changes in these two parameters.⁹

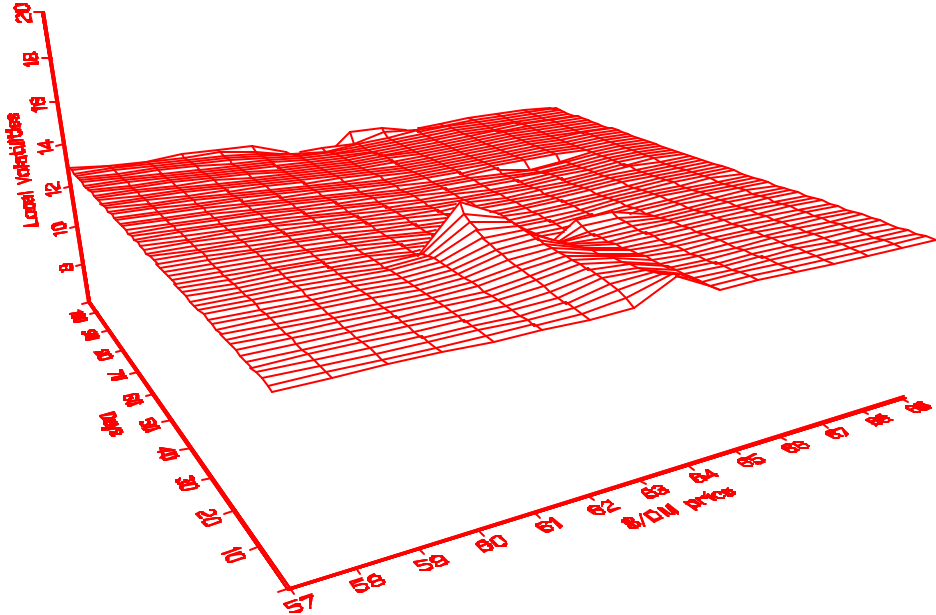
⁹ A natural way to address this optimand characteristic is to augment the regularizer with first and second differences of the parameter estimates. This procedure is fully developed in Bodurtha-Jermakyan (1999). Though we believe that imposition of these added constraints will yield a similar level of fit with much smoother estimates, the computational load of estimating that system precludes our undertaking the exercise presently.

Figure 5: PHLX DM Option Local Volatility Surface - 11/25/91

Panel A: 0.84% Bid-Ask Vol Spread-Based Discrepancy



Panel B: 1.16% Bid-Ask Vol Spread-Based Discrepancy



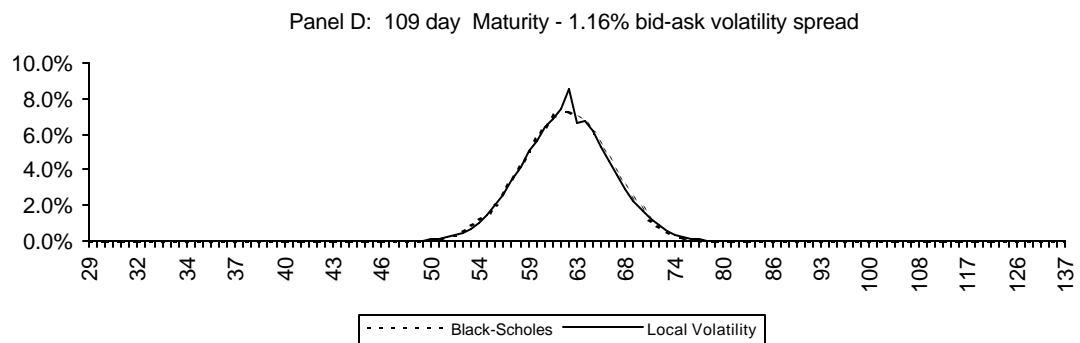
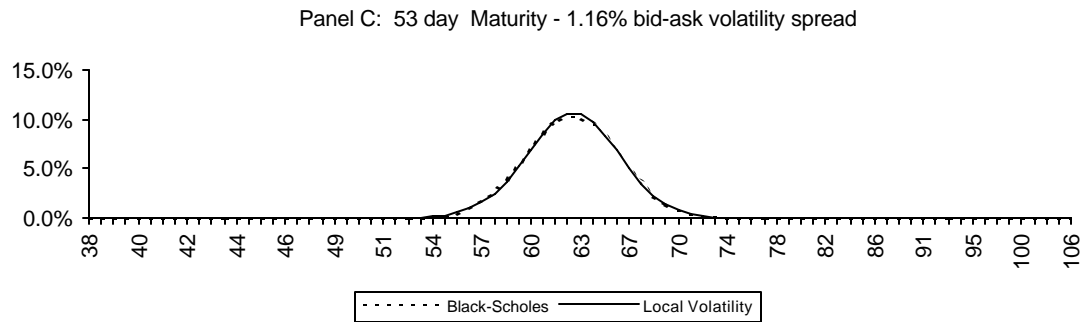
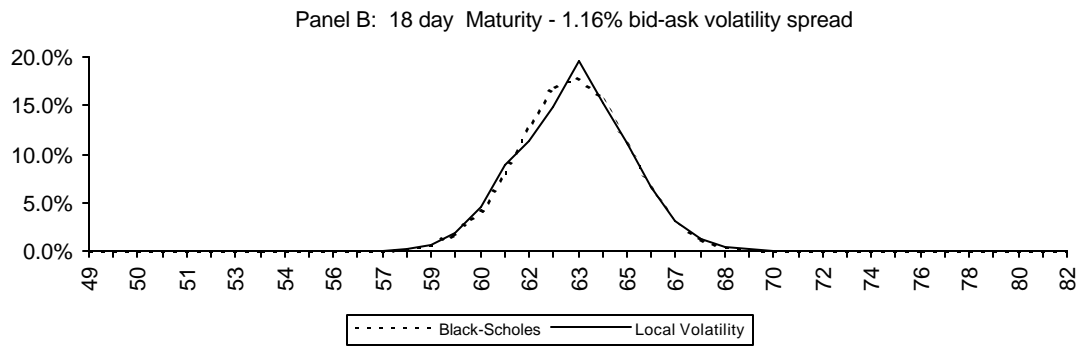
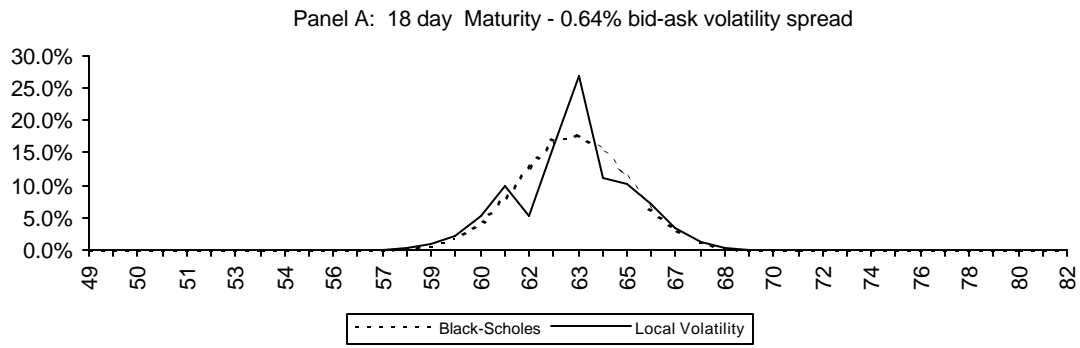
In our view, the estimates depicted in Figure 5 - Panel A are too rough. This view implies that the discrepancy value setting convergence is too low. Therefore, we also estimate the parameters with the discrepancy set by the average bid-ask volatility spread plus one standard deviation ($b=1.16\%$).

Under this higher discrepancy value, Figure 5 - Panel B depicts the estimates. The result is a smoother local volatility surface, 0.00203 least squares fit and 1.01 regularizing parameter (α). The plot has minimum and maximum local volatility estimates of 11.54 and 14.53, respectively. The general shape of the function is a saddle, with a valley running along the at-the-money strike price (Deutschemark spot at 63.19) and significant ridges running along some lower strike prices.

Another indication that the higher discrepancy value is appropriate for this data set is found in Figure 6. Figure 6 plots term probability estimates to the fitted option maturities across spot price states. An option maturity time- and state-specific term probability estimate is the sum of the path-dependent local probability products that reach the state and time. For the 18 day maturity probabilities, the Figure 6 Panel A and Panel B plots are both peaked. However, the lower bid-ask spread (0.64%) discrepancy Panel A plot is jagged while the higher (1.16%) discrepancy Panel B plot is relatively smooth. The smooth term probability curve is more reasonable.

Focusing on the 1.16% bid-ask volatility spread discrepancy plots, we see that the 18 days to maturity plot (Panel B) and 109 days to maturity plot (Panel D) clearly differ from the associated Black-Scholes-Merton log-normal plots. For both maturities, a larger chance of staying at or slightly above the current market level is implied relative to log-normal.

Figure 6: PHLX DM Option Local Volatility Surface Implied Term Probabilities - 11/25/91



For the short 18 day maturity options, the probability of small down moves is lower than log-normal, and the probability of large down moves is higher than log-normal. For the long 109 day maturity options, the probability of moderate up moves is lower than log-normal. The 53 days to maturity plot (Panel C) is quite close to log-normal.

In evaluating our estimates, it is natural to ask if seemingly simpler models may perform as well. The finance literature motivates three important alternatives:

- 1) Time-dependent volatility, $\sigma(S,t)=\sigma(t)$, introduced by Merton (1973)
- 2) State-dependent volatility, $\sigma(S,t) = \sigma(S)$, related to Cox-Ross (1976) and Cox-Rubinstein (1985)
- 3) Option maturity time- and state-dependent volatility, $\sigma(S,t) = \sigma(S,t_1,t_2,\dots,t_M)$, M = number of traded option maturities, related to Rubinstein (1994).

As illustrated in section 2.2, these alternatives may be nested within our local volatility specification as restrictions. To illustrate this nesting, we outline the first alternative specification. Since our discretized trinomial lattice has 60 time steps, the restriction imposed by Merton's time-dependent volatility reduces the parameter space to only 60 independent estimates.

The 60 associated regressors are the sums of the unrestricted regressors at a point in time across spot price states. Furthermore, the associated local variance parameters occur multiple times in the lattice. In order to keep the size of both the parameter space and the regularizer equal across specifications, the regularizer must be weighted by the number of states that a given time-dependent volatility enters the lattice.

Ordering the volatility parameters by time from zero to fifty-nine, the appropriate restricted optimand is the following:

$$\begin{aligned} \text{Min}_{\psi(a_t)} \sum_{n=1}^{N=28} \varepsilon_t(0,0,K_n,T_n)^2 + \alpha \sum_{j=1}^{60} (2j-1)a_t^2(0,j-1) \end{aligned} \quad (30)$$

where $\psi(a_t) = \{a_t(0,0), a_t(0,1), a_t(0,2), \dots, a_t(0,59)\}$

For the state-dependent volatility specification, there are 119 regressors and independent parameter estimates in a 60 time-step lattice. For the option maturity time- and state-dependent specification, there are three maturities and we discretize time so that 20 time steps are between neighboring maturity times. Therefore, this specification has 237 regressors and independent parameters. For both of these cases, the squared parameters in the regularizer portion of the optimand are weighted by their count in the trinomial lattice. As discussed in the appendix, these equation systems are solved using the singular value decomposition (SVD) method.

To determine the relative fit of the alternative specifications, we estimate each model under the discrepancy function defined by the average bid-ask volatility spread plus one standard deviation. The unconstrained local volatility model, which is depicted in and Figure 6-Panel A, has a 0.00203 optimal discrepancy and a 1.01 smoothness parameter (or α_1).

Figure 6 also contains plots of all three restrictions on our time- and state-dependent local volatility specification. Of the three restricted cases only the option maturity time- and state-dependent specification meets the discrepancy criterion. Since the time-dependent and state-dependent specifications can't meet the discrepancy criterion, the associated volatility surface plots are the minimum restricted least squares estimates.

In panel B, the time-dependent volatility function is plotted. The time-dependent volatility estimate is highest (16.09%) at the initial time, falls to a minimum (12.03%) at the shortest option maturity, 18 days, and then rises slowly to the final maturity (12.66%). In panel

C, the state-dependent volatility function forms a “W”. The maximum volatility estimate is 13.95%, which is associated with the 62.38 spot price state. The minimum volatility estimate is 11.15%, which is associated with the 61.59 spot price state.

In panel D, three option maturity time- and state-dependent volatility curves are plotted. The general form of these curves follows the state-dependent surface for the shortest maturity segment and the variability of this curve then damps out through the two subsequent maturities. The minimum local volatility estimate (8.44%) is associated with the second maturity slice, 18-53 days, and the initial spot price state, 63.19. The maximum local volatility estimate (26.70%) is associated with the first 18 day maturity curve and the 62.38 spot price.

The plots in Figure 6 are all estimated with either a 0.00203 least squares fit or the minimum least square fit if it is higher. The state-dependent and time-dependent specifications only attain least squares fits of 0.00250 and 0.00270, respectively. Therefore, both restrictions are rejected as alternative local volatility surface specifications.

Among the other two specifications, we should chose the parameterization that is the smoothest. The simplest smoothness criterion to apply is the α coefficient of the regularizing constraint. The model that attains the 0.00203 least squares fit with the largest regularizing constraint weight is “smoothest”.¹

¹ For all of our linearization-based specifications, a very important question is the following: What is the fit of the true nonlinear model with linearized parameter estimate inputs? The answer is better than the linearized fit. Convergence of the time- and state-dependent specification occurs with a slightly smoother set of estimates, $\alpha_1=1.06$. Both the time-dependent and option maturity time- and state-dependent specifications meet the discrepancy criterion with smoother estimates than in the linearized model, $\alpha_t=0.073$ and $\alpha_s=0.75$, respectively.

Figure 6: PHLX DM European Option Implied Volatility Surface Estimates - 11/25/91

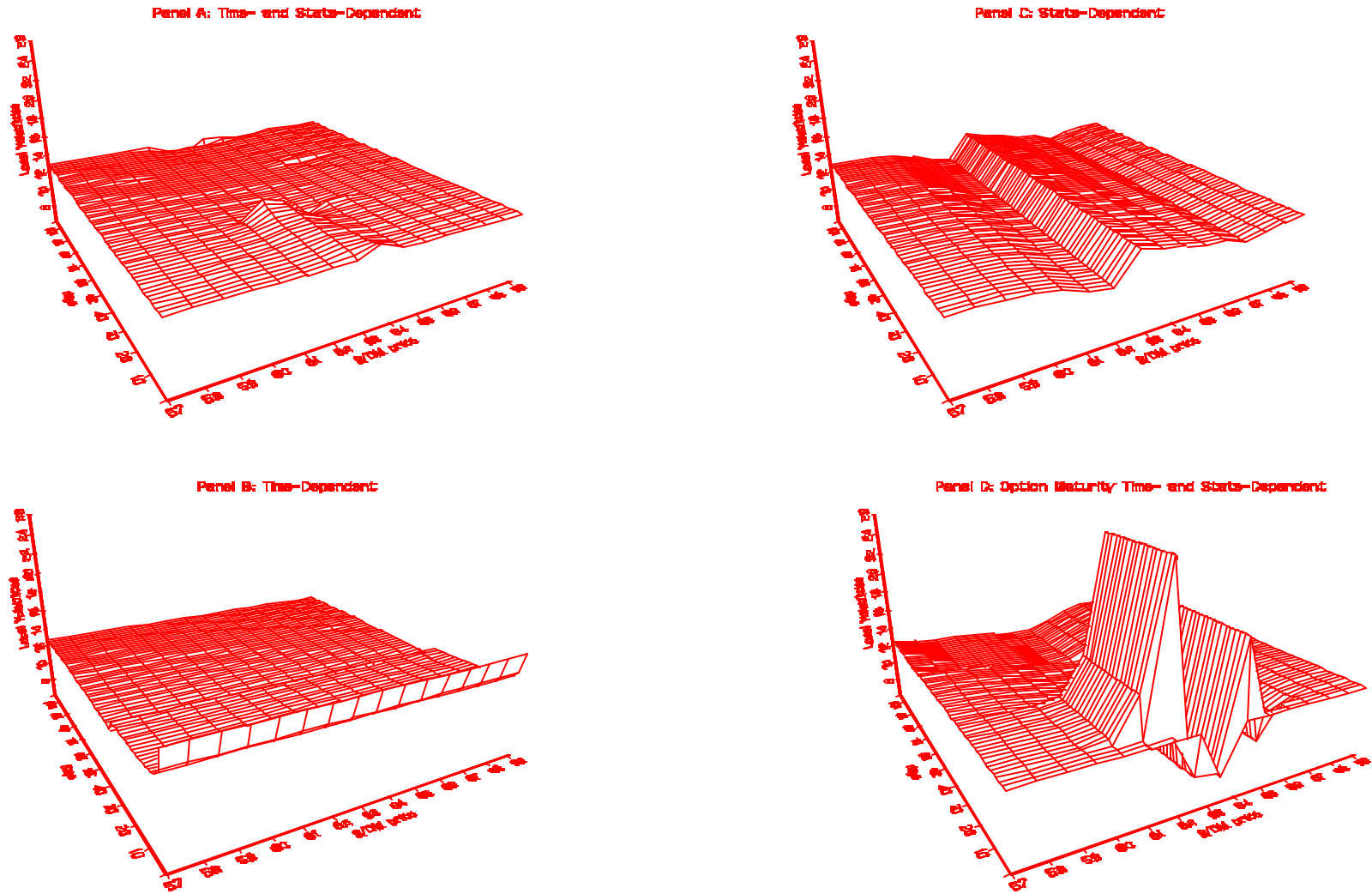


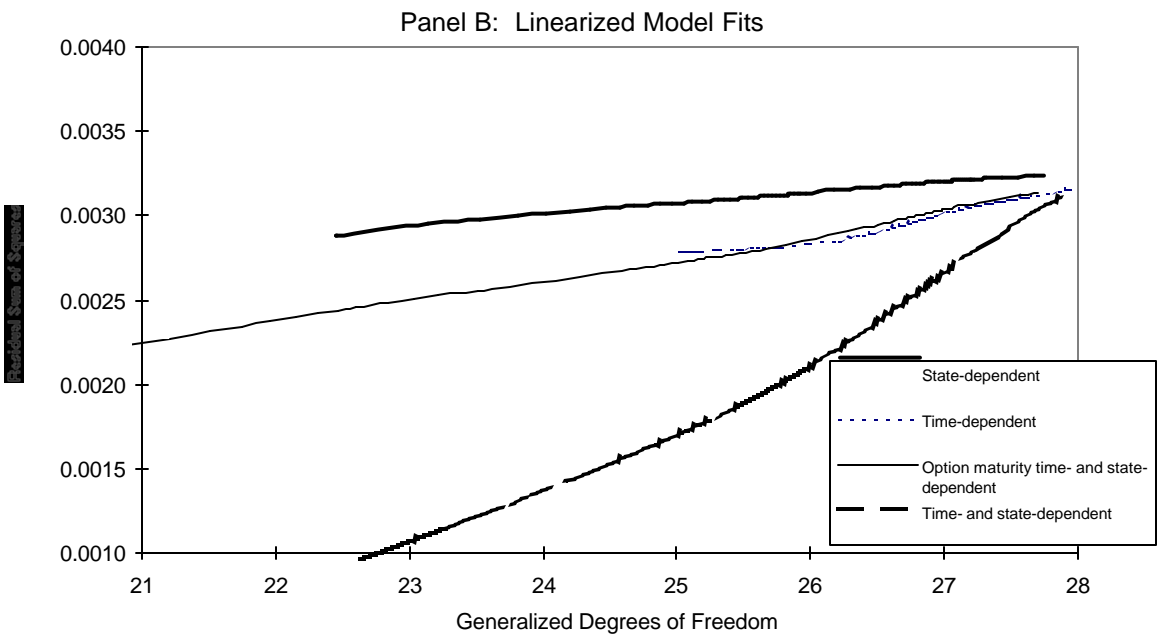
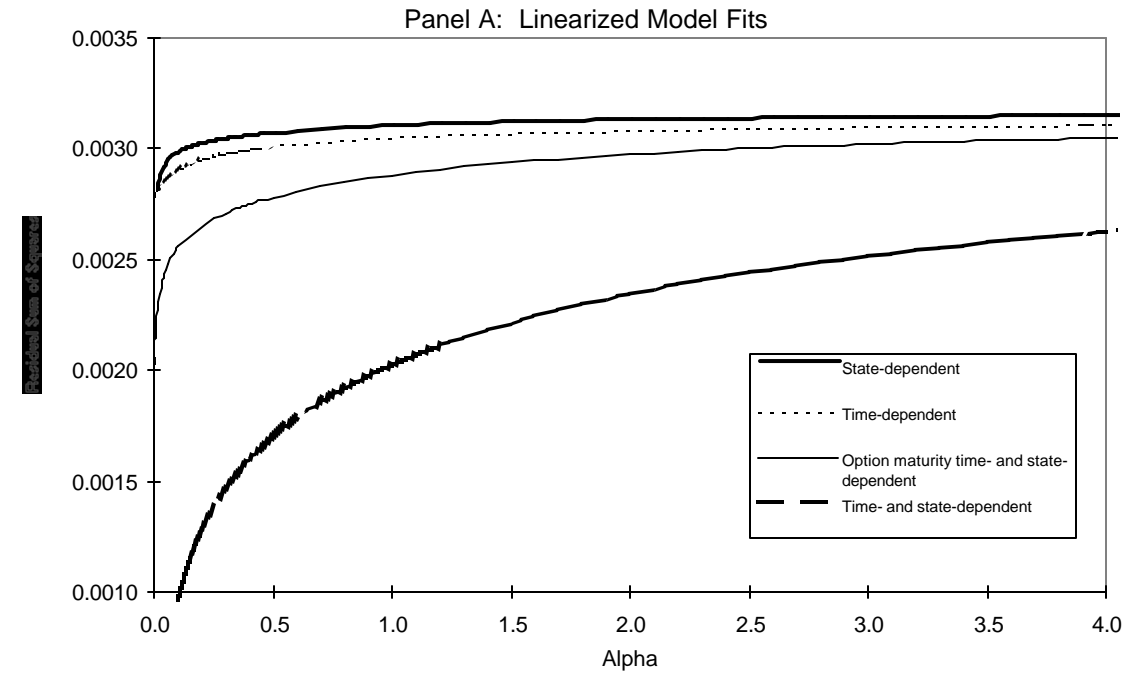
Figure 7-Panel A reports the least squares model fitting levels as functions of α_1 . Clearly, the time- and state-dependent local volatility specification is preferred. The state-dependent and time-dependent specifications fail to attain the desired discrepancy level. The option maturity time- and state-dependent specification meets the discrepancy level criterion with a 0.0001 smoothing parameter value.

Across specifications, we see that all fits improve with lower smoothing coefficient values. However, the improvement terminates for all models when alpha becomes small enough that an explicit positive local variance condition is violated. For the time- and state-dependent local volatility specification, this feasible variance violation occurs with alpha equal to 0.09 and a least squares fit of 0.00096.²

Figure 7-Panel B reports significance results based on another measure of smoothness. This measure is motivated generally by the work of Wahba (1983), and specifically follows the specification of Golub, Heath and Wahba (1979). In both works, a measure of estimator dimension is "generalized degrees of freedom" (GDF).

² Were the target discrepancy not met with the proposed (first-stage) linearization, the extended procedure of Bodurtha-Jermakyan (1999) would be applied. In this procedure, the first linearization is applied to a reasonable smoothness tolerance, e.g. only allow the first-stage estimates, $a_1(\bullet)$, to diverge from σ_0 by $\sigma_0/2$. Next, use these first stage estimates as the basis for another linearization analogous to equation 14, and solve for the second-stage value increments. I.e. $C_2(0,0;K_n,T_m)$. The sum of the three option value components is the second stage value estimate, $C_0(0,0;K_n,T_m) + C_1(0,0;K_n,T_m) + C_2(0,0;K_n,T_m)$. Another expansion could be built by keeping the second-stage parameter estimates no larger than $\sigma_0/4$, and so on. Coupling this iterative procedure with the higher-order regularizers that were outlined in footnotes four and eight should guarantee convergence in most applications.

Figure 7: PHLX DM Option Valuation - 11/25/91



The generalized degrees of freedom definition is related to the “hat” matrix of linear regression theory:

$$A_0 = X_0 \left(X_0' X_0 + \alpha_1 I \right)^{-1} X_0' \quad (31)$$

The associated generalized degrees of freedom measure (GDF) is the following:³:

$$\text{GDF} = \text{Trace} \left(I_{\# \text{options}} - A_0 \right) \quad (30)$$

Based on the 1.16% bid-ask volatility spread-induced discrepancy and the optimum 0.00203 least squares, the full time- and state-dependent local volatility specification has 25.83 degrees of freedom left out of 28 observations. To attain the same fit with the option maturity time- and state-dependent specification, we have only 19.05 degrees of freedom⁴. Both the time-dependent and state-dependent specifications have insufficient degrees of freedom to meet discrepancy criterion.

5 Conclusion

We have proposed a particular local volatility surface estimator. The local volatility or time- and state-dependent volatility specification was found to outperform three restricted versions of it, which are motivated by Merton (1973), Cox-Ross (1976) and Cox-Rubinstein (1985), and Rubinstein (1994). Furthermore, we view the model is a robust and natural point from which to analyze misspecification of diffusion-based approximations to derivative price

³ In the non-regularized (non-ridge) regression case, the degrees of freedom are integer and equal to the number of observations minus the number of explanatory variables.

⁴ In interpreting “smoothness,” other norms exist. Due to its central place in estimation theory, we have chosen to implement the GDF metric. Generally, an “eye-test” should also be consistent with most suitable measures of smoothness. In this regard, we note that option maturity time- and state-dependent volatility curves specification and surface is equivalent, in terms of fit, to another “joined” piece-wise linear alternative specification. For this piece-wise linear specification, time-related slopes of the surface would be estimated instead of a fixed set of maturity-time coefficients for a particular state price. As these two specifications are linearly related, the “fit” of both models for a given smoothness parameter value will be the same. Since this specification is dominated by the full time - and state-dependent specification, we have not implemented the visually smoother variant.

functions.

An immediate extension of this work is to consider samples across time. Before conducting this work, it will be quite helpful to implement added smoothness constraints on the estimated local volatility surface. As developed in Bodurtha-Jermakyan (1999), these additional constraints require solution of a generalized ridge regression problem (or generalized singular value decomposition). Based on some preliminary estimation, we have found that the CBOE S&P 500 European option prices, which have been the subject of much value analysis, are rougher than the currency option data that we have examined. As such, CBOE S&P 500 option implied local volatility surface estimates will behave much better with added smoothness constraints.

Another extension of this work is to broaden the set of underlying securities. Unfortunately, most active exchange-traded options are American exercise-type options. Therefore, it is important to develop implied volatility surface estimators for the American option class.

Another promising alternative is treatment of stochastic volatility and/or jump processes. The model evaluations of Bakshi, Cao and Chen (1997) and Bates (1996) provide benchmarks for valuation and hedging.⁵ An encompassing stochastic local volatility model alternative requires development of a two state variable estimation procedure that embodies both local stochastic volatility and the pricing of local volatility risk. Given local volatility risk priors, this two dimensional stochastic process inference problem may be treated in the linearization-based framework.

Finally, transaction costs and other restrictions are an obvious candidate for modeling.

⁵ Also see Dumas-Fleming-Whaley (1998), Abken-Madan-Ramasurtie (1996), Jackwerth-Rubinstein (1999), and Buraschi-Jackwerth (1998).

Though the additional model error that we incorporate in our discrepancy function is an ad-hoc method for dealing with transaction costs, the difficult treatment of transactions costs in a time- and state-dependent local (and probably stochastic) volatility setting remains for derivative valuation research.

Appendix - Singular Value Decomposition

The singular value decomposition of an $m \times n$ dimension matrix X_0 is defined by two orthogonal matrices U and V (which are dimensioned $m \times m$ and $n \times n$, respectively), and by an $m \times n$ dimension diagonal matrix W .⁶ The diagonal elements of the W matrix are ordered $w_{11} \geq w_{22} \geq \dots \geq w_{mm} > 0$, and all other W matrix elements are zero:

$$X_0 = UWV \quad \text{A-1)}$$

Based on this definition, we rearrange and substitute in the volatility estimation equation 21), which is the minimand first-order condition:

$$(V'W'U'UWV + \alpha I)a = V'W'U'Y \quad \text{A-2)}$$

The orthogonal matrices U and V have the following desirable properties: $U'U = UU' = I_{m \times m}$ and $V'V = VV' = I_{n \times n}$. Further matrix manipulation yields the following equation:

$$(W'W + \alpha I)Va = W'U'Y \quad \text{A-3)}$$

In this equation, all quantities except the volatility parameter vector, a , are known. We solve for these parameter values in two steps. First, we define two augmented vectors: $\bar{a} = Va$ and $\bar{Y} = W'U'Y$.

Substituting in equation A-3) results in an diagonal equation system that is trivial to solve:

$$(W'W + \alpha I)\bar{a} = \bar{Y} \quad \text{A-4)}$$

⁶ See Golub and Van Loan (1989).

The second calculation step recovers the desired volatility parameter vector from the diagonal equation system estimate, A-4):

$$a = V'\bar{a} \quad \text{A-5)}$$

To estimate restricted versions of our local volatility specification, we require two additional definitions: We define a diagonal weighting matrix N_r with the square root of the parameter count in the lattice on the diagonal and define $\psi(a_r)$ to be the restricted parameter set. In matrix notation, minimand 20) becomes the following:

$$\text{Min}_{\psi(a_r)} (Y - X_0 a_r)' (Y - X_0 a_r)_r + \alpha a_r' N_r' N_r a_r \quad \text{A-6)}$$

The associated first order condition is

$$(X_0' X_0 + \alpha N_r' N_r) a_r = X_0' Y \quad \text{A-7)}$$

Define a new matrix $Z_r = X_0 N_r^{-1}$ and parameter vector $\hat{a}_r = N_r a_r$. The first order condition A-7) may be written

$$(Z_r' Z_r + \alpha I) \hat{a}_r = Z_r' Y \quad \text{A-8)}$$

We calculate the singular value decomposition of matrix Z_r . The a_r parameter values are calculated by appropriate pre- and post-multiplication of the A-7) first-order condition by the Z_r SVD components and their transposes, as well as the count matrix, N_r , and its inverse.

Obviously, the equation 31) "hat" matrix used in calculating the restricted model generalized degrees of freedom (GD) differs from the unrestricted case. The restricted hat matrix is the following:

$$A_r = Z_r (Z_r' Z_r + \alpha I) Z_r' \quad \text{A-8)}$$

References

- Abken, Peter, Dilip B. Madan and Sailesh Rammamurtie, "Pricing S&P 500 Index Options Using a Hilbert Space Basis," Federal Reserve Bank of Atlanta Working Paper 96-21, December 1996
- Ait-Sahalia, Yacine and Andrew Lo, "Non-Parametric Estimation of State-Price Densities Implicit in Financial Asset Prices," Journal of Finance, 53(2), April 1998, 499-548.
- Andersen, L.B.G. and R. Brotherton-Ratcliffe, "The equity option volatility smile: an implicit finite-difference approach", The Journal Computational Finance, 1(2), Winter 1997/98, 5-37.
- Avellaneda, M, C. Friedman, R. Holmes and D. Samperi, "Calibrating Volatility Surfaces via Relative-Entropy Minimization," Applied Mathematical Finance, 4, March 1997, 37-64.
- Bakshi, Gurdip S., Charles Cao and Zhiwu Chen, "Empirical Performance of Alternative Option Pricing Models," Journal of Finance, 52(5), 1997, 2003-2049.
- Barle, Stanko and Nusret Cakici, "Growing a Smiling Tree," Risk, October 1995, 32-39.
- Bates, David S., "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," Review of Financial Studies, 9(1), 1996, 69-107.
- Bodurtha, Jr., James and Martin Jermakyan, "Non-Parametric Estimation of an Implied Volatility Surface," The Journal Computational Finance, forthcoming, 1999.
- Bouchouev, I., "Derivatives Valuation for General Diffusion Processes," Koch Industries, Inc. working paper, September 1997.
- Brown G., and K.J. Toft, "Constructing Binomial Trees from Multiple Implied Probability Distributions," University of Texas - Austin working paper, July 1996.
- Buraschi, Andrea and Jens Jackwerth, "Is Volatility Risk Priced in the Option Market? Empirical Evidence and Implications for Deterministic and Stochastic Option Pricing Models," London Business School working paper, 1999.
- Cox, John and Stephen Ross, "The Valuation of Options for Alternative Stochastic Processes," Journal of Financial Economics, 3, July 1976, 145-166.
- Cox, John and Mark Rubinstein, Options Markets, Englewood Cliffs, NJ, Prentice-Hall, 1985.
- Derman, Emanuel and Iraj Kani, "Riding on a Smile," Risk, February 1994, 32-39.
- Derman, E., Kani, I. and Zou, J., "The Local Volatility Surface: Unlocking the information in Index Options Prices," Financial Analyst Journal, July-August 1996, 25-36.
- Dumas, Bernard, Jeffery Fleming and Robert Whaley, "Implied Volatility Functions: Empirical Tests", Journal of Finance, 53(6), December 1998, 2059-2106.

- Dupire, Bruno, "Pricing with a Smile," Risk, January 1994, 18-20.
- Golub, Gene, Michael Heath and Grace Wahba, "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter," Technometrics, 21(2), May 1979, 215-223.
- Golub, Gene and Charles Van Loan, Matrix Computations, 2nd ed., Baltimore, Johns Hopkins University Press, 1989.
- Jackwerth, Jens and Mark Rubinstein, "Recovering Probability Distributions from Option Prices," Journal of Finance, 51(5), December 1996, 1611-1632.
- _____, "Recovering Stochastic Processes from Option Prices," London Business School. working paper, 1998.
- Kamrad, Bardia and Peter Ritchken, "Multinomial Approximating Models for Options with k State Variables," Management Science, 37(12), December 1991, 1640-1652.
- Lagnado, Ronald and Stanley Osher, "Reconciling Differences," Risk, 1997.
- Lagnado, R. and S. Osher, "A technique for calibrating derivative security pricing models: numerical solution of an inverse problem", The Journal Computational Finance, Volume 1, Number 1 (Fall 1997).
- Mason, Scott Phillip, "Essays in Continuous Time Finance," April 1979, MIT Ph.d. dissertation.
- Merton, Robert C., "The Theory of Rational Option Pricing," Bell Journal of Economics and Management Science, 4, Spring 1973, 141-183.
- Nychka, Douglas, "Bayesian Confidence Intervals for Smoothing Splines," Journal of the American Statistical Association, 83(4), December 1988, 1134-1143.
- Parkinson, Michael, "Option Pricing: The American Put," Journal of Business, January 1977, 20-36.
- Rubinstein, Mark, "Implied Binomial Trees," The Journal of Finance, 69(3), July 1994, 771-818.
- Schwartz, Eduardo, "The Valuation of Warrants: Implementing a New Approach," Journal of Financial Economics, January 1977, 79-93.
- Scott, Louis O., "Pricing Stock Options in a Jump-Diffusion Model with Stochastic Volatility and Interest Rates: Applications of Fourier Inversion Methods," Mathematical Finance, 7(3), July 1997, 345-358.
- Shimko, David, "Bounds of Probability," Risk, April 1993, 33-37.
- Silverman, B.W., "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting," Journal of the Royal Statistical Society, Series B, 1985, 47(1), 1-52.

Stewart, G.W. www.netlib.no/netlib/linpack/dsvdc.f, version dated 08/14/78 with correction made to shift 2/84, University of Maryland, Argonne National Lab.Routine.

Tikhonov, A.N. and V. Y. Arsenin, Solutions of Ill-Posed Problems, Washington, DC, Winston-Wiley, 1977.

Wahba, Grace, “Bayesian ‘Confidence Intervals’ for the Cross-Validated Smoothing Spline,” Journal of the Royal Statistical Society, Ser. B, 45, 1983, 133-150.