



American Finance Association

Recovering Probability Distributions from Option Prices

Author(s): Jens Carsten Jackwerth and Mark Rubinstein

Source: *The Journal of Finance*, Vol. 51, No. 5 (Dec., 1996), pp. 1611-1631

Published by: Blackwell Publishing for the American Finance Association

Stable URL: <http://www.jstor.org/stable/2329531>

Accessed: 25/11/2009 10:21

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=black>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Blackwell Publishing and American Finance Association are collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Finance*.

<http://www.jstor.org>

Recovering Probability Distributions from Option Prices

JENS CARSTEN JACKWERTH and MARK RUBINSTEIN*

ABSTRACT

This article derives underlying asset risk-neutral probability distributions of European options on the S&P 500 index. Nonparametric methods are used to choose probabilities that minimize an objective function subject to requiring that the probabilities are consistent with observed option and underlying asset prices. Alternative optimization specifications produce approximately the same implied distributions. A new and fast optimization technique for estimating probability distributions based on maximizing the smoothness of the resulting distribution is proposed. Since the crash, the risk-neutral probability of a three (four) standard deviation decline in the index (about -36 percent (-46 percent) over a year) is about 10 (100) times more likely than under the assumption of lognormality.

RECENTLY, THE INCREASING POPULARITY of derivatives and some highly publicized failures to control risk have led to increased efforts to find reasonable methods to measure the sensitivity of large institutional derivatives portfolios to extreme events. Merely because such events are rare is not sufficient to ignore them, since on the few occasions when they do occur, significant amounts of money can change hands, potentially wiping out profits accumulated over long prior periods. A key assumption behind methods of estimation is the joint probability distribution of constituent underlying asset returns. This has long been a concern of financial economists, since probability assumptions are critical to much of their research during the last quarter century. Heretofore, probability distributions of stock market returns have typically been estimated from historical time series. Unfortunately, common hypotheses may not capture the probability of extreme events, and the events of interest are rare or may not be present in the historical record, even though they are clearly possible.

Take for example the stock market crash of October 1987. Following the standard paradigm, assume that stock market returns are lognormally distributed with an annualized volatility of 20% (near its historical realization). On October 19, 1987, the two month S&P 500 futures price fell 29 percent.¹

* Jens Carsten Jackwerth is a postdoctoral visiting scholar and Mark Rubinstein is a professor of finance, both at the Haas School of Business, University of California at Berkeley. For helpful discussions, we thank Ron Lagnado, Hayne Leland, Dennis Klapacz, Steve Manaster, Stewart Mayhew, William Redfearn, and Robert Whaley.

¹ The behavior of S&P 500 futures prices probably supplies a more accurate reflection of the true drop in the S&P 500 stocks (which were only recorded as falling 20%) since the index significantly lagged due to extreme instances of illiquidity.

Table I
Standardized Kurtosis and Sample Size

This table reports standardized kurtosis of S&P 500 logarithmic returns with varying sample size over 3600 trading days from February 2, 1980, through April 28, 1995. Kurtosis is standardized to equal 0 for lognormal distributions. For example, the 3600 trading days are broken up into 144 nonoverlapping samples of 25 days each ($25 \times 144 = 3600$). The standardized kurtosis is calculated for each of the 144 samples and -0.01 is the arithmetic average of these 144 numbers.

Sample Size	Number of Samples	Average Kurtosis	Sample Size	Number of Samples	Average Kurtosis
25	144	-0.01	100	36	0.99
30	120	0.10	150	24	1.05
40	90	0.32	200	18	1.21
50	72	0.49	300	12	1.38
60	60	0.59	400	9	1.46
75	48	0.66	600	6	1.70
90	40	0.88	720	5	1.85

Under the lognormal hypothesis, this is a -27 standard deviation event with probability 10^{-160} , which is virtually impossible. Nor is October 1987 a unique refutation of the lognormal hypothesis. Two years later, on October 13, 1989, the S&P 500 index fell about 6 percent, a -5 standard deviation event. Under the maintained hypothesis, this has a probability of 0.00000027 and should occur only once in 14,756 years. In addition to this episodic evidence, it is now well known that since the 1987 crash, Black-Scholes implied volatilities for S&P 500 Index options have consistently exhibited pronounced smile effects—a fact that can perhaps be best explained by extreme departures from lognormality.

The 1987 crash also sensitizes historical sample statistics to sample size. For example, historical measurements of volatility are quite dependent on whether or not October 19, 1987, is a sample point. Using daily closing prices, S&P 500 historical annualized volatility from May 24, 1989 through April 28, 1995 is 12.1 percent, but if the sample is extended to cover the crash, the volatility from June 20, 1983, through April 28, 1995, is 15.8 percent. This places one in the uncomfortable position of deciding how much weight to place on the crash observation. What is virtually certain is that the crash should not be omitted as an outlier.

Apart from the special problems created by the stock market crash, many other difficulties are encountered sampling from an inherently nonstationary time series such as stock market prices. For example, even holding the overall sample period fixed, historical subsamples will exhibit systematic biases in sample statistics. Table I reports sample standardized kurtosis of logarithmic daily S&P 500 returns over the 3600 trading days from February 2, 1980, through April 28, 1995. At very small sample sizes of 25, the average kurtosis of the ($3600/25 =$) 144 nonoverlapping samples during the entire period is almost 0 (consistent with a lognormal distribution). Unfortunately, the sample kurtosis systematically rises as a function of sample size so that at sample size

Table II

Black-Scholes Implied and Historical Time Series Volatilities

In the first column, the table reports the implied Black-Scholes volatility for the at-the-money option on the S&P 500 index. The option with a time-to-expiration between 135 and 225 days is chosen. We use 8 dates from April 2, 1986, through April 2, 1993, and choose the date closest to April 2 for each year. The remaining four columns report the historical volatility for those dates, sampling the prior 28, 91, 364, and 1092 days, respectively.

Date	At-the-Money Implied Volatility	Historical Volatility: Prior Sampling Period			
		28 Days	91 Days	364 Days	1092 Days
04/02/86	0.193	0.137	0.133	0.108	0.118
04/02/87	0.199	0.145	0.143	0.170	0.136
04/04/88	0.262	0.149	0.220	0.348	0.233
04/03/89	0.153	0.116	0.115	0.141	0.238
04/02/90	0.194	0.108	0.136	0.137	0.231
04/02/91	0.169	0.139	0.169	0.167	0.149
04/02/92	0.153	0.073	0.095	0.127	0.145
04/02/93	0.133	0.122	0.108	0.100	0.134

720, the average kurtosis of the $(3600/720 =) 5$ nonoverlapping samples is 1.85, quite different from lognormal. This kind of systematic increase in the average sample kurtosis is, of course, exactly what one would predict from a random volatility model, first postulated over two decades ago by Barr Rosenberg.

These difficulties in dealing with historical time series can have a significant effect on option prices. Table II compares the Black-Scholes at-the-money implied volatilities of S&P 500 6-month index options at selected dates to historically computed volatilities for various prior sampling periods.

This confirms two well-known observations. First, historically measured volatility varies significantly over different time intervals; and second, it can be a poor predictor of subsequent implied volatility. But what may be more surprising is that, except for historical periods containing the 1987 crash, option-implied volatility is almost always biased upward from prior historical realizations, less so for longer sample periods than shorter sample periods. Perhaps this is a combination of flaws in the Black-Scholes formula for measuring market forecasts of volatility from at-the-money options coupled with the market's correction for the fact that extreme rare events are possible, but not usually present, in the most recent historical samples.

One recent response to these difficulties is to postulate a particular statistical time series model of returns of the autoregressive conditional heteroskedasticity/general autoregressive conditional heteroskedasticity (ARCH/GARCH) variety. In another approach, option prices are used to imply parameters of a prespecified (risk-neutral) stochastic process, usually nesting the lognormal as a special case. These parametric approaches may suffer from presupposing a particular functional relation between observable variables and statistical parameters such as volatility, skewness, and kurtosis.

In Section I, we introduce nonparametric methods that can be employed to derive risk-neutral implied probability distributions from option prices. Section II describes the data and choice of parameters, such as interest rate, dividend yield, and index level, which we use for our empirical work. Section III contains a description on how to recover the risk-neutral probability distributions. Empirical results are presented in Section IV, and Section V concludes.

I. Nonparametric Methods

This article pursues an alternative method in which the related risk-neutral probabilities are also recovered from contemporaneous market prices of associated derivatives, but which makes no prespecification of these functional relations. By presupposing less, not only can the nonparametric method reflect the possibly complex logic used by market participants to consider the significance of extreme events, but it also implicitly brings a much larger set of information (than simply the historical time series or a prespecified set of parameters in an option pricing formula) to bear on the formulation of probability distributions.

In a related article,² one of the authors introduces a new method for recovering the risk-neutral probability distribution of an underlying asset price from the contemporaneous prices of its associated options. The recommended approach is to solve the following quadratic program:

$$\min_{P_j} \sum_j (P_j - P'_j)^2 \quad (1)$$

$$\text{subject to: } \sum_j P_j = 1 \quad \text{and} \quad P_j \geq 0 \quad \text{for } j = 0, \dots, n$$

$$S^b \leq S \leq S^a \quad \text{where} \quad S = \left(d^t \sum_j P_j S_j \right) / r^t$$

$$C_i^b \leq C_i \leq C_i^a \quad \text{where} \quad C_i = \left(\sum_j P_j \max [0, S_j - K_i] \right) / r^t$$

$$\text{for } i = 1 \dots, m$$

where j indexes from lowest to highest the nodes at the end of a binomial tree:

P_j = implied (posterior) ending nodal risk-neutral probabilities

P'_j = prespecified (prior) ending nodal lognormal risk-neutral probabilities

S_j = ending nodal underlying (ex-payout) asset prices

² Rubinstein (1994).

$S^b(S^a) \equiv$ current observed bid (ask) underlying asset price

$C_i^b(C_i^a) \equiv$ current observed bid (ask) call option price with striking price

K_i

$d \equiv$ annualized payout return

$r \equiv$ annualized riskless interest return

$t \equiv$ time to expiration

The resulting P_j are the risk-neutral probabilities that the underlying asset price will be S_j on the expiration date of the options. These are, in the least squares sense, the probabilities closest to the prior that result in option and underlying asset values that fall between their respective bid and ask prices. The methodology has the virtue that general arbitrage opportunities do not exist if and only if there is a solution. In addition, if all options have bid/ask prices surrounding their values based on the prior, then $P_j = P'_j$ for all j . Finally, the more complete the set of options across striking prices, other things equal, the less sensitive P_j will be to the particular form of the minimization objective and the choice of the prior. At the extreme, the constraints themselves will completely determine the solution. The method is nonparametric because any probability distribution is a possible solution.

To some extent, the quadratic minimization criterion is arbitrary; and one could well imagine other interesting forms. A purpose of this article is to examine alternative specifications of the minimization criterion using historically observed option prices. A second purpose is to report and analyze the historical record of risk-neutral probability distributions (inferred from the nonparametric approach) for the S&P 500 index from 1986 through 1993.³

II. The Data

The empirical research in this article is based on a database that contains all reported trades and quotes covering S&P 500 European index options traded on the Chicago Board Options Exchange, S&P 500 index futures traded on the Chicago Mercantile Exchange, and intraday S&P 500 index levels from April 2, 1986, through December 31, 1993.⁴ These data are supplemented by the daily S&P 500 ex-dividend record as reported by Standard and Poor's Corporation.

A. Dividends

We assume that the dividend amount and timing expected by the market are identical to the dividends actually paid on the index. The annualized payout return d at a given date for a given option with years-to-expiration t is calculated from

$$d = [1 + (D/S)]^{1/t}$$

³ See Jackwerth and Rubinstein (1995) for more extensive and technical details.

⁴ The starting date was driven by the changeover of S&P 500 index options from American to European. The new European options first were traded on this date.

where D is the simple sum of the actual S&P 500 daily dividend amounts associated with ex-dividend dates between the given date and the given expiration date, and S is the index level.

B. Interest Rates

After experimenting with a variety of possible interest rates including T-bill rates and CD-rates, we settle on using implied interest rates imbedded in the European put-call parity relation.

Armed with option quotes, we calculate separate lending and borrowing interest returns from put-call parity using the following formulas:

$$r_l = [(P^a + Sd^{-t} - C^b)/K]^{-1/t} \quad (2)$$

$$r_b = [(P^b + Sd^{-t} - C^a)/K]^{-1/t} \quad (3)$$

where:

$C^b(C^a) \equiv$ observed call option bid (ask) price

$P^b(P^a) \equiv$ concurrent observed put option bid (ask) price.

We assign, for each expiration date, a single lending and borrowing rate to each day, which is the median of all daily observations across all striking prices.

C. Index Level

To calculate these implied borrowing and lending rates, we first use for S the S&P 500 index itself. However, since traders typically use the index futures market rather than the cash market to hedge their option positions, and since it is well-known that the cash market prices lag futures prices, we could use the corresponding future with the same time-to-expiration in formulas similar to the ones for the implied put-call parity interest rates above. In order to use all futures with different time-to-expiration, we translate the futures prices at a given minute during the day into an implied index level at that time, by solving the following two simultaneous equations for S and taking the median of all S for any minute:

$$S = F(r_f/d)^{-t} \quad \text{and} \quad d = [1 + (D/S)]^{1/t}$$

where r_f is the median implied interest return from all futures observations with years-to-delivery t across all observations for the current day. The implied interest return (r_f) is calculated from:

$$r_f = d(S/F)^{-1/t}$$

where F is the concurrent futures price and t is the time-to-delivery in years. To each day, we then assign, for each delivery date, a single implied "repo"

Table III

Volatilities of Implied Interest Rates Derived from Put-Call Parity

For the following table, we calculate implied put-call parity interest rates based on the S&P 500 index level and based on the index level derived from the futures for the first trading day of each month from May 1986 through December 1993. Since there are options with different time-to-expiration, we group them into five sets according to time-to-expiration 0–44, 45–134, 135–224, 225–314, and 315–405. For each day, we calculate the volatility across interest rates. In this table, we report the median daily volatility of interest rates for each time series.

Basis of the Index	Volatility of Implied Interest Rates in Percent				
	Days to Maturity				
	23 Days	90 Days	180 Days	270 Days	360 Days
Index based	0.0123	0.0044	0.0024	0.0020	0.0016
Future based	0.0090	0.0037	0.0021	0.0018	0.0015

rate, which is the median of all observations from 9:00 A.M. to 3:00 P.M. Central Time.

To test whether the reported index level or the concurrent implied index level is more reliable, we use both to derive minute by minute put-call parity implied interest rates, one for each option maturity, and then measure the volatility of the implied rates over a single day sampled each month. Table III reports the median volatilities of these two types of implied rates.

Since the volatility of the implied interest rates for the future-based index is lower for all times-to-expiration, we use the futures-based index levels for our research throughout. Also, we take the median put-call parity lending and borrowing rates and use the average of these two rates as the single interest rate for the day.

D. Arbitrage Violations

General arbitrage violations must be eliminated from the data for an implied risk-neutral distribution to exist. For that purpose, puts are translated into calls using European put-call parity. For all options, we check that the following condition holds:

$$Sd^{-t} \geq C_i \geq \max[0, Sd^{-t} - K_i r_f^{-t}] \quad (4)$$

First, we create subsets of options of the same time-to-expiration with quotes available during the same minute. Second, we select all “unique” options (with a call (put) but not a put (call) at the same striking price and time-to-expiration) and combine them with puts for which both a put and a call are available. Similarly, we construct the corresponding set for the unique options combined with all calls where both a put and call are available. Third, for the two sets so created, for any two options with neighboring striking prices, we check for vertical spread arbitrage violations. Fourth, for any three options with neighboring striking prices, we check for butterfly spread arbitrage

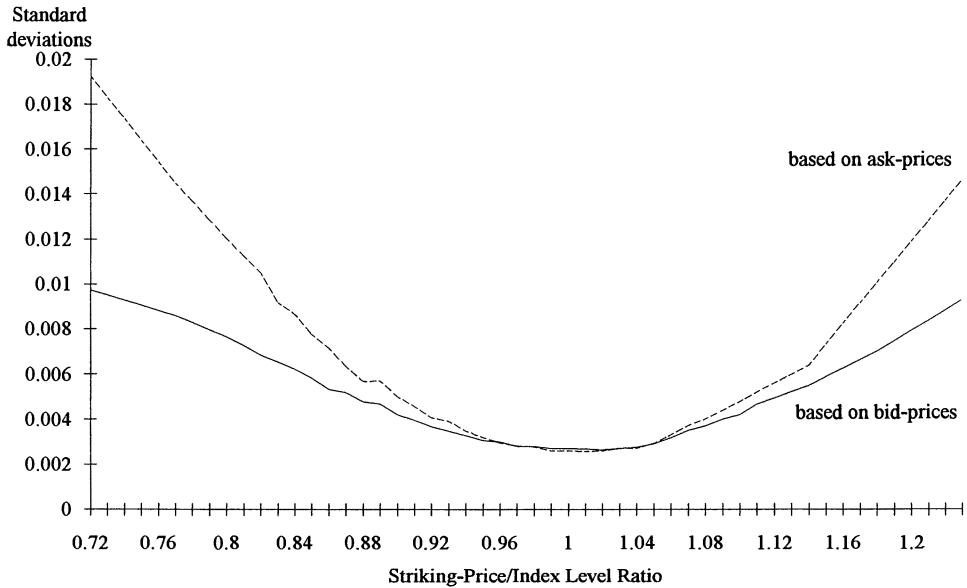


Figure 1. Standard deviations of implied volatilities across striking-price/index level ratios. For all trading days from April 2, 1986 through December 31, 1993, we use options with time-to-expiration from 135 through 225 days and calculate the annualized implied Black-Scholes volatilities for bid and ask quotes, respectively. The figure depicts the median daily standard deviation of implied volatilities grouped by striking-price/index level ratio.

violations. Fifth, all violating options are taken out of their respective subset and then reintroduced in such a way that all the above tests are not violated and that the largest number of options can be reintroduced.

E. Persistent Jaggedness of Volatility Smile

The data show rather persistent irregularities in parts or in the whole smile (implied volatility as a function of the ratio of striking price to index level) over very short periods of time. These irregularities are nonetheless below the transactions cost threshold. The reason seems to be that options are often updated by changes in their prices across the board by a fixed dollar change in the option price.

We try several methods to infer the correct smile from the data. One hopeful approach to deal with inconsistent quotes is to fit a smooth curve to the smile constructed from available options. Unfortunately, the smile can have rather idiosyncratic shapes over even prolonged periods such as a day. To assess the magnitude of this problem, for each striking price/index level ratio, over each day, we calculate the standard deviation of implied volatility. The median standard deviation for each striking price/index level ratio across all the days is shown in Figure 1.

We try to predict the shape of the smile in advance but to no avail, especially since the forecast for the far away options is unreliable. As Figure 1 suggests,

the volatilities for these options are more volatile than for at-the-money options. We experiment with weighing schemes, which place more emphasis on at-the-money options, but again, forecasts do not improve.

Next, we aggregate smiles across short periods of time during which the index level is almost constant in the hope that the underlying parameters determining option prices are sufficiently stationary and smile shape errors in the data will be cancelling. As it turns out, this only partially alleviates the jaggedness of the smile, and it causes problems if one does not observe far-in and out-of-the-money options throughout these intervals. This can be a serious problem, since the implied probability distribution is somewhat sensitive to the set of available striking prices.

We finally settle for the following approach. We use all observations of midpoint implied volatilities (average of implied volatilities based on the option bid and ask, and if there are both a put and a call available, we average across all four implied volatilities) throughout the day and calculate the median implied volatility for each striking price and expiration. We further exclude observations with Kr^{-t}/Sd^{-t} ratios less than 0.79 and more than 1.16, since the variability of the implied volatilities is large outside this range. We use an asymmetric cut-off for the strike-to-index-ratios since the low ratios seem to contain more information about the shape of the implied probability distribution than the high ratios. The options within the striking price range are then used to form a single “daily smile” for each option expiration date.

Recall that our reason for aggregating the data in this way is to eliminate noise. However, underlying this type of aggregation is the assumption that the “true” smile (implied volatility as a function of the ratio of striking price to index level) remains fixed during the day. Unfortunately, since the shape of the smile obviously changes from day to day, it remains possible that it could change nonnegligibly even during the day. To this end, during each day we calculate, for minutes with options quoted on at least 8 different striking prices, the median absolute and relative deviations for each minute’s midpoint volatility smile from the daily smile, and apply a nonparametric runs test to these time-series. Moreover, days of large index moves are also excluded if the difference between the high and low index levels for the days, scaled by $S(r/d)^{-t}$ is greater than 0.02. Using these criteria, for options maturing between 135–225 days in the future, 1172 days remain out of a total of 1850 days over the entire sample period.

III. Recovering Risk-Neutral Probability Distributions

If options existed with striking prices infinitely dense on the positive real line, given that there were no general arbitrage violations among these options, a unique implied risk-neutral probability distribution would exist and could be calculated using the techniques described by Breeden and Litzenberger

(1978).⁵ Unfortunately, since observed option prices are only available at discretely spaced striking price levels, the lowest available striking price is well above 0 and the highest is well below infinity, there are many risk-neutral distributions that can fit their market prices. There are several techniques to select among these distributions, which essentially amount to interpolating between observed striking prices and extrapolating outside their range.⁶

A. Objective Functions

As described above, a related article proposes an “optimization method” to recover risk-neutral probabilities from option prices. The method it proposes is minimizing the sum of the squared differences of prior from posterior probabilities. For this article, we examine other functions, in addition to the quadratic. One possibility is the “goodness-of-fit function”:

$$\sum_j (P_j - P'_j)^2 / P'_j \quad (5)$$

Unfortunately, the goodness of fit function suffers from numerical difficulties, since the division by P'_j causes the terms in the lower left tail of the objective function to become unacceptably large.

Another possibility is the “absolute difference function”:

$$\sum_j |P_j - P'_j| \quad (6)$$

This function creates posterior (or implied) probabilities that stick to the lognormal prior almost perfectly, but at some points leave the prior dramatically resulting in a posterior distribution that has very little smoothness and intuitive appeal.

The “maximum entropy function”:

$$-\sum_j P_j \log(P_j / P'_j), \quad (7)$$

while quite appealing from a theoretical standpoint, has unfortunate numerical difficulties, since it tends to assign very high probabilities to index levels close to zero. The optimization routine can be parameterized to deal with this problem but convergence in this case is rather slow.

Despite these difficulties, by carefully executing the optimization and monitoring the convergence, we are able to display implied probability distributions and implied cumulative probability distributions for the same dataset for

⁵ See Breeden and Litzenberger (1978) on how the second derivative ($\partial^2 C / \partial K^2$) can be used to infer the risk-neutral probability distribution.

⁶ See Shimko (1993).

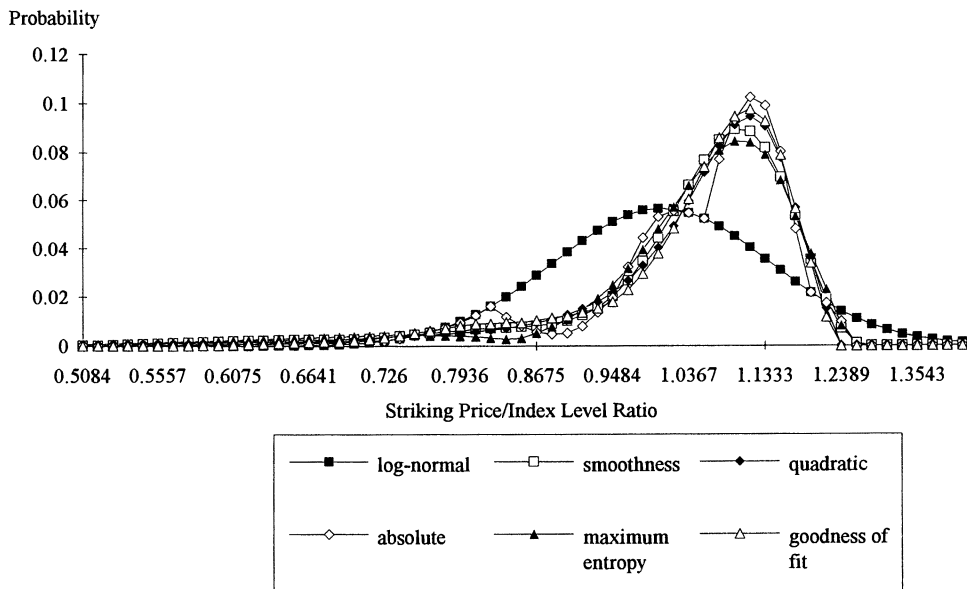


Figure 2. Implied probability distributions for different objective functions. The probability distributions are sampled at 201 discrete asset values. The data used for all distributions are dated March 16, 1990, at 11:47 A.M.

all these functions. All of them exhibit fairly consistent pictures of posterior distributions that additionally seem relatively independent of the assumed shape of the prior (Figs. 2 and 3).

While the different implied probability distributions do exhibit some differences in shape, a closer look at the cumulative probability distributions shows that all objective functions assign about the same cumulative probabilities to values near-the-money. The differences are mainly caused by the respective behavior of each objective function for stock prices that correspond to deep out-of-the-money puts. However, since those options do not trade or are omitted from our data, the objective functions can assign the associated probabilities differently. If we start with a uniform, rather than lognormal prior, convergence takes longer. But eventually the resulting posterior is very close to the posterior that results from the lognormal prior. The implied distributions are apparently rather independent of the choice of the objective function when a sufficiently high number of options is available; indeed, as few as 8 option prices seem to contain enough information in order to determine the general shape of the implied distributions.

Nonetheless, all these methods require an assumed prior distribution and occasionally lead to posterior distributions that have sufficiently little smoothness to be plausible. This suggests that an interesting approach might be to select the implied distribution with the maximum smoothness. To accomplish

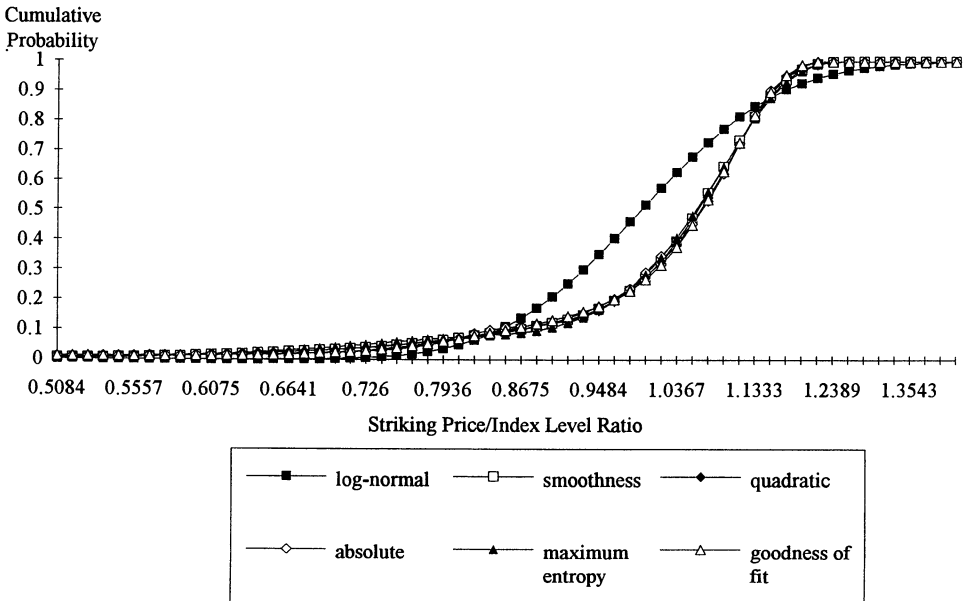


Figure 3. Implied cumulative probability distributions for different objective functions. The probability distributions are sampled at 201 discrete asset values. The data used for all distributions are dated March 16, 1990, at 11:47 A.M.

this, we select the implied probabilities P_j , which minimize the following function:⁷

$$\sum_j (P_{j-1} - 2P_j + P_{j+1})^2 \quad \text{where} \quad P_{-1} = P_{n+1} = 0 \quad (8)$$

Note that this objective does not require a prior. Its sole purpose is to find the smoothest distribution in the sense of minimizing the second derivative of P_j with respect to the underlying asset level, thereby minimizing the curvature exhibited in the implied probability distribution. Each term corresponds to the value of a butterfly option spread, which is the finite difference approximation of the second derivative $\partial^2 P_j / \partial S_j^2$ since if the S_j are equally spaced:⁸

$$\frac{(P_{j+1} - P_j)/(S_{j+1} - S_j) - (P_j - P_{j-1})/(S_j - S_{j-1}))}{1/2(S_{j+1} + S_j) - 1/2(S_j + S_{j-1}))} = \text{const} \cdot (P_{j-1} - 2P_j + P_{j+1}) \quad (9)$$

⁷ See Adams and Van Deventer (1994) for an application of the smoothness criterion to yield and forward rate curves.

⁸ The minimization of $\int_0^\infty (\partial^2 P_j / \partial S_j^2)^2 dj$ for a continuous probability distribution corresponds to fitting a cubic spline.

While we can omit the constant term, we need to square each individual contribution to the curvature, since the sum would otherwise be degenerate and, in any event, we are interested in a measure of absolute curvature.

B. Bid-Ask versus Midpoint Constraints

The optimization method proposed in Rubinstein (1994) chooses posterior probabilities so that the resulting option values all lie between their respective bid-ask quote. Unfortunately, all the objective functions discussed above give rise to option values with an implied volatility smile that is more convex than the smile based on either the bid or ask prices, approaching the bid implied volatilities for options that are at-the-money, and ask implied volatilities for in- and out-of-the-money options.

Forcing option values based on posterior probabilities to equal their values based on midpoint bid-ask quote implied volatilities corrects for this problem, but sacrifices the additional information content coming from the individual bid and ask prices. Moreover, the midpoints derived from the median daily smile are not necessarily arbitrage-free any longer, since now the single values are specified as opposed to a band between bid and ask prices. In addition, overfitting the data by following all the small wiggles in the midpoint smile also becomes a problem with this approach. Fortunately, it seems that these difficulties can be overcome by allowing for a small error between the option values based on posterior probabilities and the option values based on the midpoint bid-ask quote implied volatilities. This can be achieved by using an optimization routine that penalizes deviations from the midpoint values, but does not force the resulting posterior distribution to conform exactly to those values.

C. Optimization Method

Our problem is a nonlinear optimization with both linear and nonlinear equality and inequality constraints. For the earlier article, we used the general nonlinear optimization routine from the NAG FORTRAN library; however, it fails to exploit our private knowledge of the particular problem.

We can take advantage of the fact that some of the objective functions are quadratic and the constraints on the option prices are piecewise linear. The most powerful methods for dealing with such problems are quasi-Newton methods. After some experimentation, we use the quasi-Newton method of the Broyden-Fletcher-Goldfarb-Shanno type as an optimization method for our problem.⁹ The constraints of the problem are incorporated by using a penalty method. The idea is that one penalizes a violation of a constraint by measuring the squared violation and multiplies this with a large penalty term. Those penalized violations are then added to the original objective function. The choice of the penalty parameter is somewhat of an art, since it involves a decision with regard to the tradeoff between increased accuracy in achieving

⁹ See Luenberger (1984).

the observed option prices and avoidance of overfitting the observed prices, which leads to wiggles in the resulting probability distributions. Typical values for the penalty parameter are 10^1 for the smoothness (10^{-2} for the closed form smoothness below), 10^2 for the quadratic, 10^4 for the absolute, 10^9 for the maximum entropy, and 10^{50} for the goodness of fit objective function. The augmented objective function is thus:

$$f(x) + \alpha\{\max[0, g(x)]^2 + (h(x))^2\} \quad (10)$$

where:

$f(x) \equiv$ objective function

$g(x) \equiv$ function of violations of inequality constraints

$h(x) \equiv$ function of violations of equality constraints

$\alpha \equiv$ penalty parameter

For example, with the smoothness objective function and midpoint equality constraints on underlying asset and option prices, the augmented objective function is:

$$\begin{aligned} \min_{P_j} \sum_j (P_{j-1} - 2P_j + P_{j+1})^2 + \alpha \left\{ \sum_j [\max[0, -P_j]]^2 + \left[\sum_j P_j - 1 \right]^2 \right. \\ \left. + \left[\left(d^t \left(\sum_j P_j S_j \right) / r^t \right) - S \right]^2 + \sum_i \left[\left(\left(\sum_j P_j \max[0, S_j - K_i] \right) / r^t \right) - C_i^m \right]^2 \right\} \quad (11) \end{aligned}$$

D. Smoothness Criterion

For the smoothness objective function (which we like best), the optimization procedure can be considerably simplified, since an almost closed form solution to the optimization problem can be derived. This solution is faster by a factor of 100 to 1000 over the quasi-Newton methods.

For the following derivations, we use a little trick. Rather than choosing the ending nodal underlying asset values to be log-spaced as in the standard binomial tree, we sample S_j in 201 \$5 increments starting at zero so that they coincide with all potential striking prices and deflate these values by $S(d/r)^{-t}$. This also avoids the problem of not separating two neighboring striking prices by at least one S_j , which can happen with log-spacing. With the revised spacing, the smoothness criterion can be rewritten as follows:

$$\min_{P_j} \Omega = \sum_j (\partial^2 P_j / \partial K_j^2)^2 \quad (12)$$

where (the Breeden-Litzenberger (1978) result):

$$P_j = r^t (\partial^2 C_j / \partial K_j^2)$$

Combining these, the smoothness criterion can be rewritten in terms of C_j (where we omit the constant term r^{2t}) as:

$$\min_{C_j} \Omega = \sum_j (\partial^4 C_j / \partial K_j^4)^2$$

$$\begin{aligned} \text{subject to: } C_j &= C_i^m && \text{whenever } K_j = K_i \\ &&& \text{for } j = 0, \dots, n \quad \text{and } i = 1, \dots, m \end{aligned} \quad (13)$$

where $C_j(C_i^m)$ is the option value (quote mid-point price) at striking price $K_j(K_i)$. Since we sample K_j only at discrete intervals, we rewrite the differential in the objective function as a finite difference approximation:

$$\min_{C_j} \Omega = \sum_j (C_{j-2} - 4C_{j-1} + 6C_j - 4C_{j+1} + C_{j+2})^2$$

$$\begin{aligned} \text{subject to: } C_j &= C_i^m && \text{whenever } K_j = K_i \\ &&& \text{for } j = 0, \dots, n \quad \text{and } i = 1, \dots, m \end{aligned} \quad (14)$$

This implies that we force the option values equal to their market prices whenever their striking prices coincide with available options. Option values with striking prices that do not coincide with available options are determined freely by the optimization.

Given the solution for C_j , we can easily determine the implied probabilities from the butterfly condition:

$$P_j = r^t (C_{j-1} - 2C_j + C_{j+1}) (S(d/r)^{-t}) / 500 \quad (15)$$

Thus, solving directly for the C_j circumvents using the probabilistic constraints on option values (i.e. $\sum_i [(\sum_j P_j \max[0, S_j - K_i]) / r^t] - C_i^m]^2$). Omitting the nonnegativity constraint on the posterior probabilities ($\sum_j (\max[0, -P_j])^2$) for a moment, we can write the first order conditions of the Lagrangian for the rewritten smoothness criterion as a finite difference approximation:

$$\begin{aligned} \partial \Omega / \partial C_j &= 2C_{j-4} - 16C_{j-3} + 56C_{j-2} - 112C_{j-1} + 140C_j - 112C_{j+1} \\ &+ 56C_{j+2} - 16C_{j+3} + 2C_{j+4} = 0 \quad \text{for } j = 0, \dots, n \end{aligned} \quad (16)$$

In order to deal with the constraints that the known quote midpoint prices C_i^m from the data have to be matched, we incorporate a penalty approach where the penalized squared deviations are added to Ω :

$$\Omega' = \Omega + \alpha \sum_i (C_i - C_i^m)^2 \quad \text{for } i = 1, \dots, m \quad (17)$$

The first order conditions for the C_j that do not correspond to a known option price remain the same, and the conditions for the C_j for which the option price is known now incorporate the penalty parameter:

$$\begin{aligned} \partial \Omega' / \partial C_j = & 2C_{j-4} - 16C_{j-3} + 56C_{j-2} - 112C_{j-1} + (140 + 2\alpha)C_j \\ & - 112C_{j+1} + 56C_{j+2} - 16C_{j+3} + 2C_{j+4} = 2\alpha C_i^m \end{aligned} \quad (18)$$

For the solution of this system of equations, we need to know boundary values. Thus, we set

$$C_{n+1} = C_{n+2} = C_{n+3} = C_{n+4} = 0 \quad \text{and} \quad C_i = r^{-t} \{1 - i(500/[S(d/r)^{-t}])\}$$

for $i = -1, -2, -3, -4$

This corresponds to setting the option values with very high striking prices to zero and those with very low striking prices to $Sd^{-t} - Kr^{-t}$. In the probability space, this amounts to setting the probabilities of the index at extreme levels to zero. As a free bonus, it turns out that since the first derivative $\partial C_j / \partial S_j$ corresponds to the cumulative probability distribution, that distribution is necessarily going from -1 to 0 as the striking price increases. Thus, the requirement that the probabilities sum to one is automatically satisfied.

One problem remains. We need to add the constraint that the probabilities (second derivatives of C_j) are nonnegative, which we temporarily ignored. Fortunately, there is a quick iterative procedure to deal with this problem. Since we believe it is sensible to assume that the prices of far-away options should imply zero probabilities in the extremes, we can "clamp down" on the solution set by requiring more and more options to take on the value zero for high striking prices or the value $Sd^{-t} - Kr^{-t}$ for low striking prices. Computationally, we do not even have to calculate all the preset values of the C_j on the far sides except the two closest to the striking price range of the known option prices. This greatly speeds up computation, since the reduced set of option values can be recalculated more quickly by solving the first order conditions only. This reduces the problem to solving a system of linear equations (by, e.g., Cholesky decomposition) for which iterative procedures are clearly not needed. Even more efficient routines can be employed, since the coefficient matrix is band diagonal.

This procedure is then rerun and we recalculate the probabilities. After clamping down more and more, eventually we hope to find a reduced set of option values which translates into nonnegative probabilities only. If such a set is not reached, then the known option prices (C_i^m) are very likely to violate general arbitrage restrictions. We continue to clamp down by forcing one more extreme probability at a time to zero, until further clamping reintroduces negative probabilities.

However, even if a set of nonnegative probabilities can be found, the resulting density function may exhibit multimodalities. We reject candidate distributions with extreme multimodalities. Among all the candidate distributions,

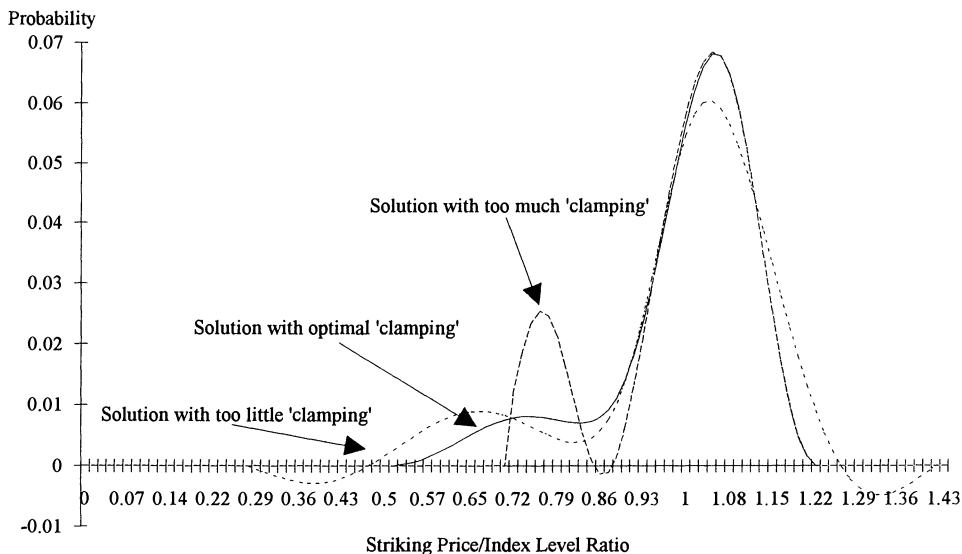


Figure 4. “Clamping down” on the solution set for the closed-form smoothness. The probability distributions are sampled at 201 discrete asset values. The data used for all distributions are dated March 16, 1990. Little (much) clamping means that few (many) probabilities at the far ends are preset to zero. Both result in negative probabilities in the middle. Optimal clamping means that there is the least number of probabilities at the far ends preset to zero so that all probabilities in the middle are nonnegative, and we do not have excessive bimodalities.

we choose the one with the largest number of nonzero probabilities if the largest “dip” between identical levels of probabilities does not span more than five index spacings. If all the “dips” are longer than 15 index spacings, we discard the day. For “dips” between 5 and 15 index spacings, we calculate the ratio of the length of the “dip” over the number of nonnegative probabilities, up to the mode of the distribution. We then select the distribution with the lowest ratio. The following diagram shows different stages of the “clamping down” procedure (Fig. 4).

IV. Empirical Results

A. Shapes of Implied Probability Distributions across Time

In order to inspect the shapes of implied probability distributions as time passes, we graph the distributions for options with 179 days-to-expiration as we observe them in June of each year. For purposes of comparison, we standardize the ending index level S_j by replacing it with:

$$S'_j = ((\log(S_j/Sd^{-t})) - \mu t)/\sigma \sqrt{t} \quad (19)$$

where

$$\mu \equiv \text{annualized risk-neutral mean of the logarithm of } S_j/Sd^{-t}$$

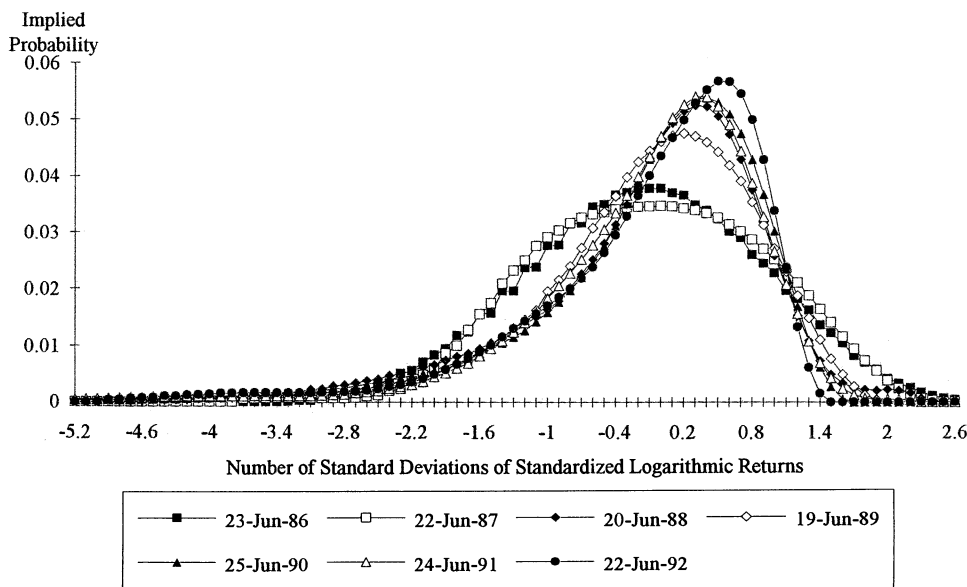


Figure 5. Implied probability distributions. The probability distributions are sampled at 201 discrete asset values. We standardize the asset values by demeaning them and dividing by $\sigma\sqrt{t}$. We graph 7 distributions with time-to-expiration of 179 days in June for each of the years from 1986 through 1992.

$$\sigma \equiv \text{annualized volatility of the logarithm of } S_j/Sd^{-t}$$

We can identify a distinct change in shape between the precrash and the postcrash distributions. While the precrash distributions resemble the lognormal distribution, the postcrash distributions exhibit leptokurtosis and left-skewness (Fig. 5).

B. Time Patterns

We analyze the time patterns of implied probability distributions by calculating the skewness and kurtosis for each day. We calculate skewness and kurtosis as:

$$\text{skewness} = \text{sign}\left(\sum_j P_j S_j'^3\right) \left| \left(\sum_j P_j S_j'^3\right) - 0 \right|^{1/3}$$

$$\text{kurtosis} = \text{sign}\left(\sum_j P_j S_j'^4\right) \left| \left(\sum_j P_j S_j'^4\right) - 3 \right|^{1/4}$$

We find that the implied probability distributions in the pre-crash period are somewhat left-skewed and platykurtic (Fig. 6). That is, the mean of the distribution tended to be to the right of the mode, and the mode tended to be

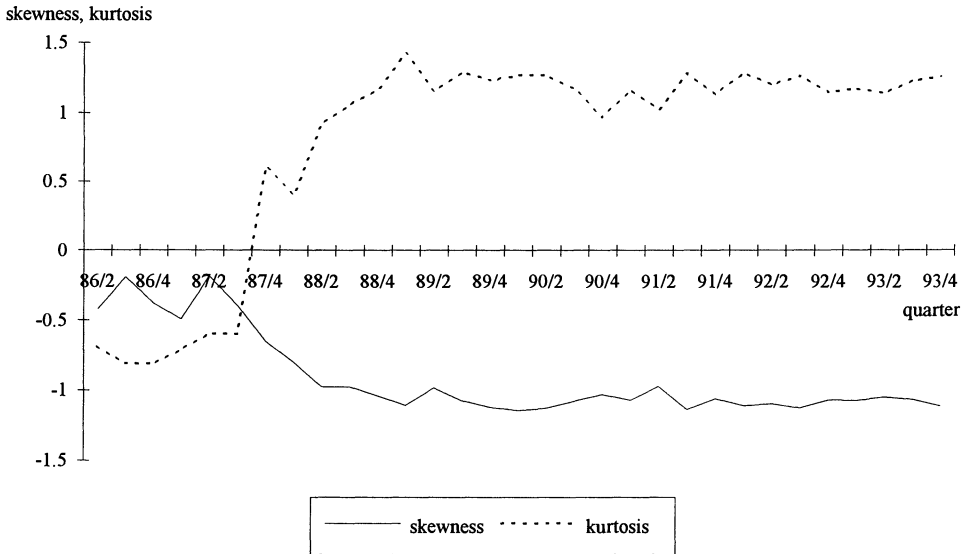


Figure 6. Skewness and kurtosis of implied probability distributions. For all dates from April 2, 1986, through December 31, 1993, we calculate the skewness and normalized kurtosis for the implied probability distributions based on options with time-to-expiration from 135 through 225 days. We graph the median daily skewness and kurtosis for each quarter.

less pronounced than the mode of the corresponding lognormal distribution. After the crash in the fourth quarter of 1987, we find a period of adjustment where the distributions become more left-skewed and change from platykurtic to leptokurtic. This adjustment is completed by mid-1988. Thereafter, we observe very consistent levels for both skewness and kurtosis. The distributions are significantly more left-skewed than in the precrash period, and the mode is persistently more pronounced than the mode of the corresponding lognormal distribution.

C. Patterns of Cumulative Probabilities

We look into the patterns of the implied probability distributions after we standardize them by replacing S_j with $S'_j = ((\log(S_j/S_d^{-t})) - \mu t)/\sigma \sqrt{t}$. This causes the lognormal distribution to be transformed into a $N(0, 1)$ normal distribution. Table IV shows cumulative probabilities for different periods where we aggregate the different times-to-expiration by taking the mean values. We find that the implied cumulative probabilities of the precrash period follow the lognormal distribution rather closely. In the precrash period, a decline in index level by 3 (−36 percent over one year¹⁰) or more standard deviations had about the same probability under the implied and the lognor-

¹⁰ We use the mean interest rate of 6.63 percent and the mean Black-Scholes at-the-money volatility of 16.95 percent for options with time-to-expiration of 135–225 days, both annualized and calculated across 89/01/03–93/12/31.

Table IV
Cumulative Probability Distributions

This table reports cumulative implied probability distributions for different periods and for the lognormal distribution. For all the available implied probability distributions based on options with time-to-expiration from 135 through 225 days, we normalize the distributions by demeaning them and dividing by $\sigma\sqrt{t}$ so that the standard deviation is set to 1. We then calculate the cumulative probability as

$$\text{cumulative probability} = \sum_j \text{Prob}\{((\log(S_j/S_d^{-t})) - \mu t)/\sigma\sqrt{t}) < X\}$$

where X = number of standard deviations of standardized logarithmic returns

		Implied Probability Distribution	Implied Probability Distribution	Implied Probability Distribution	Implied Probability Distribution
X	Log-Normal Distribution Across All Periods	04/02/86– 10/19/87	10/20/87– 12/30/88	01/03/89– 12/31/90	01/02/91– 12/31/92
–6	0.000000001	0.00000015	0.00017	0.00020	0.00016
–5	0.00000029	0.0000052	0.00054	0.00077	0.00055
–4	0.000032	0.00020	0.0023	0.0034	0.0027
–3	0.0013	0.0013	0.0095	0.014	0.014
–2	0.023	0.022	0.040	0.047	0.049
–1	0.16	0.17	0.14	0.14	0.14
0	0.50	0.49	0.42	0.41	0.41

mal distribution, but a decline by 4 (–46 percent over a year) or more standard deviations was 10 times more likely under the implied. In the postcrash period, a decline in index level by 3 (4) or more standard deviations increased to being 10 (100) times more likely under the implied than the lognormal distribution. All postcrash cumulative probabilities associate a significantly higher probability with a large drop in stock price. In addition, the cumulative probabilities in the postcrash period differ very little from each other. In comparing the implied probabilities from pre- and postcrash periods, we find that the probability of a 3 (or a 4) or more standard deviation decline was 10 times more likely after the crash than before.

V. Conclusion

Maximizing the smoothness of the resulting probability distribution seems to be a well-suited objective for nonparametric methods of recovering risk-neutral probabilities from option prices. An examination of S&P 500 Index option prices over an eight-year period shows that although implied levels of skewness and kurtosis exhibit a discontinuity across the divide of the 1987 market crash, they remain remarkably stable on either side of the divide. Moreover, the probability of another significant decline in the S&P 500 index, as implied by the recovered distribution, is far more likely now than prior to the crash.

It would seem that those investors who are particularly concerned with the probability of extreme stock market events (which would include managers of large derivatives portfolios as well as buyers and sellers of out-of-the-money options) would be advised to use methods of estimating probability distributions such as those examined here, which leave considerable flexibility in the shape of the lower left-hand tail.

It is well to keep in mind the limitations of our analysis. Essentially, the various optimization criteria one might use to recover the risk-neutral probabilities can be viewed as different ways of interpolation and extrapolation to make up for the missing options. Indeed, with an infinitely dense set of options across striking prices spanning zero to infinity, the posterior probabilities P_j will be the "correct" risk-neutral probabilities if:

- (1) prices are correctly observed and synchronous,
- (2) the market is informationally efficient,¹¹ and
- (3) the market is perfect (no trading costs).

As a result, the method is very robust. In particular, no assumptions (other than nonsatiation) are required of investors—they can be risk preferring and even irrational, and no assumptions are required of stochastic processes. We say the method is nonparametric because any probability distribution is a possible solution.

The key assumption that bothers us is the absence of trading costs. It is possible that these costs could explain much of the pronounced smile effects now observed in S&P 500 index option prices. In this article, we are essentially attributing all of the smile to shifts in probability beliefs (or investor risk aversion) since the 1987 market crash. However, those stalwarts who would place the blame on trading costs, need to explain why—given the extreme shift in the option smile—these costs were apparently of much less importance before the crash than after.

REFERENCES

- Adams, Kenneth J., and Donald Van Deventer, 1994, Fitting yield curves and forward rate curves with maximum smoothness, *Journal of Fixed Income* 4, 52–62.
- Breeden, Douglas, and Robert Litzenberger, 1978, Prices of state-contingent claims implicit in options prices, *Journal of Business* 51, 621–651.
- Jackwerth, Jens Carsten, and Mark Rubinstein, 1995, Recovering probability distributions from contemporaneous security prices. Working paper, University of California at Berkeley.
- Luenberger, David G., 1984, *Linear and Nonlinear Programming*, 2nd Ed. (Addison-Wesley).
- Rubinstein, Mark, 1994, Implied binomial trees, *Journal of Finance* 49, 771–818.
- Shimko, David, 1993, Bounds of probability, *RISK* 6, No. 4, 33–37.

¹¹ This is taken to include the presumption that investors act to eliminate riskless arbitrage opportunities.