Recovering Probability Distributions from Contemporaneous Security Prices

by

Jens Carsten Jackwerth and Mark Rubinstein*

October 16, 1995 (revised)

Abstract

Probability distributions of stock market returns have typically been estimated from historical time series. The possibility of extreme events such as the stock market crash of 1987 makes this a perilous enterprise. Alternative parametric and nonparametric approaches use contemporaneously observed option prices to recover their underlying risk-neutral probability distribution. Parametric methods assume an option pricing formula which is inverted to obtain parameters of the distribution. The nonparametric methods pursued here choose probabilities to minimize an objective function subject to requiring that the chosen probabilities are consistent with observed option and underlying asset prices. This paper examines alternative specifications of the minimization criterion using historically observed S&P 500 index option prices over an eight-year period. With the exception of the lower left-hand tail of the distribution, alternative optimization specifications typically produce approximately the same implied distributions. Most prominently, the paper introduces a new optimization technique for estimating expiration-date risk-neutral probability distributions based on maximizing the smoothness of the resulting probability distribution. Since an "almost closed-form" solution for this case is available, the smoothness method is computationally orders of magnitude faster than the alternatives. Considerable care is taken to specify such parameters as interest rates, dividends, and synchronous index levels, as well as to filter for general arbitrage violations and to use time aggregation to correct for unrealistic persistent jaggedness of implied volatility smiles. While time patterns 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, since the crash, the risk-neutral probability of a three (four) standard deviation decline in the S&P index (about -36% (-46%) over a year) is about 10 (100) times more likely than under the assumption of lognormality, and about 10 (10) times more likely than apparent in the implied distribution prior to the crash.

---

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

Recovering Probability Distributions from Contemporaneous Security Prices

Recently, the increasing popularity of derivatives and some highly publicized failures to control risk have led to increased effort 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 their historical realization). On October 19, 1987, the two month S&P 500 futures price fell 29%.[1] Under the lognormal hypothesis, this is a -27 standard deviation event with probability 10^{-160}. Even if one were to have lived through the entire 20 billion year life of the universe and experienced this 20 billion times (20 billion big bangs), that such a decline could have happened even once in this period is a virtual impossibility. 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%, a -5 standard deviation event. Under the maintained hypothesis, this has a probability of .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 volatilties 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%; but if the sample is extended to cover the crash, the volatility from June 20, 1983 through April 28, 1995 is 15.8%. 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 1 reports sample standardized kurtosis of

[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.

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 non-overlapping 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 720, the average kurtosis of the 3600/720 = 5 non-overlapping samples is 1.85, quite different from lognormal.

Table 1
Standardized Kurtosis and Sample Size
(S&P 500 logarithmic returns over 3600 trading days)
period: 80/02/02 - 95/04/28

sample size	number of samples	average kurtosis	sample size	number of samples	average kurtosis
25	144	-.01	100	36	.99
30	120	.10	150	24	1.05
40	90	.32	200	18	1.21
50	72	.49	300	12	1.38
60	60	.59	400	9	1.46
75	48	.66	600	6	1.70
90	40	.88	720	5	1.85

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 2 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.

Table 2
Black-Scholes Implied and Historical Time Series Volatilities
(for S&P 500 index options)
period: 86/04/02 - 93/04/02

date	at-the-money implied volatility	----------------- historical volatility ----------------- prior sampling period			
		28 days	91 days	364 days	1092 days
86/04/02	.193	.137	.133	.108	.118
87/04/02	.199	.145	.143	.170	.136
88/04/04	.262	.149	.220	.348	.233
89/04/03	.153	.116	.115	.141	.238
90/04/02	.194	.108	.136	.137	.231
91/04/02	.169	.139	.169	.167	.149
92/04/02	.153	.073	.095	.127	.145
93/04/02	.133	.122	.108	.100	.134

This confirms two well-known observations. First, historically measured volatility varies significantly over different time intervals; and second, this 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 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.

I. Nonparametric Methods

This paper 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 paper, 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 was to solve the following quadratic program:

$$\min_{P_j} \Sigma_j (P_j - P_j')^2 \text{ subject to:}$$

$$\Sigma_j P_j = 1 \text{ and } P_j' \geq 0 \text{ for } j = 0, \ldots, n$$

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

$$C_i^b \leq C_i \leq C_i^a \text{ where } C_i = (\Sigma_j P_j \max[0, S_j - K_i])/r^t \text{ for } i = 1, \ldots, m$$

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

$P_j \equiv$ implied (posterior) ending nodal risk-neutral probabilities

$P_j' \equiv$ prespecified (prior) ending nodal lognormal risk-neutral probabilities

$S_j \equiv$ ending nodal underlying (ex-payout) asset prices

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 which result in option and underlying asset values which 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 paper 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.

II. The Data

The empirical research in this paper is based on a database which 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.[2] Although our records were time-stamped to the second, we were worried that the clocks at the CBOE and the CME might not be synchronous. Fortunately, any indicated differences were negligible. This data was supplemented by the daily S&P 500 ex-dividend record as reported by Standard and Poor's Corporation.[3] This raw data was then used to create a subset of minute-by-minute records showing only the last transaction and quote for each security for each succeeding minute during the trading day.

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

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

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,[4] and S is the index level.[5]

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

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

---

[2] 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.

[3] Only starting on June 1, 1988 did Standard and Poor's report daily dividends on the S&P 500 index. For the two preceding years, only monthly dividends were available, which for the purpose of this study were apportioned over the days of the month in direct proportion to the daily dividends on the S&P 100 index, as calculated from the actual ex-dividend payments of its constituent stocks, which were kindly supplied by Robert Whaley.

[4] For example, for June expiration options traded on Thursday, April 2, 1992, D is the sum of the ex-dividend amounts starting on the next trading day Friday, April 3, 1992 through and including the expiration date Friday, June 19, 1992.

[5] Two questions arise at this point. First, we do not account for the discreteness and lumpiness of the dividend payments. However, we only use the payout return in order to calculate the ex-dividend stock price $Sd^{-t} = S^2/(S+D)$. Since we are dealing with European options exclusively, we do not have to be concerned about the discreteness and lumpiness other than through the impact of discounting. This relates to the second issue since we did not discount the dividend payments. However, we performed checks on the ex-dividend stock price by comparing our value with the discounted value $S - \Sigma_t D_t r^{-t}$ where r is the T-bill interest rate. The differences were all within -3 to +5 cents and on average and our value was only one cent higher than the discounted value.

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

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

where:

> C^b (C^a) ≡ observed call option bid (ask) price
> P^b (P^a) ≡ concurrent observed put option bid (ask) price

We assigned, for each expiration date, a single lending and borrowing rate to each day which was the median of all daily observations across all striking prices. Occasionally, for a given day and a given expiration date there was no minute during which both a put and a call with the same striking price were quoted. To fill in, the missing interest rates were linearly interpolated across time from the nearby observable implied rates; or, if the missing time was less or more than the expirations of the observed implied rates, the missing interest rate was set equal to either the implied rate of the lowest or highest expiration for which an implied rate was observable.

Index Level. To calculate these implied borrowing and lending rates, we first used 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 (in large part due to lags in reporting transactions of the constituent stocks in the index)[6], 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 that case, we would have to limit ourselves to calculating implied put-call parity interest rates for minutes where we concurrently observe the corresponding future. Unfortunately, we found a large number of minutes where this information was missing due to infrequent trading. This happens especially for options maturing in 135-225 days, a class of expirations which we frequently use for our analyses.

A remedy to this problem is the construction of an index level implied from all concurrent futures prices, independent of their maturity. To translate the futures prices at a given minute during the day into an implied index level at that time, we solved the following two simultaneous equations for S:

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

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) was calculated from:

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

---

[6] In earlier research we looked at a single month during 1986 and calculated the value-weighted average age of the index (excluding the first half-hour of the trading day). This was about 5 minutes. To corroborate this finding with our more recent data, but without having to look at the individual stock transactions, for each day during our sample period we regressed the index separately on each of the first twenty minute lags of the futures price. The single regression with the highest adjusted R^2 was assumed to indicate the lag for a given day. The median age of the index over the 1542 days from 1986 to 1992 was seven minutes.

where F is the concurrent futures price and t is the time-to-delivery in years. To each day, we then assigned, for each delivery date, a single implied "repo" rate which was the median of all observations from 9:00 a.m. to 3:00 p.m. Central Time.

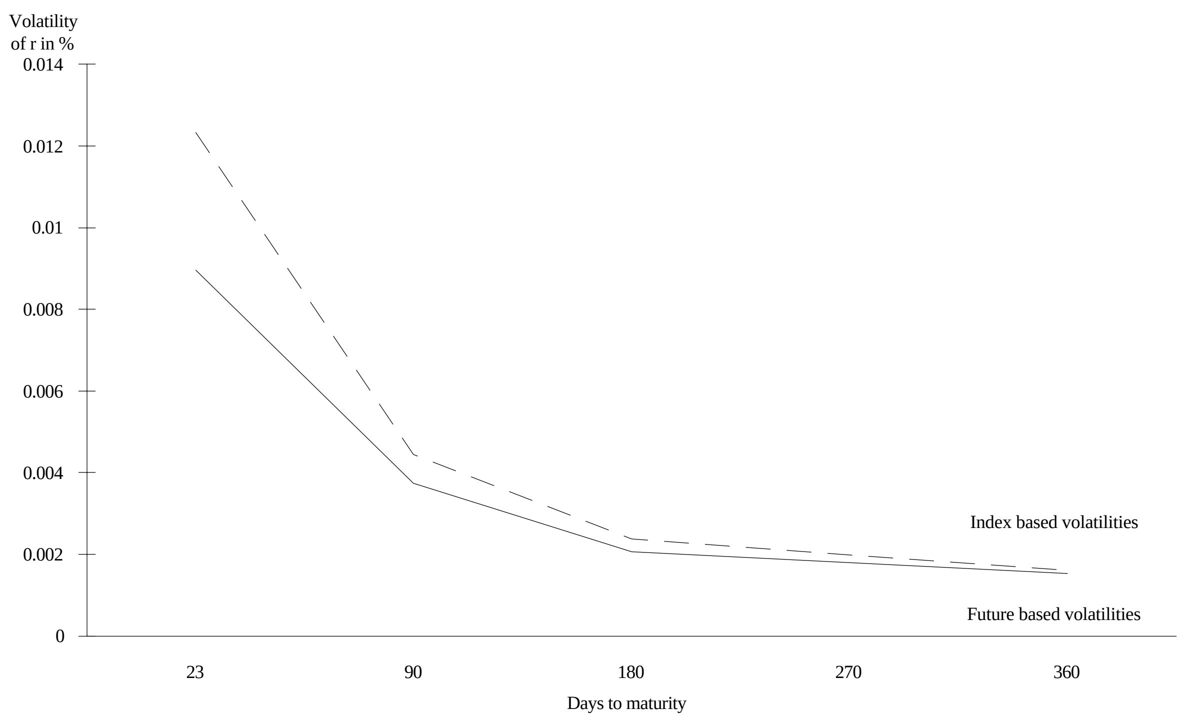
We are now able to solve our two simultaneous equations for S:

$$S = (½Fr_f^{-t}) + [(½Fr_f^{-t})^2 + DFr_f^{-t}]^{1/2}$$

For any particular minute, the implied index level can be derived from all available future bids, asks, or trades for any time-to-expiration. If there are multiple implied index levels available for a minute, we used the mean implied index level.

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

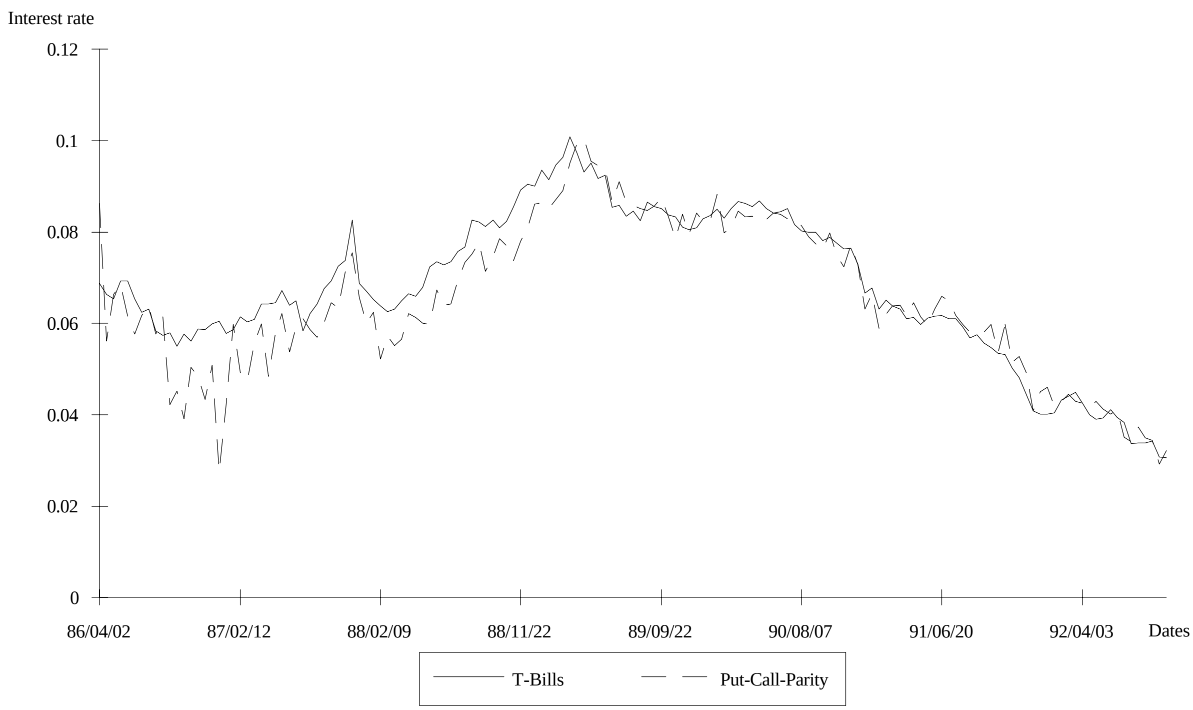
Figure 1
Volatilities of Implied Interest Rates Derived from Put-Call Parity
(daily volatility sampled monthly of the implied interest rates)
period: 86/05 - 93/12



The reduction in volatility of daily implied put-call parity interest rates from using futures-implied index levels is dramatic for short-term options which are theoretically likely to be more sensitive to the exact specification of the index (28% for short-term options with 0-45 days to expiration) but also important for the medium-term options used for this study (11% for options with 135-225 days to expiration). As a result, we use the futures-based index levels for our research throughout, and 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.

<u>T-Bill Interest Rates.</u> We also compared implied put-call parity rates to interest rates on T-bills which are used by a number of researchers. We used options with 135-225 days-to-expiration and matched them with T-bill interest rates which were calculated as the linear interpolation between the two surrounding observable T-bill rates. In figure 2 we show how the two rates behaved over time.

Figure 2
T-bill and Implied Put-Call Parity Interest Rates
(daily interest rates sampled bimonthly)
period: 86/04/02 - 92/09/24



The two interest rates are highly correlated (93.63%), with the put-call parity rate exhibiting only slightly more volatility (212 basis points) than the T-bill rate (197 basis points). On average, the implied put-call parity rates were 11 basis points higher than the T-bill rates. The T-bill rate tends to be somewhat higher for the period 86/10-89/04; but throughout our sample, the T-bill rates lie most of the time between the implied lending and borrowing rates with violations occurring around the 1987 crash. Given the similarity of the two rates and the fact that the put-call parity rate is inferred from actual option prices, we settled for the implied put-call parity interest rate for our empirical work. This decision saves us from having to introduce another security into our empirical analysis. Furthermore, the use of put-call parity implied interest rates insures that the put and call volatility smiles will align which is important for our further work. Using T-bill rates might limit us to using only puts or only calls and thus miss information on the respective out-of-the-money portion of the smile which tends to be thinly quoted.

Finally, there are two empirical results which suggest the use of implied interest rates. The first finding is that estimation errors in the dividend payout ratio are counterbalanced by the use of implied interest rates based on the faulty dividend payout ratio since they affect the implied volatility of an option in opposite directions.[7] The second finding is that when one uses the average of put and call implied volatilities, as we do where possible, then interest rate errors counterbalance again. If the chosen interest rate is too high (low), then the call implied volatility will be calculated too low (high) but the put implied volatility will be calculated too high (low), so that the average implied volatility is not significantly affected. Concluding from all these arguments, we decided to use the implied put-call parity interest rates as opposed to the T-bill rates.

<u>Arbitrage Violations.</u> In the process of setting up the database, we checked for a number of errors which were contained in the original transaction level data. We eliminated obvious errors as well as securities with excessive spreads -- more than 200 cents for options and 20 cents for futures.

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

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

First, we created subsets of options of the same time-to-expiration with quotes available during the same minute. Second, we selected all "unique" options (with a call (put) but not a put (call) at the same striking price and time-to-expiration) and combined them with puts for which both a put and a call were available. Similarly, we constructed the corresponding set for the unique options combined with all calls where both a put and call were available. Third, for the two sets so created, for any two options with neighboring striking prices, we checked for vertical spread arbitrage violations. Fourth, for any three options with neighboring striking prices, we checked for butterfly spread arbitrage violations. Fifth, all violating options were taken out of their respective subset and then reintroduced in such a way that all the above tests were not violated and that the largest number of options could be reintroduced. A violating unique option thus affects both subsets; a violating option with a corresponding put or call only affects its own subset. Finally, the remaining two subsets were merged to form the filtered database.

---

[7] See Steward Mayhew, 1995, "Calculating Implied Parameters from Option Prices", working paper, Haas School of Business, University of California at Berkeley.

Non-Simultaneous Observations. A critical requirement for the methods suggested below for recovering risk-neutral probability distributions is simultaneity of option price observations. It turns out that most of the quotes for S&P 500 index options are generated by the CBOE's Auto-Quote system. These automated quotes occur within the same second for whole sets of puts and/or calls. One of about 60 market makers is assigned to update a deep-in-the-money driver option with a delta of about 1. This `driver' is used to infer the new level of the index (in a new version, the corresponding futures contract is used instead of the index itself). The market maker in charge, with different volatilities for each option and this index level, uses the Black-Scholes formula to set prices for many of the other options. For options with more than two months to expiration, a similar system is in place but is managed by outside securities companies. The quality of the data is somewhat assured by rule 851 which forces all market makers to fill at least 10 contracts for the nearest two month options with prices of less than $20 at the displayed quote. This insures that the market makers will follow the quotes carefully and suggest updates to the parameters for the system. Even for options not covered by rule 851, there seems to be a good business code in place which makes it likely that market makers will fill subsequent orders at the quoted prices.

Thus, rather than using data for a prespecified minute or trying to utilize data from nearby minutes, we only examined minutes with at least 8 quoted calls and 8 quoted puts. This approach saves us from having to find an updating procedure in order to achieve simultaneity in option prices.

Persistent Jaggedness of Volatility Smile. The data shows shifts in parts or the whole smile (implied volatility as a function of the ratio of striking price to index level) over very short periods of time. The reason seems to be that options which are misaligned according to the smile are often updated by changes in their prices across the board by a fixed dollar change in the option price. The adjustment is often 1/8th of a dollar but can be larger for options with high prices or when the index is changing rapidly. Moreover, a down- or up-tick of all quotes can happen even without the index level changing. Thus, the resulting smile will be somewhat jagged and deviations from reasonable pricing (which are nonetheless below the transactions cost threshold) often persist until they are large enough to warrant a specific realignment of particular options.

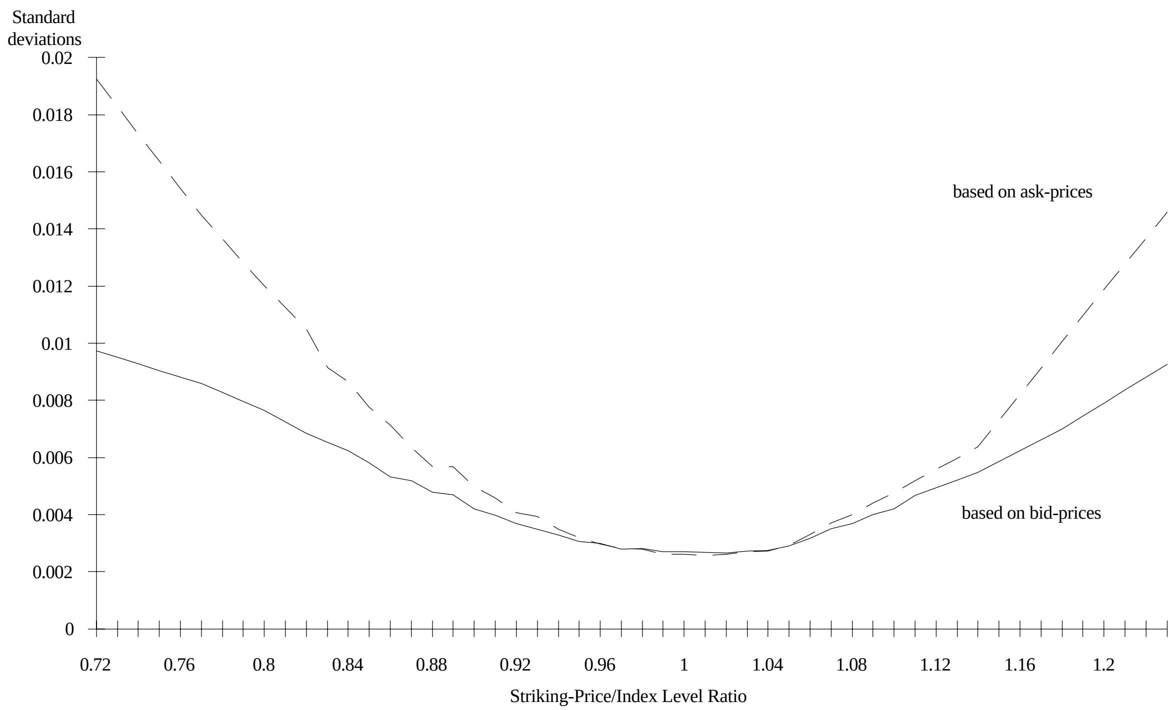
We tried several methods to infer the correct smile from the data. First, we aggregated smiles across short periods of time during which the index level is almost constant in the hope that the underlying parameters determining option prices are sufficiently stationary and smile shape errors in the data 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 do not want to conclude that the implied distribution changes from time to time merely because the set of available striking prices changes. To see how difficult this problem can be, we measured the range of available striking prices over each day. Over our entire sample period, the median minimum option Kr^{-t}/Sd^{-t} ratio for a day is .79 with a standard deviation of about 8%, and the median maximum option Kr^{-t}/Sd^{-t} ratio for a day is

1.11 with a standard deviation of about 9%.

One hopeful approach to deal with missing quotes is to fill in the missing options by fitting a smooth curve to the smile constructed from available options. Unfortunately, the smile can have rather idiosyncratic shapes over even prolonged periods of time such as a day. To assess the magnitude of this problem, for each striking price/index level ratio, over each day, we calculated 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 3.

Figure 3
Standard Deviations of Implied Volatilities across Striking-Price/Index Level Ratios
(median daily standard deviation of annualized implied volatility)
period: 86/04/02 - 93/12/31 time-to-expiration: 135-225 days



We tried 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 3 suggests, the volatilities for these options are more volatile than for at-the-money options. We experimented with weighing schemes which place more emphasis on at-the-money options, but again, forecasts did not improve.

We finally settled for the following approach. We used all observations of mid-point 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 averaged across all four implied volatilities) throughout the day and calculated the median implied volatility for each striking price and expiration. We further excluded observations with Kr^{-t}/Sd^{-t} ratios less than .79 and more than 1.16 since the variability of the implied volatilities was large outside this range. We used 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 were then used to form a single `daily smile' for each option expiration date.

One concern was that the observed option prices might be so low that small absolute price changes might cause rather large relative changes. This concern is caused by the fact that option prices have to be quoted in units of 1/8th of a dollar (1/16th of a dollar for option prices less than 3 dollars). However, the furthest out-of-the-money puts and calls which we considered with time-to-expiration between 135 to 225 days had median prices of $2.51 and $4.04 respectively.

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 non-negligibly even during the day. We tried to detect this by measuring the extent to which the calculated smile during the day changed in a systematic manner which would indicate a fundamental change in the functional form of smile. To this end, during each day we calculated the median absolute and relative deviations for each minute's mid-point volatility smile for minutes with options quoted on least 8 different striking prices from the daily smile and applied a nonparametric runs test to these time-series. The null hypothesis is that the deviations are randomly distributed around their daily medians. We rejected the null hypothesis if the test for both the relative and absolute deviations resulted in critical values of less than 1%. In addition, we were also concerned that our daily aggregation approach would not be reliable for days during which the underlying index moved over a large range. Thus, days were also excluded if the difference between the high and low index levels for the days, scaled by $S(r/d)^{-t}$ was greater than .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 can be calculated using the techniques described by Breeden and Litzenberger.[8] 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.

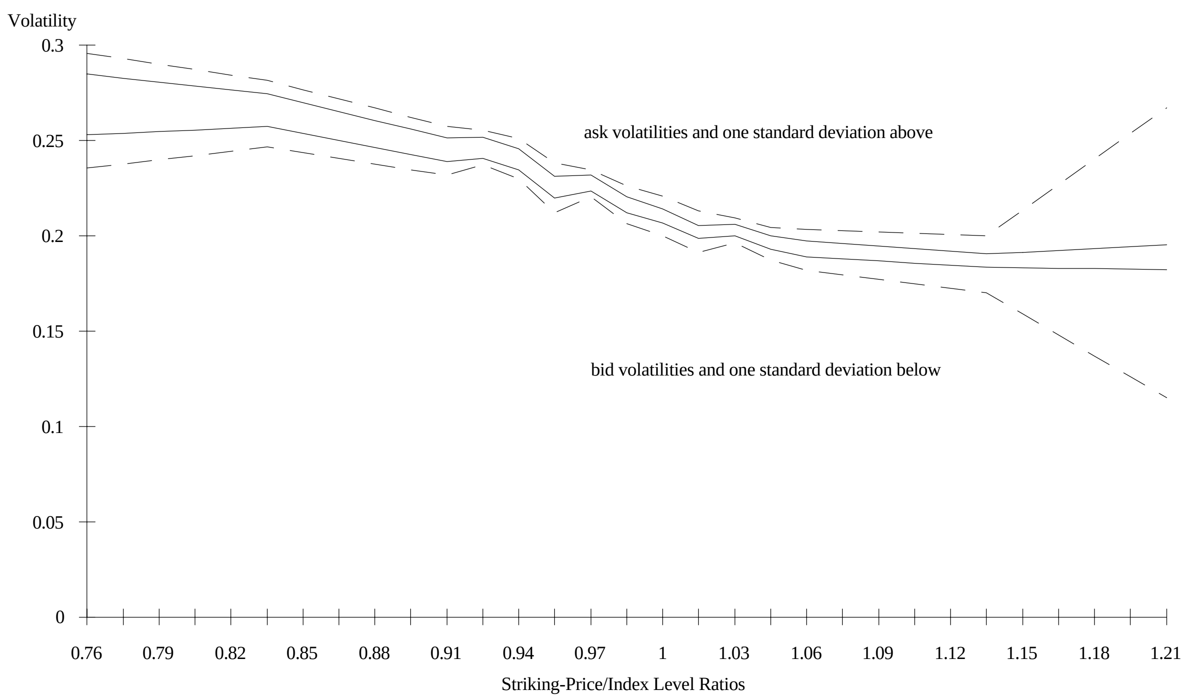
For example, Shimko interpolates by fitting a quadratic function to the Black-Scholes implied volatility smile and then fills in the missing options by using the Black-Scholes formula as a

---

[8] See Douglas Breeden and Robert Litzenberger, 1978, "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business* 51, 621-651.

function of the interpolated volatility.[9] The second derivative ($\partial^2 C/\partial K^2$) can then be used to infer the risk-neutral probability distribution. Shimko extrapolates by grafting lognormal tails to the implied probability distribution. Unfortunately, actual S&P 500 option prices tend not to fit a quadratic curve when there are striking prices far away from the money. The smile can then actually look more like a ~-shaped curve where the low striking price options have lower implied volatilities than a quadratic fit would predict and the high striking price options can have lower or higher implied volatilities than a quadratic fit would predict. Such deviations can persist throughout the day and seem thus to be inherent to the market's perception of risk rather than merely temporary aberrations from the norm. The following diagram shows this behavior for an arbitrary day.

Figure 4
Median Daily Implied Volatility Smile
(bid-ask implied volatilities with + and - one standard deviation range)
date: December 4, 1990



Objective Functions. As described above, a related paper proposed an alternative "optimization method" to recover risk-neutral probabilities from option prices. While other possibilities were proposed, the paper provided an example based on minimizing the sum of the squared differences of posterior from prior probabilities. For this paper, we examined other functions, in addition to the quadratic. One possibility is the "goodness-of-fit function":

$$\Sigma_j (P_j - P_j')^2 / P_j'$$

9 David Shimko, 1993, "Bounds of Probability," *RISK* 6, 33-37.

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":

$$\Sigma_j |P_j - P_j'|$$

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

The "maximum entropy function":

$$- \Sigma_j P_j \log(P_j/P_j')$$

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 all these functions. All of them, including the quadratic suggested in Rubinstein (1994), exhibit fairly consistent pictures of posterior distributions which additionally seem relatively independent of the assumed shape of the prior.

Figure 5
Implied Probability Distributions for Different Objective Functions
date: March 16, 1990 time: 11:47 am

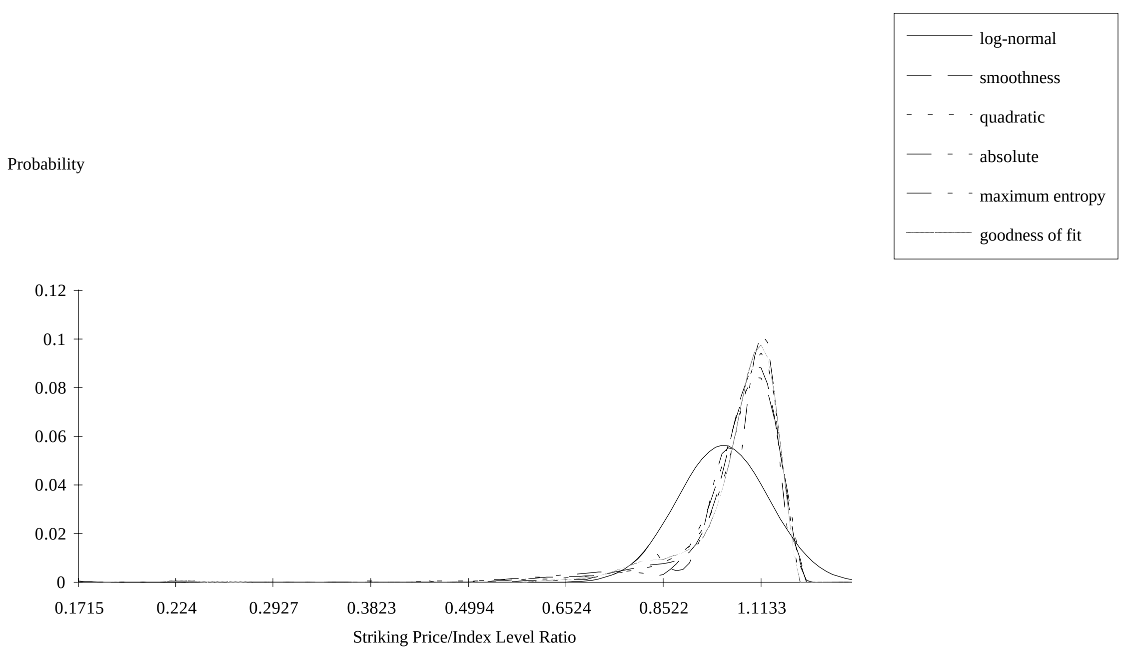
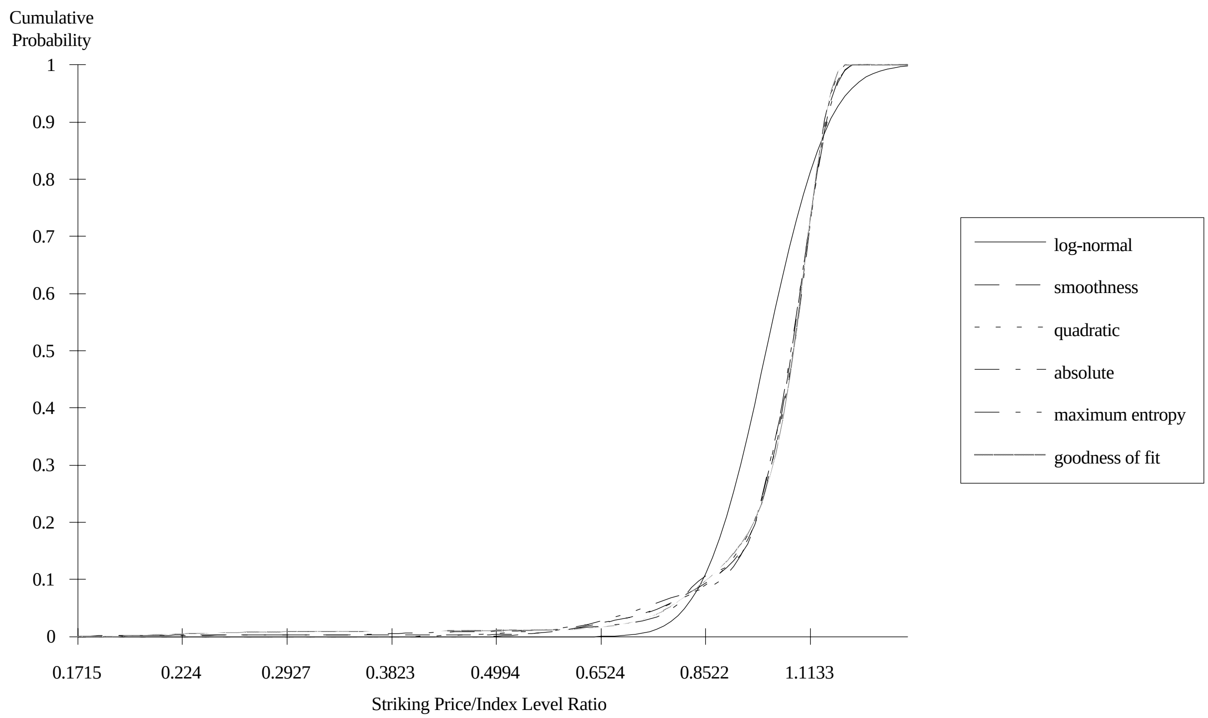


Figure 6
Implied Cumulative Probability Distributions for Different Objective Functions
date: March 16, 1990   time: 11:47 am



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 which 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. Even starting with a uniform, rather than lognormal prior, although convergence takes longer, eventually the resulting posterior is very close to the posterior which results from the lognormal prior. These comparisons suggest that in the cases we examined there were a sufficient number of options for the option prices themselves to constrain within narrow bounds the implied distribution consistent with them.[10]

Nonetheless, all these methods require an assumed prior distribution and occasionally lead to posterior distributions that have sufficiently little smoothness to be implausible. This suggests that an interesting approach might be to select the implied distribution with the maximum smoothness. To accomplish this, select the implied probabilities $\{P_j\}$ which minimize the

[10] See Hiroto Kuwahara and Terry Marsh, 1994, "Why Doesn't the Black-Scholes Model Fit Japanese Warrants and Convertible Bonds?," *Japanese Journal of Financial Economics* 1, 33-65, for related findings with regard to the quadratic, the goodness of fit, and the maximum entropy function for Japanese equity warrants. See Gilbert Basset, 1995, "A Note on Probability Bounds Based on Option Prices", working paper, Department of Economics, University of Illinois at Chicago, for bounds on the probabilities rather than for exact values as in our approach.

following function[11]:

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

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:[12]

$$\frac{\dfrac{P_{j+1} - P_j}{S_{j+1} - S_j} - \dfrac{P_j - P_{j-1}}{S_j - S_{j-1}}}{\frac{1}{2}(S_{j+1} + S_j) - \frac{1}{2}(S_j + S_{j-1})} = const.(P_{j-1} - 2P_j + P_{j+1})$$

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.

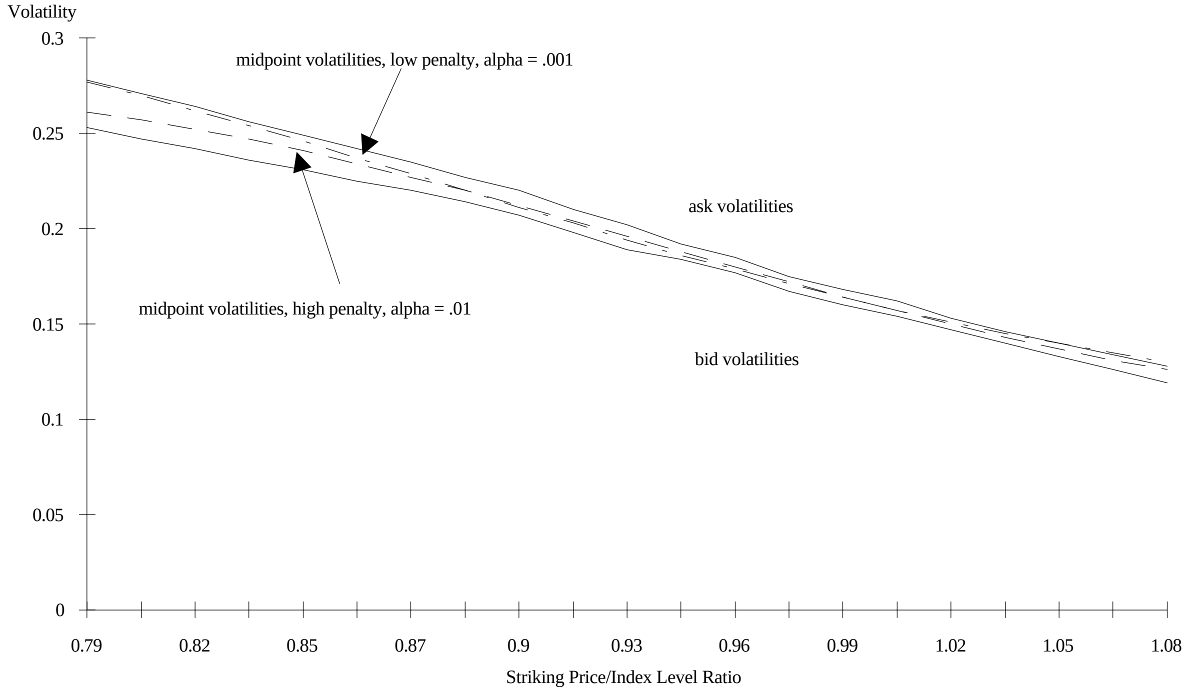
Bid-Ask vs. Midpoint Constraints. The optimization method proposed in Rubinstein (1994) chooses posterior probabilities so that the resulting option values all lie between their respective bid-ask quote. Unfortunately, all the objective functions discussed above give rise to option values with an implied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

---

11 See Kenneth Adams and Donald Van Deventer, 1994, "Fitting Yield Curves and Forward Rate Curves with Maximum Smoothness", *Journal of Fixed Income* 4, 52-62, for an application of the smoothness criterion to yield and forward rate curves.

12 The minimization of the integral from 0 to infinity $(\partial^2 P_j/\partial S_j^2)^2$ dj for a continuous probability distribution corresponds to fitting a cubic spline.

Figure 7
Implied Volatilities across Striking Prices Based on Bid-Ask Prices and Optimized Values
(calculated with penalties $\alpha = .01$ and $\alpha = .001$)
date: March 16, 1990



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 which penalizes deviations from the midpoint values but does not force the resulting posterior distribution to conform exactly to those values.

Underlying Asset Price Spacing. In the optimization method proposed in the earlier paper, the ending nodal underlying asset values of the implied binomial tree were preset equal to the same ending values that would have appeared in a standard binomial tree. In this case, the natural logarithms of the ending nodal values are equally spaced. A potential problem with ending nodal values spaced in this way occurs if two striking prices of observed options fall in between adjoining ending nodal values. In that case, the binomial tree cannot distinguish properly between those two options. Thus, the difference in option values will be set unrealistically. One solution is to increase the number of steps in the tree, say from 200 to 500; but such a tree significantly raises the computational burden.

A better way to assure that all striking prices are separated from each other by at least one underlying asset price is to create an equally spaced index at the end of the tree. Since we express all values at the end of the tree in terms deflated by $S(d/r)^{-t}$, we picked a set of 201 $5 increments starting at zero and deflated these values. The resulting set corresponds to all potential striking prices.[13]

Optimization Method. Our problem as we set it up is a nonlinear optimization with both linear and nonlinear equality and inequality constraints. For the earlier paper, we used the general nonlinear optimization routine from the NAG FORTRAN library. That, however, fails to exploit our private knowledge of the particular problem. To utilize this knowledge, we designed a specific optimization routine with the advantage that we can exert greater control over the optimization.

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 used the quasi-Newton method of the Broyden-Fletcher-Goldfarb-Shanno type as an optimization method for our problem.[14] The constraints of the problem were 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 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:

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

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 mid-point equality constraints on

---

[13] For optimization techniques requiring a lognormal prior, we can calculate it based on a 500-step tree, calculate the cumulative probability, and then resample using the revised new spacing. Alternatively, we can assign intervals of +/-$2.50 times the deflator factor around our new spacing, calculate the cumulative probability of the lognormal straight away, and then differentiate once to obtain the density.
[14] See David Luenberger, 1984, *Linear and Nonlinear Programming*, 2nd Edition, Addison-Wesley.

underlying asset and option prices, the augmented objective function is:

$$\min_{P_j} \quad \Sigma_j(P_{j-1}-2P_j+P_{j+1})^2 + \alpha\{\Sigma_j[\max[0,-P_j]]^2 + [\Sigma_j P_j-1]^2 + [(d^t(\Sigma_j P_j S_j)/r^t)-S]^2 + \Sigma_i[((\Sigma_j P_j \max[0,S_j-K_i])/r^t)-C_i^m]^2\}$$

We have now a formally unconstrained problem with a nonlinear, non-quadratic objective function. After obtaining a starting value (the lognormal distribution in our case), we search in the direction of steepest descent (which is the negative of the first derivative at that solution) for a minimum of the objective function. This step is called the line search. Those steepest descent steps tend to converge slowly. A very powerful method for quadratic functions is Newton's method which yields the optimal solution in one single step by using the matrix-product of the inverse of the Hessian and the negative gradient as a search direction. The Hessian is the matrix of second derivatives of the objective function with respect to the probabilities. However, the inverse of the Hessian is difficult to compute. Thus, quasi-Newton methods offer help by starting off the Hessian as an identity matrix and then updating it by incorporating information as the optimization progresses. A final concern is that the our problem is not quadratic. However, reasonably close to the solution, a quadratic approximation of our problem will behave well enough. To assure that we are dealing with a decent quadratic approximation, we restarted the Newton steps of the optimization after at most 20 steps with a steepest decent step.

Since any penalty approach will cause the eigenvalue structure of the Hessian to be badly scaled which slows convergence, we implemented a self-scaling version of the quasi-Newton method. The non-negativity constraints were implicitly dealt with by resetting any negative probabilities to zero throughout the line search. The line search was implemented to be inaccurate and based upon a quadratic fit. This insures rapid progress in the optimization and that the optimization is unlikely to converge to local optima. Furthermore, in order to assure global convergence, we dynamically increased the penalty parameter during the optimization thereby insuring that the resulting density function tends to be smooth. As a final touch, to speed up computations considerably, we exploited the fact that for days with downwards-sloping smiles, all probabilities for index levels beyond a certain cutoff point were set to zero by any optimization procedure. We thus reduced the probability space in such cases and later reset these probabilities to zero.

<u>Smoothness Criterion.</u> This approach is able to deal with any of the above objective functions. However, 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.

Since the S_j are chosen to coincide with all possible striking prices K_j, the smoothness criterion can be rewritten as follows:

$$\min_{P_j} \Omega = \Sigma_j (\partial^2 P_j/\partial K_j^2)^2$$

where (the Breeden-Litzenberger 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:

$$\min_{C_j} \Omega = \Sigma_j (\partial^4 C_j / \partial K_j^4)^2 \quad \text{subject to:}$$

$$C_j = C_i^m \text{ whenever } K_j = K_i \text{ for } j=0,...,n \text{ and } i=1,...,m$$

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:

$$\min_{C_j} \Omega = \Sigma_j (C_{j-2} - 4C_{j-1} + 6C_j - 4C_{j+1} + C_{j+2})^2 \quad \text{subject to:}$$

$$C_j = C_i^m \text{ whenever } K_j = K_i \text{ for } j=0,...,n \text{ and } i=1,...,m$$

This implies that we force the option values equal to their market prices whenever their striking prices coincide with available options. Option values with other striking prices corresponding to other underlying asset prices are determined freely by the optimization.

The formulation uses the Breeden-Litzenberger result that the second derivative of C_j with respect to K is the discounted density function of the underlying asset. That is, given the solution for C_j, we can easily reverse the procedure and determine the implied posterior probabilities from the butterfly condition:

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

Thus, solving directly for the C_j circumvents using the probabilistic constraints on option values (i.e. $\Sigma_i[((\Sigma_j P_j \max[0, S_j - K_i])/r^t) - C_i^m]^2$). Since the index spacing is chosen as we discussed, we sample at all available striking prices and at a set of hypothetical striking prices, one for every potential level of the ending underlying asset price. Omitting the non-negativity constraint on the posterior probabilities ($\Sigma_j[\max[0,-P_j]]^2$) for a moment, we can write the first order conditions of the Lagrangian for the rewritten smoothness criterion as a finite difference approximation:

$$\partial\Omega/\partial C_j = 2C_{j-4} - 16C_{j-3} + 56C_{j-2} - 112C_{j-1} + 140C_j - 112C_{j+1} + 56C_{j+2} - 16C_{j+3} + 2C_{j+4} = 0$$

$$\text{for } j = 0,...,n$$

In order to deal with the constraints that the known quote mid-point 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:

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

The first order conditions for the C_j which 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:

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

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

$$C_{n+1} = C_{n+2} = C_{n+3} = C_{n+4} = 0$$

and

$$C_i = r^{-t}\{1 - i(500/[S(d/r)^{-t}])\} \quad \text{for} \quad i = -1, -2, -3, -4$$

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.

One problem remains. We need to add the constraint we temporarily ignored that the probabilities (second derivatives of C_j) are nonnegative. 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.[15]
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.

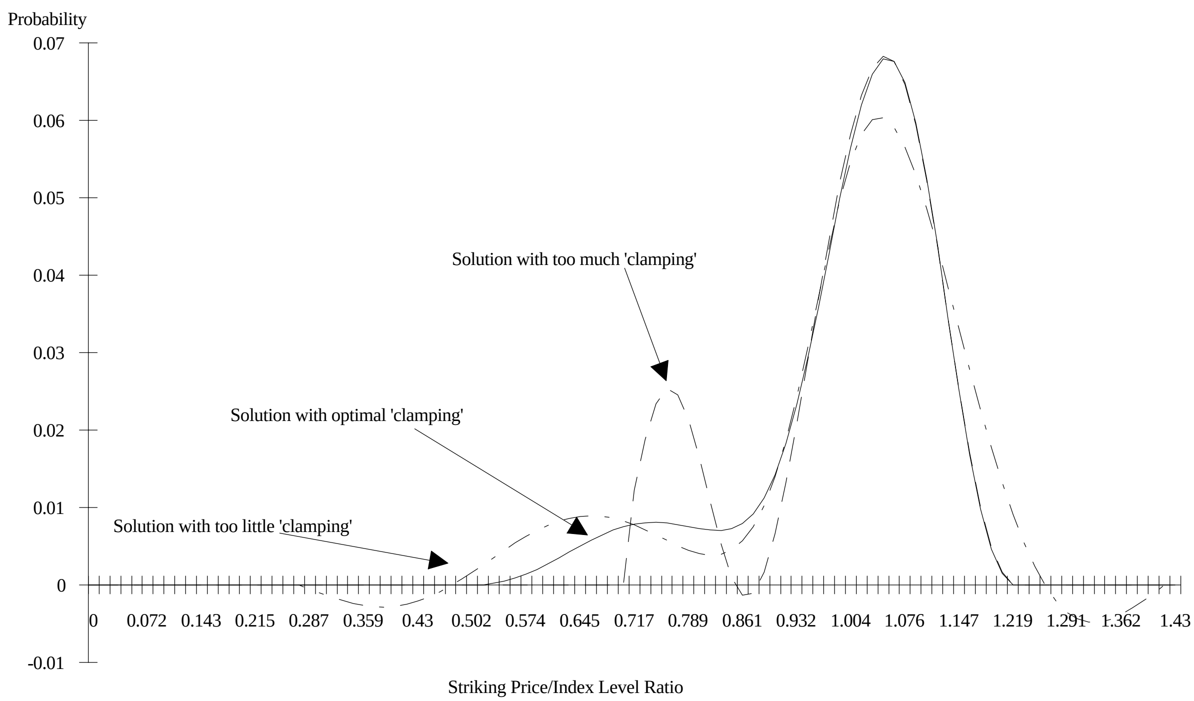
However, even if a set of nonnegative probabilities can be found, the resulting density function may exhibit multimodalities. We decided to reject candidate distributions with extreme

[15] See William Press et al., 1992, *Numerical Methods*, 2nd edition, University Press.

multimodalities.

Among all the candidate distributions, we chose the one with the largest number of non-zero probabilities if the largest `dip' between identical levels of probabilities did not span more than five index spacings. If all the `dips' were longer than 15 index spacings, we discarded the day. For `dips' between 5 and 15 index spacings, we calculated the ratio of the length of the `dip' over the number of nonnegative probabilities up to the mode of the distribution. We then selected the distribution with the lowest ratio.[16] The following diagram shows different stages of the `clamping down' procedure.

Figure 8
`Clamping Down' on the Solution Set for the Closed-Form Smoothness
date: March 16, 1990



In practice, it is even faster to preset all option values outside the striking price range of available options to 0 or $Sd^{-t} - Kr^{-t}$. Then the preset values can gradually be relaxed until all probabilities become nonnegative. Again, we continue in this way until non-negativities start recurring. The best distribution among those resulting in nonnegative probabilities is then chosen according to the above procedure. 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

---

[16] An alternative method would be to calculate the value of the objective function associated with any nonnegative probability distribution and select the one with the lowest value for the objective function. This normally yields approximately the same probability distribution.

probabilities sum to one is automatically satisfied.

IV. Empirical Results

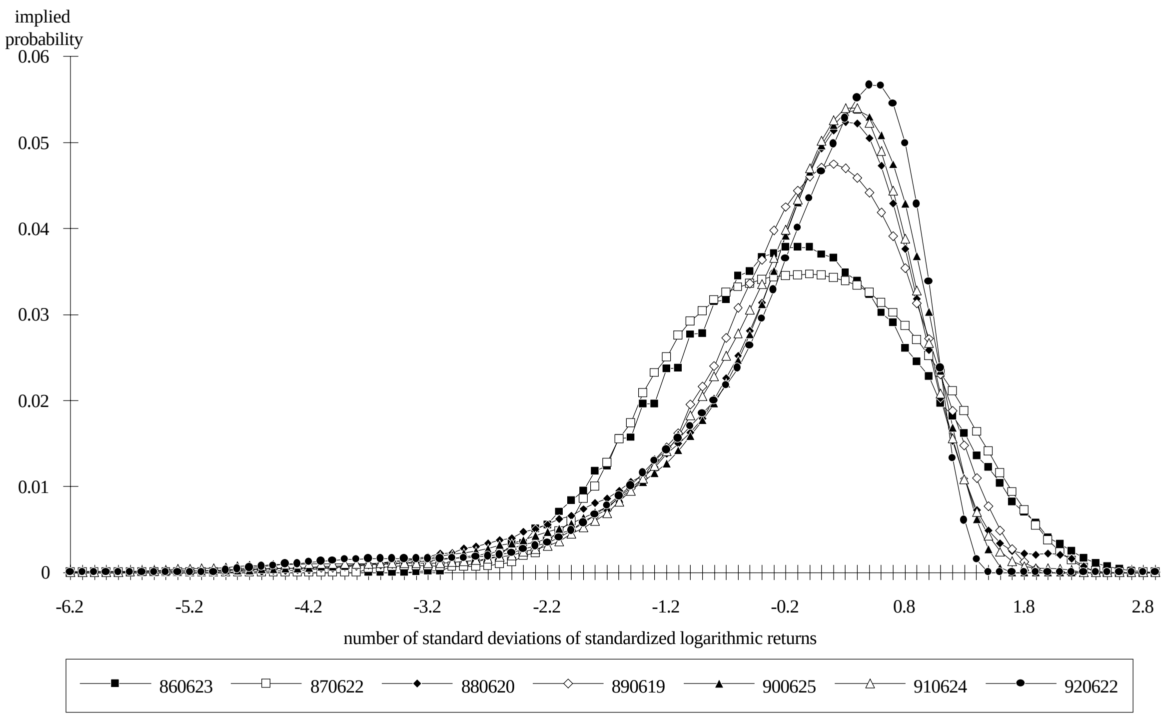
<u>Shapes of Implied Probability Distributions across Time.</u> In order to inspect the shapes of implied probability distributions as time passes, we graphed the distributions for options with 179 days-to-expiration as we observe them in June of each year. For purposes of comparison, we standardized the ending index level S_j by replacing it with:

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

where $\mu \equiv$ annualized mean of the logarithm of S_j/Sd^{-t} and $\sigma \equiv$ annualized volatility of the logarithm of S_j/Sd^{-t}

We can identify a distinct change in shape between the pre-crash and the post-crash distributuions. While the pre-crash distributions resemble the lognormal distribution, the post-crash distributions exhibit leptokurtosis and left-skewness.

Figure 9
Implied Probability Distributions
(time-to-expiration: 179 days)



Time patterns. We analyzed the time patterns of implied probability distributions by calculating the skewness and kurtosis for each day. We calculated skewness and kurtosis as:

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

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

Figure 10
Skewness and Kurtosis of Implied Probability Distributions
(median daily skewness and kurtosis across time, sampled quarterly)
period: 86/04/02 - 93/12/31 time-to-expiration: 135-225 days

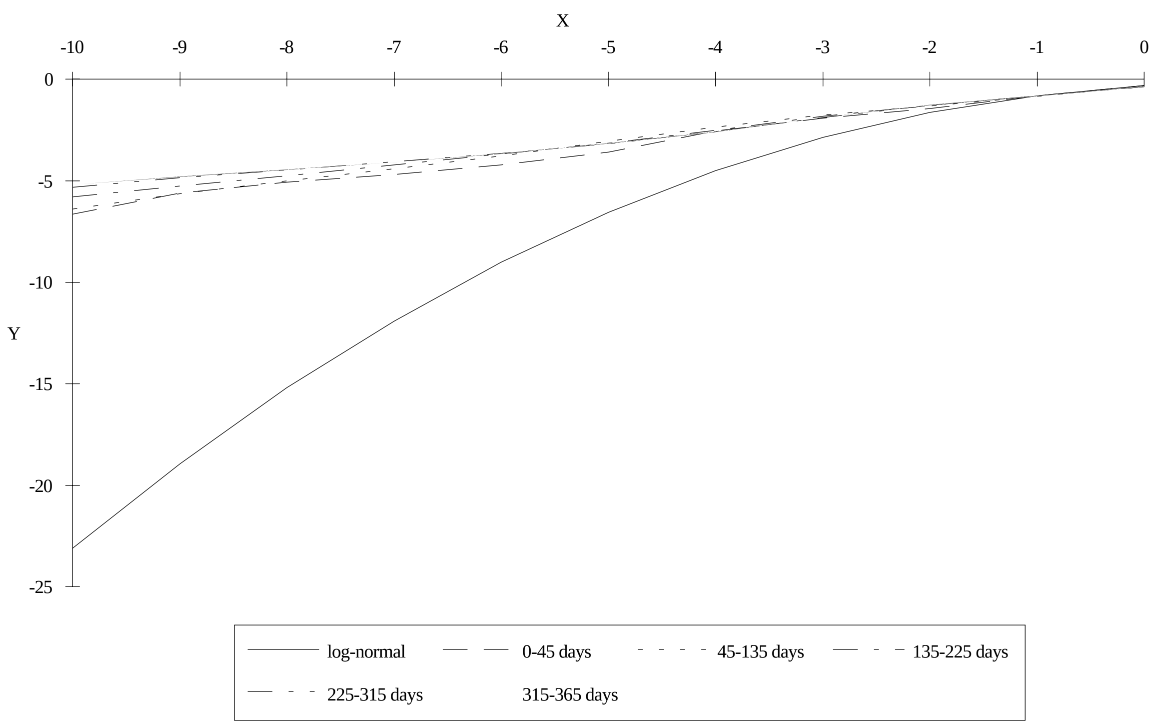
We find that the implied probability distributions in the pre-crash period were somewhat left-skewed and platykurtic. That is, the mean of the distribution tended to be to the right of the mode and the mode tended to be less pronounced than the mode of the corresponding lognormal distribution. After 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 pre-crash period and the mode is persistently more pronounced than the mode of the corresponding lognormal distribution.

Table 3
Skewness and Kurtosis of Implied Probability Distributions
(median daily skewness and kurtosis across time, sampled quarterly)
period: 86/04/02 - 93/12/31 time-to-expiration: 135-225 days

Quarter	skewness	kurtosis
86/2	-0.4213	-0.6925
86/3	-0.1908	-0.8099
86/4	-0.3773	-0.8130
87/1	-0.4936	-0.7104
87/2	-0.1834	-0.5973
87/3	-0.3966	-0.5976
87/4	-0.6501	0.6075
88/1	-0.8002	0.3990
88/2	-0.9741	0.9220
88/3	-0.9728	1.0579
88/4	-1.0362	1.1652
89/1	-1.1049	1.4232
89/2	-0.9804	1.1535
89/3	-1.0707	1.2901
89/4	-1.1182	1.2301
90/1	-1.1410	1.2645
90/2	-1.1259	1.2661
90/3	-1.0731	1.1673
90/4	-1.0280	0.9679
91/1	-1.0673	1.1612
91/2	-0.9695	1.0221
91/3	-1.1340	1.2758
91/4	-1.0564	1.1302
92/1	-1.1082	1.2804
92/2	-1.0952	1.1972
92/3	-1.1244	1.2595
92/4	-1.0665	1.1442
93/1	-1.0706	1.1678
93/2	-1.0474	1.1388
93/3	-1.0630	1.2222
93/4	-1.1105	1.2544

Patterns of cumulative probabilities. Here, we looked into the patterns of the implied probability distributions after we standardized them by replacing S_j with $S_j' = ((\log (S_j/Sd^{-t}))-\mu t)/\sigma\sqrt{t}$. This causes the lognormal distribution to be transformed into a N(0,1) normal distribution. Several interesting features should be noted. Figure 11 shows that the implied probability distributions associate a cumulative probability with a large drop in stock price by 10 standard deviations (using the mean interest rate of 6.63% and the mean Black-Scholes at-the-money volatility of 16.95% for options with time-to-expiration of 135-225 days, both annualized and calculated across 89/01/03 - 93/12/31, this translates into a drop in S by 80% over one year) which is about 10^{16} times higher than the cumulative probability for the lognormal distribution. For a drop in stock price by 4 standard deviations (drop in S by 46%) over one year, the cumulative probability of the implied distributions is still 10^{2} times higher than the cumulative probability for the lognormal distribution and 10^{1} times higher than the cumulative implied probability for the pre-crash period. Moreover, the graph of cumulative probabilities shows little differences across time-to-expiration after standardizing the implied probability distributions.

Figure 11
Cumulative Probability Distributions
(cumulative implied probability distributions for different times-to-expiration and for the lognormal distribution)
period: 87/10/19 - 93/12/31



X ≡ number of standard deviations of standardized logarithmic returns
$Y \equiv \log_{10} \Sigma_j \text{Prob}\{(((\log(S_j/Sd^{-t}))-\mu t)/\sigma\sqrt{t}) < X\}$

Table 4 shows cumulative probabilities for different periods where we aggregated the different times-to-expiration by taking the mean values. We find that the implied cumulative probabilities of the pre-crash period follow the lognormal distribution rather closely. In the pre-crash period, a decline in index level by 3 (-36% over one year) or more standard deviations had about the same probability under the implied and the lognormal distribution, but a decline by 4 (-46% over a year) or more standard deviations was 10 times more likely under the implied. In the post-crash 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 post-crash cumulative probabilities associate a significantly higher probability with a large drop in stock price. In addition, the cumulative probabilities in the post-crash period differ very little from each other. In comparing the implied probabilities from pre- and post-crash periods, we find that the probability of a 3 (or a 4) or more standard deviation decline was 10 times more likely after than before the crash.

Table 4
Cumulative Probability Distributions

(cumulative implied probability distributions for different periods and for the lognormal distribution)

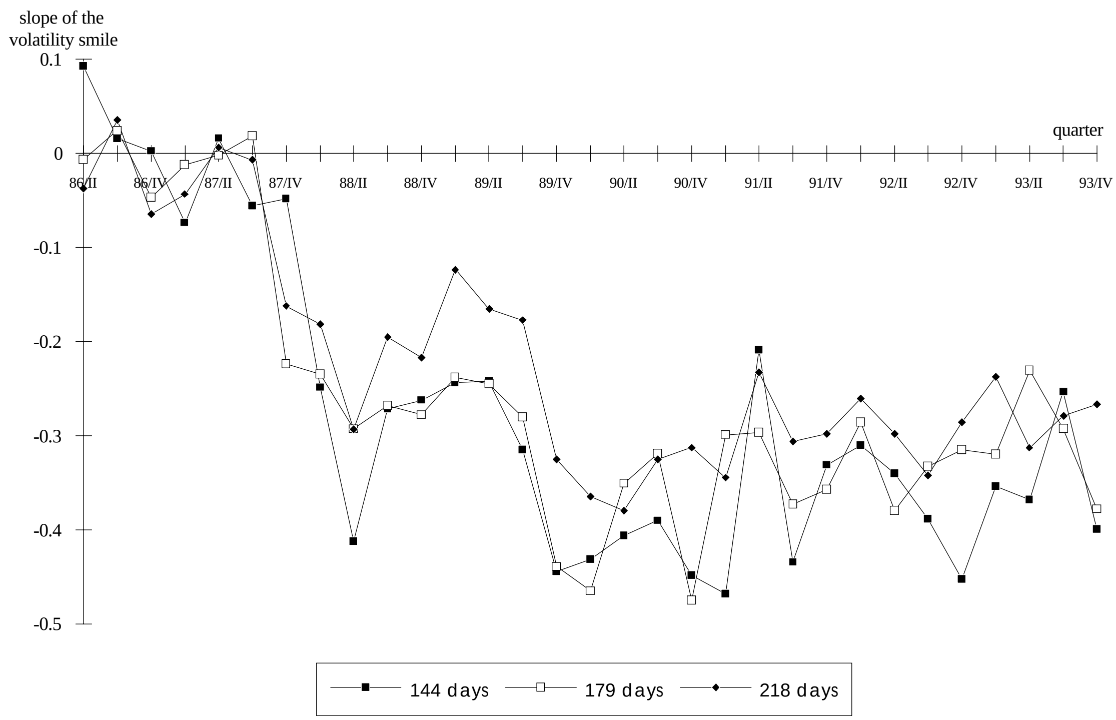
cumulative probabilities = Σ_j Prob { $(((\log(S_j/Sd^{-t}))-\mu t)/\sigma\sqrt{t}) < X$}

where X = number of standard deviations of standardized logarithmic returns

X	log-normal distribution across all periods	implied probability distribution 86/04/02 - 87/10/19	implied probability distribution 87/10/20 - 88/12/30	implied probability distribution 89/01/03 - 90/12/31	implied probability distribution 91/01/02 - 92/12/31
-10	.0000000000000000000000078		.0000066	.0000050	.0000020
-9	.00000000000000000011		.000013	.000010	.0000052
-8	.00000000000000063		.000030	.000024	.000016
-7	.0000000000013		.000069	.000063	.000049
-6	.000000001	.00000015	.00017	.00020	.00016
-5	.00000029	.0000052	.00054	.00077	.00055
-4	.000032	.00020	.0023	.0034	.0027
-3	.0013	.0013	.0095	.014	.014
-2	.023	.022	.040	.047	.049
-1	.16	.17	.14	.14	.14
0	.50	.49	.42	.41	.41

Smile Patterns across Time. Finally, we wanted to investigate the changes in volatility smiles across time. For that purpose, we measured the slope of the smile for options with 144, 179, and 218 days-to-expiration. Since we had learned that the smile can move considerably for far-out-of-the-money options, we limited our samples to striking price/index level ratios between .91 and 1.09. The result is shown in figure 12.

Figure 12
Slope of the Volatility Smiles
(striking price/index level ratio: .91-1.09)
dates: 86/04 - 93/12



We find a rapid adjustment to a negatively sloped smile after the crash of 1987. However, the small crash of 1989 seems to have some futher impact on the slope which becomes more negative before it somewhat rebounds towards the end of the sample.

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 a 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,[17] 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 paper, 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 than after the crash.

[17] This is taken to include the presumption that investors act to eliminate riskless arbitrage opportunities.

Bibliography

Adams, Kenneth J; Van Deventer, Donald. (1994). Fitting Yield Curves and Forward Rate Curves with Maximum Smoothness, *Journal of Fixed Income* 4, 52-62.

Basset, Gilbert W. (1995). A Note on Probability Bounds Based on Option Prices, working paper, Department of Economics, University of Illinois at Chicago.

Bookstaber, Richard M.; McDonald, James B. (1987). A General Distribution for Describing Security Price Returns, *Journal of Business* 60, 401-424.

Breeden, Douglas; Litzenberger, Robert (1978). Prices of State-Contingent Claims Implicit in Options Prices, *Journal of Business* 51, 621-651.

Cox, John; Ross, Stephen; Rubinstein, Mark (1979). Option Pricing: A simplified approach, Journal of *Financial Economics* 7, 229-263.

Derman, Emanuel; Kani, Iraj (1994). Riding on the Smile, *RISK* 7, 32-39.

Dupire, Bruno (1994). Pricing with a Smile, *RISK* 7, 18-20.

Kuwahara, Hiroto; Marsh, Terry (1994). Why Doesn't the Black-Scholes Model Fit Japanese Warrants and Convertible Bonds?, *Japanese Journal of Financial Economics* 1, 33-65

Longstaff, Francis (1990). Martingale Restriction Tests of Option Pricing Models, working paper, University of California at Los Angeles.

Luenberger, David G. (1984). *Linear and Nonlinear Programming*, 2nd edition, Addison-Wesley.

Mayhew, Stewart. (1995). Calculating Implied Parameters from Option Prices, working paper, Haas School of Business, University of California at Berkeley.

Press, William H.; Teukolsky, Saul A.; Vetterling, William T.; Flannery, Brian P. (1992). *Numerical Recipes in FORTRAN*, 2nd edition, University Press.

Rubinstein, Mark (1994). Implied Binomial Trees, *Journal of Finance* 49, 771-818.

Shimko, David (1993). Bounds of Probability, *RISK* 6, 33-37.

Siegel, Sidney (1956). *Nonparametric Statistics*, McGraw-Hill, New York.