

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 706

August 2001

INTERPRETING THE VOLATILITY SMILE:
AN EXAMINATION OF THE INFORMATION CONTENT OF OPTION PRICES

Steven A. Weinberg

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at www.federalreserve.gov/pubs/ifdp/.

INTERPRETING THE VOLATILITY SMILE:
AN EXAMINATION OF THE INFORMATION CONTENT OF OPTION PRICES

Steven A. Weinberg*

Abstract: This paper evaluates how useful the information contained in options prices is for predicting future price movements of the underlying assets. We develop an improved semiparametric methodology for estimating risk-neutral probability density functions (PDFs), which allows for skewness and intertemporal variation in higher moments. We use this technique to estimate a daily time series of risk-neutral PDFs spanning the late 1980's through 1999, for S&P 500 futures, U.S. dollar/Japanese yen futures and U.S. dollar/deutsche mark futures, using options on these futures. For the foreign exchange futures, we find little discernable additional information contained in the estimated PDFs beyond the information derived from the Black-Scholes model, a fully parametric specification. For S&P 500 futures, we find that the risk-neutral distribution implied by the volatility smile better fits the realized returns than the Black-Scholes model, although this better overall fit is not exhibited in the second and third moments.

JEL Classification: F31, G13, G15

Keywords: foreign exchange; derivative asset pricing; probability density functions

* The author is an economist in the International Finance Division of the Federal Reserve Board and can be reached by e-mail at steven.a.weinberg@frb.gov. Helpful comments were provided Doug Bernheim, Alain Chaboud, Neil Ericsson, Linda Kole, Mico Loretan and Will Melick. The views in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

1 Introduction

Part of the job of risk management is assessing the likelihood of a large adverse price move of a financial asset. It would be useful to market participants in making investment decisions to know how their individual assessment of the risk of holding an asset compares with the market's assessment of this risk. Policy makers would also benefit from knowing the market's assessment of risks. The Black-Scholes option pricing model (1973) allows for the computation of an estimate of market expectations of the volatility of the assets underlying the options. These estimates have generally been found to be good predictors of future volatility. However, researchers have noted that a maintained assumption of the Black-Scholes model, that the price path of the underlying asset is fully described by geometric Brownian motion, tends not to hold in practice. This is reflected in the presence of a "volatility smile,"¹ the empirical fact that implied volatilities for options with strike prices that are far out of the money tend to exceed those of at-the-money options. These differences in implied volatilities across strike prices can be exploited to obtain estimates of the implied risk-neutral probability density functions (PDF) of expected returns. The purpose of this paper is to examine the information content of the PDFs estimated from options prices, in particular to determine how helpful risk-neutral PDFs are in predicting the future distribution of returns.

A unique element of this study is that we not only examine how the risk-neutral distributions implied by options prices compare with realized distributions, but also test the information content of the second and third moments of the risk-neutral distribution against the corresponding moments of daily returns. In practice, a risk manager is generally most concerned about large price movements – high volatility – and is likely to be even more sensitive to large moves that decrease the total value of assets under his control – high skewness. Having reliable measures for predicting volatility and skewness would help a risk manager better assess when his portfolio is exposed to adverse risks, as he then might mitigate this exposure.

Soon after publication of the Black-Scholes option pricing model (1973), stock return volatility derived from option prices was compared with realized volatility. Early on, Latané and Rendelman (1976) used prices of options listed on the Chicago Board Options Exchange to compare the realized volatility of a cross-section of stocks with the volatility implied by the Black-Scholes model. They found that forecasts using implied volatility did better in predicting realized volatility than those based on historical volatility. Later studies examining the time-series properties of Black-Scholes implied volatilities confirmed, by and large, that these are more useful than historical volatility for forecasting realized future volatility. More

¹See Derman and Kani (1994) for an overview of the volatility smile.

recently, economic and financial researchers have developed a number of techniques to estimate the option-implied probability density function (PDF) of an underlying security's price movements. Much of the body of research in this field has, thus far, focused on developing and comparing estimation techniques for implied PDFs. However, only a handful of studies have examined whether the information derived from option prices can be used to forecast more than just the realized volatility, i.e., to predict further characteristics of the realized distribution. This study attempts to address this gap.

Financial assets were selected for this study based upon their liquidity, range of strike prices, and the existence of sufficiently long series of historical data. Options on futures traded on the Chicago Mercantile Exchange (CME) were chosen because they provide synchronized daily settlement prices that are matched with the settlement price for the underlying futures contract. We formalize the range of strikes criterion by requiring that for each day at least ten strike prices be actively traded. The three underlying financial assets that best met the selection criterion were the S&P 500 index futures (SP), the U.S. dollar/Japanese yen exchange rate futures (JY) and the U.S. dollar/deutsche mark futures (DM).²

For each day in the sample, we estimate a risk-neutral PDF of the underlying futures price. To obtain our PDF estimates, we extend a well-known nonparametric estimation technique by applying a variation of the generalized cross-validation method to determine the smoothing parameter for a cubic smoothing spline.

We initiate the investigation by utilizing a number of different tests of the goodness of fit to determine whether the risk-neutral PDF obtained by the volatility smile smoothing method is a reasonable representation of the true stochastic distribution. We also compare the fit of the PDF obtained by the volatility smile smoothing method against a lognormal distribution obtained by the parametric Black-Scholes model. We find that both models fit the realized returns of the JY and DM series reasonably well, but surprisingly, the lognormal distribution outperforms the more flexible semiparametric distribution in terms of goodness of fit. Neither model, however, fit realized returns of the SP series well, as risk neutrality is inadequate for modeling equity price movements. To correct for risk neutrality we shift the distributions by the average excess return of the S&P 500 futures over the period. For the risk-adjusted SP case, we find that the distribution obtained using the volatility smile smoothing method dominated the lognormal distribution. In fact, we cannot reject at the 10% level the hypothesis that the distribution of realized returns is statistically equivalent to that derived from the volatility smile smoothing method.

²The instrument on the CME with the highest options and futures volumes is the three-month eurodollar contract. However, for much of the time that options on this contract has traded, the number of actively traded strike prices has been quite limited.

We then utilize the estimated PDFs to revisit the literature on testing the information content of implied volatility. We examine the predictive ability of three measures of volatility: a moving average of historical volatility, the implied volatility from the at-the-money option, and the implied volatility of the distribution obtained using the smoothed volatility smile method. We find that both measures of implied volatility contain significant information for predicting realized volatility beyond that of the moving average measure. However, we also find that implied volatility of the distribution obtained using the complete set of traded strike prices is highly correlated with the implied volatility of just the at-the-money option, with neither measure of implied volatility dominating the other for the three data series we examine.

This paper then examines the next-higher moment, utilizing two measures of skewness to determine how well implied skewness performs in predicting realized skewness. We find that implied skewness does not help in predicting realized skewness, though historical skewness is not a good predictor of realized skewness either. We find that in situations of asymmetric PDFs, we could have profited from a strategy of selling an option with greater implied volatility than on the option on the opposite side of the PDF, and buying the less expensive option. This suggests that investors either consistently misassessed the probability of large moves, or more likely, that investors were willing to pay more than the actuarial cost for protection against large adverse price movements.

The remainder of this paper is organized as follows. We review previous studies that examined the role of options prices in predicting various characteristics of the distribution of price movements in section 2. Our methodology for estimating risk-neutral PDFs is presented in section 3. Section 4 discusses the data and gives summary statistics for the data and estimated PDFs. Section 5 examines the goodness of fit of the overall distribution. The information content of the second moment estimates is presented in Section 6. The information content of the third moment is addressed in Section 7, and the final section offers some conclusions.

2 Literature Review

We review previous studies that have empirically examined the information content of options prices so that the results of our study can be more readily compared. This also helps demonstrate the more novel empirical contributions of this paper.

2.1 Implied Volatility Studies

Research examining the information content of implied volatility addresses only a subset of what we study in this paper, since option-implied volatility is merely the second moment of the PDF. Findings of previous empirical studies have been mixed. Canina and Figlewski (1993) examined the most actively traded options contract in the United States, on the S&P 100 index.³ Canina and Figlewski (CF) regressed realized volatility over the remaining life of the contract on implied volatility and found that implied volatility contained little information for predicting realized volatility. In contrast, CF reported that a moving average of volatility over the previous 60 trading days had much greater predictive power. When realized volatility was regressed on both implied volatility and historical volatility, CF found that the contribution of implied volatility was insignificant. Christensen and Prabhala (1998) reexamined the findings of CF, using a longer sample period that included nearly eight years of data after the 1987 stock market crash, as well as the four years of pre-crash data used by CF. Christensen and Prabhala's empirical findings differed from those of CF, in that implied volatility was found to contain substantially more information and historical volatility had much less explanatory power.

Jorion (1995) also addressed the CF findings, examining implied volatility from options on foreign exchange futures traded on the CME. Jorion's main finding also ran counter to that of CF; when realized volatility is regressed on implied volatility⁴, historical volatility becomes insignificant and implied volatility has substantial predictive power. This informational superiority of implied volatility over historical volatility in predicting future volatility was called into question by Neuhaus (1995), who considered options on German Bund futures traded on the London International Financial Futures and Options Exchange (LIFFE). Neuhaus concluded that while implied volatility was useful in predicting realized volatility, historical volatility outperformed implied volatility in prediction.

2.2 Information Content of the Probability Density Function

There have only been a few studies until now that have examined the predictive capabilities of the wealth of information contained in time series estimates of PDFs. One reason is that the density function obtained from options prices are based on a risk-neutral pricing kernel which may differ from the subjective density that market participants place on future returns. Since the subjective density cannot be derived from options

³Our data set differs somewhat, as we used options on S&P 500 futures, whereas Canina and Figlewski used options on a cash index.

⁴Jorion defined historical volatility as the average volatility over the preceding 20 trading days.

prices without additional assumption,⁵ the risk-neutral density has been used instead. Bates (1991, 1997) links the risk neutral stochastic process to the actual stochastic process generating returns and derives a measure to assess abnormal expectations of return skewness from the risk-neutral PDF. Bates (2000) found that the high degree of negative skewness implied in S&P 500 index futures option prices is inconsistent with two option pricing models that extend the Black-Scholes model — stochastic volatility models, such as Hull and White (1987) and Heston (1993), and jump diffusion models such as Merton (1976) and Bates (1991). Thus far, models that attempt to represent the underlying stochastic process in asset price returns have not been successful in generating options prices that match market prices. This lack of fit offers the potential for methods that simply use observed option prices to extract risk-neutral PDFs to have better predictive power than methods based on estimating the underlying stochastic process.

Fackler and King (1990) developed a method, which they called the empirical calibration function, to evaluate the fit between the option-based risk-neutral distribution and the actual returns. Their method does not explicitly differentiate between the risk-neutral density and the subjective density. The calibration function is based on transforming a series of independent observations of realized prices to cumulative distribution function values under the hypothesized distribution. Fackler and King applied their calibration function to various agricultural commodity options and found that the lognormal distribution obtained from parametric Black-Scholes model fit most of the price distributions they assessed. Silva and Kahl (1993) looked at more recent data for some of the same markets that Fackler and King investigated, and found that the fit of the lognormal risk-neutral distribution improved in the later years, suggesting that prices in immature or illiquid markets tend to be less efficient.

Dumas, Fleming and Whaley (1998) defined a statistical measure, the hedge portfolio error, to compare the forecasting performance of various option pricing techniques. The hedge portfolio error is the difference between the change of the market quoted option price and the change in its model-derived theoretical value. For S&P 500 index options from June 1988 to December 1993, Dumas, Fleming and Whaley examined a series of one week intervals and found that the Black-Scholes model had a smaller hedge portfolio error than techniques that took into account the shape of the volatility smile. However, in examining the hedge portfolio error on FTSE-100 index options trading on the LIFFE over one day intervals between 1987 and 1997, Gemmill and Safekos(1999) found, using an implied distribution based on a mixture of lognormals, that the error was smaller when taking into account the volatility smile, as compared to the Black-Scholes

⁵Jackwerth (1996), Aït-Sahalia and Lo (2000) and Coutant (1999) have studied the relationship between the risk neutral density, the subjective density and the risk aversion function.

model. However, the smaller hedge portfolio errors only translated to improved pricing of about the same magnitude as the bid/ask spread on the options.

Most of the quantitative applications of PDFs have related the skewness of the risk-neutral distribution to the returns of the underlying asset. A number of studies have noted the high degree of negative skewness implied by equity options. Bates (1991) found that the PDF of the S&P 500 index futures was negatively skewed between October 1986 and August 1987, which he interpreted as caused by crash fears. However, the implied distribution was not negatively skewed for the two months before the October 1987 crash. Jackwerth and Rubinstein (1996) examined options on the S&P 500 index from 1986 through 1993 and found that since the 1987 crash the risk-neutral probability of a large decline in the index is far greater than under the assumption of lognormality. Nakamura and Shiratsuka (1999) looked at options on the Nikkei 225 index between 1989 and 1996 and found that skewness moved in the opposite direction to changes in the index, which they interpreted as a lag by market participants in adjusting their expectations to new market levels.

A risk-neutral PDF may better represent the actual distributions for foreign exchange returns than for equity index returns. One of the reasons posited for the excess skewness found in risk-neutral PDFs of equities is the higher marginal utility of wealth under a state where the market experienced a large downturn. While the probability of a crash may be low, the market will value a unit of wealth much more highly in this state, blurring the ability to distinguish between changes in market's perception of the probability of a crash and the value of wealth in this state. Following this line of reasoning, a change in a foreign exchange rate is not as likely to generate the same change in the marginal utility of wealth as a change in an equity index. A change in the exchange value of a currency is less likely to have a noticeable impact on an individual's wealth, given the home bias in asset allocation. In addition, movements in exchange rates among the major currencies likely generate little net change in aggregate global wealth.

Campa, Chang and Reider (1998) looked at implied skewness in one- and three-month over-the-counter options on a number of different exchange rates between April 1996 and March 1997. They found that the direction of skewness was positively correlated with returns over the remaining length of the option. They interpreted this as indicating positive momentum, i.e., a strong currency is likely to get stronger, rather than revert to an earlier level. Malz (1997b) also examined over-the-counter foreign exchange rates, using one-month options from April 1992 through June 1996. Malz found that investors can earn excess returns (in a CAPM sense) by holding currencies whose option prices indicate positive skewness.

The implied skewness of PDFs from foreign exchange options has also been used in event studies. Leahy and Thomas (1996) examined the PDF of the U.S. dollar-Canadian dollar exchange rate around the time of the Quebec independence referendum in 1995. Galati and Melick (1999) looked at the effect of market perceptions of foreign currency intervention operations by the Federal Reserve and the Bank of Japan on the distribution of the dollar/yen exchange rate. Galati and Melick found that while there was no statistically significant effect on the skewness of the PDFs, the market perception of intervention was associated with a higher variance of future spot rates.

3 Estimation of Implied PDFs

3.1 Review and Comparison of Estimation Methods

Employing the taxonomy of methods presented in Chang and Melick (1999), the techniques used to recover the risk-neutral probability distribution function of an asset on the date the option contract expires fall into two broad categories. The first semiparametric category is based on the observation – presented by Breeden and Litzenberger (1978) – that the risk-neutral PDF, $f(S)$, of the asset price, S , can be determined by taking the second derivative of the price of a European-style call, C , with respect to its strike price, X :

$$f(S) = e^{rT} \frac{\partial^2 C(S, T, X)}{\partial X^2}, \quad (1)$$

where T is the time to expiry of the option. Putting this technique into operation has been far from straightforward, as the range of available strikes generally falls far short of the continuity in $C(\cdot, X)$ needed to produce reasonably smooth PDFs.

Based on a result shown by Cox and Ross (1976) that the price of a call option is equal to its discounted value under a risk-neutral basis:

$$C(T, X) = e^{-rT} \int_X^\infty (S - X) f(S) dS, \quad (2)$$

a second category of techniques imposes some structure on the function $f(S)$ in order to guarantee that a proper probability distribution will be found. In empirical work, the most frequently used functional form for $f(S)$ has been the mixture of lognormal densities.⁶ Some other functional forms that have been applied

⁶Melick and Thomas (1997)

include the Burr III polynomial,⁷ a Hermite polynomial expansion around the lognormal density,⁸ and a Bayesian maximum entropy estimate using the lognormal density as the prior.⁹

A number of papers have compared the performance of some of the functional forms used to estimate the risk-neutral PDF based on equation (2). Coutant, Jondeau and Rockinger (1998) compared the mixture of lognormals with the Hermite polynomial expansion and the method of Bayesian maximum entropy, and found that the three methods produced similar PDFs, though the Hermite expansion was found to be more robust. McManus (1999) also compared the mixture of lognormals, the Hermite polynomial expansion and the Bayesian maximum entropy method in terms of their accuracy in fitting the options data. McManus found that the mixture of lognormals performed best.

Bliss and Panigirtzoglou (2000) compared the mixture of lognormals technique with a method, to be discussed below, based on equation (1). The basis for Bliss and Panigirtzoglou's comparison was the robustness and stability of the estimated PDFs to small errors in the quoted option prices. Bliss and Panigirtzoglou randomly perturbed the quoted prices by up to one-half of the contract's tick size,¹⁰ so that the perturbed prices would still round to the quoted prices if they were constrained (as they would be when traded on the exchange) to only move in tick increments. Bliss and Panigirtzoglou found strong evidence in favor of greater robustness of the volatility smoothing method over a mixture of lognormals method.

Given Bliss and Panigirtzoglou's finding that the volatility smoothing method produced more robust PDF estimates than those based on equation (2), an extension of the volatility smoothing method was developed for this paper.

3.2 Volatility Smoothing Method

The idea of smoothing the "volatility smile" to obtain continuous option prices was introduced in Shimko (1993). The volatility smile is obtained by applying the Black-Scholes options pricing formula to listed options at each strike price to calculate the implied volatility. Shimko then fit the implied volatilities as a parabolic function of strike price; outside the range of listed strikes, Shimko assumed that implied volatility was constant. By obtaining continuous call option prices, the PDF can be directly found by twice differentiating call prices with respect to strike price, as given by equation (1). It should be noted that this technique does not require that the maintained assumptions of the Black-Scholes option price model

⁷Sherrick, Garcia and Tirupattur (1996)

⁸Madan and Milne (1994)

⁹Buchen and Kelley (1996)

¹⁰The tick size is the smallest increment that an exchange quoted price may move.

hold; the Black-Scholes model merely serves as a means of transforming option price information into a more tractable form. Campa, Chang and Reider (1997) improved upon the fit of the implied volatility function by employing a cubic smoothing spline instead of the parabolic function.

While exchange-listed options are priced in terms of specific strike prices, the pricing convention in over-the-counter (OTC) foreign exchange options is represented in terms of delta, where $\delta \equiv \partial C / \partial S$ is the derivative of the Black-Scholes call option price with respect to the underlying asset price.¹¹ Utilizing the OTC pricing convention, Malz (1997a) smoothed Black-Scholes implied volatilities as a parabolic function of delta. The method developed by Bliss and Panigirtzoglou entails smoothing Black-Scholes implied volatilities via a cubic spline using delta rather than strike price as the independent variable.

There are some operational advantages to using delta rather than the strike price as the independent variable. First, as shown in Clews, Panigirtzoglou and Proudman (2000), working in delta space, where possible values range from 0 to e^{-rT} , is more conducive to examining the term-structure of volatility as opposed to working in strike space, where the distribution of strike prices widens as time to maturity, T , increases. Second, the transformation from strike space into delta space gives prominence to the most liquid contracts, the at-the-money-contracts, which trade near the current spot price of the asset. Near the spot price, a given change in the strike price translates into a relatively large change of delta, while a given change in the strike price farther from spot translates into a smaller change in delta. This property of the transformation from strike space into delta space played an important role in the greater robustness of Bliss and Panigirtzoglou's PDF estimates over those obtained using a mixture-of-lognormals technique. Bliss and Panigirtzoglou further emphasized the more liquid contracts by weighting, in the smoothing spline, each observation in terms of the parameter vega, ν , where $\nu \equiv \partial C / \partial \sigma$ is defined as the derivative of the Black-Scholes call price with respect to the Black-Scholes implied volatility. The value of an option becomes less sensitive to volatility the further an option's strike price is from the spot price.

While the cubic smoothing spline technique used by Bliss and Panigirtzoglou produced more robust results than the mixture of lognormals, for general usage, it remained unresolved how to optimally choose the appropriate amount of smoothing to apply. For a given λ , the smoothing spline is determined by minimizing:

$$\min_{f(\delta)} \sum_{i=1}^N w(i) \left(y(i) - f(\delta(i)) \right)^2 + \lambda \int_0^{e^{-rT}} f''(\delta)^2 d\delta, \quad (3)$$

¹¹It is easily shown that $0 \leq \delta \leq e^{-rT}$.

where N is the number of unique strike prices and f is a piecewise cubic polynomial with knot points $\delta(i)$ at the observed deltas. The weight of each observation, $w(i)$, is set equal to $\nu(i)$, and $y(i)$ is the Black-Scholes implied volatility for observation i . The cubic polynomial f is constrained so that it has continuous first and second derivatives at the knot points. The smoothness is determined by the parameter $\lambda \geq 0$, which controls how much to penalize departures from smoothness in the spline function, f :

$$\int_0^{e^{-rT}} (f''(\delta))^2 d\delta. \quad (4)$$

Varying the value of λ changes the effective parameters of the curve, which can vary from a linear function (with a least-squares fit) as λ approaches infinity, to an interpolation spline when λ is zero. In their study, Bliss and Panigirtzoglou set λ so that the mean-squared fitted option price error across dates and strikes approximately equaled the errors from those using the method of equation (2), with a mixture of two lognormals as the underlying functional form.

One problem with using a fixed value for λ is that the number of strikes can vary across days for the same option contract. Even holding constant the number of strikes, the amount of smoothing that a fixed value of λ induces will depend on the observed data. Therefore, by allowing λ to vary, we can treat observations in a more evenhanded manner. A number of techniques have been developed to allow a data-driven method of choosing the amount of smoothing. One such technique that has been widely used in scientific applications, is the method of generalized cross validation (GCV), described in Wahba (1990). The idea behind GCV is to find a value of λ that minimizes the error:

$$\sum_{k=1}^N w(i) \left(y(k) - f_{\lambda}^{[k]}(\delta(k)) \right)^2 \quad (5)$$

where $f_{\lambda}^{[k]}$ is the minimization of equation (3), for a given λ , with data point k omitted. GCV in effect serves to find an optimal λ by lowering the influence of outlying data points on the curve.¹² Fisher, Nychka and Zervos (1994) used GCV to find the smoothing parameter each day when fitting the term structure of interest rates. Fisher, Nychka and Zervos (1994) found in a comparison with other interpolation methods that GCV smoothing splines produced the most stable, accurate and least-biased results when fitting forward interest rates.

¹²If a point on the curve differs too much from its immediate neighbors, it will be eliminated because of no-arbitrage considerations, as will be discussed in Section 4.

In the process of developing the PDF estimation method for this study, we found that spline methods using the GCV algorithm tended to compute a value of λ that resulted in choppy PDFs that often contained negative probabilities. While the splines are optimal in a GCV sense in delta-implied volatility space, after transformation into strike space and then after twice-differentiation to extract the underlying PDF, the resulting PDFs tended not to be particularly smooth. Apparently, the GCV-determined smoothness penalty was too low. However, our GCV computational framework also allowed us to compute a value of λ corresponding to a specified number of effective parameters. For example, fixing $\lambda = 0$ would give as many effective parameters as there are strikes, while a value of λ near infinity would give only two effective parameters.

For this study we want the semiparametrically estimated PDFs to have roughly the same number of effective parameters as in the parametric methods.¹³ Previous studies using the mixture of lognormal distribution used either a mixture of two lognormals, which has five free parameters¹⁴, or a mixture of three lognormals, which has eight. While allowing more than eight effective parameters allows a closer fit of the observed implied volatilities, this generally comes at a cost of less smoothness and higher sensitivity to noisy observations. Data limitations led us to fix the effective number of parameter at six. We want to ensure that there are sufficient degrees of freedom (equal to the number of strikes minus the number of effective parameters) to still have some capability to mitigate the effects of data errors. We were not able to use foreign exchange options data in the initial years of exchange trading (1984 and 1985) because the degrees of freedom, even for six effective parameters, were close to zero on many days, as it took several years before enough interest in these options developed to allow a relatively large number of strikes to be traded.

In Figure 1 we graph the PDF of the SP series on June 24, 1998 using 6 effective parameters. For comparison we also plot the same data using the average value of λ for the SP series, which gave 9.2 effective parameters. The additional effective parameters placed fewer constraints on the shape of the PDF, but in the presence of noisy data this can lead to curves that seem less likely to represent the true underlying risk-neutral distribution. The sharp wiggles of the dashed PDF in Figure 1 suggest that extra effective parameters can amplify rather than dampen the effects of noisy data.

¹³In practice, for the same number of effective parameters, the semiparametric volatility smoothing method tends to have the ability to reflect a wider variety of shapes than a mixture of lognormals.

¹⁴The free parameters for a mixture of two lognormals are the mean and variance parameters for each of the lognormal distributions as well as a mixing parameter.

4 Data

4.1 Procedures for Data Selection

Given the purpose of this paper – to examine the information content of option-implied PDFs – a premium is placed on utilizing observations with the most information content. For most exchange-listed futures and their associated options, volume and open interest begin to increase shortly after a particular contract becomes the second-to-expiration contract, i.e., with an expiry of approximately six months. The next-to-expiration futures contract, with an expiry of less than three months, is generally the most actively traded. The pattern of trading activity for exchange-listed options on futures closely follows the futures themselves. With exchange-listed option strike prices available only at fixed intervals, (usually whole numbers), there is a tradeoff between liquidity and information. As the time to expiry gets closer, options with strike prices farther away from the current futures price become both less valuable and less liquid.

Besides the liquidity considerations, which favor using dates close to expiration, a competing consideration is to allow as much time as possible until expiry, so that realized values of volatility and skewness would be less influenced by any single day's return. In addition, as the time to expiration diminishes, the number of effective strike prices decreases since options relatively far out of the money become essentially worthless.¹⁵ Therefore, the selection rule for each trading date is to select the next-to-expiration contract until fewer than 45 calendar days remained until expiration, at which point the second-to-expiration contract was selected. Hence, option contract times to expiry in this study range from 45 to 139 calendar days.

Data on settle prices¹⁶ for futures and options on futures were obtained from the CME. The contracts examined are on the S&P 500 futures (SP), the Japanese yen/U.S. dollar futures (JY), and the deutsche mark/U.S dollar futures (DM). For each series, the options on futures were American-style. For the SP series, we chose to study post-crash observations, in view of the results of Canina and Figlewski (1993). Therefore, the sample period of daily observations went from January 4, 1988 to September 30, 1999, covering all quarterly futures contracts from 1988 through 1999. Futures trading in the JY began in early 1986. We use daily observations covering the quarterly futures contracts that expired from September 1986 through December 1999. Unfortunately, for the JY options, the CME was unable to provide data for 1987, as some of the data that year was corrupted. For the DM, we selected futures contracts between 1986 and 1998, with

¹⁵Options far in the money also have little information content, as their value essentially becomes the difference between the strike price and the futures price.

¹⁶Settle prices are determined at the end of each day by a settlement committee comprised of market participants. The use of settle prices serves to ensure the synchronization of the option prices and the associated futures price.

the exception of 1993, for which the CME was unable to provide uncorrupted data. With the introduction of the euro at the beginning of 1999, DM futures trading activity began to dry up, in favor of activity in the euro.¹⁷

The raw option price data are filtered to eliminate observations that allowed for arbitrage opportunities:

$$-X \leq \frac{\partial C(T, X)}{\partial X} \leq 0 \quad (6)$$

$$X \geq \frac{\partial P(T, X)}{\partial X} \geq 0, \quad (7)$$

where P is the price of a put. The first constraint states that the price of a call option should not become more expensive for higher strike prices, while the second constraint states that a put should not become less expensive for higher strike prices. The first constraint also states, for example, that a call with a strike price of \$45 cannot be worth more than \$5 above a call option with a strike price of \$50, all else constant. For a given strike price, both a put option and a call option may be available. For our data series, put-call parity did not generally hold, thus the contracts were not redundant. An examination of the CME options data confirms the conventional wisdom that for a given strike price, an out-of-the-money option is more liquid than its in-the-money counterpart. Therefore, in the case where both puts and calls were traded for a given strike price, the out-of-the-money option was selected.

The methodology of equation (1), described in the previous section, is based on differentiating European-style call options. Because of the value of early exercise, an American-style option is always at least as valuable as a European-style option. If volatility is extracted from American option prices, this implied volatility is biased upward.¹⁸ Thus, all American-style option prices are converted to European-style prices by applying the quadratic approximation of Barone-Adesi and Whaley (1987).

4.2 Variable definitions

For each day of options data we estimate a risk-neutral PDF¹⁹ of the futures for the date on which the option contracts expire. As we move along time, the PDF for a particular option contract is for a fixed expiry date. For each date, we compute the implied volatility of the estimated PDF. To convert the moments of the volatility-smoothed PDF into an implied volatility measure, we use the following equations, presented in

¹⁷The conversion rate between the deutsche mark and euro was permanently fixed at the end of 1998, giving the euro contract the same capabilities for hedging and speculating on the deutsche mark/U.S. dollar exchange rate as DM contracts.

¹⁸Most other researchers have ignored this issue.

¹⁹For the remainder of the paper we will use “PDF” as our shorthand notation for the distribution obtained by using the smoothed volatility-smile method.

Jarrow and Rudd (1982), to compute σ , the implied volatility:

$$q^2 = m^2/\mu^2 \quad (8)$$

$$\sigma = \sqrt{\log(q^2 + 1)/T} \quad (9)$$

where m^2 is the second moment of the PDF, μ is the first moment of the PDF and T is the time to expiration. We also calculate the Black-Scholes implied volatility by averaging the implied volatilities of the least-out-of-the-money put and call. The realized future volatility is computed over the remaining trading days of the contract, τ , by

$$\hat{\sigma}_{ft} = \sqrt{\frac{1}{\tau} \sum_{i=t}^{t+\tau} (R_i - \bar{R})^2}, \quad (10)$$

where R_i is the return on day i and \bar{R} is the mean return over the remaining days of the contract. Historical volatility is computed as

$$\hat{\sigma}_{ht} = \sqrt{\frac{1}{60} \sum_{i=1}^{60} (R_{t-i+1} - \bar{R})^2}, \quad (11)$$

a moving average of returns over the 60 trading days preceding date t .

Historical and realized future skewness necessitates computing the third central moment, with the realized third moment given by

$$\hat{m}_{ft}^3 = \frac{1}{\tau} \sum_{i=t}^{t+\tau} (R_i - \bar{R})^3, \quad (12)$$

and the historical third moment by

$$\hat{m}_{ht}^3 = \frac{1}{60} \sum_{i=1}^{60} (R_i - \bar{R})^3. \quad (13)$$

Skewness is then $\hat{m}^3/\hat{\sigma}^3$. From the PDF we extract two measures of skewness. The first measure is computed by directly calculating the second and third moments from the PDF. Because there is generally less liquidity as strike prices move away from the current futures price, the resulting imprecision in estimating the shape of the extreme tails of the PDF may decrease the reliability of the first measure of implied skewness. For this reason we also compute, from the PDF, the Pearson median skewness measure:

$$\frac{\hat{\mu} - \hat{Z}_{50}}{\hat{\sigma}}, \quad (14)$$

where $\hat{\mu}$ is the mean of the PDF (which equals the current futures price) and \hat{Z}_{50} is the median of the PDF. As will be discussed in Section 7, we adjust the two skewness measures by removing the skewness that is already inherent in the lognormal distribution to obtain measures of excess skewness.

4.3 Descriptive statistics

Table 1 presents descriptive statistics for the three data series. Comparing one-day returns, average returns on the SP are substantially larger than the foreign exchange series, but only slightly more volatile. For each series we see that Black-Scholes implied volatility is less than the PDF-implied volatility, indicative of implied volatility forming either a smile pattern, where implied volatility increases in either direction away from the center of the curve, or a smirk/sneer pattern, where volatility rises on one side of the curve. For DM and JY, average implied volatility is reasonably close to realized volatility. For SP, implied volatility is too high, which indicates that SP options are relatively expensive. Figures 2a–2c plot the realized future volatility, σ_{ft} , against implied volatility for each of the time series.

Observing the summary statistics for realized skewness provides some insights into the difficulty of accurately forecasting skewness. For each of the three data series, the standard deviation of realized skewness greatly exceeds its mean. Estimates based on the two measures of implied excess skewness show that the range of market expectations of skewness is generally much narrower than what is actually realized, consistent with the view that predicting the timing of a large move is quite difficult. The direction of skewness implied by the PDFs is, on balance, the same as that of realized skewness. For DM and JY, implied excess skewness tended to remain near zero. For SP, both implied skewness measures consistently predicted negative realized skewness. The time series of implied skewness is plotted in Figure 3.

The kurtosis implied by the risk-neutral PDFs exceeds the level of kurtosis found in a lognormal distribution, which indicates that market participants may expect excess kurtosis or they may be willing to pay a higher premium for protection against large price changes. The daily returns show that the realized kurtosis is far greater than would be observed under geometric Brownian motion.

5 Goodness of Fit

We begin our investigation of the information content of risk-neutral PDFs by testing whether a sequence of realized values of futures prices — taken at the expiration of the corresponding option contract — come from the risk-neutral distribution implied by option prices. We compare the goodness of fit of PDFs obtained

from the volatility smoothing method with the goodness of fit of a lognormal distribution with the same mean and standard deviation (i.e., as implied by the Black-Scholes model with matching implied volatility).

Our methodology for measuring the goodness of fit uses statistics based on the empirical distribution function (EDF), as described in Stephens (1974). The standard EDF test compares an empirical distribution function $F_n(x)$ obtained from a random sample of observed values x_1, x_2, \dots, x_n , with a hypothesized distribution function $F(x)$. In our framework, daily changes in option prices are transformed into a new $F(t, x)$ each day. Because EDF tests require independent realizations, we are restricted to using only one realized value x , per contract, with n equal to the total number of contracts in the sample for the financial variable in question. We also must eliminate the distributions, $F(t, x)$ having greater than 90 days until expiration to ensure that there is no overlap between contracts. Stephens showed that the Anderson-Darling A^2 statistic, the Cramér-von Mises W^2 statistic and the Watson U^2 statistic are the most powerful goodness of fit tests. The Anderson-Darling and Cramér-von Mises test are more powerful in detecting differences in location between the hypothesized and realized distributions. The W^2 statistic is sensitive to the observations across the entire distribution, while the A^2 statistic is more sensitive to the tails (cumulative probabilities near 0 and 1). The U^2 statistic is more powerful in detecting differences in variance.

Using the same framework as in EDF tests, Fackler and King (1990) developed a graphical technique, the empirical calibration function, that helps in assessing differences between $F(x)$ and $F_n(x)$. Both the calibration function and the EDF tests utilize a sequence of cumulative probabilities of the realized futures value at maturity for the various hypothesized distribution functions $F(t, x)$. The EDF tests and the empirical calibration function compare this sequence of cumulative probabilities with the uniform distribution function on the unit interval $[0, 1]$. Silva and Kahl (1993) succinctly describe how to form the calibration function for options on futures:

[T]he empirical calibration function, $C(U_T)$, is specified as

$$C(U_T) = T/n \quad \text{for } T = 1, 2, \dots, n$$

where $U_T = F(P_T^a)$ ($T = 1, 2, \dots, n$) = the cumulative probability assessment for P_T^a , the actual futures price at the maturity date T , and the values of U_T have been placed in ascending order (i.e., $U_1 \leq U_2 \leq \dots \leq U_n$); and T/n = the estimator of the “real” cumulative probability. (p. 768)

The calibration function can be viewed by plotting $(U, C(U))$ on the unit square. The closer the calibration curve lies to the 45° line, the better the fit between the hypothesized distribution function $F(x)$ and the realized distribution function $F_n(x)$. The calibration function also helps in detecting deviations between

the hypothesized distribution function and the realized distribution function. If a calibration curve lies consistently below the diagonal, this implies that the location of the hypothesized distribution function is below that of the realized distribution. This could occur, for example, if futures prices tended to be downward-biased predictors of spot prices, as the mean of the risk-neutral PDF would then be too low. If futures prices tended to exceed realized prices, then the mean of the PDF would be too high and the calibration function would lie above the diagonal. An S-shaped calibration curve suggests that the hypothesized distribution has too much dispersion; this would be the case, for example, if implied volatility consistently exceeded realized volatility. A reverse-S-shaped curve is consistent with too little dispersion in the hypothesized distribution.

We create calibration curves for each of the financial variables by obtaining the CDF value $F(x, t)$ for all contracts, holding time to expiry, t , constant. We plot calibration curves for the minimum and maximum values of t . In the upper panels of Figure 4 we examine the calibration curves for the DM contract. At 45 days to expiration, the volatility smoothing method generates a calibration curve that is slightly closer to the 45° line than the lognormal distribution. At 90 days to expiration the lognormal appears to be closer to the diagonal, with both curves displaying the reverse-S-shape consistent with smaller dispersion in either of the risk-neutral distributions (RND's) than in the realized distribution.

For the JY contract (lower panels of Figure 4), the calibration curve for the lognormal distribution appears to be closer to the diagonal at both maturities. The reverse-S-shape again suggests that the dispersion is underestimated for both of the RND's.

The calibration functions of the SP contract are examined in the upper panels of Figure 5. Both curves lie primarily below the 45° line, which indicates that means of both RND's are located to the left of the realized distribution. By a no-arbitrage argument, the mean of a risk-neutral distribution is tied to the risk-free rate of return. Historically, the return on equities has tended to exceed the risk-free rate. Hence, it comes as no surprise that the mean of the risk-neutral distribution lies below the true distribution. To correct for this effect, a second set of distributions was computed, with the same shape as the RND's, but with the means adjusted to equal the average excess return of equities over the risk-free rate for the years 1988 through 1999. The risk-adjusted SP calibration curves are plotted in the lower panels of Figure 5. At 45 days to expiration, the volatility smoothing method clearly dominates the lognormal. The difference between the two calibration curves is less at 90 days until expiration. At the longer maturity, both calibration curves are clearly S-shaped, suggesting that the hypothesized distributions overestimated dispersion (i.e., implied volatility exceeded realized volatility). This is consistent with the results in Table 1, which showed for SP that implied volatility, on net, exceeded realized volatility.

Table 2 displays the results of the EDF goodness of fit tests for the financial instruments in our study. We compute the goodness of fit at 45 and 90 days until expiration, to quantitatively measure the fit shown by the calibration curves in Figures 4 and 5. We also compute EDF statistics for the entire sample by using a Monte-Carlo technique. For each contract, we randomly selecting one date (allowing time to expiry to randomly vary from 45 and 90 days). We repeat this exercise 1000 times, taking the average of the EDF statistics over the 1000 trials. In addition to the EDF statistics, we want to compute a measure that would give some insight into whether the volatility-smoothing method is a better predictor of realized values than the Black-Scholes model. Therefore we also calculate the log-likelihood of the realized value, x , for each distribution, $F(t, x)$, weighted by the standard deviation of the distribution. In Table 2, we report the average value over the 1000 iterations of the four statistics, for both the PDF using the volatility-smoothing method and the matching lognormal distribution. Underneath the sample means we report the number of times that one distribution fits the data better than the other for the 1000 iterations.

We first examine the goodness of fit for the DM contract. At the shortest time horizon, 45 days to expiration, the volatility smoothing method slightly outperforms the Black-Scholes method. At 90 days to expiration, the Black-Scholes outperforms the volatility smoothing method by a greater margin than at 45 days to expiry. For both methods the fit is much better at the shorter maturity. These statistical results can be visually verified in Figure 4. For the full sample, the lognormal distribution clearly outperforms, winning more than 90% of the time when measured by the three EDF statistics. For the A^2 and W^2 statistics, we cannot reject the hypothesis at the 10% confidence level that the realized data came from either distribution. For the more stringent U^2 statistic we can reject the volatility-smoothing method at the 90% confidence level.

For JY, the lognormal distribution outperforms the volatility-smoothing method for both the shorter and longer maturities. Both methods have a better fit at 90 days to expiry than at 45 days. For the full sample, the lognormal distribution wins more than 95% of the time using the three EDF measures. We cannot reject the hypothesis that the realized data came from either distribution at the 10% confidence level. Using the U^2 statistic, we can reject the volatility-smoothing method at the 95% confidence level and the lognormal distribution at the 90% confidence level.

As discussed above and also shown in Figure 5, for the SP series the risk-neutral mean understates the realized mean. Not surprisingly, the EDF statistics that are most sensitive to location of the distribution, A^2 and W^2 , reject both methods at the 99% confidence level. Even the U^2 test, which is less sensitive to first moment differences, rejects both methods at the 99% confidence level for the full sample, although at 45 days to expiration, the volatility smoothing method is rejected at the 95% confidence level. When the distribution

means are adjusted to compensate for risk-neutrality, the improvement in the EDF statistics is striking. The volatility-smoothing method performs much better at 45 days to expiration than at 90 days. At 45 days, by the A^2 and W^2 statistics, we cannot reject the hypothesis at the 10% confidence level that the realized data came from either distribution. For the more stringent U^2 statistic we can reject the volatility-smoothing method at only the 90% confidence level. At 90 days, we can reject the volatility-smoothing method at the 99% confidence level by the U^2 statistic. For the full sample, the volatility-smoothing method won every trial versus the lognormal distribution. Whereas we can reject the lognormal distribution at the 95% confidence level for the A^2 and W^2 tests, we cannot reject the distribution given by the volatility-smoothing method at the 10% confidence level. The log-likelihood measure clearly favors the volatility-smoothing method over the Black-Scholes model.

In summary, the Black-Scholes model appears to fit the realized data very well for the DM series and nearly as well for the JY series. The volatility-smoothing method slightly underperforms for both series in terms of statistical fit. This underperformance is consistent across days to expiration, since the lognormal wins over 90% of the time in head-to-head comparisons. For the adjusted-SP series, in contrast, we obtain the opposite the result. The volatility-smoothing method seems to fit the realized data well in terms of location, as suggested by the A^2 and W^2 statistics. In head-to-head comparisons, the lognormal distribution is dominated by the volatility smoothing method.

The results of these goodness of fit tests are quite surprising. For the two series that, a priori, seemed to better fit the assumption of risk-neutrality, the DM and JY, the extra information contained in the distribution obtained from smoothing implied volatilities was expected to lead to a better fit than the lognormal distribution. For the SP, in which richly-valued out-of-the-money puts generate a consistently negative skew in the volatility-smoothing method, this distribution was not expected to do as well as the lognormal distribution (which is unencumbered by the negative skew), especially after both distributions were adjusted for excess returns. This result is puzzling, and suggests that for the SP contract, a set of options contained information about the realized distribution of price movements beyond what was found from just utilizing at-the-money options.

6 Implied Volatility

In the previous section we saw that the lognormal distribution fit the realized data a little better for the DM and JY series, while the distributions generated by the volatility smile smoothing method fit the SP series

better. In this section we investigate how well the models do in predicting realized volatility; we relate these results to the qualitative implications we found from the calibration curves. To measure the ability of our models to predict volatility, we use the same regression framework that was used in many earlier studies:

$$\sigma_{ft} = \alpha + \beta \hat{\sigma}_t + e_t, \quad (15)$$

where $\hat{\sigma}_t$ is either the Black-Scholes implied volatility for day t , the implied volatility of the smoothed volatility smile distribution (labeled as PDF in Table 3), or the historical volatility. If the estimate of β is significant and positive, then implied volatility contains some useful predictive information about future volatility. The hypothesis that implied volatility is an unbiased forecast of future volatility may be expressed as $\alpha = 0$ and $\beta = 1$.

The efficiency of implied volatility, which Canina and Figlewski found to be particularly poor, is tested with

$$\sigma_{ft} = \alpha + \beta_1 \hat{\sigma}_{It} + \beta_2 \hat{\sigma}_{ht} + e_t, \quad (16)$$

regressing realized volatility on both implied and historical volatility. If implied volatility contains all information conveyed by historical volatility, β_2 should be 0.

Given that relatively few quarterly observations could occur over the sample period, and since we want to take advantage of all of the information available, every daily observation is used. However, this induces serial correlation in the error term. Furthermore, there is no reason to believe that the error term is homoscedastic. Therefore, the standard errors were corrected for serial correlation and heteroscedasticity using the procedure of Newey and West (1987), using a lag length of 10.

The results of these information content regressions are shown in Table 3. In the information content regressions for the three series, β for both Black-Scholes and PDF implied volatility is significantly greater than zero, confirming that implied volatility provides useful information in forecasting future volatility. Comparing the Black-Scholes β coefficient for the DM and JY series with the results of Jorion (1995), the magnitude and the statistical significance of the regression coefficient is greater in our study²⁰ – we benefit from having seven years of additional data. Another difference between the studies is that we converted the American-style option prices to European-style prices, which provides a better estimate of implied volatility. Given that using American-style option prices will produce an upward biased estimate of volatility, it is

²⁰For DM we had a slope coefficient of 0.592 with standard error of (0.061), whereas Jorion had a slope coefficient of 0.547 with standard error of (0.138). For JY we had a slope coefficient of 0.667 with standard error of (0.080) vs. 0.496 and (0.181) for Jorion.

