

## Option-Implied Risk Aversion Estimates

by

**Robert R. Bliss**

Research Department  
Federal Reserve Bank of Chicago  
230 South La Salle Street  
Chicago, IL 60604-1413  
U.S.A.

(312) 322-2313

(312) 322-2357 Fax

Robert.Bliss@chi.frb.org

and

**Nikolaos Panigirtzoglou**

Monetary Instruments and Markets Division  
Bank of England  
Threadneedle Street  
London EC2R 8AH  
U.K.

+44-207-601-5440

+44-207-601-5953 Fax

nikolaos.panigirtzoglou@bankofengland.co.uk

April 11, 2003

First Draft: November 2, 2001

JEL Classifications: G13, C12

---

We are particularly grateful for helpful discussions with Lars Hansen, for comments by Jeremy Berkowitz, Avi Bick, Peter Christopherson, George Kapetanios, Jesper Lindé, David Marshall, Marti Subrahmanyam, and seminar participants at the Bank of England, the Federal Reserve Bank of Chicago, Indianapolis University–Purdue University—Indianapolis, McGill University, the Sveriges Riksbank, the University of Georgia, DePaul University, the 2002 Derivatives Securities Conference, 2002 Bachelier Finance Society Congress, the 2002 European Financial Management Association Annual Meeting, the 2002 European Finance Association Annual Meeting, and the Warwick Business School, Financial Options Research Center 2002 conference on *Options: Recent Advance*, for the guidance and suggestions of the editor Richard Green, and the referee. We thank Darrin Halcomb for his excellent research assistance. Any remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Chicago or the Bank of England. This paper was previously titled “Recovering Risk Aversion from Options.”

# Option-Implied Risk Aversion Estimates

ROBERT R. BLISS and NIKOLAOS PANIGIRTZOGLOU\*

## ABSTRACT

Using a utility function to adjust the risk-neutral PDF embedded in cross-sections of options, we obtain measures of the risk aversion implied in option prices. Using FTSE 100 and S&P 500 options, and both power and exponential utility functions, we estimate the representative agent's relative risk aversion at different horizons. The estimated coefficients of relative risk aversion are all reasonable. The relative risk aversion estimates are remarkably consistent across utility functions and across markets for given horizons. The degree of relative risk aversion declines broadly with the forecast horizon and is lower during periods of high market volatility.

Estimating the representative agent's or the market's degree of risk aversion from securities prices has a long history. However, it is only recently that scholars have begun using options data to do so. Options provide a particularly promising context for studying risk preferences. Stocks are infinitely lived and so inferences must be drawn from the discounted stream of cash flows over an indefinite horizon. Usually this involves additional assumptions as to how those cash flows evolve (e.g., constant growth of dividends). Since only one value, the discounted present value of all cash flows, is known, no inferences are possible about variations in preferences over different horizons. Options on the other hand have a fixed expiry date at which payoffs are realized.<sup>1</sup> Furthermore, options contracts exist for different investment horizons. Options thus permit studying preferences over specific horizons and simultaneously over multiple horizons. Futures contracts also share this fixed-horizon characteristic. Options, however, provide a spectrum of observations for each expiry date on any given observation date—one for each quoted strike price—while futures provide only a single statistic for each expiry date/observation date pair. The multiplicity of prices for different payoffs on the same underlying asset provided by options allows us to construct a density function for

---

<sup>1</sup>American options present a somewhat more complicated investment horizon, but only to the extent that early exercise is optional. Still, even in that case, American options allow a greater specificity of investment horizon than stocks do.

the distribution of possible values of the underlying asset. In contrast, single-datum stock and futures prices allow inference only about the mean of the distribution, unless additional assumptions are made linking the observed time-series to a stochastic process or density function functional form.

This paper uses the informativeness of options, together with a new method of inferring risk aversion implied by security market prices, to present unique evidence of the term structure of risk preferences. We confirm this across markets. We also present evidence that the implied risk preferences are volatility dependent.

Cross-sections of option prices have long been used to estimate implied probability density functions (PDFs). These PDFs represent forward-looking forecasts of the distributions of prices of the underlying asset. Option-derived distributions have the distinct advantage of (usually) being based on data from a single point in time, rather than being taken from an historical time-series. As a result, these PDFs are theoretically much more responsive to changing market expectations than are density forecasts estimated from historical time-series data using statistical density estimation methods or deriving density forecasts from parameterized time-series models.

Unfortunately, theory also tells us that the PDFs estimated from options prices are risk-neutral PDFs. If the representative investor who determines options prices is not risk-neutral, these PDFs need not correspond to the representative investor's (that is, the market's) actual forecast of the future distribution of underlying asset values.

If investors are rational, their subjective density forecasts should correspond, on average, to the distribution of realizations; that is, their subjective density forecasts should coincide with the objective or physical densities from which realizations are in fact drawn. Thus, one way to test whether risk-neutral densities reflect market expectations is to test whether they provide accurate density forecasts. If risk-neutral PDFs do not forecast accurately, we may infer that the difference between the risk-neutral and the accurate or objective forecast arises from the risk aversion of the representative agent. We can then use this difference to infer the degree of risk aversion of the representative investor.

A number of papers have examined the density forecast accuracy for different option-derived risk-neutral PDFs.<sup>2</sup> Most of these studies have rejected the hypothesis that option-derived risk-neutral PDFs are accurate forecasts of the distribution of future values of the underlying asset. Thus, evidence suggests that implied PDFs cannot reliably be used to infer market expectations concerning the future distribution of the underlying asset. This is not entirely surprising as there is a large literature establishing the existence of risk premia in market prices, particularly equity markets. Nonetheless, numerous other papers have proceeded to interpret risk-neutral PDFs as representative of market expectations.<sup>3</sup>

The theoretical relation between the state prices and state probabilities is well understood (see Huang and Litzenberger (1988) eq. 6.2.1), as is the extension of this idea to continuous distributions and the relation between risk-neutral and objective density functions.<sup>4</sup> Subject to certain conditions such as complete and frictionless markets and a single asset, the risk-neutral density function,  $q(S_T)$ , is related to the objective density function,  $p(S_T)$ , by the representative investor's utility function,  $U(S_T)$ , as follows:

$$\frac{p(S_T)}{q(S_T)} = \lambda \frac{U'(S_T)}{U'(S_t)} \equiv \zeta(S_T), \quad (1)$$

where  $\lambda$  is a constant. The function  $\zeta(S_T)$  is the pricing kernel. Thus, knowing any two of the three functions—the risk-neutral density, the objective density, and/or the pricing kernel (equivalently the utility function)—permits us to infer the third.

The methodology in most previous studies of options and risk aversion has been to separately estimate the risk-neutral density from options prices and the objective (or statistical) density function from historical prices of the underlying asset, use these two separately derived functions to infer the pricing kernel, and then draw conclusions from the implied relative risk aversion function.<sup>5</sup> These papers have typically imposed an

---

<sup>2</sup>Anagnou et al. (2002) provide an excellent review of previous papers before adding their own results.

<sup>3</sup>See Bliss and Panigirtzoglou (2002) for a partial list.

<sup>4</sup>The origins of the development of this result have proven difficult to trace. The formulation used here is taken from Ait-Sahalia and Lo (2000).

<sup>5</sup>See, for example, Ait-Sahalia and Lo (2000), Ait-Sahalia, Wang, and Yared (2001), Coutant (2001), Jackwerth (2000), Weinberg (2001), Pérignon and Villa (2002), and Rosenberg and Engle (2002).

assumption of stationarity on the statistical density function or the parameters of the underlying stochastic process to facilitate estimating the objective density function from historical data. The degree of stationarity assumed varies. Some studies, such as Ait-Sahalia and Lo (2000) and Jackwerth (2000), assumed that the distribution of returns is constant over a long period. Other studies such as Ait-Sahalia, Wang, and Yared (2001) and Rosenberg and Engle (2002) fit a stochastic process with parameters assumed to be constant over long periods; in other words, that the conditional densities are time-invariant. Some also pool option cross-sections from different observation dates under the assumption that the risk-neutral PDF is also stationary.<sup>6</sup> This ignores evidence to the contrary.<sup>7</sup> These stationarity assumptions are not implied or required by the theory, they are made for practical econometric reasons. Unfortunately, the resulting risk-aversion functions are somewhat inconsistent with theory: either U-shaped or generally declining but not monotonically so.

We can directly estimate time-varying risk-neutral PDFs without imposing strong *a priori* structures on the data by using single-observation cross-sections of options prices. However, one cannot independently estimate a time-varying statistical density from a time-series of prices without imposing an *a priori* structure, for instance by assuming that prices follow a particular stochastic process. Unfortunately, the statistical density and/or stochastic process stationarity assumptions made in doing so are subject to several criticisms. Estimated risk-neutral PDFs are rarely consistent with the simple functional forms implied by simple one-factor diffusion models, nor are changes in PDFs over time consistent with simple shifts in the mean of the stochastic process. Furthermore, when faced with changing risk-neutral PDFs, any assumed stationarity of the objective PDF becomes questionable. The assumption made in the previously

---

<sup>6</sup>See, for example, Ait-Sahalia and Lo (2000), Ait-Sahalia, Wang, and Yared (2001), Coutant (2001), and Pérignon and Villa (2002).

<sup>7</sup>Daily changes in implied PDFs are readily observable in the data and have been used in numerous event studies (for example: Campa, Chang, and Reider (1998), Melick and Thomas (1997), Campa, Chang, and Refalo (2002), Galati and Melick (1999), Gemmill and Saflekos (1999), Shiratsuka (2001), and Söderlind (2000)); furthermore, several central banks track monthly changes in implied PDFs to infer changes in market sentiment. Though to be absolutely fair, this evidence is indirect and not necessarily conclusive. We know of no study that has directly tested whether it is possible to reject the stationarity of implied risk-neutral PDFs over various time intervals.

discussed papers that the true statistical distribution is constant begs the question of why then the risk-neutral distributions clearly are not. A stationary statistical distribution requires either that the pricing kernel is time-varying or that investors are irrational; that is, they do not account for the supposedly stationary distribution of prices, in order to explain the clearly time-varying risk-neutral distributions that we observe. Directly testing the stationarity of the statistical distribution requires more data than is usually available, though volatility clustering, price and volatility spikes, and the frequent application of time-varying volatility models and regime-switching models to describe financial time-series points to the strong possibility that true underlying statistical distributions are time-varying.<sup>8</sup>

An alternative to assuming statistical-distribution or stochastic-process stationarity is to assume risk-aversion function stationarity. We can do this by assuming some well-behaved functional form for the underlying utility function, consistent with most non-options-based studies of market risk aversion.<sup>9</sup> Thus, rather than imposing stationarity restrictions on the underlying statistical processes to permit estimating the objective density from a time-series of historical prices, we impose an alternative restriction on the pricing kernel and permit the objective density to time-vary. We assume a parametric form for the utility function, estimate the appropriate risk aversion parameter under the assumption that this value is stationary over the sample period, and then, using time-varying risk-neutral density functions estimated from option prices, derive the time-varying implied objective density functions. Our goal is to find implied subjective density functions that are consistent with both utility theory and rational expectations.

We investigate these questions using FTSE 100 and S&P 500 options and two different utility functions to adjust risk-neutral PDFs. We find, as others have, that risk-neutral PDFs are poor forecasters of the distribution of future values of the underlying indices. We then find the optimal value for the parameters of the utility functions used to

---

<sup>8</sup>Rosenberg and Engle (2002) capture some of these effects through the use of a stochastic volatility model for the return generating process.

<sup>9</sup>Bartunek and Chowdhury (1997) combine this approach with a stationary model for the true return generating process to generate a risk-neutral density function which they then use to price options.

construct the subjective PDFs and show that these subjective PDFs are improved forecasters of the distribution of future values of the underlying indices. In most cases these utility-adjusted PDFs can no longer be rejected as “good” forecasts of the distributions of the underlying asset.<sup>10</sup> The measures of risk aversion implicit in these adjustments are well behaved, of reasonable magnitude, remarkably consistent across the two markets and the two utility functions considered. We also examine relative performance of alternative sets of PDFs to determine whether one competing alternative is superior to another.

The remainder of the paper proceeds as follows. Following a brief description of the Data in Section I, the Methodology section (Section II) outlines the theory underlying the comparison of risk-neutral and objective densities, details how we estimate risk-neutral PDFs, adjust them to get the subjective PDFs, and then test these subjective density forecasts to see if these conform to the objective densities from which realizations are drawn. The empirical results are presented and analyzed in the Results section (Section III), and the Conclusion follows in Section IV. The Appendix discusses alternate methods of testing density forecasts, together with the Monte Carlo tests we conducted to select the method used in this paper.

## **I. Data**

Two sets of equity options contracts are used in this study—S&P 500 options traded on the Chicago Mercantile Exchange (CME) and FTSE 100 options traded on the London International Financial Futures Exchange (LIFFE)—together with data on the underlying asset and the risk-free interest rates needed to price options.<sup>11</sup> Data included options expiries from February 18, 1983 through June 15, 2001 for the S&P 500 options and June 19, 1992 through March 16, 2001 for the FTSE 100 index options.

The CME S&P 500 options contract is an American option on the CME S&P 500 futures contract. S&P 500 options trade with expiries on the same expiry dates as the futures contracts, which trade out to one year with expiries in March, June, September,

---

<sup>10</sup>Of course, “failure to reject” does not mean we should “accept.” Our use of the term a “good forecast” is merely an expositional convenience and should be understood as such.

<sup>11</sup>Short Sterling options were also examined but failed to produce enough usable cross-sections for meaningful analysis.

and December. In addition, there are monthly serial options contracts out to one quarter. Thus, at the beginning of January, options are trading with expiries in January, February, March, June, September, and December; at the beginning of February options trade with expiries in February, March, April, June, September, and December. Options expire on the third Friday of the expiry month, as do the futures contracts in their expiry months. Prior to March 1987, the S&P 500 futures settled to the value of the S&P 500 index at the close on Friday. Beginning in March 1987, the futures settled to an exchange-determined Special Opening Price on the expiry Friday. For serial months there is no futures expiry and the options settle to the closing price on the option expiry date of the next maturing S&P 500 futures contract. The S&P 500 realizations used in this study to compute options payoffs are the Special Opening Quote for quarterly contracts beginning in March 1987 and the S&P 500 futures settlement price for serial contracts and all contracts prior to 1987. Option quotations used to compute PDFs are the closing prices; the associated value of the underlying is the settlement price of the S&P 500 futures contract maturing on or just after the option expiry date.

The LIFFE FTSE 100 option contract used in this study is a European option on the FTSE 100 equity index. Options are traded with expiries in March, June, September, and December. Additional serial contracts are introduced so that options trade with expiries in each of the nearest three months. FTSE 100 options expire on the third Friday of the expiry month. FTSE 100 options positions are marked-to-market daily based on the daily settlement price, which is determined by LIFFE and confirmed by the Clearing House. The FTSE 100 options prices used in this study are the LIFFE-reported settlement prices.

The quarterly FTSE 100 futures contracts expire on the same date as the options and therefore will have the same value as the index when the option expires. The European-style FTSE 100 contract may thus be viewed as an option on the futures, if one assumes that mark-to-market effects are insignificant. LIFFE reports the futures prices as the value of the underlying asset in their options data. For serial months, LIFFE constructs a theoretical futures price based on a fair value spread over the current futures front quarterly delivery month. In computing FTSE 100 implied volatilities, the value of the underlying asset corresponding to each cross-section of option quotes used in this



study is the actual or theoretical futures price reported by LIFFE for that contract. At expiry the options settle to the Exchange Delivery Settlement Price determined by LIFFE by taking the average level of the FTSE 100 index sampled every 15 seconds between 10:10 and 10:30 on the last trading day, after first discarding the highest and lowest 12 observations. This series was used to compute FTSE 100 option payoffs for this study.

The risk-free rates used in this study are the British Bankers Association's 11 a.m. fixings of the three-month EuroDollar and Short Sterling London Inter-Bank Offer Rate (LIBOR) rates reported by Bloomberg. While this does not provide a maturity-matched interest rate, it can nonetheless be justified by necessity and lack of materiality. Overnight rates (Fed Funds, repos) are not representative of the borrowing costs faced by options traders and are subject to distortions arising from central bank activities. Short-maturity Treasury bills or LIBOR rates are illiquid or are subject to price distortions due to their use by central banks for reserve management purposes.<sup>12</sup> The three-month LIBOR market has the dual advantages of liquidity and approximating the actual market borrowing and lending rates faced by options market participants. Furthermore, Treasury rates represent lending, not borrowing, rates. In any case, the choice of interest rate has little effect on the methodology. Interest rates are used as an input when converting option prices to implied volatilities (for smoothing) and back again. A 100 basis point (bp) change in the assumed interest rate will produce approximately a two bp change in the measured at-the-money implied volatility for a one-month contract, increasing to five bp at the six-month horizon. Use of three-month interest rate as a proxy for the one-month rate is unlikely to misstate the correct (if unobservable) rate by anything approaching 100 bp.

A target observation date was determined for horizons of one, two, three, four, five, six weeks, one, two, three, four, five, six, nine months, and one year, by subtracting the appropriate number of days (weekly horizons) or months (monthly and one-year horizons) from the option expiry date. If no options traded on the target observation date, the nearest options trading date was determined. If this nearest trading date differed from the target observation date by no more than three days for weekly horizons or four days

---

<sup>12</sup>Duffee (1996) provides evidence that short maturity U.S. Treasury securities exhibit idiosyncratic variations that makes them unsuitable proxies for the U.S. risk-free rate. The U.K. does not have a liquid Treasury bill market.

for monthly and one-year horizons, that date was substituted for the original target date. If no sufficiently close options trading date existed, that expiry was excluded from the sample for that horizon.

Options quotes for the target dates were then filtered. Because trading in options markets is asymmetrically concentrated in at- and out-of-the-money strikes, and because the spline algorithm will not accommodate duplicate strikes in the data, we discard in-the-money options. Options for which it was impossible to compute an implied volatility (usually far-away-from-the-money options quoted at their intrinsic value) or options with implied volatilities of greater than 100 percent were also discarded. If there were fewer than five remaining usable strikes in a given cross-section, the entire cross-section was discarded. Table I presents the resulting cross-section counts and the range and mean of the strikes per cross-section of the remaining data. In practice, too few cross-sections leads to insufficient power to conduct meaningful tests. Horizons greater than two months were found to have too few usable cross-sections for our study.

Another problem we encounter is overlapping data. Serial options, those with expiries of less than three months, expire at monthly intervals. Forecasts and realizations for horizons of less than or equal to one month can therefore be reasonably expected to be independent from one observation/realization interval to the next since the intervals share no common innovations in the price path of the underlying asset. However, for forecasts beyond a one month horizon the time paths from forecast date to realization for consecutive forecasts begin to overlap and thus contain some common innovations in the price path of the underlying asset. Since our tests, see below, are predicated on the null hypothesis that the data are independent, serial dependence arising from overlapping observations undermines the informativeness of the tests. The actual degree of the problem is an empirical question, which we test for. Overlapping data produced serious autocorrelation problems for maturities longer than five weeks. Our final sample therefore consists of filtered cross-sections for weekly horizons of between one and six weeks, with the six-week included to show the effects of overlapping data.<sup>13</sup>

---

<sup>13</sup>Serial dependence arising from overlapping data could be addressed by increasing the sampling to, for instance, quarterly expiries. This, however, reduces the sample size to the point that tests lack sufficient power to draw meaningful conclusions. In the presence of four- and five-

## II. Methodology

Our approach to studying the risk premium implicit in options prices involves looking at the ability of risk-neutral and risk-adjusted or subjective PDFs to forecast future realizations of the underlying asset. Our assumption is that investors are rational and perhaps risk-averse. If we were interested only in point forecasts this would mean that the degree of bias in the price forecast could be interpreted as an indication of the degree of market risk aversion, provided the bias is of the correct sign, rather than an indication that investors are irrational.

In this study, we are interested in forecasts of distributions rather than of a single point estimate. We will therefore examine whether the realizations over time are consistent with the PDFs implicit in options prices observed at some horizon prior to the respective realizations. Option prices embed risk-neutral PDFs. If these risk-neutral PDFs provide good forecasts of the distribution of future realizations, then we must conclude that there is no evidence of risk premia in the pricing of options. On the other hand, if risk-neutral PDFs are not good forecasters, we can test whether risk-adjusted PDFs provide better forecasts. If this is the case, the relative risk aversion of the utility function used to adjust the risk-neutral PDF provides a measure of the degree of risk aversion of the representative investor.

To execute our study we need to be able to:

1. Compute risk-neutral PDFs from option prices,
2. Test the forecast ability of PDFs, both risk-neutral and subjective, and
3. Adjust risk-neutral PDFs to derive subjective PDFs.

### *A. Estimating the Risk-Neutral Probability Density Function*

Breeden and Litzenberger (1978) showed that the PDF for the value of the underlying asset at option expiry,  $f(S_T)$ , is related to the European call price function by

$$f(S_T) = e^{r(T-t)} \frac{\partial^2 C(S_t, K, T, t)}{\partial K^2} \Big|_{K=S_T}, \quad (2)$$

---

weeks horizons, the one-month horizon is redundant. Results for one-month are consistent with the patterns found in the weekly horizon data.

where  $S_t$  is the current value of the underlying asset,  $K$  is the option strike price, and  $T-t$  is the time to expiry. Unfortunately, available option quotes do not provide a continuous call price function. To construct such a function, we must fit a smoothing function to the available data.

In this paper, we employ a refinement of the smoothed implied volatility smile method developed by Panigirtzoglou and presented in Bliss and Panigirtzoglou (2002).<sup>14</sup> The essence of the Panigirtzoglou and related methods is to smooth implied volatilities rather than option prices and then convert the smoothed implied volatility function into a smoothed price function, which can be numerically differentiated to produce the estimated PDF.

The Black-Scholes formula is used to extract implied volatilities for European options (FTSE 100) and the Barone-Adesi and Whaley (1987) formula is used for American options (S&P 500). At the same time, strike prices are converted into deltas using the Black-Scholes delta and the appropriate at-the-money implied volatility, thus producing a series of transformed raw data points in implied volatility/delta space.<sup>15</sup> It is important to note that the use of the Black-Scholes and Barone-Adesi-Whaley formulae is solely to convert data from one space (price/strike) to another (implied volatility/delta) where smoothing can be done more efficaciously. Doing so does not presume that either formula correctly prices options.

A weighted natural spline is used to fit a smoothing function to the transformed raw data. The natural spline minimizes the following function:

$$\min_{\theta} \sum_{i=1}^N w_i (IV_i - IV(\Delta_i, \theta))^2 + \lambda \int_{-\infty}^{\infty} g''(x; \theta)^2 dx, \quad (3)$$

---

<sup>14</sup>Numerous methods have been developed for extracting PDFs from option prices. Bliss and Panigirtzoglou (2002) provide a review of many of these. The Panigirtzoglou method itself derives from previous work as discussed in Bliss and Panigirtzoglou. The Panigirtzoglou method was selected for this paper because Bliss and Panigirtzoglou found it to be relatively robust and the method permits calibrating the desired smoothness of the extracted PDF.

<sup>15</sup>Smoothing the implied volatilities in delta-space rather than strike-space was introduced by Malz (1997). Moreover using delta rather than strike groups the away-from-the-money implied volatilities more closely together than near-the-money. This permits a greater variation in shape near the center of the distribution where the data are more reliable. Malz shows that this achieves better results than smoothing in strike space.

where  $IV_i$  is the implied volatility of the  $i$ th option in the cross-section;  $IV(\Delta_i, \theta)$  is the fitted implied volatility which is a function of the  $i$ th option's delta,  $\Delta_i$ , and the parameters,  $\theta$ , that define the smoothing spline,  $g(x; \theta)$ ; and  $w_i$  is the weight applied the  $i$ th option's squared fitted implied volatility error. In this paper we use the option vegas,  $v \equiv \partial C / \partial \sigma$ , to weight the observations.<sup>16</sup> The parameter  $\lambda$  is a smoothing parameter that controls the tradeoff between goodness-of-fit of the fitted spline and its smoothness measured by the integrated squared second derivative of the implied volatility function. In our preliminary tests we used values of  $\lambda$  ranging from 0.99 to 0.9999 to check the sensitivity of our results to the degree of smoothness we impose on the estimated PDF. These tests indicated that forecast results were insensitive to the choice of  $\lambda$ . We therefore report results based on  $\lambda = 0.99$ .

When fitting a PDF it is necessary to extrapolate the spline beyond the range of available data.<sup>17</sup> Since we rarely observe extreme realizations of the underlying asset (outcomes beyond the range of available strikes at horizons of less than two months), we have little information as to the appropriate shape to impose on the tails of the density function. Fortunately, the scarcity of tail outcomes also means that the results are not critically dependent on the choice so long as it is economically reasonable. The natural spline is linear outside the range of available data points and can thus result in negative or implausibly large positive fitted implied volatilities. To prevent this happening, we force the spline to extrapolate smoothly in a horizontal manner. We do this by introducing two pseudo-data points spaced three strike intervals<sup>18</sup> above and below the range of strikes in the cross-sections and having implied volatilities equal to the implied volatilities of the respective extreme-strike options. These pseudo-data points are added to the cross-sections before the above transformations and spline-fitting take place. Extrapolating the

---

<sup>16</sup>Vega weighing is consistent with homoscedastic pricing errors, such as those resulting from discrete tick size. Furthermore, vega weighing places less weight on away-from-the-money strikes, which is also consistent with observed lower liquidity of away-from-the-money options.

<sup>17</sup>Anagnou et al. (2002) use PDFs truncated to the range of available strikes and then rescaled. This unusual procedure avoids extrapolating the tails of the PDF, but cannot handle realizations falling outside the range of strikes available when the PDF was constructed.

<sup>18</sup>Strike intervals refers to the interval between adjacent quoted option strikes.

implied volatility function in this manner has the effect of smoothly pasting log-normal tails onto the implied density function beyond the range of strikes.

Once the spline,  $g(x; \theta)$ , is fitted, 5,000 points along the function are converted back to price/strike space using the Black-Scholes formula. The delta-to-strike conversion uses the same at-the-money implied volatility used for the earlier strike-to-delta conversion, thus preserving the consistency in the initial data transformation and its inverse. The implied volatility-to-call price conversion uses the implied volatility provided by the fitted implied volatility function to produce a fitted European call price function. The 5,000 points are selected to produce equally spaced strikes over the range where the PDF is different from zero. This range varies with each cross-section, primarily as the price level of the underlying changes. Finally, we use the 5,000 call price/strike data points to numerically differentiate the call price function to obtain the estimated PDF for the cross-section.

### *B. Testing PDF Forecast Ability*

Each option cross-section produces an estimated PDF,  $\hat{f}_t(\cdot)$ , for a single option expiry date. Our goal is to test the hypothesis that the estimated PDFs,  $\hat{f}_t(\cdot)$ , are equal to the true PDFs,  $f_t(\cdot)$ . The time-series of PDFs generated for a given forecast horizon are all different. Only one realization,  $X_t$ , is observed for each option observation/expiry date pair.

Under the null hypothesis that the  $X_t$  are independent and that estimate PDFs are the true PDFs; that is,  $\hat{f}_t(\cdot) = f_t(\cdot)$ , the inverse probability transformations of the realizations,

$$y_t = \int_{-\infty}^{X_t} \hat{f}_t(u) du, \quad (4)$$

will be independently and uniformly distributed:  $y_t \sim \text{i.i.d. } U(0,1)$ .<sup>19</sup> The range of the transformed data is guaranteed by the inverse probability transformation itself, but the uniformity need obtain only if the estimated PDF equals the true PDF. Independence should also be established as most distributional tests assume independence and would generate incorrect inferences if this were not the case, though independence is not always verified in practice.

Several non-parametric methods have been proposed for testing the uniformity of the inverse probability transformed data, including the Kolmogorov-Smirnov, Chi-squared, and Kupier tests. None of these methods provides a joint test of the assumption that the  $y_t$  are i.i.d.

Berkowitz (2001) has proposed a parametric methodology for jointly testing uniformity and independence. He first defines a further transformation,  $z_t$ , of the inverse probability transformation,  $y_t$ , using the inverse of the standard normal cumulative density function,  $\Phi(\cdot)$ :

$$z_t = \Phi^{-1}(y_t) = \Phi^{-1}\left(\int_{-\infty}^{x_t} \hat{f}_t(u) du\right), \quad (5)$$

under the null hypothesis,  $\hat{f}_t(\cdot) = f_t(\cdot)$ ,  $z_t \sim \text{i.i.d. } N(0,1)$ . Berkowitz tests the independence and standard normality of the  $z_t$  by estimating the following model

$$z_t - \mu = \rho(z_{t-1} - \mu) + \varepsilon_t, \quad (6)$$

using maximum likelihood and then testing restrictions on the estimated parameters using a likelihood ratio test.<sup>20</sup> Under the null, the parameters of this model should be:  $\mu = 0$ ,

---

<sup>19</sup>Kendal and Stuart (1979), section 30.36, discusses the case where the  $X_i$  are i.i.d. and the estimated densities do not depend on the  $X_i$ . Where the estimated densities do depend on the  $X_i$ , problems may ensue and the inverse probability transform need not be independent or uniform. Diebold, Gunther, and Tay (1998) show that for a special case (arising in GARCH processes) where the true densities depend only on past values of  $X_i$  (and no other conditioning information), the i.i.d. uniform result holds. However, in the problem addressed in this paper, the PDFs are estimated from option prices and values of the underlying, which do not include the  $X_i$ . We therefore rely on Kendal and Stuart.

<sup>20</sup>The log-likelihood function for this model is given in Hamilton (1994), equation (5.2.9). This test does not test the normality of the transformed data per se, but rather that the data are standard normal under the assumption that they are normally distributed.

$\rho = 0$ , and  $\text{Var}(\varepsilon_t) = 1$ . Denoting the log-likelihood function as  $L(\mu, \sigma^2, \rho)$ , the likelihood ratio statistic,  $\text{LR}_3 = -2[L(0, 1, 0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})]$ , is distributed  $\chi^2(3)$  under the null hypothesis.

In practice, it is sometimes necessary to test overlapping forecasts, for example two-month-ahead forecasts of monthly realizations. In this case, if the above test rejects, it is possible that the rejection arises from the overlapping nature of the data, which may induce autocorrelation, rather than from problems with the estimated PDFs. This is also true for non-overlapping, but serially correlated, data. Berkowitz (2001) therefore tests the independence assumption separately by examining  $\text{LR}_1 = -2[L(\hat{\mu}, \hat{\sigma}^2, 0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})]$ , which has a  $\chi^2(1)$  distribution under the null.

If  $\text{LR}_3$  rejects the hypothesis that the  $z_t \sim \text{i.i.d. } N(0, 1)$ , failure to reject  $\text{LR}_1$  provides evidence that the estimated PDFs are not providing accurate forecasts of the true time-varying densities. On the other hand, if both  $\text{LR}_3$  and  $\text{LR}_1$  reject, we cannot determine whether the problem arises from a lack of forecast ability or serial correlation in the data. Failure to reject both  $\text{LR}_3$  and  $\text{LR}_1$  is consistent with forecast power, though as in all statistical tests failure to reject the null hypothesis does not necessarily mean that the null hypothesis is true.

The simple AR(1) model used in the above Berkowitz test captures only a specific sort of serial dependence in the data, though this is the dependence most likely to occur in this case. Berkowitz (2001) shows how to expand the model and associated tests to higher order AR(p) processes. However, this results in increasing numbers of model parameters and reduced power.

The LR test is uniformly most powerful only in a single-sided hypothesis test. However, as we show in Appendix A in Monte Carlo simulations, the Berkowitz test is more reliable than the Chi-squared and Kupier tests in large and small samples under the null hypothesis, and is additionally superior to the Kolmogorov-Smirnov test in small samples when the data are autocorrelated because it is a joint test of uniformity and independence. We therefore use the Berkowitz test in this paper.



### C. Estimating the Subjective Density Function

To compute and test the forecast ability of a subjective density function, it is first necessary to hypothesize a utility function for the representative agent and then, following Ait-Sahalia and Lo (2000), use this to convert the estimated risk-neutral density function into a subjective density function. The forecast ability of the resulting subjective density function is then tested in the same manner as the risk-neutral density function.

Given an estimated risk-neutral density function and a utility function, equation (1) can be transformed to solve for the implied subjective density function. The resulting subjective density function must be normalized to integrate to one. Thus,

$$p(S_T) = \frac{\frac{q(S_T)}{\zeta(S_T; S_t)}}{\int \frac{q(x)}{\zeta(x; S_t)} dx} = \frac{\frac{U'(S_t)}{\lambda U'(S_T)} q(S_T)}{\int \frac{U'(S_t)}{\lambda U'(x)} q(x) dx} = \frac{\frac{q(S_T)}{U'(S_T)}}{\int \frac{q(x)}{U'(x)} dx}. \quad (7)$$

In the normalization process the parameter  $\lambda$  disappears, however any parameters of the utility function itself must be estimated.

In this paper, we test subjective density functions derived using two representative agent utility functions: (1) the power utility function and (2) the exponential utility function. In both cases the utility functions, and thus the resulting subjective density functions, are conditioned on the value of the single parameter,  $\gamma$ . In testing the subjective density functions, we first selected the value of  $\gamma$  to maximize the forecast ability of the resulting subjective PDFs by maximizing over  $\gamma$  the p-value of the Berkowitz LR<sub>3</sub> statistic.

Table II provides the functional forms of the power and exponential utility functions and the marginal utility function used to transform the risk-neutral density into the corresponding subjective density, together with the measure of relative risk aversion (RRA) for each utility function. The power utility function has constant relative risk aversion, and the measure of RRA is simply equal to the parameter  $\gamma$ . However, the exponential utility function exhibits constant absolute risk aversion, the parameter  $\gamma$ , rather than constant relative risk aversion. For exponential utility, the RRA is dependent

on both  $\gamma$  and the realization  $S_T$ , which is time varying. Therefore, for exponential utility RRA, we report the distribution of the RRA across the sample observations.

#### *D. Significance of the RRA Estimates*

The Berkowitz likelihood ratio statistic has a  $\chi^2_3$  distribution for a fixed  $\gamma$ . However, the process of searching for the optimal level of  $\gamma$  alters the distribution of the test statistic, biasing the likelihood ratio towards unity and thus overstating the p-value. (This does not affect the tests of the risk-neutral densities.) Furthermore, the process of maximizing the Berkowitz statistic over values of  $\gamma$  does not provide a measure of whether the resulting  $\gamma$  is significantly different from zero, only whether the resulting adjusted PDFs can be rejected as correct forecasts of the distribution of future outcomes. Therefore, we need to find a way of correcting the Berkowitz p-values and determining the significance of the estimated values for  $\gamma$ . Unfortunately, the  $\gamma$ -estimation methodology is complex and does not lend itself to simple analysis. Certainly analytic expressions for standard errors are impossible as the likelihood depends on empirical rather than analytical PDFs.

In such situations, resampling and Monte Carlo methods provide straightforward approaches to investigating the properties of the estimation procedure and the resulting estimates. We use Monte Carlo to investigate the properties of our methodology under the null hypothesis that the representative investor is risk-neutral. To determine whether the properties of the estimator vary with the level of  $\gamma$ , we also conduct the same tests under assumed levels of risk aversion comparable to those we estimate using actual data. Bootstrap and cross-validation yield evidence of the sampling variations of the estimated RRAs using actual, rather than simulated data, and thus provide a cross-check of the Monte Carlo results. Cross-validation also provides a systematic test for (single) influential observations.

##### *D. 1. Monte Carlo Tests*

The Monte Carlo tests use the same risk-neutral PDFs employed in our other tests. However, actual outcomes are replaced with pseudo-outcomes repeatedly drawn from

risk-neutral PDFs. This process produced a set of outcomes where the true value of  $\gamma$  was known (that is, zero).<sup>21</sup> One thousand replications were employed for each contract type/expiry-horizon combination. For each simulated set of outcomes, the Berkowitz p-value-maximizing value of  $\gamma$  was estimated as described in the previous section. The resulting distributions of p-values and  $\gamma$ s give us an idea of the standard error of the estimates which our methodology produces under known conditions.

Figure 1 presents the distribution of the p-values and  $\gamma$ s from simulations based on risk-neutral PDFs by horizon, utility-adjustment method, and contract type. The box portion of each Tukey plot encompasses the inter-quartile range of the estimates (p-values or  $\gamma$ s) derived from the Monte Carlo simulations. The lines dividing the box are the mean (dotted) and median (solid), respectively. The “whiskers” are the 10th and 90th percentiles of the distributions and the end points correspond to the 5th and 95th percentiles.

The top four panels of Figure 1 present the distribution of the p-values produced by the simulations under the assumption that the Berkowitz statistic  $LR_3$  statistic is distributed  $\chi^2_3$ . If the distribution of the Berkowitz  $LR_3$  produced by the search process which maximized the  $LR_3$  statistic over values of  $\gamma$  was indeed  $\chi^2_3$ , the various percentiles of the distributions of the simulated p-values would appear at the corresponding percentiles indicated on the vertical axes: Thus, the mean and median should coincide at 0.5, and the box should extend from 0.25 to 0.75, and so forth. This is clearly not the case; the distribution of Monte Carlo generated p-values show significant upward bias, and the distribution is no longer uniform or even symmetric. Therefore, we cannot evaluate the Berkowitz  $LR_3$  statistic derived from the utility-adjusted PDFs under the assumption that it is distributed  $\chi^2_3$ . However, we can use the Monte Carlo distribution of p-values to adjust the  $\chi^2_3$  p-values estimated using actual data—we use the

---

<sup>21</sup>In preliminary Monte Carlo experiments, we found that the distributions of p-values and  $\gamma$  estimates are virtually identical when outcomes are drawn from risk-neutral PDFs and from subjective (utility function adjusted) PDFs constructed using fixed values for  $\gamma$ , excepting for the expected mean shift in the estimated values of  $\gamma$ . This finding was confirmed using rank sign tests for differences in paired distributions. We therefore base our analysis on Monte Carlo simulations using risk-neutral PDFs.

frequency with which the Monte Carlo p-values fall short of our actual-data p-values.<sup>22</sup> Using the one-week horizon power utility-adjusted simulations as an example, a p-value of less than 0.5 occurs 30.5 percent of the time. We therefore adjust the p-value for the Berkowitz likelihood ratio statistic downwards from the 0.5 it has under the incorrect  $\chi^2_3$  distribution to the 0.305 we find in the Monte Carlo simulations under the null hypothesis that the outcome data are drawn from the risk-neutral PDFs.<sup>23</sup>

The distributions of simulated  $\gamma$  estimates, by horizon, utility-adjustment method and contract type, produced by the same simulations are shown in the bottom four panels of Figure 1. The distributions show a slight, but not significant, positive bias (recall that the true value for  $\gamma$  is zero). The biases are also small compared to the values of  $\gamma$  estimated from actual outcome data (vide infra). We can use these distributions of  $\gamma$ -values obtained under the assumption that  $\gamma = 0$ , to provide a rough proxy for the statistical significance of the actual-data  $\gamma$  estimates against the null hypothesis that the true value of  $\gamma$  is zero. That is, we can test whether the differences between the risk neutral and utility-adjusted PDFs are significant, even when tests of the forecast ability based on the p-values of the LR<sub>3</sub> statistic fail to reject the hypothesis that the risk-neutral PDFs are “good” forecasts of the distribution of future outcomes of the underlying index.<sup>24</sup>

#### *D. 2. Bootstrap Tests*

Monte Carlo provides us with information about the distribution of test results when the data are drawn from a known model—that is when the null hypothesis is indeed

---

<sup>22</sup>Because the mapping from the LR<sub>3</sub> statistic to the corresponding  $\chi^2_3$  p-value is monotonic, we can work with the Monte Carlo distribution of either statistic to derive the adjusted p-value for our empirical tests.

<sup>23</sup>Because there is a monotonic mapping from the Berkowitz likelihood ratio statistic to the  $\chi^2_3$  p-value, adjusting the p-value using the distribution of simulated p-values is equivalent to computing the adjusted p-value of the Berkowitz statistic directly from the distribution of the simulated likelihood ratios.

<sup>24</sup>Efron and Tibshirani (1993) argue that precise estimation of confidence intervals (at say the 95 percent level) requires several thousand replications since the relevant tail outcomes are by definition infrequent. This is impractical in this instance. Therefore, while a few extreme outcomes in 1,000 replications is suggestive of an extremely low probability, we do not wish to assert that the exact confidence level has been determined.

true. Complex estimation methodologies involving non-linear optimization techniques which appear to be well behaved under the assumed data structure can sometimes become ill behaved under actual data which invariably differs from the null hypothesis to some degree. It is important, therefore, to confirm the Monte Carlo results with actual data, to show that the sampling variation of the estimator is not materially different than the sampling variation derived from the Monte Carlo simulations (particularly if the Monte Carlo results are to be used to infer significance in later tests). Bootstrapping captures the impact of the actual data and potential model misspecification on the reliability of parameter estimates. We applied the bootstrap using the two representative contracts—five-week S&P 500 and three-week FTSE 100—again with 1,000 replications in each case. Each replication consisted of drawing with replacement a random sample of pairs of densities and associated outcomes from the original sample. Each bootstrap sample was the same size as the original samples (183 and 108, respectively). Each bootstrap sample was then used to estimate power- and exponential-utility  $\gamma$ s and p-values. Since bootstrapping destroys the independence assumption underlying computation of p-values, the distribution of bootstrap p-values is uninformative. However, the distribution of  $\gamma$ s, or equivalently RRAs, provides an indication of the sampling variation of these estimates when the full sample is used. We report the distribution of RRAs in Table III, rather than  $\gamma$ s, to facilitate comparison across utility functions. RRAs and  $\gamma$ s are identical in the case of the power utility function, and differ by a fixed scalar (the average level of the index at expiry) in the case of exponential utility.

The standard deviations of the bootstrap  $\gamma$ -estimates are comparable to the standard deviation of the Monte Carlo  $\gamma$ -estimates (last row of Table III), providing additional support for using the Monte Carlo  $\gamma$  distributions to estimate the significance levels of the actual-data  $\gamma$ -values. If used to compute t-statistics for the observed full-sample estimated values (top row of Table III), these standard deviations suggest significance levels exceeding conventional levels against the null hypothesis that the true RRAs are zero. This is confirmed in the low incidence of resampled datasets that produce estimated RRAs of less than zero.

### D.3. Cross-Validation Tests

Cross validation is useful in checking for sampling error and individual influential observations. Figure 2 presents the results of cross validations of the four-week S&P 500 contract and the three-week FTSE 100 contract. The FTSE 100 shows much greater sensitivity to the data than does the S&P 500 contract. This is not surprising as the S&P 500 data set has a larger number of observations and therefore dropping a single data point will naturally have less effect. In both cases no single data point stands out as an outlier, and the sampling variation in the estimated RRAs is consistent with the true RRA being different from zero. In addition, we checked to see if the relative performance of the two utility function adjustments was sensitive to individual outliers. The superiority of the exponential utility adjustment over the power utility adjustment, evidenced by a higher p-value, was not altered by dropping any individual observation.

## III. Empirical Results

The analysis of the empirical results consists of three sequential steps. We first examine the risk-neutral PDFs to determine whether there is evidence that they adequately capture the distribution of *ex post* realizations. We next risk adjust the risk-neutral PDFs and then test these subjective PDFs in the same manner. Conditional on the subjective PDFs providing a better forecast of the distributions of future realizations, we examine the measures of RRA implicit in these risk-adjusted PDFs.

Table IV provides the evidence on our first two questions. We cannot reject the hypothesis that the risk-neutral PDFs provide accurate forecasts of the distributions of future realizations for the FTSE 100 contracts at the one-week horizon. With a p-value of 23 percent, we find no support for the hypothesis that the one-week horizon risk-neutral PDFs forecast the FTSE 100 densities poorly. However, in the remaining 13 cases the Berkowitz test rejects the hypothesis that risk-neutral PDFs are good forecasts of the distribution of future values of the underlying index. With the exception of the two six-week horizon contracts, we cannot reject the hypothesis that the probability integral transforms are uncorrelated (by examining the p-values of the  $LR_1$  statistic). This leads us to conclude that the rejection of the null hypothesis of good forecast ability arises from a poor forecast rather than a violation of the independence assumption underlying the test

statistic. This result is consistent with our priors that risk-neutral PDFs are unlikely to provide adequate forecasts—there is simply too much evidence that equity markets price risk. This result confirms the evidence found in most previous studies.

These results also demonstrate that the Berkowitz test has sufficient power to reject the good-forecast null. This observation becomes important when we examine the forecast ability of the risk-adjusted PDFs and find very different results. Having previously established that our tests are able to reject in the risk-neutral case, we are more secure in interpreting the failure of the same test to reject in the subjective cases as arising from superior performance of the risk-adjusted PDFs rather than lack of power in our test methodology.

The second stage of our analysis asks whether power and/or exponential utility-adjusted PDFs can improve the forecast ability in the sense of producing PDFs that can no longer be clearly rejected as good forecasts of the distribution of future values of the underlying index. We begin with the six-week contracts. In all four cases—FTSE/S&P; power/exponential adjusted—the subjective PDFs continue to be strongly rejected as good forecasts. However, when we examine the tests for autocorrelation, the  $LR_1$  p-values, we reject the null hypothesis that the underlying probability integral transforms are uncorrelated. This is consistent with the overlapping nature of the data—six-week-ahead forecasts of monthly observations. This autocorrelation may or may not be driving the rejection of the  $LR_3$  statistics. In this situation no inference can be drawn from the rejection of the  $LR_3$  statistic. This limitation applies equally to the rejection of risk-neutral and subjective PDFs at the six-week horizon.<sup>25</sup> For the remaining horizons, the  $LR_1$  statistic p-values all fail to be rejected, almost always by comfortable margins.

At the two-week horizon for both contract types and both utility functions, the subjective PDFs are rejected as good forecasts. The same is true for the power utility-adjusted PDFs at the five-week horizon for both contract types, and the exponential utility-adjusted PDFs barely fail rejection at the five percent level. The power-utility-adjusted PDFs for the one-week S&P 500 contract are also rejected, but the exponential-utility-adjusted PDF cannot be rejected as providing good forecasts. For the one-week

---

<sup>25</sup>The overlapping horizon problem is even more severe for longer horizons. We include the six-week results in our analysis to illustrate the problem and the consequences, and exclude longer horizons because they are similarly uninformative.

FTSE 100 contract, neither of the subjective PDFs was rejected. For the remaining horizons, both power and exponential utility-adjusted PDFs fail to be rejected at conventional levels of significance. We conclude that overall adjusting risk-neutral PDFs using utility functions results in subjective PDFs that for the most part are reasonable forecasts of future distributions of the underlying index, in the sense that they cannot be rejected.

Casual inspection reveals that the power-utility adjusted PDFs are more likely to be rejected or be marginally significant when compared with the exponential-utility-adjusted PDFs. Inspection of the p-values reveals that in eight cases out of 12 the exponential-utility-adjusted PDFs produce a better goodness-of-fit than do the power-utility-adjusted PDFs, with equality obtaining in one of the remaining four cases. This invites a comparison of the relative forecast ability improvement provided by the two utility functions. However, direct comparison on a case-by-case (horizon/contract-type) basis is not possible. The two models are not nested and even if a parametric test could be constructed it is unlikely to have power in such small samples. We can however use non-parametric tests to check whether the apparent superiority of the exponential utility adjustment is significant in an overall sense, across contract types and horizons. The null hypothesis for such a test is that the two utility functions produce equally good forecasts on average, and that therefore we expect to see the exponential-utility-adjusted PDF p-values exceed the power-utility-adjusted PDF p-values approximately half the time, rather than eight of 12 times. Assessing the statistical significance of the eight of 12 outcomes result is however complicated by the fact that while the PDFs used to compute the probability integral transforms differ across horizons, the same realizations are involved for all horizons. This causes the probability integral transforms to be correlated across horizons. For this reason, we cannot apply a simple binomial distribution to determine the significance of the eight of 12 result.

To solve this problem, Monte Carlo simulations were used to determine the probability of observing eight instances of exponential utility PDFs achieving higher Berkowitz statistics out of 12 cases under the assumption that the data were drawn from identically distributed, but cross-horizon correlated data. To mimic the two parallel sets of inverse probability transforms resulting from adjusting the risk-neutral PDFs using



power and exponential utility functions, 14 paired sets (A and B) of uniformly distributed data were generated. To capture the correlation resulting from common expiry-date realizations of the underlying index, we imposed the same correlation structure on the pseudo-data as the actual inverse probability transforms had, and we matched the series lengths to the data. While replicating the actual inverse probability transforms in sample size and correlation structure, each of the paired sets of pseudo-inverse probability transforms were constructed to have identical distributions, so the null hypothesis that the pairs of Berkowitz statistics had the same expectation was true by construction. Pairs of Berkowitz statistics were then computed for each of the 12 pairs of constructed series, and the frequency of  $\text{Berkowitz}(A) > \text{Berkowitz}(B)$  was noted. The process was repeated 10,000 times. These simulations show that, given the correlation structure in the data, if the power- and exponential-utility adjustments were equally efficacious, we would observe one adjustment method beating the other (higher p-values) eight times out of 12 possibilities only 9.0 percent of the time. We can therefore reject at the 10 percent significance level the hypothesis that the power-utility function does equally as good a job as the exponential-utility function in improving the forecast ability of risk-neutral PDFs.<sup>26</sup> This indicates that, at least for this market, the exponential-utility function provides a better fit of the representative investor's utility function.<sup>27</sup> Many studies assume the representative investor has power utility because of the mathematical tractability of this utility function and the convenience of constant relative risk aversion. This finding suggests a tradeoff is involved in such a choice. Nonetheless, we will continue to examine both methods of risk adjustment in the proceeding analysis.

---

<sup>26</sup>Given the small sample size, 12 comparisons, a 10 percent level of significance is a reasonable threshold.

<sup>27</sup>The same analysis can be used to ask whether subjective PDFs improve forecast ability relative to risk-neutral PDFs on an overall basis. While the subjective PDFs do involve an additional parameter, this need not necessarily result in a higher bias-adjusted p-value. Indeed the estimated risk aversion parameter need not be positive, though it always is. The power-utility adjusted PDFs resulted in higher p-values than the risk-neutral PDFs in nine of 12 cases, and the exponential-utility function p-values were higher in 10 cases. The probabilities of these outcomes under the null hypothesis of "same forecast ability" are 8.8 and 7.5 percent, respectively. We can, therefore, conclude that even though the adjusted PDFs are occasionally themselves rejected as good forecasts, utility-adjusting risk-neutral PDFs does produce better overall forecasts of the distributions of the futures values of the underlying index.

The estimated values of  $\gamma$  for each contract type, horizon, and utility function are presented in Table V, along with an indication of their approximate levels of significance against a null hypothesis that the true value is zero based on the Monte Carlo experiments described in the previous section. Excepting only the longer-horizon FTSE cases, we can reject the hypothesis that the estimated gammas are equal to zero. Even in cases where the adjusted PDFs were rejected as good forecasts of the distribution of future values of the underlying asset, for instance both five-week power-utility-adjusted cases, the forecast performance maximizing values of  $\gamma$  were significantly different from zero. These results provide confirmation of the non-parametric tests in footnote 27 that showed that utility adjusting PDFs improves the overall forecast ability of risk-neutral PDFs.

The goal of our estimation methodology is to obtain estimates of the market's risk aversion, which we measure by the relative risk aversion of the representative investor. For the power utility function this is simply  $\gamma$  itself. For the exponential utility function, we need to multiply  $\gamma$  by the level of the underlying asset. The top panel of Table VI presents the "all observations" RRAs corresponding to the results just discussed.

There is close agreement across horizons and contract types between the power utility RRAs and the mean exponential utility RRAs. Furthermore, the RRAs for FTSE 100 and S&P 500 are similar for matched horizons.<sup>28</sup> This is not an artifact of the methodology as the samples are entirely distinct, and we see variation between RRAs for different horizons. The median exponential-utility RRAs are slightly lower than the mean, reflecting the positive skew in the distribution of index values. Thus, while the exponential-utility adjustment appears to produce somewhat superior forecasts of density, the (mean) measured RRAs are broadly consistent between the two risk adjustment functions. However, the range of RRAs permitted by the exponential-utility adjustment, which is quite substantial, coupled with the evidence of better fit, suggests that the constant relative risk aversion inherent in the power-utility adjustment may be unduly restrictive, and that constant absolute risk aversion (a characteristic of the exponential-utility function) seems to be more consistent with the data.

---

<sup>28</sup>This is true for the power utility and mean and median values for the exponential utility. The S&P exponential-utility RRA ranges are greater than the corresponding FTSE 100 ranges because of the greater range found in the values of the S&P index, which in turn arises from the longer time-series available for S&P data.

In all cases, the RRAs are consistent with the moderate values found in most of the other studies shown in Table VII. There is no evidence in our results of the extreme, “puzzle” values found in Mehra and Prescott (1985) and Cochrane and Hansen (1992). Our study thus adds to the accumulating evidence that the risk aversion of the representative agent is not always or necessarily extreme and therefore that the observed equity premium puzzle may be idiosyncratic.

These results, taken together, demonstrate that the risk premium adjustment needed to move from risk-neutral to objective densities is more subtle than a simple mean shift. Both utility functions produce nearly the same (average) measured levels of risk aversion and hence the same average shift in the means of the risk-neutral densities. Our finding of differential performance of power- and exponential-utility function adjustments demonstrates that the manner in which the entire density changes matters. This is consistent with Anagnou et al. (2002) who tested the predictive power of a mean-shifted log-normal density versus a power-utility-adjusted density and found that the latter did better. Even if crude adjustments to the mean of a density may eliminate a significant portion of the forecast error, to give economic content to the idea of a risk premium requires relating the necessary changes in the density function to preferences.<sup>29</sup>

Our methodology necessarily imposes the assumption that the utility function’s coefficient of risk aversion parameter,  $\gamma$ , is constant across the sample. With only one realization per observation it is difficult, if not impossible, to estimate time-varying values for these parameters. However, we can examine the robustness of this constant parameter assumption by dividing the sample into sub-samples and re-estimating the parameters for each sub-sample. Wald or similar tests of differences across sub-samples on a case-by-case basis are unlikely to be statistically significant, recalling that the data did not reject the full sample risk-adjusted PDFs. However, the patterns across horizons and markets of sub-sample differences are consistent and instructive.

Rather than divide the two samples by time-period, we elected to divide each into two equal-sized sub-samples corresponding to periods of high and low volatility as measured by the implied volatility of at-the-money options. Our rationale is that risk

---

<sup>29</sup>In any case, an empirical density function estimated using non-parametric methods cannot be simply “moved over” in order to shift the mean without changing the range over which the PDF is defined.

aversion is more likely to vary with the degree of risk than with time. The middle and lower panels in Table VI present the RRAs measured over these two sub-samples. The results are marked and consistent. For every horizon, and for both FTSE 100 and S&P 500, the low-volatility RRAs exceed the high-volatility RRAs by a factor of approximately three to five times. Taken as a whole and applying the Monte Carlo-based non-parametric test described above, these differences are statistically significant at the one percent level if we consider all horizons and both markets and at the 5.5 percent level if we consider only horizons of three through five weeks.

This volatility dependence of the risk aversion measure can also be observed in equity markets. For instance, under assumptions similar to those employed earlier, it can be shown that equity returns should be roughly proportional to changes in the volatility of returns, where the constant of proportionality equal the representative agents' RRA:<sup>30</sup>

$$\frac{dP_t}{P_t} \approx -\gamma d\text{Var}(r_t). \quad (8)$$

Table VIII presents the slope coefficients obtained by regressing the daily percentage changes in the S&P 500 index against changes in the variance of the index computed using the CBOE's implied volatility index. From the above model the slope equals the negative of  $\gamma$ , which under the assumed power-utility function equals the RRA.<sup>31</sup> Again we see that periods of high volatility correspond to low estimates of RRA. The full sample results are dominated by a few observations where implied volatility spiked sharply with little change in the S&P 500 index. Dropping these outliers reduces but does not eliminate the disparity. This provides independent confirmation of inverse relation between the volatility and risk aversion found in Table VI using different data and a completely different method of estimating RRA.<sup>32</sup>

---

<sup>30</sup>See Appendix B for the derivation of this result.

<sup>31</sup>The order of magnitude of these values differs from those in Table VI due to differences in scaling of the input data.

<sup>32</sup>A few other studies have examined the question of time-varying risk aversion. Rosenberg and Engle (2002) find evidence that measured risk aversion is correlated to macroeconomic factors. Guo and Whitelaw (2001) on the other hand find no evidence that risk aversion is time varying. Normandin and St-Amour (1998) find that taste shocks effect risk premia but that market risk does not. Han (2002) finds that the market risk premium is negatively related to volatility risk.

A possible explanation for this inverse relation between equity risk and measures of risk aversion lies in our proxy for consumption risk. If consumption risk is more stable than equity risk, as seems likely, then periods of high equity volatility will overstate consumption risk and the representative investor will appear correspondingly less risk averse.<sup>33</sup> Similarly, when equity volatility is low it will more closely approximate consumption volatility, and the representative investor will appear to be correspondingly more risk averse. Thus, use of equity returns as a proxy for consumption may induce the observed volatility dependence in the derived measures of risk aversion. An alternative explanation is that the representative investor is changing systematically as volatility changes. This might happen if during periods of high volatility the more risk-averse investors left the market resulting in a lower average level of risk aversion amongst the remaining investors. The final hypothesis is that in fact the representative investor has volatility-dependent risk aversion.

The first, the model-error-due-to-proxy hypothesis, suggests that the higher risk aversions measured during periods of low volatility are closer to the mark. The second hypothesis suggests potential problems that arise in aggregating across investors with dissimilar risk aversions. If a time-varying mix of market participants changes the characteristics of the representative agent—that is the representative investor himself changes—then time-invariant representative investor models may be insufficiently rich to capture the empirical regularities.<sup>34</sup> The last hypothesis suggests a more fundamental problem of identifying the link between volatility and risk aversion. While it is not inconceivable that a representative investor might become more risk averse as risk declines, intuition suggests exactly the opposite. Deciding amongst these hypotheses is beyond the scope of this paper. Our methodology is necessarily wedded to the use of

---

<sup>33</sup>Numerous studies have documented that equity volatility exceeds consumption volatility and that equity volatility is itself volatile. We know of no study, however, of the volatility of consumption volatility. Given the comparative long sampling intervals of consumption data and resulting few observations to work with, this is not surprising. Our conjecture depends only on the assumption that during periods of high(low) equity volatility the differences between equity volatility and consumption volatility increase(decrease), which seems reasonable.

<sup>34</sup>Constantinides' (2002) criticism of pricing models based on aggregate rather than idiosyncratic consumption is another challenge to the heretofore dominant representative investor paradigm.

equity risk to proxy for consumption risk, and the development of theoretical pricing models to aggregate time-varying heterogeneous agents or a single representative agent with volatility-dependent risk aversion is apt to prove a challenge to theoreticians. Nonetheless, our volatility-dependent risk aversion estimates, confirmed as they are in stock returns, provide an intriguing insight into the determinants of asset prices.

Returning to Table VI, the RRAs estimated over the full samples are generally declining with the forecast horizon, excepting the consistently anomalous two-week horizon. If we focus on the three to five week results, which show the clearest contrast between risk-neutral and risk-adjusted forecast performance, the RRAs decline by a factor of slightly less than two over that range of forecast horizons. In the at-the-money implied volatility-based sub-samples, the overall tendency of the estimated RRAs to decline with increasing horizon persists, though it is less consistently monotonic. Again errors resulting from the use of equity index to proxy for wealth, and indirectly consumption, are a possible explanation. Excess volatility in equity prices may attenuate at longer horizons, resulting in a better proxy. This plausible explanation is unfortunately inconsistent with the fact that over the one- to six-week range of horizons under consideration the observed annualized volatility is essentially flat. An alternative explanation is that skewness is priced and we do indeed see that annualized skewness declines over the range of horizons in this study. Lastly, the strong horizon-dependence in estimated RRAs may result from short-horizon investors being more risk-averse. Long-horizon investors can take steps to recover from adverse shocks, including smoothing consumption and increasing non-investment income (working harder), while short-horizon investors may have less flexibility. It is therefore plausible that risk aversion may be horizon dependent.

Our full-sample results show that, at least for forecast horizons of three to five weeks, the risk-neutral distributions provide poor forecasts of future densities while the subjectively adjusted densities provide reasonably good (i.e., not statistically rejectable) forecasts. An obvious question is how much the risk-neutral and subjective densities differ. One measure of this is to look at the tail percentile points under the risk-neutral and subjective distributions. The estimation of tail-percentile points is of particular importance in risk management where value-at-risk is widely used. Suppose we were to

compute the one-percentile value under the (rejectable) risk-neutral density forecast each period. These values of the underlying will have different percentile values under the (not rejectable) subjective densities, and the corresponding subjective percentiles of the risk-neutral one-percentile values may vary from observation to observation. For instance, at the two-week horizon, the values of the FTSE 100 corresponding to the one-percentile of the risk-neutral density measured each observation period have subjective cumulative probabilities (percentiles) ranging between 0.2 and 0.8 percent for the power-utility-adjusted densities and between 0.2 and 0.9 percent for the exponential-utility-adjusted densities.<sup>35</sup> In all cases, specific loss levels have lower probabilities under the utility-adjusted densities than under the risk-neutral densities. Thus, reliance on risk-neutral densities to estimate and hold capital against a one percent value-at-risk would be unduly conservative (and expensive) for long equity positions, and would understate the risk and required capital for short positions. Whether these differences are material depends on the particular application. These differences may be economically unimportant for an unlevered equity portfolio, while for a highly levered or equity derivative portfolios these differences could be crucial to the sound management of risk.

These are, of course, average results. The high-low implied volatility results presented in Table VI show that the reliance on risk-neutral densities would be less problematical during periods of high volatility and more problematical during periods of low volatility.

The difference between the mean of the risk-neutral and the subjective PDFs, normalized by one of the means (we use the risk-neutral PDF mean), is an approximate measure of the equity risk premium. Figure 3 plots the time-series of the four-week horizon risk premia for the S&P 500 contract. The same data is presented in both panels with differing scales for clarity. Until 1997, the exponential-utility-estimated risk premium was less than that estimated using a power-utility adjustment. Since 1997, this relation has been reversed. Changes in the risk premia appear to be correlated across risk-

---

<sup>35</sup>The differences between the mean two-week horizon FTSE 100 risk-neutral one-percentile point (3,975) and the corresponding power and exponential-utility-adjusted one-percentile points (4,010 and 4,015) is a small percentage of the mean level of the index. However, when compared to the mean absolute change in the index level over the two-week horizon (85), the one-percentile point differences (35/45) are large. Comparisons for other horizons/contracts are similar.

adjustment methods, as one would expect. However, differences in estimated risk premia can be large. For instance, during the 1987 stock market crash, the power-utility-adjusted PDF suggested a risk premium nearly three times as large as that estimated using an exponential utility function to adjust PDFs. This spike results from the subjective PDFs having markedly higher variances during the 1987 crash (power: 0.33; exponential: 0.31) than the corresponding risk-neutral PDF (0.27).

Figure 4 compares the standard deviations and skewness coefficients implied by the subjective PDFs against those from the risk-neutral PDFs for one contract/horizon (S&P500; four-weeks). Results for other contracts and horizons are similar. Figure 4 shows that for most observations second and third moments do not differ substantially between risk-neutral and subjective PDFs. The exception is the September 16, 1987 observation which shows up as an outlier on the scatter plots. Nonetheless, the differences are sufficient to induce a statistically significant difference in the forecast ability of the subjective and risk-neutral PDFs and a time-varying equity risk premium of around 10 percent per annum for most of the 1983 to 2001 period.

## **IV. Conclusions**

Options prices embed market expectations of the distribution of futures values of the underlying asset. This can provide potentially useful information for risk managers and analysts wishing to extract forecasts from security market prices. However, the risk-neutral density forecasts that are produced from options prices cannot be taken at their face value. We have shown, consistent with the work of others, that risk-neutral PDFs estimated from S&P 500 and FTSE 100 options do not provide good forecasts of the distribution of future values of the underlying asset, at least at the horizons for which we can obtain unambiguous results.

Theory tells us that if investors are risk averse and rational, the subjective density functions they use in forming their expectations will be linked to the risk-neutral density functions used to price options by a pricing kernel. Theory also suggests certain properties this pricing kernel might be expected to have. We have employed two widely used, and theoretically plausible, utility functions to infer the unobservable subjective densities by adjusting the observed risk-neutral densities. Our criterion in making this



adjustment is to choose the risk aversion parameter that produces subjective densities that best fit the distributions of realized values. That is, we assume that investors are rational forecasters of the distributions of future outcomes and thus the risk aversion parameter value that best fits the data is most likely to correspond to that of the representative agent.

In applying this methodology, we assume that investors' utility functions are stationary. This contrasts with the assumption made in previous papers that the statistical distribution was stationary. The subjective density functions derived under our assumption cannot be rejected as good forecasters of the distributions of future outcomes (unlike the unadjusted PDFs), and so this assumption appears to be validated on a practical level, subject to the caveat that there is some evidence of volatility dependence in risk aversion estimates.

The coefficient-of-risk-aversion estimates obtained by our methodology are comparable to those obtained in most previous studies. There is little evidence of risk aversions so high as to constitute a puzzle. We have also been able to establish, we believe for the first time, that the risk aversion estimates are surprisingly robust to differences in the specification of the representative investor's utility function and to the data set used. We also show that the estimated coefficients of risk aversion decline with the forecast horizon and are higher during periods of low volatility; both results imply that theoretical models may need to evolve to capture these effects.

## Appendix A: Density Forecast Evaluation

Testing whether a series of estimated time-varying density functions,  $\hat{f}_t(\cdot)$ , equals the true underlying density functions,  $f_t(\cdot)$ , when we observe a series of single outcomes,  $X_t$ , for each density, reduces to testing whether the inverse probability density transforms,  $y_t$ ,

$$y_t \equiv \int_{-\infty}^{X_t} \hat{f}_t(u) du, \quad (\text{A1})$$

are uniformly distributed. The test statistics used for making this determination all assume that the  $y_t$  are independently and identically distributed (i.i.d.); therefore, independence is a necessarily joint hypothesis with uniformity.

Several non-parametric methods have been proposed for testing the uniformity of the inverse-probability-transformed data. The Chi-squared test is based on dividing the  $[0,1]$  interval into a number of buckets and then counting the number of times the inverse probability transform falls into each bucket. The result is a series of counts  $n_i, i = 1, \dots, K$ ,

where  $K$  is the number of buckets and  $N = \sum_{i=1}^K n_i$  is the number of observations. Under

the null hypothesis that  $y_t \sim \text{i.i.d. } U(0,1)$ , each bucket is expected to contain

$\bar{n}_i \equiv E(n_i) = N/K$  observations. The Chi-squared test then uses the statistic

$$\chi^2 \equiv \sum_{i=1}^K \frac{(n_i - \bar{n}_i)^2}{\bar{n}_i} \text{ which is distributed } \chi^2 \text{ with } K-1 \text{ degrees of freedom under the null}$$

hypothesis.

The Kolmogorov-Smirnov and Kupier tests are based on the difference between the observed and theoretical cumulative density functions  $D(y_t) = \hat{F}(y_t) - F(y_t)$ . In this case the theoretical cumulative density is the uniform, so  $F(y_t) = y_t$ . The observed cumulative density is just the rank order divided by the number of observations,

$$F(y_t) = \frac{\text{rank}(y_t)}{T}. \text{ The Kolmogorov-Smirnov test is the maximum absolute difference}$$

between the observed and theoretical cumulative densities:  $D_{KS} = \max_{1 \leq t \leq T} |\hat{F}(y_t) - F(y_t)|$ .

The significance level for an observed value of  $D_{KS}$  under the uniformly distributed null is given by

$$\Pr ob(D_{KS} > observed) = Q_{KS} \left( \left( \sqrt{T} + 0.12 + 0.11/\sqrt{T} \right) D_{KS} \right), \quad (A2)$$

where

$$Q_{KS}(x) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 x^2}, \quad (A3)$$

which we approximate by summing the first 1,000 terms.

The Kupier test sums the maximum positive and negative differences between the observed and theoretical cumulative densities:

$$D_K = \max_{1 \leq t \leq T} \left( \hat{F}(y_t) - F(y_t) \right) + \max_{1 \leq t \leq T} \left( F(y_t) - \hat{F}(y_t) \right). \quad (A4)$$

The significance level for an observed value of  $D_K$  under the uniformly distributed null is given by

$$\Pr ob(D_K > observed) = Q_K \left( \left( \sqrt{T} + 0.155 + 0.24/\sqrt{T} \right) D_K \right), \quad (A5)$$

where

$$Q_K(x) = 2 \sum_{j=1}^{\infty} (4j^2 x^2 - 1) e^{-2j^2 x^2}, \quad (A6)$$

which we again approximate by summing the first 1,000 terms.

None of these methods provides a test the joint assumption that the  $y_t$  are i.i.d. Diebold, Gunther, and Tay (1998) suggest testing the independence and uniformity separately using the correlogram for the  $y_t$  to test for independence and, subject to not rejecting the hypothesis that the data were independent, using a Chi-squared test to test the hypothesis that the probability integral transforms are uniformly distributed. However, as Diebold, Tay, and Wallis (1999) point out, to separate fully the desired  $U(0,1)$  and i.i.d. properties of  $y_t$ , “we would like to construct confidence intervals for histogram bin heights that condition on uniformity but that are robust to dependence of unknown form,” and “confidence intervals for the autocorrelations that condition on independence but are robust to non-uniformity.” (p. 81) Unfortunately, since there is no serial-correlation-adjusted Chi-squared test of known small sample properties for the

uniformity hypothesis, Diebold, Gunther, and Tay (1998) and Diebold, Tay, and Wallis (1999) are unable to conduct a simultaneous joint test of the i.i.d. and uniformly distributed properties. Both papers use the Kolmogorov-Smirnov (K-S) and the Chi-squared statistics for testing for uniformity but, as Diebold, Tay, and Wallis (1999) note, the impact of departures from randomness on the performance of these non-parametric tests is not known.

To test the significance of autocorrelations, Diebold, Tay, and Wallis (1999) construct finite-sample confidence intervals that condition on independence but are robust to deviations from uniformity by sampling with replacement from the series of the probability integral transforms and building up the distribution of sample autocorrelations. The drawback of their methodology is that they separate the joint hypothesis test into two different tests. The use of the binomial distribution that they mention in their paper is also controversial since the numbers of observations in each bin are not independent and actually follow a multinomial distribution. The Chi-squared test is the appropriate test for the uniform-distribution null hypothesis in this case.

Berkowitz (2001) proposes a density evaluation methodology that does provide a joint test of independence and normality. Furthermore, unlike the non-parametric Chi-squared, Kolmogorov-Smirnov, and Kupier tests, which discard sample information either by bucketing or by selecting single observations (maximum deviations), the Berkowitz test utilizes all observations. The Berkowitz joint hypothesis test,  $LR_3$ , described in the body of this paper, tests both uniformity and independence. For diagnostic purposes, restricted forms of the Berkowitz test are possible; for instance, in  $LR_1$  tests independence under the assumption that the data are uniform.

To choose between alternative methods for testing the uniformity of the inverse-probability transformed data, we used Monte Carlo simulations to test the small sample properties of these different statistical tests, both under the null hypothesis of independently distributed uniform random variates and when the simulated data were autocorrelated. To do this we needed to generate autocorrelated uniformly distributed random numbers. Beginning with a series of random standard normal numbers,  $x_t \sim N(0,1)$ , we construct autocorrelated normally distributed random variables with first order correlations equal to  $\rho$ , by creating the MA(1) variables  $y_t = x_t + \theta x_{t-1}$ , where

The  $y_t \sim N(0, 1 + \theta^2)$ , with autocorrelation  $\rho$ . To create uniformly distributed numbers  $u_t$  we transform the  $y_t$  using the inverse cumulative normal distribution, that is,  $u_t = \Phi^{-1}(y_t; 0, 1 + \theta^2)$ , where  $\Phi^{-1}(x; \mu, \sigma^2)$  is the inverse of the normal cumulative density function with parameters  $\mu$  and  $\sigma^2$ , evaluated at  $x$ . These uniform random numbers also have a first order autocorrelation of  $\rho$ .

We test three small sample sizes—50, 100, and 200— and three autocorrelations coefficients—0, 0.1, and 0.2—using 10,000 replications for each size/autocorrelation pair. To validate our simulations, we also ran large sample simulations using 10,000 data points in each simulation. For each size/autocorrelation pair, we computed the number of times each test statistic exceeded its theoretical 90 and 99 percent levels. The results are presented in Table A1.

In large samples ( $T = 10,000$ ) and when the null hypothesis is true, i.e.,  $\rho = 0$ , all four tests perform well. However, when the null hypothesis is true, but sample sizes are small, only the Berkowitz and Kolmogorov-Smirnov tests perform well. The Chi-squared test does slightly worse; while the Kupier test is quite unreliable. In large samples with autocorrelated data, the Berkowitz test rejects with near certainty, while the Chi-squared, Kolmogorov-Smirnov, and Kupier tests reject at approximately the same frequency as with uncorrelated data. In small samples, the Berkowitz test rejects slightly more frequently than with uncorrelated data with the rejection rate increasing in the degree of autocorrelation. For the same data, the Kolmogorov-Smirnov tests rejects only trivially more frequently than for uncorrelated data.

Thus, we conclude that the Kupier test is wholly inadequate for small-sample analysis. The Chi-squared test, while perhaps adequate, is dominated by the Berkowitz and Kolmogorov-Smirnov tests for small sample analysis. While both the Berkowitz and the Kolmogorov-Smirnov tests appear to do well under the null hypothesis in large and small samples, the Berkowitz test has an edge when the data is in fact autocorrelated. Since some of our actual data are from overlapping observations (five- and six-week horizons), we are concerned about potential autocorrelation. For this reason, and because the Berkowitz test is the only one of the four tests to jointly test independence and normality, we choose to use the Berkowitz test in this paper.

## Appendix B: How Volatility Changes Impact Prices

The impact of changes in volatility on the risk premium, and hence on the price of assets, involves several steps. An increase in equity volatility generally leads to an increase in the risk premium though the expected change is model dependent.<sup>36</sup> An increase in the risk premium in turn has two effects: An immediate decrease in prices need to provide a greater expected return in the future, assuming the increase in risk is not accompanied by any information to change the expected level of future cash flows. To make this relation explicit, consider the following simple model.

Campbell, Lo, and MacKinlay (1997, p. 307) show that under the certain (usual) assumptions regarding the structure of markets, log-normality of asset returns and a representative investor with a power utility function, then the risk premium can be expressed as follows:

$$\log E_t \left( \frac{1 + R_{i,t+1}}{1 + R_{f,t+1}} \right) = \gamma \sigma_{ic}, \quad (\text{B1})$$

where  $R_{i,t+1}$  is the return on the risky traded asset,  $R_{f,t+1}$  is the riskless return,  $\sigma_{ic}$  is the covariance between the risky asset and consumption, and  $\gamma$  is the coefficient of risk aversion, or under the assumed power utility the representative investor's RRA. In our analysis, we use the return on the equity index to proxy for wealth and therefore changes in consumption. Under this assumption, and to acknowledge that variance may change, we replace  $\sigma_{ic}$  with  $\sigma_{it}^2$ . We can also replace  $1/(1 + R_{f,t+1})$  with the price of the riskless bond  $B_t$  and take it outside the expectation. Lastly, we replace  $(1 + R_{i,t+1})$  with the equivalent in terms of the current and end-of-period prices of the risky asset,  $P_{i,t+1}/P_{it}$ , where  $P_{i,t+1}$  is the cum-dividend value of the risky asset at time  $t + 1$ . The previous expression for the risk premium thus becomes:

---

<sup>36</sup>Han (2002) argues that the risk premium is positively related to volatility and negatively related to volatility risk, that is the risk the volatility will change. Depending on the relation between the level of volatility and volatility risk, the total effect of an increase in volatility on the risk premium may be less than expected.

$$\log B_t E_t \left( \frac{P_{i,t+1}}{P_{it}} \right) = \gamma \sigma_{it}^2. \quad (\text{B2})$$

We next examine the effect of an instantaneous change in volatility of the risky asset by differentiating both sides by  $\sigma_{it}^2$ .

$$\frac{\partial}{\partial \sigma_{it}^2} \gamma \sigma_{it}^2 = \gamma \quad (\text{B3})$$

$$= \frac{\partial}{\partial \sigma_{it}^2} \log B_t E_t \left( \frac{P_{i,t+1}}{P_{it}} \right) \quad (\text{B4})$$

$$= \frac{1}{B_t E_t \left( \frac{P_{i,t+1}}{P_{it}} \right)} B_t \frac{\partial}{\partial \sigma_{it}^2} E_t \left( \frac{P_{i,t+1}}{P_{it}} \right) \quad (\text{B5})$$

$$= \frac{1}{E_t \left( \frac{P_{i,t+1}}{P_{it}} \right)} \left\{ \frac{1}{P_{it}} \frac{\partial E_t (P_{i,t+1})}{\partial \sigma_{it}^2} - \frac{E_t (P_{i,t+1})}{P_{it}^2} \frac{\partial P_{it}}{\partial \sigma_{it}^2} \right\}. \quad (\text{B6})$$

If we assume that the instantaneous change in volatility does not affect future expected cash flows, then as a first approximation<sup>37</sup>

$$\frac{\partial E_t (P_{i,t+1})}{\partial \sigma_{it}^2} \approx 0 \quad (\text{B7})$$

and

$$\frac{1}{P_{it}} \frac{\partial P_{it}}{\partial \sigma_{it}^2} \approx -\gamma. \quad (\text{B8})$$

Thus, a change in volatility can be expected to have a coincident change on prices as follows:

$$\frac{\Delta P_{it}}{P_{it}} \approx -\gamma \Delta \sigma_{it}^2. \quad (\text{B9})$$

---

<sup>37</sup>There is a large literature showing that volatility shocks are persistent, at least over shorter intervals. On the other hand, an examination of the time-series of implied volatilities shows that volatility spikes tend to be of short duration. For purposes of this “back of the envelope” analysis, it is sufficient if the effects of volatility changes on future prices at some horizon are less than that on current prices.

This is consistent with our expectations that as volatility increases asset prices fall, and states that under the representative investor power utility assumption the constant of proportionality is the coefficient of risk aversion.



## REFERENCES

- Ait-Sahalia, Yacine, and Andrew W. Lo, 2000, Nonparametric risk management and implied risk aversion, *Journal of Econometrics* 94, 9–51.
- Ait-Sahalia, Yacine, Yubo Wang, and Francis Yared, 2001, Do option markets correctly price the probabilities of movement of the underlying asset? *Journal of Econometrics* 102, 67–110.
- Anagnou, Iliana, Mascia Bedendo, Stewart Hodges, and Robert Tompkins, 2002, The relation between implied and realised probability density functions, Working paper, University of Technology, Vienna.
- Arrow, Kenneth J., 1971, *Essays in the Theory of Risk Bearing* (North Holland, Amsterdam).
- Barone-Adesi, Giovanni, and Robert E. Whaley, 1987, Efficient analytic approximation of American option values, *Journal of Finance* 42, 301–320.
- Bartunek, Kenneth S., and Mustafa Chowdhury, 1997, Implied risk aversion parameter from option prices, *The Financial Review* 32, 107–124.
- Berkowitz, Jeremy, 2001, Testing density forecasts with applications to risk management, *Journal of Business and Economic Statistics* 19, 465–474.
- Bliss, Robert R., and Nikolaos Panigirtzoglou, 2002, Testing the stability of implied probability density functions, *Journal of Banking and Finance* 26, 381–422.
- Breeden, Douglas T., and Robert H. Litzenberger, 1978, Prices of state-contingent claims implicit in options prices, *Journal of Business* 51, 621–651.
- Campa, Jose M., P. H. Kevin Chang, and R. L. Reider, 1998, Implied exchange rate distributions: Evidence from OTC option markets, *Journal of International Money and Finance* 17, 117–160.
- Campa, Jose M., P. H. Kevin Chang, and James F. Refalo, 2002, An options-based analysis of emerging market exchange rate expectations: Brazil's real plan, 1994–1997, *Journal of Development Economics* 69, 227–53.
- Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).
- Cochrane, John H., and Lars P. Hansen, 1992, Asset pricing explorations for macroeconomics, in 1992 *NBER Macroeconomics Annual* (NBER, Cambridge, MA).
- Constantinides, George M., 2002, Rational asset prices, *Journal of Finance* 57, 1567–1591.
- Coutant, Sophie, 2001, Implied risk aversion in options prices, in *Information Content in Option Prices: Underlying Asset Risk-neutral Density Estimation and Applications*, (Ph.D. thesis, University of Paris IX Dauphine).

- Diebold, Francis X., Anthony S. Tay, and Kenneth F. Wallis, 1999, Evaluating density forecasts of inflation: The survey of professional forecasters, in *Cointegration, causality, and forecasting: A festschrift in honour of Clive W. J. Granger*, edited by Robert F. Engle and Halbert White, Oxford University Press, New York, 76–90.
- Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay, 1998, Evaluating density forecasts with applications to financial risk management, *International Economic Review* 39, 863–883.
- Duffee, Gregory R., 1996, Idiosyncratic variation of Treasury bill yields, *Journal of Finance* 51, 527–551.
- Efron, Bradley, and Robert J. Tibshirani, 1993, *An Introduction to the Bootstrap* (Chapman & Hall, New York).
- Epstein, Larry G., and Stanley E. Zin, 1991, Substitution, risk aversion and the temporal behaviour of consumption and asset returns: An empirical analysis, *Journal of Political Economy* 99, 263–268.
- Ferson, Wayne E., and George M. Constantinides, 1991, Habitat persistence and durability in aggregate consumption: Empirical tests, *Journal of Financial Econometrics* 29, 199–240.
- Friend, Irwin, and Marshall E. Blume, 1975, The demand for risky assets, *American Economic Review* 65, 900–922.
- Galati, Gabriele, and William R. Melick, 1999, Perceived central bank intervention and market expectations: An empirical study of the yen/dollar exchange rate 1993–96, Working Paper No. 77, Bank of International Settlements.
- Gemmell, Gordon, and Apostolos Saffekos, 1999, How useful are implied distributions? Evidence from stock-index options, Working paper, City University Business School, London.
- Guo, Hui, and Robert Whitelaw, 2001, Risk and return: Some new evidence, Working Paper No. 2001–001A, Federal Reserve Bank of St Louis.
- Hamilton, James D., 1994, *Time Series Analysis* (Princeton University Press, Princeton, NJ).
- Han, Yufeng, 2002, On the relation between the market risk premium and volatility, Working paper, Washington University, St. Louis.
- Hansen, Lars P., and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269–1286.
- Hansen, Lars P., and Kenneth J. Singleton, 1984, Errata: Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 52, 267–268.
- Huang, Chi-Fu, and Robert H. Litzenberger, 1988, *Foundations for Financial Economics* (North-Holland, New York, NY).
- Jackwerth, Jens Carsten, 2000, Recovering risk aversion from option prices and realized returns, *Review of Financial Studies* 13, 433–467.

- Jorion, Philippe, and Alberto Giovannini, 1993, Time series test of a non-expected utility model of asset pricing, *European Economic Review* 37, 1083–1100.
- Kendal, Sir Maurice, and Alan Stuart, 1979, *The Advanced Theory of Statistics* (Macmillan Publishing Co., Inc., New York, NY).
- Malz, Allan M., 1997, Estimating the probability distribution of the future exchange rate from options prices, *Journal of Derivatives* 5, 18–36.
- Mehra, Rajnish, and Edward C. Prescott, 1985, The equity premium: A puzzle, *Journal of Monetary Economics* 15, 145–161.
- Melick, William R., and Charles P. Thomas, 1997, Recovering an asset's implied PDF from option prices: An application to crude oil during the Gulf crisis, *Journal of Financial and Quantitative Analysis* 32, 91–115.
- Normandin, Michel, and Pascal St-Amour, 1998, Substitution, risk aversion, taste shocks and equity premia, *Journal of Applied Econometrics* 13, 265–281.
- Pérignon, Christophe, and Christophe Villa, 2002, Extracting information from options markets: Smiles, state-price densities and risk aversion, *European Financial Management* 8, 495–513.
- Rosenberg, Joshua V., and Robert F. Engle, 2002, Empirical pricing kernels, *Journal of Financial Economics* 64, 341–372.
- Shiratsuka, Shigenori, 2001, Information content of implied probability distributions: Empirical studies of Japanese stock price index options, *Monetary and Economic Studies* 19, 143–170.
- Söderlind, Paul, 2000, Market expectations in the UK before and after the ERM crisis, *Economica* 67(265), 1–18.
- Weinberg, Steven A., 2001, Interpreting the volatility smile: An examination of the information content of options prices, International Finance Discussion Paper No. 706, Board of Governors of the Federal Reserve System.

**Table I**  
**Summary Statistics for Samples of Option Cross-Sections**

Description of option cross-sections after constructing matched sets option prices and interest rates for constructing forecast densities and realizations of the underlying asset for testing. Option cross-sections containing less than five strikes with positive time value were eliminated. Option observation dates were selected to have expiries nearest to target horizon with a maximum permissible variation of three days for weekly horizons and four days for monthly horizons. Horizons out to two months include serial contracts expiring at monthly intervals. Beyond two months, only quarterly expiries are available.

Forecast Horizon	FTSE 100				S&P 500			
	Number of Cross- Sections	Strikes per Cross-Section			Number of Cross- Sections	Strikes per Cross-Section		
		Min.	Mean	Max.		Min.	Mean	Max.
1 week	99	5	10.8	28	169	5	14.8	64
2 weeks	108	5	14.9	43	172	5	20.0	63
3 weeks	108	7	18.5	54	178	5	23.7	70
4 weeks	108	7	21.7	59	184	5	25.3	77
1 month	108	9	22.8	56	183	5	26.0	76
5 weeks	108	10	24.4	62	184	5	27.0	83
6 weeks	108	11	26.5	66	182	5	28.0	81
2 months	108	7	33.8	81	172	5	27.5	89
3 months	47	8	35.0	91	74	7	30.4	98
4 months	36	9	24.4	87	72	5	28.4	77
5 months	36	9	23.4	91	72	6	24.5	60
6 months	35	8	21.6	56	67	5	20.3	56
9 months	34	8	20.5	59	38	5	18.1	30
1 year	4	11	11.3	12	2	5	6.0	7

**Table II**  
**Utility Functions and Associated Formulae**

Functional forms of the two utility functions used to adjust risk-neutral density functions, together with the marginal utility and the measure of relative risk aversion:

$$\text{RRA} = -\frac{S_T U''(S_T)}{U'(S_T)}.$$

Utility Function	$U(S_T)$	$U'(S_T)$	RRA
Power	$\frac{S_T^{1-\gamma} - 1}{1-\gamma}$	$S_T^{-\gamma}$	$\gamma$
Exponential	$-\frac{e^{-\gamma S_T}}{\gamma}$	$e^{-\gamma S_T}$	$\gamma S_T$

**Table III**  
**Distribution of Bootstrap Estimates of the Coefficient of Relative Risk Aversion**

Coefficients of relative risk aversion (RRA) estimated from 1,000 bootstrap samples for each contract type and representative horizon. Each bootstrap sample was of the same size as the original sample (see Table I) and was constructed by sampling with replacement from the original sets of matched option cross-sections, interest rates, and realizations. Percent of  $RRA < 0$  provides a bootstrap test of the hypothesis that  $RRA = 0$ , that is that the representative investor in these securities is risk neutral, based on the sampling variation in the data. Standard deviation of Monte Carlo RRA estimates, constructed under the assumption the investors are risk neutral (see Section II.D.1), are provided for comparison with the standard deviation of the bootstrap estimates. Point estimates of the RRA using the original sample are provided for comparison with the mean and median of the bootstrap estimates.

	FTSE 100 (3-week)		S&P 500 (5-week)	
	Power Utility	Exponential Utility (Mean)	Power Utility	Exponential Utility (Mean)
Bootstrap RRA Estimates				
Minimum	-1.25	-1.90	-1.34	-1.17
Mean	5.02	5.11	3.72	3.95
Median	5.05	5.08	3.68	3.95
Maximum	11.76	11.03	8.17	8.56
RRA<0	0.4%	0.4%	0.5%	0.5%
Standard Deviation	1.99	1.92	1.37	1.41
Monte Carlo RRA Estimates				
Standard Deviation	2.24	2.17	1.34	1.28
Original Sample RRA Estimates				
Point estimates	5.10	4.99	3.53	3.26

**Table IV**  
**Berkowitz Statistic P-Values for Risk-Neutral and Power- and Exponential-Utility-Adjusted PDFs**

This table presents the results of tests of the ability of risk neutral and subjective PDFs to forecast the future distribution of the prices of the underlying asset. Tests use a modified Berkowitz test. Power and exponential utility adjusted PDFs are constructed by adjusting the risk-neutral PDF using the appropriate utility function and equation (1). The utility function risk aversion parameters were selected to maximize the Berkowitz statistic  $LR_3$ . The reported  $LR_3$  value is the p-value of the Berkowitz likelihood ratio test for i.i.d. normality of the inverse-normal transformed inverse-probability transforms of the realizations (equations (2) and (3)):  $LR_3 = -2 \left[ L(0,1,0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho}) \right]$ . The power- and exponential-utility p-values have been adjusted for the bias resulting from maximizing the  $LR_3$  p-value over values of the risk aversion coefficient  $\gamma$ . Adjustments are based on Monte Carlo simulations (see Section II.D.1). The  $LR_1$  statistic is the p-value of the Berkowitz likelihood ratio test for independence of the same transformed data:  $LR_1 = -2 \left[ L(\hat{\mu}, \hat{\sigma}^2, 0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho}) \right]$ . Rejection of the test for independence (low  $LR_1$  values) means that rejection of the “good” forecast null hypothesis (low  $LR_3$  values) may be due serial correlation rather than poor density forecasts.

Forecast Horizon	PDF	FTSE 100			S&P 500		
		N	$LR_3$	$LR_1$	N	$LR_3$	$LR_1$
1 week	Risk-neutral	99	0.233	0.578	168	0.003	0.295
	Power		0.740	0.506		0.038	0.319
	Exponential		0.643	0.368		0.159	0.763
2 weeks	Risk-neutral	108	0.006	0.581	171	0.003	0.845
	Power		0.009	0.442		0.010	0.973
	Exponential		0.010	0.324		0.033	0.486
3 weeks	Risk-neutral	108	0.044	0.596	177	0.001	0.441
	Power		0.207	0.920		0.096	0.591
	Exponential		0.269	0.854		0.274	0.727
4 weeks	Risk-neutral	108	0.035	0.466	183	0.022	0.678
	Power		0.114	0.713		0.204	0.916
	Exponential		0.143	0.863		0.326	0.712
5 weeks	Risk-neutral	108	0.021	0.659	183	0.007	0.258
	Power		0.041	0.447		0.051	0.184
	Exponential		0.046	0.319		0.048	0.113
6 weeks	Risk-neutral	108	0.018	0.057	181	0.000	0.019
	Power		0.018	0.033		0.003	0.009
	Exponential		0.016	0.025		0.003	0.002

**Table V**  
**Estimates of the Risk Aversion Parameter  $\gamma$**

Values of the risk aversion parameter  $\gamma$  obtained by maximizing the forecast ability of the adjusted PDFs measured using the Berkowitz LR<sub>3</sub> statistic. \*, \*\*, and \*\*\* indicate statistically that the values are significantly different from zero at the 10, five, and one percent levels of significance respectively, based on Monte Carlo simulations described in Section II.D.1.

<b>Forecast Horizon</b>	<b>FTSE 100</b>		<b>S&amp;P 500</b>	
	<b>Power Utility</b>	<b>Exponential Utility</b>	<b>Power Utility</b>	<b>Exponential Utility</b>
1 week	7.91**	1.52**	9.52***	15.97***
2 weeks	4.44*	1.00*	5.38**	8.44**
3 weeks	5.10**	1.11**	6.85***	10.38***
4 weeks	4.05**	0.91**	4.08***	6.33**
5 weeks	3.04**	0.66**	3.53***	5.22***
6 weeks	1.97	0.37	3.37***	4.36**



**Table VI**  
**Measures of Relative Risk Aversion Implied by PDFs Adjusted Using**  
**Power and Exponential Utility Functions**

Values for the representative investors relative risk aversion (RRA) obtained by maximizing the forecast ability of the adjusted PDFs measured using the Berkowitz LR<sub>3</sub> statistic. For the power utility the RRA is  $\gamma$ , while for the exponential utility the RRA is  $\gamma S_t$ , and therefore varies with the level of the index. The high and low at-the-money (ATM) volatility results were obtained by first dividing the sample in two equal halves based on the mean of the implied volatility of the two nearest-the-money strikes for each cross-section, and then re-estimating the Berkowitz LR<sub>3</sub>-maximizing value of  $\gamma$  for each sub-sample.

Forecast Horizon	Power Utility	FTSE 100			Power Utility	S&P 500		
		Exponential Utility Range	Mean	Median		Exponential Utility Range	Mean	Median
All Observations								
1 week	7.91	3.60–10.09	7.00	6.89	9.52	3.709–23.972	10.56	7.39
2 weeks	4.44	2.37–6.64	4.47	4.05	5.38	1.959–12.660	5.54	3.91
3 weeks	5.10	2.64–7.41	4.99	4.52	6.85	1.633–15.574	6.64	4.75
4 weeks	4.05	2.17–6.08	4.09	3.71	4.08	0.947–9.502	3.96	2.88
5 weeks	3.04	1.57–4.39	2.96	2.68	3.53	0.777–7.826	3.26	2.37
6 weeks	1.97	0.89–2.50	1.68	1.52	3.37	0.650–6.543	2.74	1.98
High ATM Implied Volatilities Observations								
1 week	4.84	2.22–6.19	4.96	5.45	8.35	2.88–18.63	9.34	8.70
2 weeks	1.90	1.22–3.42	2.72	3.00	1.53	0.61–3.96	2.07	2.10
3 weeks	3.20	1.72–4.50	3.59	3.95	5.57	1.80–11.59	6.31	6.94
4 weeks	2.64	1.26–3.54	2.84	3.14	2.32	0.46–4.63	2.52	2.80
5 weeks	1.48	0.83–2.18	1.70	1.91	2.41	0.46–4.58	2.46	2.74
6 weeks	0.95	0.62–1.63	1.29	1.45	2.33	0.41–4.14	2.24	2.48
Low ATM Implied Volatilities Observations								
1 week	12.76	7.50–20.28	11.88	11.19	11.14	5.58–32.79	12.84	10.42
2 weeks	7.90	5.36–14.21	8.01	7.61	11.03	5.46–32.68	11.62	10.07
3 weeks	8.25	5.26–14.44	8.09	7.73	8.92	3.06–28.80	8.94	8.70
4 weeks	6.37	4.37–11.59	6.43	6.14	6.90	2.56–12.39	7.04	7.14
5 weeks	5.62	3.50–9.61	5.58	5.31	5.66	2.04–14.87	6.06	6.10
6 weeks	4.03	2.41–6.62	3.77	3.64	5.70	2.06–15.04	6.13	6.17

**Table VII**  
**Coefficient of Relative Risk Aversion Estimates from Previous Studies**

This table is an updated version of Ait-Sahalia and Lo (2000) Table 5. Most studies assume constant relative risk aversion. An exception is Ait-Sahalia and Lo (2000). The CRRA value of 12.7 reported in their paper and below is an average value. However they informally reject constant relative risk aversion in favor of a broadly U-shaped relative risk aversion function that varies between zero and 60.

Study	CRRA Range
Arrow (1971)	1
Friend and Blume (1975)	2
Hansen and Singleton (1982, 1984)	0–1
Mehra and Prescott (1985)	55
Epstein and Zin (1991)	0.4–1.4
Ferson and Constantinides (1991)	0–12
Cochrane and Hansen (1992)	40–50
Jorion and Giovannini (1993)	5.4–11.9
Normandin and St-Amour (1998)	<3
Ait-Sahalia and Lo (2000)	12.7
Guo and Whitelaw (2001)	3.52

**Table VIII**  
**Relation Between Stock Returns and Changes in Options-Implied Variance**

Slope coefficients obtained by regressing the daily percentages in the S&P 500 index level against daily changes in the VIX implied volatility index level. The high- and low-volatility sub-samples were constructed by dividing the sample into two halves based on the level of the VIX index. The slope coefficients provide an alternative indication of changes in the representative agents risk aversion,  $\gamma$ , during periods of high and low volatility:

$$\frac{dP_t}{P_t} \approx -\gamma dVar(r_t).$$

To examine the effects of outliers, the largest absolute changes in volatility were trimmed from each sample. All coefficients are significant at the one percent level.

		All Observations	High Volatility	Low Volatility
Complete Sample		0.40	0.38	4.41
Delete $n$ largest volatility changes: $n$	5	0.41	0.35	1.26
	10	0.53	0.48	0.94
	20	0.55	0.49	0.87
	50	0.52	0.47	0.76
	100	0.49	0.44	0.71

**Table A1**  
**Monte Carlo Tests of Various Methods of Testing Density Forecasts**

Monte Carlo tests consisted of generating 10,000 samples of indicated size (N) of U(0,1) random numbers. Tests included both uncorrelated and correlated samples. Numbers reported in the table are the frequency with which the Berkowitz, Chi-squared, Kolmogorov-Smirnov, and Kupier statistics, computed for each Monte Carlo sample, exceeded their theoretical 90- and 99-percent values. Thus, in small samples the Chi-squared test fails to reject as frequently as it should (0.839 & 0.984 vice 0.900 & 0.990). Tests using correlated Monte Carlo data provide a test of the robustness of these tests under conditions when the data are U(0,1), but not i.i.d., as may occur with overlapping observations.

$\rho$	N	Berkowitz		Chi-square	Kolmogorov-Smirnov	Kupier
		LR <sub>3</sub>	LR <sub>1</sub>			
$\alpha = 0.90$						
0 (i.i.d.)	50	0.901	0.905	0.839	0.890	0.487
	100	0.896	0.903	0.846	0.893	0.621
	200	0.898	0.904	0.858	0.893	0.718
	10,000	0.898	0.899	0.889	0.901	0.881
0.1	50	0.918	0.915	0.850	0.896	0.503
	100	0.933	0.931	0.858	0.899	0.639
	200	0.961	0.957	0.858	0.899	0.725
	10,000	1.000	1.000	0.892	0.906	0.885
0.2	50	0.961	0.951	0.862	0.904	0.521
	100	0.985	0.984	0.867	0.903	0.657
	200	0.999	0.999	0.871	0.908	0.736
	10,000	1.000	1.000	0.896	0.920	0.894
$\alpha = 0.99$						
0 (i.i.d.)	50	0.990	0.989	0.984	0.987	0.744
	100	0.991	0.990	0.981	0.988	0.857
	200	0.991	0.991	0.984	0.989	0.919
	10,000	0.991	0.990	0.987	0.989	0.985
0.1	50	0.992	0.992	0.985	0.989	0.751
	100	0.993	0.994	0.984	0.987	0.864
	200	0.995	0.996	0.984	0.987	0.917
	10,000	1.000	1.000	0.988	0.989	0.984
0.2	50	0.997	0.996	0.985	0.988	0.764
	100	0.999	0.999	0.982	0.989	0.865
	200	1.000	1.000	0.984	0.990	0.927
	10,000	1.000	1.000	0.990	0.992	0.987

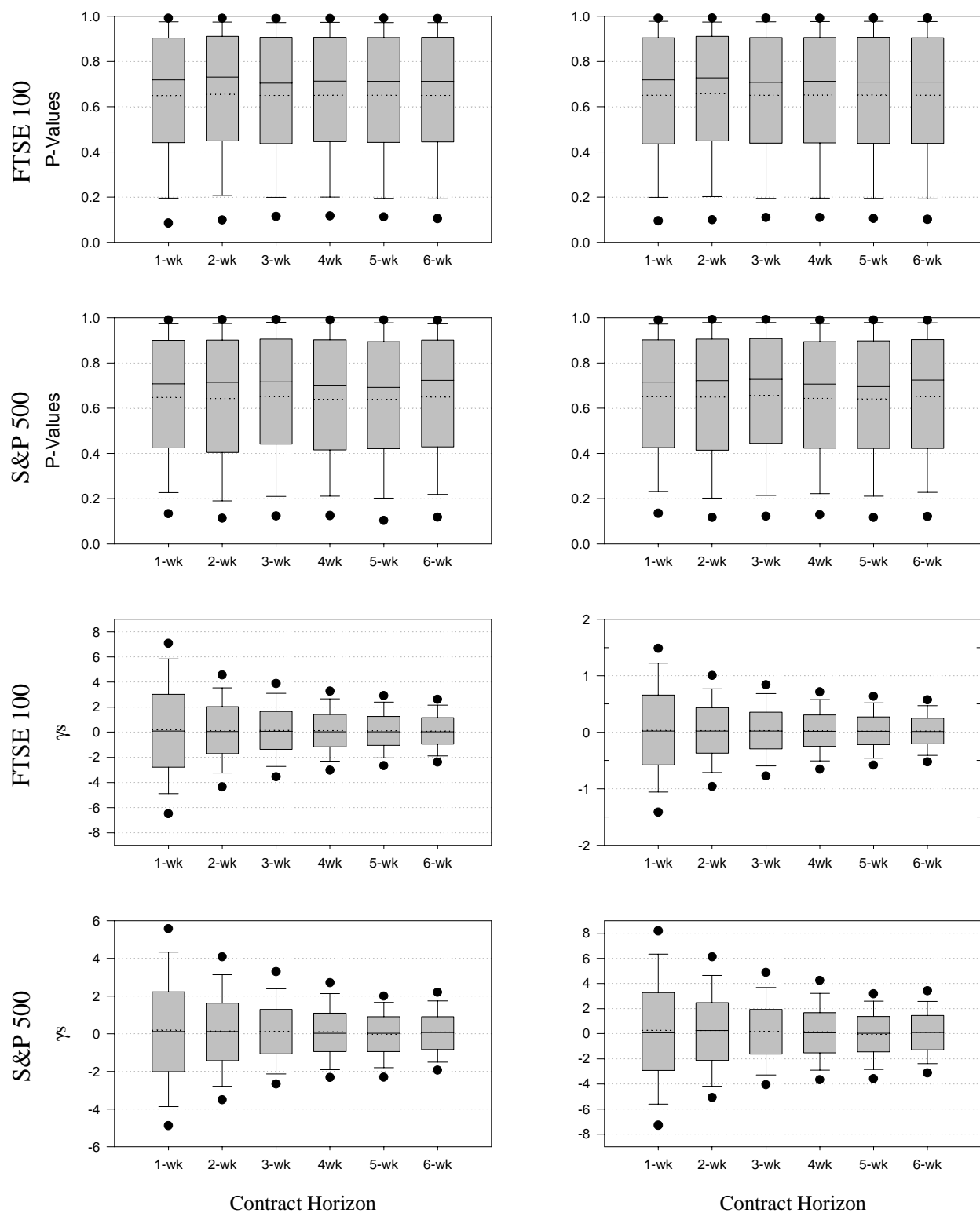
## **Description of Figures**

**Figure 1. Distribution of estimated p-values and  $\gamma$ s from Monte Carlo simulations using risk-neutral PDFs.** Underlying data are the actual FTSE 100 and S&P 500 risk-neutral PDFs and pseudo-realizations drawn from those PDFs—implying the true  $\gamma = 0$ . The p-values and  $\gamma$ s were obtained by maximizing the Berkowitz LR<sub>3</sub> statistic over values of  $\gamma$  to obtain the best density forecast performance. Simulations were repeated 1,000 times for each contract/horizon. The box portion of each Tukey plot encompasses the inter-quartile range of the estimates (p-values or  $\gamma$ s) derived from the Monte Carlo simulations. The lines dividing the box are the mean (dotted) and median (solid), respectively. The “whiskers” are the 10th and 90th percentiles of the distributions and the end points correspond to the 5th and 95th percentiles.

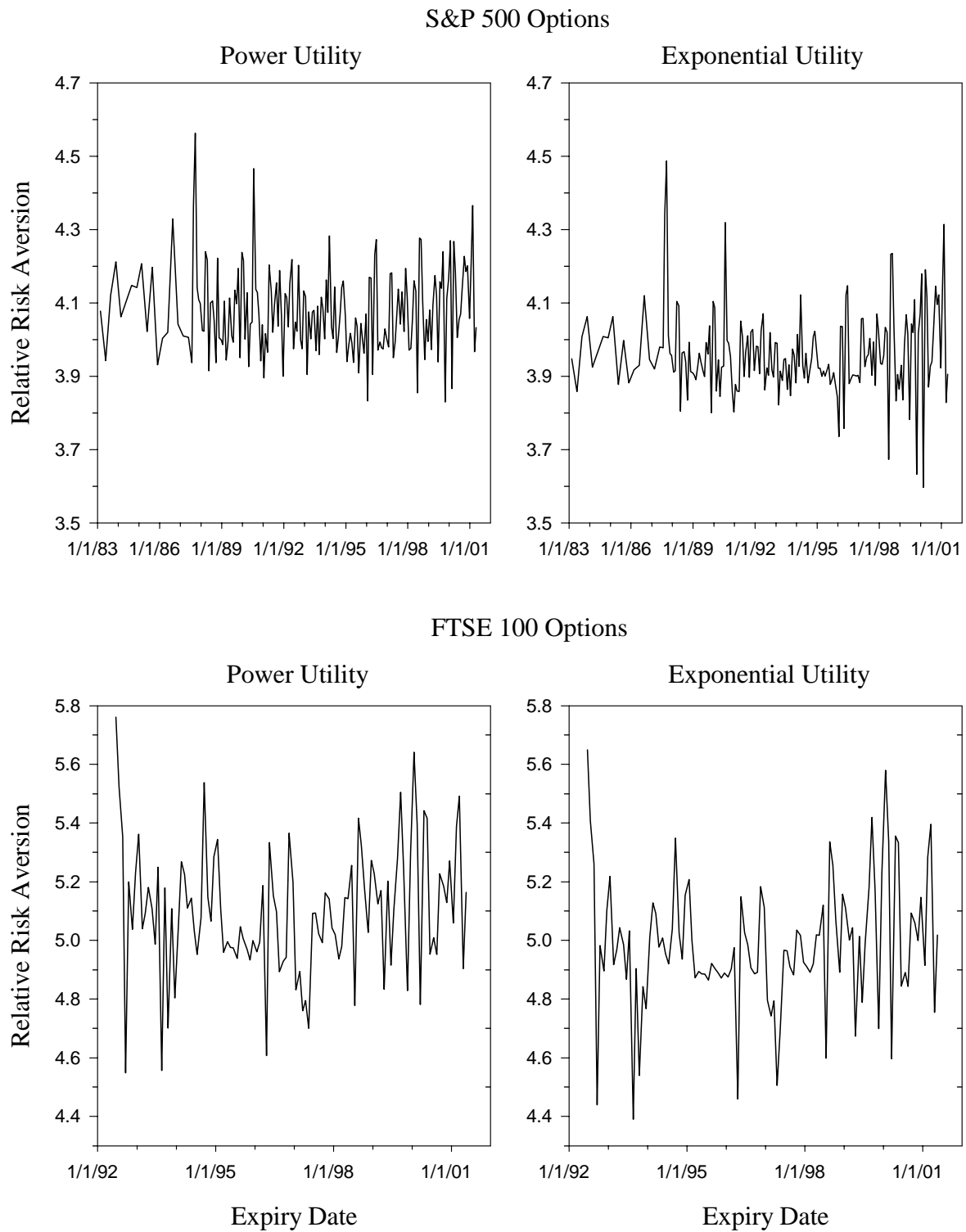
**Figure 2. Cross-validation results.** The time-series of relative risk aversion (RRA) estimates was obtained by cross-validation using the four-week S&P 500 and three-week FTSE 100 options contracts. Cross-validation consists of removing a single observation (option cross-section and realization pair), estimating the RRA, replacing the omitted observation and the removing the next one, and so forth. The time-index of each RRA corresponds to the observation date of the omitted observation. For each cross-validation subset, the RRA was estimated by maximizing the Berkowitz LR<sub>3</sub> statistic over values of  $\gamma$  to obtain the best density forecast performance, and then converting the risk-aversion parameter  $\gamma$  into an RRA. The mean RRA is plotted for the exponential utility adjustment.

**Figure 3. Plot of risk premia implied by risk-neutral and subjective PDFs.** The risk premia are measured using the four-week ahead S&P 500 risk-neutral and subjective PDFs. The subjective PDFs were constructed using the relevant utility function and equation (1), and then maximizing the Berkowitz LR<sub>3</sub> statistic over values of  $\gamma$  to obtain the best density forecast performance. The risk premium observed each observation date is the difference between the means of the subjective and risk-neutral PDFs, normalized by the mean of the risk-neutral PDF. The data are plotted twice on different vertical scales for convenience.

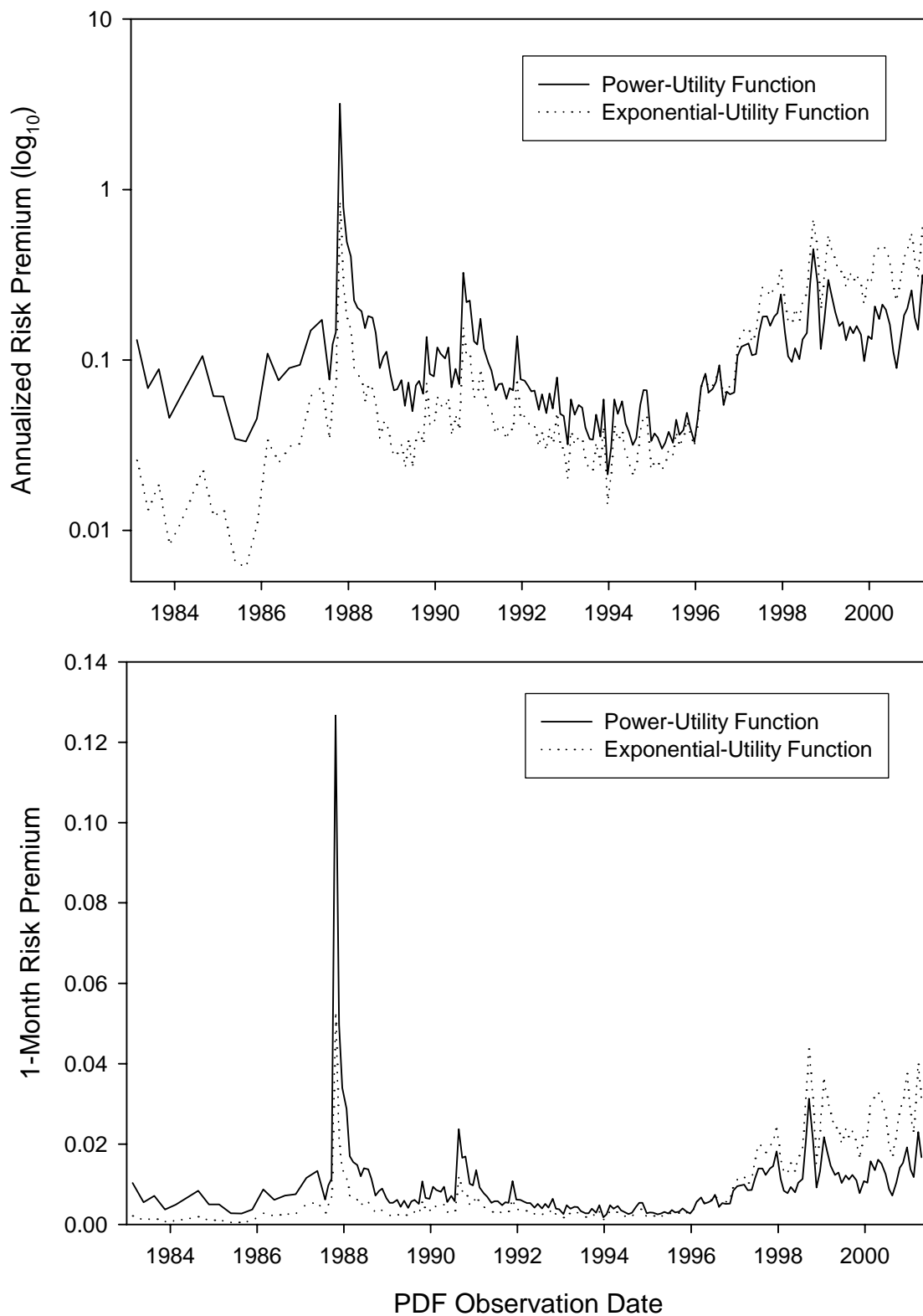
**Figure 4. Comparison of standard deviations and skewness coefficients from risk-neutral and subjective PDFs.** Underlying data are 183 risk-neutral and subjective PDFs for the S&P500 one-month options. The subjective PDFs were constructed using the relevant utility function and equation (1), and then maximizing the Berkowitz LR<sub>3</sub> statistic over values of  $\gamma$  to obtain the best density forecast performance. The standard deviation and skewness of the resulting subjective PDFs are then plotted against the corresponding statistics for the unadjusted risk-neutral PDFs.



**Figure 1. Distribution of estimated p-values and from Monte Carlo simulations using risk-neutral PDFs.**

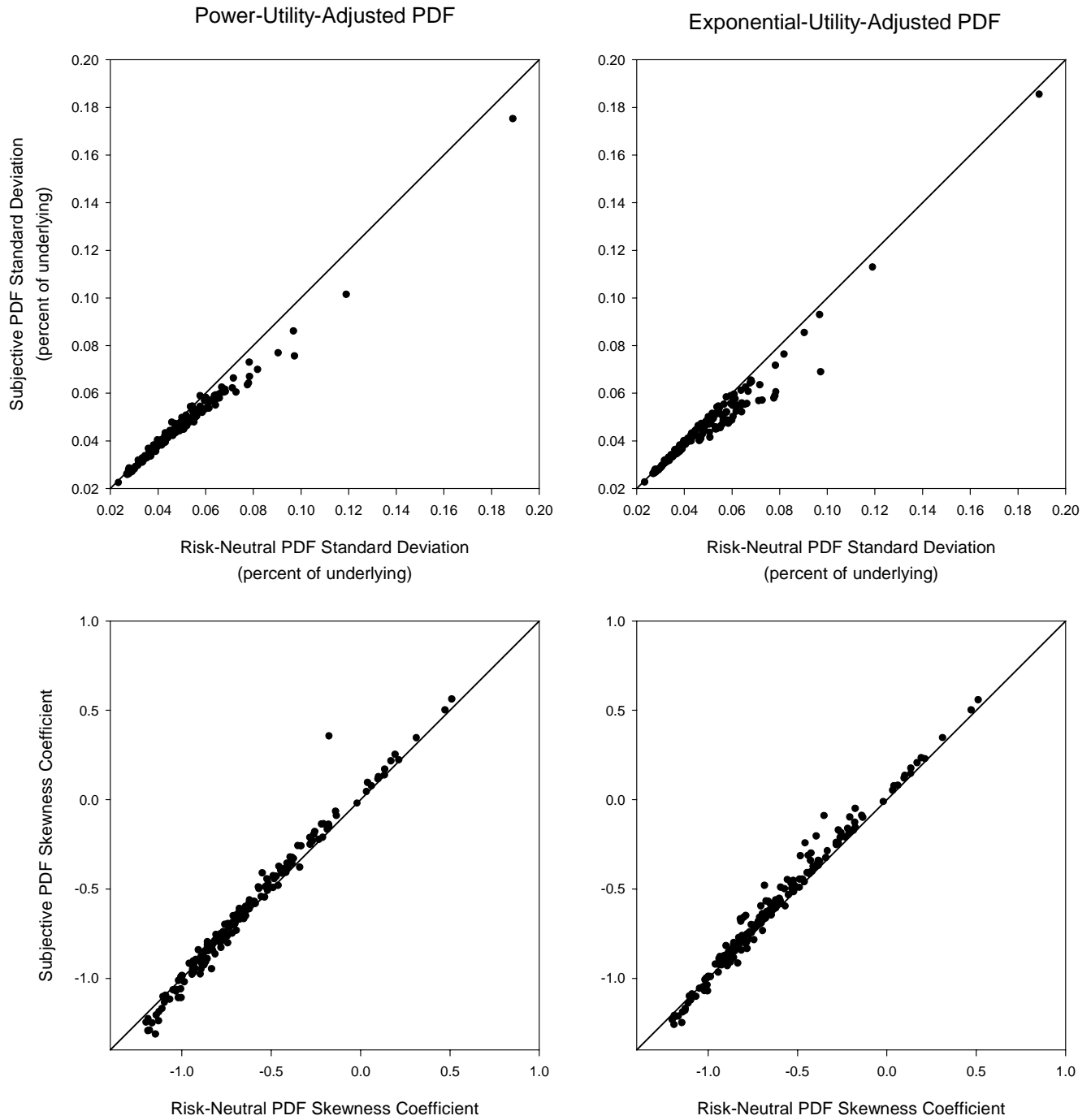


**Figure 2. Cross-Validation Results.**



**Figure 3. Plot of risk premia implied by risk-neutral and subjective PDFs.**





**Figure 4. Comparison of standard deviations and skewness coefficients from risk-neutral and subjective PDFs.**