Forecasting Future Variance From Option Prices

Allen M. Poteshman

Allen M. Poteshman
Department of Finance
University of Illinois at Urbana-Champaign

OFOR Paper Number 00-07
September 2000

Forecasting Future Variance from Option Prices

Allen M. Poteshman*

March 2000

* Department of Finance, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign. Helpful discussions with Jegadeesh Narasimhan and Neil Pearson are gratefully acknowledged. The author bears full responsibility for all errors. Address correspondence to Allen M. Poteshman, Department of Finance, University of Illinois at Urbana-Champaign, 304E David Kinley Hall, 1407 West Gregory Drive, Urbana, IL 61801; Tel: (217) 244-6651; Fax: (217) 244-9867; E-mail: poteshma@uiuc.edu.

Forecasting Future Variance from Option Prices

Abstract

Although it is widely believed that option prices provide the best possible forecasts of the future variance of the assets which underlie them, a large body of empirical evidence concludes that option prices consistently yield biased forecasts of future variance. The prevailing interpretation of these findings is that option investors may be forming unbiased forecasts of the future variance of underlying assets but that these unbiased forecasts fail to get impounded into option prices because of either (1) the difficulty of carrying out the necessary arbitrage strategies that would force the prices to their proper levels, or (2) the availability to market makers of lucrative alternative strategies in which they simply profit from the large bid-ask spreads in the options markets. This interpretation has significant consequences for nearly the entire range of option pricing research, since it implies that non-continuous trading, bid-ask spreads, and other market imperfections substantially influence option prices. This implication is important, both because incorporating these types of market imperfections into option pricing models is much more difficult than, for example, altering the dynamics of the underlying asset and also because it suggests that researchers cannot learn about option investor expectations by filtering option prices through available option pricing models.

The present paper studies the variance forecasting ability of SPX option prices against the backdrop of the prevailing interpretation of the findings in the variance forecasting literature. The paper presents two main empirical findings. First, approximately one third of the usual bias is eliminated when high frequency futures data rather than daily closing data is used to construct measures of realized variance. Second, roughly another third of the bias disappears when forecasts of future variance are extracted from option prices via an option pricing model that – unlike the commonly employed model – permits a non-zero market price of variance risk and a non-zero correlation between innovations to the level and variance of the SPX index. Furthermore, the remaining bias is no longer significant. In addition to the empirical results, Monte Carlo simulations are performed to study the impact on the results of model misspecification and errors in the futures and options data. The simulations indicate that failure to account for a non-zero market price of variance risk produces a forecasting bias similar to that found in the previous literature when the conventional option pricing model is employed but that errors in the variables do not produce appreciable bias.

It is widely believed that option prices provide the best forecasts of the future variance of the assets which underlie them. One reason for this belief is that option prices have the ability to impound all publicly available information – including all information contained in the history of past prices – about the future variance of the underlying assets. A second related reason is that option pricing theory maintains that if an option price fails to embody an optimal forecast of the future variance of the underlying asset, a profitable trading strategy would be available whose implementation would push the option price to the level that reflects the best possible forecast of future variance.

A large empirical literature, however, provides evidence that option prices do a poor job of forecasting future variance. In an important study Canina and Figlewski (1993) report that in the OEX (i.e. S&P 100 index) options market

> Implied volatility is an inefficient and biased forecast of realized future volatility that does not impound the information contained in recent historical volatility. In fact, the statistical evidence shows little or no correlation at all between implied volatility and subsequent realized volatility.

While other studies have not reached the extreme conclusion that option implied variances contain essentially no information about future realized variances, they almost uniformly find that option prices provide irrational, biased predictions of future realized variance. This is the finding, for example, of Lamoureux and Lastrapes (1993) for options on individual stocks, of Fleming (1998) in a more recent study of the OEX options market, and of Jorion (1995) for options on foreign currency futures.

The full range of possible explanations for the failure of option prices to provide accurate forecasts of future variance will be reviewed in the next section of the paper. The prevailing standard explanation for this failure is presented by Figlewski (1997) in his authoritative monograph on variance forecasting. He argues that investors may well be making optimal forecasts of future variance but that these forecasts are not reflected in option prices because of frictions in the options markets. Commenting on the results of Canina and Figlewski (1993), Figlewski (1997) writes that

> Investors could be making appropriate use of available information about the volatility of the underlying stock index in forming their expectations, but these expectations do not translate properly into option prices in the market. This is our preferred explanation for the failure of the rationality test regression to show any predictive power for OEX implied volatilities. (p.64)

This interpretation maintains that the translational failure results from obstacles to executing the arbitrage strategies that would profit from the optimal variance forecasts that investors are presumed to possess. Figlewski (1997) identifies the likely obstacles by asking us to

> Consider the difficulty a trader would face in trying to profit from a belief that an OEX option's price did not accurately reflect the best forecast of future volatility. In theory, an investor who knows an option to be mispriced should lock in a riskless excess return by forming a hedged position with the option and the underlying asset, and rebalancing the proportions frequently (continuously) over the option's lifetime. But following such a trading strategy when the underlying is a large portfolio of 100 stocks is very expensive in terms of transaction costs, and is also exposed to considerable execution risk. (p. 64)

The expense and risk inherent in trading strategies that attempt to realize arbitrage profits from options whose prices do not reflect accurate variance forecasts may be sufficient to discourage investors from attempting to implement them. In addition, there may be simpler, less expensive, less risky ways for option market makers to realize profits. Figlewski (1997) suggests that

> Trading options by betting that one's volatility estimate is a more accurate forecast than that contained in implied volatility is unlikely to be an optimal strategy for an active options market maker. Rather, basing option bids and offers on the current implied volatility (whether or not it is an accurate prediction) will lead to transactions that are largely balanced between buying and selling, and the resulting turnover should produce an ongoing flow of profits from the bid-ask spread. (p. 64)

Based on this interpretation of the Canina and Figlewski (1993) results, Figlewski (1997) hypothesizes that the forecasting performance of implied variance across options markets will be inversely related to the difficulty of executing the arbitrage trading strategy between the options and underlying asset. A survey of the predictive ability of implied variance in different markets finds some support for this hypothesis.

Figlewski (1997)'s interpretation of the variance forecasting literature attributes the poor variance forecasting performance of option prices largely to violations of the underlying option pricing model's assumptions of continuous trading and no transactions costs. A recent paper by Christensen and Prabhala (1998) reinforces the Figlewski interpretation. Christensen and Prabhala (1998) study the variance forecasting performance of OEX options. As in previous studies, a straightforward comparison of the realized variance with the implied variance indicates that implied variance is a biased predictor. It is shown, however, that an instrumental variable correction for error in the predicted variance variable that is orthogonal to (i.e., uncorrelated with) investors's true variance predictions substantially improves the performance of option price based predictions of future realized variance.[1] This finding suggests that options market imperfections which produce unsystematic option price deviations from model prices substantially influence the option prices from which forecasts of future variance are extracted.

From the point of view of option pricing *theory* the Figlewski interpretation of the variance forecasting results has dire consequences. It has been known for some time that the Black-Scholes formula does not do a very good job of pricing options. While it is obvious that

[1] In the Christensen and Prabhala (1998) paper an F-test is performed on the regressions where the dependent variable is the uncorrected option based prediction of future variance. This F-test rejects the null hypothesis that the option based predictions are rational forecasts. No such test is performed for the instrumental variable regression where the dependent variable has been corrected for error orthogonal to the true predicted variance. Consequently, although it appears that the instrumental variable correction substantially improves the option price based predictions, there is no formal statistical tests of whether the null hypothesis of rational variance forecasts is still rejected.

each of the assumptions underlying the Black-Scholes formula is false, the vast majority of theoretical research has focused on relaxing the assumption that the dynamics of the underlying asset obey geometric Brownian motion. The reasons for this focus are both the belief that violations of this assumption are the most important causes of the empirical shortcomings of the Black-Scholes model and also the fact that the model remains tractable under a number of richer specifications of the dynamics of the underlying asset. The Figlewski interpretation of the results in the variance forecasting literature, on the other hand, suggests that violation of some other set of assumptions of the Black-Scholes model causes substantial differences between observed market option prices and their theoretical values which results in the poor performance of option price implied forecasts of future variance. In particular, the assumptions of continuous trading, zero bid-ask spreads, and perfect liquidity are identified as the culprits. If the Figlewski interpretation is correct, then the theoretical option pricing literature's focus on generalizing the dynamics of the underlying asset may be misplaced. At the least, it suggests that more effort should be applied to generalizing the assumptions of continuous trading, no bid-ask spreads, and perfect liquidity. Since relaxing these assumptions introduces a number of thorny problems, it is important to make sure that the results from the literature that forecasts future variance from option prices necessitates this step.

This paper addresses the question of whether violation of option pricing model assumptions other than that on the dynamics of the underlying asset do, in fact, produce biased forecasts of future variance from option prices. This question is investigated in three steps. First, the impact of constructing the forecast and realized variance variables in progressively cleaner ways is examined. Second, the effect of using the Heston (1993) model to forecast future variance from option prices is determined. The Heston (1993) model is more general than the option pricing models used in previous studies insofar as it allows a non-zero market price of variance risk and a non-zero correlation between the level of the underlying asset and its instantaneous variance. Finally, Monte Carlo studies are performed under alternative assumptions on the market price of variance risk and the errors in the futures and options prices that are used to form the forecasting and realized variance variables in order to assess the significance of the various results.

SPX (i.e. S&P 500 index) options will be used in the study for two reasons. First, since these options are on an equity index, the Figlewski hypothesis maintains that these options are among the most likely to have their forecasts of future variance adversely impacted by violations of option pricing model assumptions other than the assumption on the dynamics of the underlying asset. Second, since unlike OEX options, SPX options are European and do not have a wildcard feature the possible distortionary effects of these option characteristics are completely absent.

The main findings are as follows. First, when the standard forecasting regressions are run using specifications and data processing procedures like those in previous research, the size and significance of the forecasting bias from SPX options is similar to that found for OEX options (e.g. in Fleming (1998) or Christensen and Prabhala (1998).) Second, cleaner methods for constructing the forecasted variance and realized variance variables eliminate about one third of the bias. Third, approximately another third of the bias is eliminated by extracting variance predictions from option prices using the Heston (1993) model which generalizes the assumptions on the dynamics of the underlying assets relative to the model used in other studies of variance forecasting while leaving the previous non-dynamical assumptions (e.g. continuous trading, no bid-ask spreads, perfect liquidity) in place. The remaining bias is not significant. Fourth, Monte

Carlo simulations indicate that failure to account for a non-zero market price of variance risk produces a forecasting bias similar to that found in the previous literature when the conventional option pricing model is employed but that errors in the variables do not produce appreciable bias.

The remainder of the paper is organized as follows. The first section sets out the standard approach to testing the variance forecasting performance of option prices, reviews some of the important empirical findings, and identifies all possible sources of the bias that has been reported in the literature. Section two describes the data. The third section reports the results of variance forecasting tests under the conventional option pricing model both when standard data processing procedures are followed and when attempts to remove further error from both the predicted variance and the forecast variance variables are made. Section four tests the variance forecasting ability of option prices under an option pricing model that permits a non-zero market price of variance risk and a non-zero correlation between innovations to the level and the variance of the underlying asset. The fifth section performs a Monte Carlo simulation to study the impact of model misspecification and errors in the futures and options data on the results. Section six concludes.

I. Variance Forecasting from Option Prices: Standard Procedures and Previous Results

Studies of the variance forecasting performance of option prices typically run the following two regressions

$$f\left(Var_{Realized}\left(t\right)\right) = \alpha + \beta f\left(Var_{Implied}\left(t\right)\right) + \varepsilon\left(t\right) \tag{1}$$

and

$$f\left(Var_{Realized}\left(t\right)\right) = \alpha + \beta f\left(Var_{Implied}\left(t\right)\right) + \gamma f\left(Var_{Historical}\left(t\right)\right) + \varepsilon\left(t\right) \tag{2}$$

where $Var_{Realized}(t)$ is a measure of the *ex-post* average realized variance of the underlying asset over the period *t* to $t+T$, $Var_{Implied}(t)$ is the average variance forecast over the period *t* to $t+T$ implied from the price of an option observed at time *t* that expires at time $t+T$, $Var_{Historical}(t)$ is the average variance forecast over the period *t* to $t+T$ from some subset of the information available to the market at time *t* (usually from the history of past returns on the underlying asset), $\varepsilon(t)$ is a zero mean forecast error that is uncorrelated with the forecasting variables, and $f(\bullet)=\sqrt{\bullet}$. Some researchers have used other specifications for $f(\bullet)$.

The results of these regressions are used to draw conclusions about three main issues. First, it is claimed that option prices are *informative* about future variance if and only if the estimate of β is significantly greater than zero. Second, it is claimed that option prices contain *biased* forecasts of future variance if and only if the joint hypothesis of α being equal to zero and β being equal to one is rejected. Finally, it is claimed that option prices contain *informationally inefficient* estimates of future variance if and only if the estimate of γ is statistically distinguishable from zero.

The extant studies on forecasting variance from option prices are usefully broken into two periods. In the earlier period options had not been trading on exchanges for a long enough period of time to conduct time-series studies. Consequently, these earlier studies (Black and Scholes (1972), Latané and Rendleman (1976), Chiras and Manaster (1978), Beckers (1981)) examine a cross-section of options written on different underlying securities. More recent studies (Canina and Figlewski (1993), Christensen and Prabhala (1998), Day and Lewis (1992), Fleming (1998), Jorion (1995), Lamoureux and Lastrapes (1993), Day and Lewis (1993)), on the other hand, run the above regressions in a time-series framework using options written on only one underlying asset. Both types of studies generally find that option-implied forecasts of future variance are *informative* but also that they are *biased*. The results on *informational efficiency* are mixed. The estimates from regression (1) and (2) for some of the more recent time-series studies are presented in Table I. On the whole, the bias increases with the difficulty of performing arbitrage trades in a market to take advantage of mispricing with the bias greatest for options on stock indices and least for options on crude oil futures which trade side by side with the underlying futures in the same market. This observation provides some evidence in favor of the Figlewski hypothesis discussed in the introduction. Note, however, that the relationship between the degree of bias and the difficulty of performing the arbitrage trades is somewhat loose.

The bias that is consistently found in studies of the variance forecasting ability of option prices constitutes a rejection of the joint hypothesis of market efficiency and the option pricing model that is used to extract variance forecasts from option prices. The logical space for explaining this rejection is circumscribed by four possibilities: (1) the market under study is inefficient; (2) improbable events are observed; (3) the econometric method utilized is flawed; and/or (4) the presumed option pricing model is misspecified. These possibilities will be discussed in turn.

According to the market inefficiency explanation for the bias in the variance forecasting studies, the finding that β is statistically less than one indicates that option market forecasts of future variance are more variable than actual realized future variance. Consequently, in order to get the best fit in the regressions of realized variance to forecast variance, the forecasts need to be dampened with a coefficient of less than one. Such forecasts are irrational since the time series of *ex ante* expected values of a quantity should always be less variable than the time series of *ex post* realized values of the quantity.[2] It is possible that investors systematically make these types of irrational forecasts across a wide variety of options markets. However, given the high liquidity and large number of professional traders present in some of the markets that have been studied, a market inefficiency interpretation should be entertained only after every possible alternative explanation has been exhausted.

The second possible explanation for the poor quality of variance forecasts derived from option prices is simply that improbable events are observed in the various markets. According to this interpretation of the findings the *ex ante* variance forecasts are unbiased and efficient, but the *ex post* observed variances are unusual (relative to their true distribution) so that it appears that the *ex ante* forecasts are biased or inefficient. Although this possibility cannot be excluded, it is highly unlikely that improbable events have been observed in all of the different markets that have been studied.

The third possible explanation for the forecasting bias documented in the literature is econometric problems. These problems could take a number of forms, but the two most likely culprits are inaccurate standard errors for the regression coefficients because of overlapping observations and measurement error in the forecasting variable, $f\left(Var_{Implied}\left(t\right)\right)$. Most of the studies use daily data and options with horizons of one to three months. As a result, there is overlap in the error term which biases the usual OLS standard errors downward (although this problem does not cause inconsistent point estimates). Beginning with the Canina and Figlewski (1993) study, researchers have addressed this problem by computing standard errors according to the GMM method of Hansen (1982). Fleming (1998) develops a modified GMM estimator to

[2]This interpretation of the evidence would directly parallel the interpretation given to the fact that the variability of stock prices is greater than the variability of the future realized dividend stream by Shiller (1981) and LeRoy and Porter (1981) in the literature on the excess variability of stock prices.

account for the fact that the option maturity shrinks over time as the option approaches maturity. Even Fleming's modified GMM estimator, however, corrects the standard errors only asymptotically. This may present a problem, because it is not clear how "large" the samples in the forecasting studies are. Although some studies use thousands of options, the number of truly independent forecasts is only equal to the number of years of data times twelve (if one month options are used). Canina and Figlewski (1993) and Jorion (1995) run simulations that construct confidence intervals for OLS point estimates of the coefficients of regressions (1) and (2) for the actual sample sizes used in their empirical work. These simulations show that in the absence of error in the variables, it is unlikely that the literature's conclusion that option-based forecasts of future variance are biased is caused by faulty standard error calculations.

A second econometric problem that could spuriously produce the forecasting bias seen in the literature is error in the forecasting (i.e. independent) variable in regressions (1) and (2). For simplicity, focus on specification (1). If there is measurement error in the $Var_{Implied}(t)$ variable that is uncorrelated with its true value (henceforth, referred to as *orthogonal* error), then it is well-known that the β estimate will be biased downward and that the bias can be corrected by an instrumental variable procedure (Johnston (1984),Chapter10). The possibility that such *orthogonal* error causes the forecasting bias is pursued by Christensen and Prabhala (1998). They argue that measurement error could be present in the $Var_{Implied}(t)$ variable that is uncorrelated with its true value, because of nonsynchronous measurement of option prices and index levels, early exercise and dividends, bid-ask spread, the wild-card option in the OEX market, and/or misspecification of the process governing index returns. When Christensen and Prabhala (1998) estimate equation (1) using data from the OEX option market and a sampling procedure in which the forecast intervals do not overlap, they get estimates of

$\alpha = -0.56$ $(t_{\alpha=0} = -3.47)$ and $\beta = 0.76$ $(t_{1-\beta=0} = 2.99)$ (where the *t*-statistic is derived from the reported *t*-statistic for $\beta = 0$.) When past implied volatility is used as an instrument, the instrumental variable estimates are given by $\alpha = -0.15$ $(t_{\alpha=0} = -0.63)$ and $\beta = 0.97$ $(t_{1-\beta=0} = 0.25)$. Similar results holds for regression equation (2).

Do the Christensen and Prabhala (1998) results then demonstrate that orthogonal error in the forecasting variable explains the mystery of option prices inability to provide unbiased forecasts of future variance as hypothesized by Figlewski (1997)? The answer to this question is no for a number of reasons. First of all, it is unlikely that the error in $Var_{Implied}(t)$ is uncorrelated with its true value. By setting $Var_{Implied}(t)$ equal to the Black-Scholes implied variance of a close to at-the-money (henceforth, ATM) option, Christensen and Prabhala (1998) assume that the dynamics of the OEX index obey a stochastic variance model in which innovations to the level of the stock price and innovations to the level of the variance are uncorrelated and the market price of variance risk is zero. It is well known, however, that there is a substantial negative correlation between the level and variance of the OEX index. In addition, it is likely that the market price of variance risk is nontrivial in this market. (In the SPX market – which most probably has similar properties to the OEX market – Pan (1999) and Poteshman (1999) find evidence of a substantial market price of variance risk.) Furthermore, the true process for the OEX index may have jumps or be non-Markovian. Any of these types of misspecification of the dynamics of the underlying asset may well produce error in the $Var_{Implied}(t)$ variable which is correlated with its true value. In the presence of such errors the assumptions of instrumental variable estimation are violated. In particular, the instrument chosen by Christensen and Prabhala (1998) (last period's implied variance) will be correlated with the error in the current period's implied variance. Consequently, it cannot be concluded from the fact that the instrumental variable procedure appears to remove the bias from the regression coefficients that orthogonal error in the $Var_{Implied}(t)$ variable causes the forecasting bias documented in the literature.[3]

A second reason to question whether the Christensen and Prabhala (1998) results show that orthogonal error in the $Var_{Implied}(t)$ variable produces the poor performance of variance forecasts from option prices is that orthogonal error in the $Var_{Implied}(t)$ variable biases the estimate of the intercept in (the uncorrected) regression (1) upward. (See Johnston (1984), pp. 428-430). As noted above, however, the Christensen and Prabhala (1998) estimate of the coefficient in regression (1) is $\alpha = -0.56$ $(t_{\alpha=0} = -3.47)$. Since this estimate is statistically less than zero, it provides direct evidence that the forecasting bias is not the result of orthogonal error in $Var_{Implied}(t)$. Further evidence that orthogonal error in the $Var_{Implied}(t)$ variable is not the main source of the forecasting bias documented in the literature comes from the simulations presented in Jorion (1995) which show that plausible orthogonal errors in this variable do not produce a large enough downward bias in the estimate of the β coefficient.

The fourth possible explanation for the biased and sometimes inefficient nature of variance forecasts from option prices is that the presumed option pricing model is misspecified. It will be helpful to break model misspecification into two categories. The first category is misspecification of the dynamics of the underlying asset, and the second category is violation of

---

[3] In addition, Christensen and Prabhala (1998) choose the specification $f(\bullet) = \ln(sqrt(\bullet))$. It will be seen below that this choice significantly changes the results of the estimation relative to the choices $f(\bullet) = I(\bullet)$ or $f(\bullet) = \sqrt{\bullet}$. Consequently, their findings may result from their specification as $f(\bullet)$ rather than from their orthogonal error in the $V_{Implied}(t)$ variable.

any of the model's other assumptions (e.g. continuous trading, perfect markets). Misspecification in the second category will tend to introduce orthogonal error into the forecasting variable which has already been considered in the discussion of econometric explanations. Consequently, the focus here will be on misspecification of the dynamics of the underlying asset.

The early studies and the Canina and Figlewski (1993) study use the Black-Scholes model to extract variance forecasts from option prices. Accordingly, these studies assume that the underlying asset follows a geometric Brownian motion with constant variance. The more recent time-series studies, however, use a more sophisticated stochastic variance model. The model that these studies employ is described by the following equations:

$$\frac{dS_t}{S_t} = \mu\left(S_t, V_t, t\right)dt + \sqrt{V_t}\,dW_t^S \tag{3}$$

$$dV_t = \phi\left(V_t, t\right)dt + \xi\left(V_t, t\right)dW_t^V \tag{4}$$

$$Corr\left(dW_t^S, dW_t^V\right) = 0 \tag{5}$$

$$\lambda\left(S_t, V_t, t\right) = 0 \tag{6}$$

$$r = constant \tag{7}$$

where S_t is the level of the underlying asset, V_t is the instantaneous variance of the underlying asset, $\lambda\left(S_t, V_t, t\right)$ is the market price of variance risk, and r is the instantaneous risk-free rate. Feinstein (1989) shows that under this model the Black-Scholes implied variance of an ATM option is almost exactly equal to the average variance expected over the life of the option. The more recent time-series papers exploit this fact to extract variance forecasts from model (3)-(7) by using the Black-Scholes implied variance of closest to ATM options as the forecast for the average variance over the life of the option.

The question then is whether misspecification contained in equations (3)-(7) might be producing systematic errors in the predictions of future variance from option prices that results in the forecasting bias. Some researchers have dismissed this possibility. Christensen and Prabhala (1998) write that

> Following Cox, Ingersoll, and Ross (1985) and Lamoureux and Lastrapes (1993) the Black-Scholes implied volatility is approximately equal to expected future return volatility for at-the-money options (used in our study) even when returns follow the (non-Black-Scholes) stochastic volatility model of Hull and White (1987). Thus, misspecification error should be small and there is little reason for it to be correlated across time.

Fleming (1998), on the other hand, states that his results "should not be viewed as a test of market efficiency" largely because of a concern that model (3)-(7) might be misspecified. Finally, Lamoureux and Lastrapes (1993) are clearly concerned that misspecification in model (3)-(7) may be influencing their results. Consequently, they run simulations to test whether the zero-correlation assumption contained in equation (5) and the slight non-linearity in variance of the Black-Scholes formula for close to ATM options has the potential to explain the bias that they find in option implied forecasts of future variance. Although they conclude that these two factors are insufficient to explain the bias, they allow that the zero market price of variance risk assumption contained in equation (6) could play an important role in explaining their results. Indeed, they write that

> One possible reason for the rejection of the null is that volatility risk is priced. Therefore, further attempts to learn from the data should explicitly model a risk premium on the variance process. (pp. 296-297)

There are a number of ways that dynamics (3)-(7) might be misspecified that may have an important impact on the ability to forecast future variance from option prices. Attempts to check the impact of the zero correlation assumption in equation (5) indicate that it is not an important factor in producing the forecasting bias. An important open question, however, is whether the zero market price of variance risk assumption contained in equation (6) plays an important role in producing the forecasting bias.

This section of the paper has argued that market inefficiency and unusual *ex post* draws out of the realized variance distribution are not promising candidates for explaining the observed bias in option price based forecasts of future realized variance. It has also suggested that there is not clear evidence that the forecasting bias comes from orthogonal error in the forecasting variable which is the preferred explanation of Christensen and Prabhala (1998) and would be consistent with the Figlewski hypothesis. The remainder of this paper will investigate the extent to which the bias can be explained by errors in the variables that can be removed through more careful empirical procedures and an option pricing model that allows for a more general specification of the dynamics of the underlying asset.

II. Data

In order to determine whether the difficulty in executing the arbitrage trades envisioned in standard option pricing models leads to the biased forecasts of future variance from option prices that is consistently found in the literature, the SPX option market will be studied. This market is studied, because it is one in which the arbitrage trades are likely to be among the most difficult. Figlewski (1997) claims that with respect to the difficulty of executing the arbitrage trades required by option pricing theory

> Stock index options represent a polar case, where the arbitrage trade is complicated to execute at the outset, and the resulting position is both costly and risky to hedge over time. (p. 78)

SPX options are studied rather than OEX options, because SPX options are European and do not have a wildcard option. Consequently, there is no possibility that the inability to properly account for early exercise or the wildcard feature will influence the results. It is not obvious whether hedging SPX options is more or less difficult than hedging OEX options. On the one hand, the SPX index contains five times as many stocks as the OEX options and its average stock is much smaller. As a result, hedging the SPX option with its underlying asset is surely more difficult than hedging the OEX option with its underlying asset. On the other hand, both of these types of options could be hedged with SPX futures. It is clear that SPX option can be better hedged with SPX futures than can OEX options. SPX futures, however, only have expirations in March, June, September, and December. Accordingly, on most months neither SPX nor OEX options can be hedged with futures that have the same expiration date. Furthermore, SPX and OEX options trade at the Chicago Board Options Exchange while the SPX futures trade at the Chicago Mercantile Exchange. Despite the fact that no definitive statement can be made about whether SPX or OEX options are more difficult to hedge, both markets should be among those where it is most difficult to implement option pricing theory's arbitrage strategy.

SPX options trade at the Chicago Board Options Exchange (CBOE) with expiration dates in the three near term months along with the following three months from the March expiration cycle (March, June, September, December). The options expire on the third Friday of the contract month. Strike price intervals are 5 points for near months and 25 points for far months. The minimum tick for options trading below $3.00 is 1/16 and 1/8 for options trading at higher prices.

Bid-ask price quotes that are time-stamped to the nearest second were obtained directly from the CBOE for the period June 1, 1988 through August 29, 1997. Data are available from

October 2, 1985 through August 29, 1997. The data from October 2, 1985 through May 31, 1988 is not used because of the evidence presented in Jackwerth and Rubinstein (1996) that there is a structural break in the SPX market at the time of the October 1987 stock market crash. In addition, the market was considerably less liquid during its earlier years, so the data may not be as reliable. Consequently, The main data period for test that use the Black-Scholes model to extract predictions of future variance from option prices is June 1, 1988 through August 29, 1997. The Heston (1993) model was published in the summer of 1993. As a result, the data period for tests which use the Heston model to make predictions of future variance from option prices is June 1, 1993 through August 29, 1997.

On each trade date, for each strike price and time-to-expiration information is extracted on the last bid-ask quote reported prior to 3:00 PM Central Standard Time.[4] For each of these bid-ask quotes, the bid-ask midpoint, the time-to-expiration, the exercise price, and the type of option (i.e. call or put) are recorded. Before August 24, 1992 the time-to-expiration is measured as the number of calendar days from the trade date to the Friday before the third Saturday of the expiration month. After August 23, 1992 the time-to-expiration is measured as the number of calendar days from the trade date to the Thursday immediately preceding the Friday before the third Saturday of the expiration month. This method of computing time-to-expiration is followed, because prior to August 24, 1992 SPX options expired at the close of trading and since August 24, 1992 SPX options have expired at the open of trading (see Dumas, Fleming, and Whaley (1998).) The risk-free rate of interest associated with each option is the one month LIBOR rate on the day the option bid-ask quote is observed. Table II provides descriptive

[4] The options trade on the CBOE from 8:30 AM to 3:15 PM Central Standard Time. Bakshi, Cao, and Chen (1997) also use the last reported bid-ask quote before 3:00 PM Central Standard Time. This choice is also in line with Dumas, Fleming, and Whaley (1998) who use the options in the 2:45 PM to 3:15 PM Central Standard Time window in their tests.

statistics on the options data over the period June 1993 through August 1997 which is the period of greatest of interest for this paper. The descriptive statistics are for ATM and OTM (i.e. out-of-the-money) options which are the ones that will be used in this paper. In particular, the descriptive statistics describe those options that (1) have time-to-expiration of greater than or equal to six calendar days and less than or equal to 7/12 of a year, (2) have a bid price of greater than or equal to $3/8 and a bid-ask spread of less than or equal to $1, (3) a Black-Scholes implied volatility greater than zero and less than or equal to 0.70, and (4) are ATM or OTM where ATM is defined as the strike price being equal to the CRSP closing value of the SPX index. ATM is defined this way only for purposes of inclusion or exclusion in Table II and to select options for estimating the Heston model in Section IV below. Throughout the rest of the paper ATM is defined as the strike price being equal to the futures price.

The fact that SPX options are European and do not have a wildcard feature considerably simplifies empirical work with them. Nonetheless, two serious challenges remain. The first challenge is matching observations on option prices to observations on the underlying index level. Even if a quote on the underlying index can be exactly matched temporally to an option price, the quoted index level will not be the proper underlying value for the option, since all 500 underlying stock prices will not correspond to trades that occurred at the quoted time. The second challenge is determining the *expected* future rate of dividend payments by the stocks that compose the index until the expiration of an option. Of course, dividend rates can always be calculated after the fact from the actual dividends paid out by the SPX stocks. The *ex-post* rate, however, may not match the *ex-ante* expectation at the time the option is priced.

When addressing these two challenges, it is not necessary to determine the contemporaneous underlying value and dividend rate separately. It is sufficient to determine the quantity $Se^{-\delta T}$, where δ is the dividend rate paid by the index and T is the time to expiration of the option. If an accurate value for $Se^{-\delta T}$ can be determined, then the option can be priced under the assumption that the underlying value is equal to $Se^{-\delta T}$ and that the dividend rate is zero. (This is because a stock with a continuous dividend yield δ which grows from S at time 0 to S_T at time T will grow from $Se^{-\delta T}$ at time 0 to S_T at time T if it pays no dividend.) The quantity $Se^{-\delta T}$ will be determined from transactions data on SPX futures which trade on the Chicago Mercantile Exchange (CME).

Data on SPX futures were obtained from the Futures Industry Institute which is the official data supplier for the CME. The SPX futures trade with delivery dates in March, June, September, and December. On a given trade date there is typically one delivery date which is extremely liquid (i.e. thousands of transactions per day) and a second delivery date which is moderately liquid (i.e. on the order of thirty transactions per day). Consequently, the quantity $Se^{-\delta T}$ associated with an option bid-ask quote will be determined from the futures data in one of two ways depending on whether for a given trade date the delivery date of the most liquid futures contract is the same as the expiration date of the option.

The SPX futures contracts expire at the open of the trading day throughout the time period studied in this paper. Taking this into account, when a given option bid-ask quote observation corresponds to an option that expires at the time of delivery of the most liquid futures contract, the transaction price of the futures contract which trades at the time closest to the observation of the option bid-ask quote is used in conjunction with spot-futures parity to determine the quantity $Se^{-\delta T}$. In particular, spot-futures parity is used to determine $Se^{-\delta T}$ from

$$Se^{-\delta T} = Fe^{-rT}$$

where F is the futures prices, r is the one month LIBOR rate, and T is the time to expiration of the option (or, equivalently, the time to delivery of the futures) computed as specified above.

When, on the other hand, an option corresponding to a bid-ask quote expires at a time other than that of the most liquid futures contract, the following procedure is used to compute $Se^{-\delta T}$. First, three futures are identified. Let F_1 be the transaction price of the futures contract with the most liquid delivery date that transacts closest in time to the option bid-ask quote. Let F_2 be the transaction price of the futures contract with the second most liquid delivery date that transacts closest in time to the option bid-ask quote. Let F_3 be the transaction price of the futures contract with the most liquid delivery date that transacts closest in time to F_2. Finally, let T_1, T_2, and T_3 be the time to delivery of, respectively, F_1, F_2, and F_3. Then the dividend rate is computed via spot-futures parity from F_2 and F_3 (which are typically observed within a few seconds of each other) by

$$\delta = r - \frac{\log\left(F_3/F_2\right)}{\left(T_3 - T_2\right)} \tag{8}$$

This dividend rate is then used in conjunction with F_1 (which is typically observed within a few seconds of the option bid-ask quote) and spot-futures parity to compute

$$Se^{-\delta T} = F_1 e^{\left(-(r-\delta)*T_1 - \delta*T\right)}$$

which is associated with the bid-ask quote for the option that expires at a time T in the future.[5]

---

[5] To get the final equation, begin with $Se^{-\delta T_1} = F_1 e^{-rT_1}$ from spot-futures parity. Then multiply both sides by $e^{\delta(T_1 - T)}$. Aï t-Sahalia and Lo (1998) provide an alternative method for computing $Se^{-\delta T}$ that employs put-call parity and requires only data on a call and a put option, rather than data on an option and a future. The Aï t-Sahalia and Lo (1998) method has the advantage of requiring no auxiliary assumption of integration between the options and futures markets which are housed at different exchanges. The disadvantage of their technique is that options markets tend

As was noted above, the Black-Scholes implied variance of an ATM option is almost exactly equal to the expected average variance over the life of the option under the model described by equations (3)-(7). Tests that use (3)-(7) as the option pricing model proceed first by determining the quantity $Se^{-\delta T}$ from the appropriate futures prices. Once the quantity is in hand, the Black-Scholes implied variance is determined by numerically inverting the Black-Scholes formula under the assumptions that the level of the underlying asset is $Se^{-\delta T}$ and that the underlying asset pays no dividends.

III. Forecasting Variance from Option Prices using the Standard Option Pricing Model

Most of the recent literature on the quality of variance forecasts from option prices uses an option pricing model in which the dynamics of the underlying asset are described by equations (3)-(7). This section of the paper will report the results of forecasting variance from SPX options under this model. Different aspects of the regression specification and data processing procedure will be varied one at a time in order to determine their impact on the variance forecast regression results.

The sampling procedure of Christensen and Prabhala (1998) is adopted throughout the paper. SPX options expire on the Friday before the third Saturday of each month. Following Christensen and Prabhala (1998) prices of options that expire the next calendar month are observed on the Wednesday after each expiration date. The September 1992 observation date is missing, because short term option prices are missing from the CBOE data for the required

---

to be less liquid than the futures markets and to have larger bid-ask spreads. Experimentation with the data indicated that in the present context the method employed here produces cleaner values for $Se^{-\delta T}$.

Wednesday date and for several trade dates on either side of the required date. The option prices are used to predict the future variance until the options expire – three and a half to four and a half weeks later. The main virtue of this sampling procedure is that there is absolutely no overlap in the data. As a result there is no need to account for overlapping data when making statistical inferences. Of course, a good argument can be made for using option prices from every trade date and then correcting the standard errors for overlap in the data using GMM type adjustments in order to obtain the statistically most powerful tests. This procedure was not followed, however, because the GMM corrections are only valid asymptotically and the number of independent observations in the variance forecasting regressions is at most the number of months in the sample. Since GMM is known to sometimes have poor performance in small samples, it would be difficult to have confidence in GMM type corrections in the present context. Instead, the non-overlapping sampling procedure is adopted and Monte Carlo simulations are performed to construct the distributions of the coefficient estimates in regressions (1) and (2). In principle, such simulations could be performed using an overlapping methodology. In the present context, however, this is not feasible, because of the computational cost of estimating the Heston (1993) model which will be used to forecast future variance from option prices in the next section of the paper. From the point of view of computational efficiency, the chosen sampling procedure can be viewed as the one that maximizes the amount of information extracted from the option prices while holding fixed the number of time points at which option prices are observed.

The realized variance of the SPX index over some time period from t to $t+T$ is computed as follows. Let $\Delta = T/N$ be a fixed interval of trading time (measured in years) and let $\left\{S_t, S_{t+\Delta,} S_{t+2\Delta}, ..., S_{t+N\Delta}\right\}$ be the $N+1$ evenly spaced (in trading time) observed levels of the SPX index over the time period from t to $t+T$. Define the return from time $t+(i-1)\Delta$ to time

$t+i\Delta$ to be $R_{t+i\Delta} = \ln\left(S_{t+i\Delta} / S_{t+(i-1)\Delta}\right)$ for $i = 1,...,N$. The annualized realized variance over the period from t to $t+T$ is then defined as

$$Var_{Realized}\left(t\right) \equiv \frac{1}{\Delta}\left[\frac{1}{N}\sum_{i=1}^{N}\left(R_{t+i\Delta}\right)^2\right]. \tag{9}$$

The quantity inside of the square bracket is a measure of the variance of the returns over the period t to $t+T$, and multiplying by the $1/\Delta$ factor (which is equal to the number of price observations in a year) annualizes it. Equation (9) constrains the mean return to be zero. This constraint is imposed, because the sample average return is a very noisy estimate of the true mean return. The results reported below, however, were also computed using a measure of realized variance which sets the mean return to the sample average of the returns. This change produced no appreciable difference in the results.

For the first set of tests the realized variance variable, $Var_{Implied}\left(t\right)$, is set equal to the Black-Scholes implied variance of the closest to ATM call at observation time *t* where ATM is defined as the strike price being equal to the futures price. For purposes of finding the closest to ATM call the futures price is taken to be the quantity $Se^{-\delta T}$ which is derived from the futures prices as described in the previous section multiplied by e^{rT} where r is the one month LIBOR rate. When implying the Black-Scholes variance from the option price, the level of the SPX index is set equal to the quantity $Se^{-\delta T}$ which is derived from the futures prices as described in the previous section, and it is assumed that the underlying index pays no dividends. (As explained above, this is exactly equivalent to implying variance using the correct index level and the correct expected dividend rate.)

The initial set of tests use three different time-series of SPX levels to compute the realized variance variable via equation (9). The first time-series of SPX levels for computing

$Var_{Realized}(t)$ is the daily closing values for the SPX index from the CRSP indices files. The second times-series is the daily 3:00 PM SPX level implied from the SPX futures transaction prices. The third time-series is the five minute trading time (i.e., on each trade the 8:35 AM CST, 8:40AM CST, …, 3:15 PM CST) SPX index levels implied from SPX futures transaction prices. For the second and third time-series, the SPX level for a particular target time is inferred from the SPX futures prices as follows. First a dividend rate, δ, for the SPX index is computed from equation (8) above where as before F_2 is the transaction price of the futures contract with the second most liquid delivery date that transacts closest in time to the target time, and F_3 is the transaction price of the futures contract with the most liquid delivery date that transacts closest in time to F_2. Next F_1 is chosen to be the transaction price of the SPX futures with the most liquid delivery date that transacts closest in time to the target time. Then spot-futures parity is used to infer the SPX level at the target time from the futures transaction data by

$$S = F_1 e^{-(r-\delta)T_1}$$

where T_1 is the time to delivery of F_1 and r is the one month LIBOR rate.

In order to minimize Jensen inequality type biases, the function $f(\bullet)$ in equation (1) will be set equal to the identity function, $I(\bullet)$. Since this is not the most common choice, the impact of changing $f(\bullet)$ to other specifications that have been used in the literature will be examined later. The results of estimating equation (1) using the three different time-series of SPX levels to construct $Var_{Realized}(t)$ and setting $Var_{Implied}(t)$ equal to the Black-Scholes implied variance of the closest to ATM call are presented in Table III. The main data period used in this study will be June 1993 through August 1997, because one of the focuses of the paper will be to determine the impact of forecasting variance from option prices using the Heston (1993) model which was published in the summer of 1993. The results for the June 1993 through August 1997 data period are presented in Panel A. Since the current regressions use model (3)-(7) (which has been available since before 1988) to forecast variance, results for the entire data period of June 1988 through August 1997 are included in Panel B for comparison. It can be seen that the results for the longer and shorter data periods are similar.

Focusing on Panel A which covers the June 1993 through August 1997 data period, it can be seen that as the time-series of SPX index levels used to construct the $Var_{Realized}(t)$ variable moves from (1) daily SPX closing levels to (2) daily 3:00 PM SPX levels implied from futures prices to (3) five minute interval SPX levels implied from futures prices the α estimates move monotonically toward zero and the β estimates move monotonically toward 1. Altogether the α estimate moves about one third of the way toward zero and the β estimate moves almost half of the way toward one when the realized variance is computed from five minute futures implied SPX index levels rather than from SPX daily closing levels. In addition, the adjusted R^2 increases from 0.35 to 0.47 and the *F*-statistic for testing the hypothesis that $\alpha = 0$ and $\beta = 1$ decreases from 23.14 to 5.94. The *F*-statistic, nonetheless, still rejects the hypothesis $\alpha = 0$ and $\beta = 1$ at conventional levels. An improvement in forecasting performance when moving from daily SPX closing prices to daily 3:00 PM SPX levels is to be expected, because the daily closing prices contain positive serial correlation from infrequent trading of some components of the index which is not present in the SPX levels inferred from futures prices. Similarly, the improvement in the forecasting performance when moving from daily 3:00 PM SPX levels inferred from futures prices to five minute trading time SPX levels inferred from futures prices is reasonable, since it is well know that more frequent sampling of a time-series improves estimates of its variance (but not its mean). Bevan, Poon, and Taylor (1999) also find gains to using high frequency index returns in a variance forecasting context.

If the remaining bias in the variance forecasts from option prices is produced by orthogonal error in the forecast (i.e. independent) variable, then constructing the $Var_{Implied}(t)$ variable in a cleaner manner should reduce the bias. Although there are a number of possible sources of orthogonal error in $Var_{Implied}(t)$, the large bid-ask spreads in the SPX options market are most likely an important source of such error. (For a one month close to ATM SPX option the bid-ask spread is typically more than five percent of the value of the option.) Consequently, the next experiment will examine the impact of attempting to mitigate the orthogonal error in $Var_{Implied}(t)$ produced by the bid-ask spread by computing $Var_{Implied}(t)$ from several rather than one close to ATM options.

Table IV presents the results of estimating equation (1) when the $Var_{Implied}(t)$ variable is constructed from the closest to ATM call, the closest to ATM put, the two closest to ATM calls, the two closest to ATM puts, and the four closest to ATM options (i.e. the two closest to ATM calls and the two closest to ATM puts.) In all cases, the $Var_{Implied}(t)$ variable is set equal to the Black-Scholes implied variance of the option(s). When there is more than one option, this variance is defined as the real number that minimizes the sum of the squared pricing errors of the options (i.e. it is *not* the average of the individual Black-Scholes implied variances.) Here and henceforth, unless otherwise noted, the $Var_{Realized}(t)$ variable is constructed from five minute trading time observations on the SPX level inferred from futures prices. From Panel A of Table IV which covers the period June 1993 through August 1997 it can be seen that there is very little difference in the estimates when puts are used instead of calls or when the implied variance is extracted from multiple option prices rather than just one. For example when $Var_{Implied}(t)$ is computed from the four closest to ATM options, the α estimate and R^2 are the same as when it is computed from just the closest to ATM call, and the β estimate changes only from 0.76 to 0.78. This finding suggests that orthogonal error in $Var_{Implied}(t)$ may not be an important source of the forecast bias. There is some indication in Panel B of Table IV that over the entire data period from June 1988 through August 1997 call prices provide better forecasts of future variance than put prices. However, there appears to be no improvement in the regression results in Panel B when two rather than one ATM calls are used to construct $Var_{Implied}(t)$ or when two rather than one ATM puts are used to construct $Var_{Implied}(t)$. This lack of improvement also suggests that orthogonal error in $Var_{Implied}(t)$ may not play a substantive role in biasing the option price based forecasts of future variance.

This paper sets the function $f(\bullet)$ in regressions (1) and (2) to the identity function, $I(\bullet)$, in order to eliminate Jensen inequality type distortions. In order to assess the impact of this choice on the regressions, Table V presents the results of estimating equation (1) for various specifications of the function $f(\bullet)$ that have been used in the variance forecasting literature. The specifications that will be examined are $f(\bullet) = I(\bullet)$, $f(\bullet) = \sqrt{\bullet}$, $f(\bullet) = \ln(\bullet)$, and $f(\bullet) = \ln(\sqrt{\bullet})$. The $f(\bullet) = \ln(\bullet)$ specification is included for comparative purposes even though it is not used in any of the papers summarized in Table I. Here and henceforth, unless otherwise noted, $Var_{Implied}(t)$ is set equal to the real number whose square-root minimizes the sum of the Black-Scholes pricing errors for the two closest to ATM calls and two closest to

ATM puts. It is immediately obvious from Table V that varying the specification for $f(\bullet)$ while holding everything else fixed produces non-trivial changes in the coefficient estimates, the R^2 statistic, and the *F*-statistic. For example, focusing on the β estimate, changing $f(\bullet)$ from $I(\bullet)$ to the most common specification, $\sqrt{\bullet}$, increases the β estimate from 0.78 ($t_{1-\beta=0} = 1.94$) to 0.88 ($t_{1-\beta=0} = 0.98$). Changing $f(\bullet)$ to $\ln(\bullet)$ or $\ln\left(\sqrt{\bullet}\right)$ increases the estimate of β to 1.02 ($t_{1-\beta=0} = -0.18$). The last of these specifications is the one used in Christensen and Prabhala (1998). An in-depth analysis of the choice of $f(\bullet)$ function is beyond the scope of this paper. It should be noted, however, that if the results of the type presented in Table V extend to other markets and time periods, some of the findings in the variance forecasting literature that use specifications other than $f(\bullet) = I(\bullet)$ may need reinterpretation.

The next issue that will be addressed is the efficiency of the variance forecasts extracted from option prices via model (3)-(7). The efficiency tests will be carried out by running the "encompassing" regression described by equation (2). Before running these regressions, however, it will be useful to establish the forecasting ability of variance predictions obtained from historical data on the SPX index by estimating the following equation:

$$Var_{Realized}(t) = \alpha + \beta Var_{Historical}(t) + \varepsilon(t) \tag{10}$$

As before, $Var_{Realized}(t)$ is the realized variance from time *t* to time $t+T$ where *T* is the time to expiration of the option observed at time *t* according to the sampling scheme described above. $Var_{Realized}(t)$ is still computed by applying equation (9) to five minute trading time observations on the SPX index derived from futures prices. $Var_{Historical}(t)$ is a measure of the historical variance of the SPX index leading up to time *t*. For example, if the historical data period is one month, then $Var_{Historical}(t)$ is calculated via equation (9) from the five minute interval SPX index levels inferred from SPX futures prices from twenty-two trade dates before time *t* to one trade date before time *t*.

The results of estimating equation (10) for various historical data periods are presented in Table VI. When historical variance is computed from one month to six months of past five minute SPX index levels inferred from SPX futures prices, its ability to forecast future variance appears to be comparable to or better than that of option price forecasts based on the model described by equations (3)-(7). Indeed, the *F*-statistics in these cases do not reject the hypothesis $\alpha = 0$ and $\beta = 1$ at conventional levels. Historical variance computed from one or two years of past data does a poor job of predicting variance over the next three and a half to four and a half weeks. A GARCH(1,1) model was also estimated for the past 500 trade dates of past returns and used to predict future variance. In unreported results, the variance forecasts from the GARCH(1,1) models performed very poorly. However, no conclusion about the forecasting ability of GARCH models should be drawn from this finding, because no effort was made to determine either the optimal GARCH specification or the optimal amount of past data to estimate the GARCH model.

Table VII contains the estimates of equation (2) for the specification $f(\bullet) = I(\bullet)$ when the historical variance, $Var_{Historical}(t)$, is computed from 1, 2, 3, or 6 months of past data. The coefficient estimates on the $Var_{Implied}(t)$ variable drop slightly as compared to the values in the univariate regressions while the coefficient estimates on $Var_{Historical}(t)$ drop dramatically from their values in the univariate regressions. In fact, for the data period June 1993 through August 1997 all of the regressions have an intercept that is not statistically different from zero, a coefficient on $Var_{Implied}(t)$ that is not statistically different from one, and a coefficient on $Var_{Historical}(t)$ that is not statistically different from zero. Hence, the individual coefficient estimates are consistent with market efficiency given the model of market equilibrium described by equations (3)-(7). Nonetheless, the joint hypothesis of market efficiency and model (3)-(7) is rejected at conventional levels, because the probability that the *F*-statistics are as large as observed is only one percent if $\alpha = 0, \beta = 1,$ and $\gamma = 0.$ Furthermore, in the longer data period from June 1988 through August 1997 reported in Panel B, most of the β estimates are significantly different from one.

The tests presented in this section of the paper indicate that a nontrivial portion of the bias found when using standard procedures to forecast the future variance of the SPX index from SPX option prices under model (3)-(7) can be eliminated by more carefully constructing the realized variance variable. A statistically significant bias remains, however, even when the cleaner empirical techniques are used in the variance forecasting exercise. The next section investigates the impact on the forecasting regressions of using the Heston option pricing model to construct the implied variance variable.

IV. Forecasting Variance from Option Prices using the Heston Model

The extant literature on forecasting future variance from option prices uses either the Black-Scholes model or model (3)-(7) to extract forecasts of future variance from option prices. This section of the paper will examine the impact of extracting future variance forecasts from the Heston (1993) model instead. The Heston model is chosen because it generalizes equations (3)-

(7) in two important ways and has been available to option market participants since the summer of 1993.

The dynamics for the underlying asset presumed by the Heston model are described by the following set of equations:

$$\frac{dS_t}{S_t} = \mu(S_t, V_t, t)\,dt + \sqrt{V_t}\,dW_t^S \tag{11}$$

$$dV_t = k(\theta - V_t)\,dt + \eta\sqrt{V_t}\,dW_t^V \tag{12}$$

$$Corr\left(dW_t^S, dW_t^V\right) = \rho \tag{13}$$

$$\lambda(S_t, V_t, t) = \lambda V_t \tag{14}$$

$$r = \delta = constant \tag{15}$$

where $k, \theta, \eta, \rho, \lambda, r,$ and δ are constants. As before, S_t and V_t are, respectively, the time *t* level and instantaneous variance of the underlying asset. dW_t^S and dW_t^V are increments of standard Weiner processes with constant correlation ρ. $\lambda(S_t, V_t, t)$ is the market price of variance risk, *r* is the constant rate for risk free borrowing and lending, and δ is the constant dividend payout rate of the underlying asset.

The Heston model is more flexible than the model described by equations (3)-(7) insofar as it allows a non-zero correlation between the Weiner processes driving the system and a non-zero market price of variance risk. This greater flexibility is potentially significant. The non-zero correlation may be important, because there is known to be a substantial negative correlation between the level and variance of the SPX index (i.e. the leverage effect). The non-zero market price of variance risk may be important, because a non-zero market price of variance risk will bias the forecasting tests in the observed direction (presuming that investors need to be compensated for taking on variance risk.) Recently, models have appeared with even more realistic dynamical assumptions (e.g. models that also include jumps in the return process) than the Heston model. These models are not used, however, because they were not available to investors during the data period.

Pan (1999) develops a sophisticated implied state generalized method of moments technique to estimate the Heston model, and applies the procedure to SPX index option and SPX index level data over the period January 1989 through December 1996. The resulting point estimates for the parameters of the model are $k = 7.1$, $\theta = 0.0137$, $\eta = 0.32$, $\rho = -0.53$, and $\lambda = -7.6$. As an initial experiment, equation (1) will be re-estimated assuming that the investors subscribed to the Heston model and believed that the Pan estimates are the true values of the parameters throughout the period June 1993 through August 1997. It should be emphasized that this exercise should *not* be regarded as a variance forecasting test for two reasons. First, the model has been estimated through most of the test period (e.g. from June 1993 through December 1996) using at least some data that was not available to SPX market participants at the time forecasts are made. Second, the Pan implied state generalized method of moments estimation technique was not available to investors during any of the test period. Nonetheless, the initial experiment is interesting in that it provides a baseline in which the Heston model is given more than a fair chance of succeeding. The results of this baseline case will then be compared to procedures that use the Heston model in a way that actual investors could have implemented it.

When running the baseline Heston model test, $Var_{Realized}(t)$ will be computed as before from five minute trading time observations on the SPX index derived from futures prices. The $Var_{Implied}(t)$ variable will be computed as follows. First, on each observation date an instantaneous variance, $Var_{Instantaneous}(t)$, will be derived from the same four closest to ATM options as before by finding the level of instantaneous variance which minimizes the sum of their squared pricing errors under the Heston model given the Pan parameters. The forecast variable, $Var_{Implied}(t)$, will then be set to the expected average instantaneous variance over the remaining life of the options given $Var_{Instantaneous}(t)$, the Pan parameters, and the assumption that the instantaneous variance follows the process

$$dV_\tau = k(\theta - V_\tau)dt + \eta\sqrt{V_\tau}dW_\tau^V.$$

It is shown in the Appendix that under these conditions, $Var_{Implied}(t)$ is equal to

$$Var_{Implied}(t) = \theta + \left[\frac{1-e^{-kT}}{kT}\right]\left[Var_{Instantaneous}(t) - \theta\right]. \tag{16}$$

Equation (1) is re-estimated over the period June 1993 through August 1997 with the $Var_{Implied}(t)$ variable computed from equation (16). The results are presented in the first row of Table VIII. The α estimate is 0.0009 $(t_{\alpha=0} = 0.37)$, and the β estimate is 0.97 $(t_{1-\beta=0} = 0.19)$. Furthermore, the *F*-statistic for the hypothesis $\alpha = 0$ and $\beta = 1$ is only 0.11 which has a *p*-value of 0.89. Consequently, the hypothesis is not rejected at conventional levels. The conclusion is that when the Heston model is estimated with Pan's method using data from the period January 1989 through December 1996, it produces unbiased variance forecasts from option prices over the period June 1993 through August 1997.

The natural question to ask at this point is how well future variance is forecast from option prices under the Heston model when using a forecasting method that would have been implementable (and might have been implemented) by actual SPX market participants. In order to answer this question, it will be assumed that on each monthly observation date from June 1993 through August 1997 SPX option market investors estimate the $k, \theta,$ and η parameters of the variance process given by equation (12) from the daily closing levels of the SPX index in the CRSP indices file over the period July 3, 1962 (the first date available in the CRSP file) through the end of the month preceding the observation date. Over this period an instantaneous variance is estimated for each calendar month from equation (9) (which constrains the mean return to be zero) using the SPX daily closing levels for that month. The resulting time-series of monthly instantaneous variance estimates is then used to estimate the parameters in equation (12) via maximum likelihood. The necessary likelihood function is known, for example, from equations (18) and (20) in Cox, Ingersoll, and Ross (1985). The month of October 1987 is omitted from the time-series, because any diffusion based model (i.e. any model that does include jumps in its dynamics) cannot reasonably accommodate the October 1987 crash. Estimating the parameters in this way yields $k = 6.87,$ $\theta = 0.0170,$ and $\eta = 0.44$ for the period July 1963 through May 1993. These estimates do not change by much as the months June 1993 through July 1997 are added to the estimation period.

Once the $k, \theta,$ and η parameters have been determined for a given observation date, ρ and λ are chosen to minimize the sum of the squared pricing errors of all out-of-the-money SPX options on that observation date that meet several criteria. The out-of-the money SPX options must have a Black-Scholes implied volatility greater than zero and less than 0.70, a bid price of greater than 3/8, a bid-ask spread that is less than or equal to one dollar, and a time to expiration of greater than six calendar days and less than or equal to seven months. The price of each option is taken to be the midpoint of the last bid-ask quote prior to 3:00 PM CST. When determining ρ and λ the instantaneous variance is also allowed to vary in order to minimize the sum of squared pricing errors of the out-of-the money options. Once ρ and λ are determined, however, this instantaneous variance is discarded. Finally, a new instantaneous variance is determined by finding the level of instantaneous variance which minimizes the sum of squared pricing errors of the four closest to ATM options on the observation date (i.e. the same two closest to ATM calls and two closest to ATM puts as above) given the estimated values of $k, \theta, \eta, \rho,$ and λ. $Var_{Instantaneous}(t)$ is then set equal to this final level of instantaneous variance and $Var_{Implied}(t)$ is computed from equation (16). $Var_{Realized}(t)$ is computed as before from five minute trading time observations on the SPX index level derived from futures prices.

The second entry in Table VIII gives the results of estimating equation (1) with $f(\bullet) = I(\bullet)$ when $Var_{Implied}(t)$ is computed from the Heston model using the method just describe that would have been implementable by SPX market participants. The estimate of α is -0.0001 $(t_{\alpha=0} = -0.03)$, and the estimate of β is 0.88 $(t_{1-\beta=0} = 0.90)$. The *F*-statistic is 2.25 which corresponds to a *p*-value of 0.12 for the hypothesis $\alpha = 0$ and $\beta = 1$. Consequently, the hypothesis that option price based forecasts of future variance are unbiased is not rejected at conventional levels.

Table IX presents the results of estimating equation (2) with $f(\bullet) = I(\bullet)$ when $Var_{Implied}(t)$ is computed under the Heston model and the historical variance prediction is computed from one, two, three, or six months of past data. Panel A contains the results when using the Pan estimates of the parameters of the model while Panel B contains the results when using the implementable method for estimating the parameters of the model. In both of the panels the α estimates are never statistically different from zero, the β estimates are never statistically different from one, and the γ estimates are never statistically different from zero. In addition, in Panel A the *p*-values for the *F*-statistics are around 0.8 and in Panel B the *p*-values for the *F*-statistics are around 0.1 and never less than 0.08. Consequently, when the Pan estimates are used the joint hypothesis of $\alpha = 0, \beta = 1,$ and $\gamma = 0$ is never rejected and when the implementable estimates are used the joint hypothesis is never rejected at the 5% level. When the cleaner data processing techniques used in this paper are combined with the use of the Heston model to predict future variance from option prices, there is no longer reliable evidence that SPX option prices entail either biased or inefficient forecasts of future variance.

V. Monte Carlo Simulations

All of the point estimates and standard errors (i.e. *t*-statistics) in the tests presented above were computed using OLS. Although the non-overlapping sampling procedure eliminates the most important source of inaccuracies in the OLS standard errors, the point estimates or standard errors may, nonetheless, be biased by errors in the variables or persistence in the variance process. The potential impact of these factors will be assessed by running Monte Carlo simulations in order to construct the sampling distributions of the OLS estimates under various assumptions about the dynamics of the underlying asset and the errors present in the options and the futures data that are used to generate the $Var_{Realized}(t)$, $Var_{Implied}(t)$, and $Var_{Historical}(t)$ variables.

The true (unobserved) dynamics of the underlying SPX index level and SPX variance are assumed to be described by equations (11)-(15). The drift of the index return is set to $\mu(S_t, V_t, t) = 0.12,$ and the parameters of the variance process and the correlation between the Weiner processes are taken from Pan (1999) and have the values $k = 7.1,$ $\theta = 0.0137,$ $\eta = 0.32,$ and $\rho = -0.53.$ For each experiment, 1000 sample paths are simulated from model (11)-(15).

Each sample path begins at an SPX level of 210 and a variance level of 0.0137. The SPX level on January 2, 1986 was approximately 210 (it closed at 209.59 that day) and 0.0137 is the long-run mean of the variance process. Taking account of holidays, late openings, and early closings, there were 238,869 five-minute trading intervals in the SPX futures market from the opening of trading on January 2, 1986 through the close of trading on September 18, 1997 which is the expiration time for the final option that is observed in August 1997. Consequently, the sample paths are obtained by simulating equations (11)-(13) from the initial values for 238,869 steps using a bivariate Euler scheme where each step corresponds to five minutes of trading time. In order to assess the impact of the discretization bias introduced by the use of the Euler scheme, several of the experiments reported below were repeated simulating equations (11)-(13) for 2,388,690 steps using a bivariate Euler scheme where each step corresponded to 30 seconds of trading time. The simulated paths were then sampled at a frequency of ten steps to obtain five minute trading time paths. The results were not distinguishable from those reported below.

Throughout the simulations the interest rate is set to $r = 0.05$ and the dividend rate is set to $\delta = 0.025$. Error was introduced into both the SPX index levels from which the realized variances were computed and also the option prices from which the implied variances were computed. Error was introduced into the SPX index levels as follows. For each of the 238,869 steps on each simulated path, spot-futures parity was used to convert the SPX level into a futures price for each of the three delivery times that were actually used at the corresponding time during the tests on the real data. A band with width of either \$0.00, \$0.10, or \$0.20 was then centered on each of the spot-futures parity derived futures prices, and a price for each of the three futures was drawn uniformly from its band. These three futures prices were then used to derive an SPX level in the manner described above. Error was introduced into the option prices as follows. On each observation date (i.e. the Wednesday after an option expiration) the ATM strike price level was determined from the current (true) level of the SPX index, the time to expiration of the options, r, and δ. The strike prices of the two closest to ATM calls and two closest to ATM puts were then set equal to the integer multiple of five less than and greater than the ATM strike price level. Next, the true price of these options was set to their values under the Heston model given the assumed parameters values. A band with width of either zero, the bid-ask spread of the corresponding call or put in the real data observed on the same trade date (i.e. the call or put whose strike price was also just below or just above the ATM strike price), or twice that bid-ask spread was then centered on each of the options prices, and a price for each of the four options was drawn uniformly from its band.

Table X reports fractiles, means, standard deviations, and empirical *p*-values for the regression coefficients, adjusted R^2, and *F*-statistics for the OLS estimation of equation (1) with $f(\bullet) = I(\bullet)$ for a simulated period covering June 1993 through August 1997. The different panels of the table correspond to different assumptions about the errors in the futures prices, the errors in the options prices, and the market price of variance risk. The empirical percentile is the percentage of times the observed statistic exceeds the simulated statistic under the panel's assumptions. The empirical 1-sided *p*-value is the percentage of times that the simulated statistic is (1) on the same side of the statistic's null value as the observed statistic and (2) further away from the statistic's null value than the observed statistic. The empirical 2-sided *p*-value is the percentage of times that the simulated statistic is further away (in absolute value) from the statistic's null value than the observed statistic. The null values of the statistics are $\alpha = 0$, $\beta = 1$, $R^2 = 1$, and $F = 0$.

Panels A-D of Table X report the simulation results when the market price of variance risk is set to zero under various assumptions about the errors in the futures and option prices. Since the SPX futures prices move in increments of $0.05 and the market is very liquid, it is guessed that observed SPX futures prices for the most part have errors in the range -$0.05 to $0.05 which corresponds in the simulations to a futures spread of $0.10. In Panels A-D the futures spread is set either to $0.00 or $0.10. In the SPX options market the bid-ask spreads tend to be fairly wide. It is guessed that for the most part observed SPX option prices have errors within the range of their bid-ask spreads. (In other words, in the empirical work when the observed option price is set to the bid-ask midpoint, it is assumed that the true option price without error usually falls somewhere within the bid-ask spread.) This corresponds in the simulations to an options spread of 1*x* the actual bid-ask spread. In Panels A-D the options spread is set either to 0 or 1*x* the actual bid-ask spread.

In Panel A the futures spread and the options spread are both set equal to zero. This results in simulated mean values of $\alpha = 0.0008$, $\beta = 0.97$, $R^2 = 0.69$, and $F = 1.52$. Unsurprisingly, when there is no error in the data and the market price of variance risk is set equal to zero the simulated forecasts outperform the real variance forecasts (whose results are repeated in the observed statistic column) on each of these statistics. The deviation of the mean α and β values from their null values of, respectively, 0 and 1 presumably results from a combination of the non-linearity of the Black-Scholes formula in variance for ATM options and the fact that the options used are not exactly ATM. (In unreported results, simulations were run in which equation (16) was used directly to forecast future variance from the current true level of variance, and the mean simulated regression coefficients were almost exactly equal to their null values.)

Panel B adds error into the options data equal to a uniform draw out of the bid-ask spread of the options used in the actual empirical work. Adding this option error has almost no impact on the means or standard errors of the simulated statistics. The small impact that it does have, however, is in the direction that would be expected. The mean β estimate decreases from 0.97 to 0.96, the mean R^2 estimate decreases from 0.69 to 0.68, and the mean *F*-statistic estimate increases from 1.52 to 1.55. The standard deviations of the statistics also increase slightly. These simulation results make it seem unlikely that orthogonal error in the independent variable is the main source of the variance forecasting bias observed in the literature as suggested by Christensen and Prabhala (1998). They also raise questions for the Figlewski interpretation of the variance forecasting bias insofar as the option pricing errors produced either by the difficulty of executing an arbitrage strategy or by market makers being focused on making a profit from the bid-ask spread are unsystematic. Panel C presents the results of simulations where there is error only in the futures prices, and Panel D presents the results of simulations where there is error in both the options and futures prices. The main results are much the same. These errors in the variables make very little difference to the means or standard errors of the regression estimates. There is no evidence that they can account for the bias in the estimates of equation (1) observed in this paper or in the literature. In particular, the estimates of the α and β coefficients from the real SPX data were, respectively, 0.0013 and 0.78 with empirical *p*-values in all four panels of about 0.35 and 0.05. These empirical *p*-values indicate that if the market price of variance risk is zero it is not unlikely that actual α estimate will be observed under any of the various assumptions on the options and futures errors but that there is only about a 5% chance that the actual β estimate will be as small as is observed. Likewise the empirical *p*-values for the R^2 statistic and *F*-statistic are around, respectively, 0.03 and 0.05 indicating that there is only about a 3% chance of observing a R^2 statistics as small as the 0.47 seen in the actual data and only about a 5% chance of seeing a *F*-statistic as large as the 4.82 seen in the actual data.

Panels E-G report the results of simulations under various assumptions about the options and futures errors when the market price of variance risk is assumed to be $\lambda = -7.6$ which is the value estimated in Pan (1999). When the options and future errors are assumed to be zero in Panel E, the non-zero market price of variance risk has very little impact on the distribution of the α estimate. The mean value of the simulated β coefficient, on the other hand, drops from 0.97 to 0.76 which is very close to the observed β estimate of 0.78. As a result, the one-sided empirical *p*-value increases from 0.06 to 0.59. Similarly, the mean value of the simulated *F*-statistic increases from 1.52 to 18.76 which increases its one-sided *p*-value from 0.05 to 0.99. In other words, under the assumptions that the market price of variance risk is $\lambda = -7.6$ and that there are no errors in the options or futures prices, 99% of the time an *F*-statistic as large as the actual value of 4.82 will be observed. The distribution of the R^2 statistic is virtually unchanged as a result of the inclusion of the non-zero market price of variance risk. It should be noted, however, that a higher R^2 is not necessary for and does not imply a better forecast of future variance (see Figlewski (1997), pp. 73-74.)

Panel F reports the results of simulations in which the market price of variance risk is set equal to $\lambda = -7.6,$ the futures spread is set equal to \$0.10, and the options spread is set equal to 1*x* the actual bid-ask spread of the options. As in the case where the market price of variance risk is zero, adding these errors makes very little difference to the distribution of the estimates. Panel G reports the results of simulations in which the market price of variance risk is still set equal to $\lambda = -7.6$ but the futures spread is set equal to \$0.20 and the options spread is set equal to 2*x* the actual bid-ask spread of the options. The only appreciable difference that results from increasing the errors in this way is to increase the mean estimate and standard deviation of the α estimate. The empirical 1-sided *p*-value for the actual observed α estimate, however, is still 0.63 in this case. As noted above, orthogonal error in the independent variable will bias the α estimate upward and the β estimate downward. It appears that in the present context, the biasing of the α estimate is more pronounced.

Table XI reports fractiles, means, standard deviations, and empirical *p*-values for the regression coefficients, adjusted R^2, and *F*-statistics for the OLS estimation of equation (2) with $f(\bullet) = I(\bullet)$ for a simulated period covering June 1993 through August 1997. The different panels make the same assumptions about the market price of variance risk and the futures and options errors as the corresponding panels in Table X. The results are much the same. In Panels A-D where the market price of variance risk is assumed to be zero, the observed β value is smaller than all but 1% or 2% of the simulated values and the *F*-statistic is larger than 90% to 95% of the simulated values regardless of the assumptions made on the futures and options errors. Hence, the observed values tend to reject model (11)-(15) when the market price of variance risk is constrained to be zero. It should be noted, however, that the two-sided *p*-values for the observed γ estimate is around 30% which indicates that the simulated γ estimates are as far away from zero as the observed statistic approximately 30% of the time. Consequently, there is no evidence that the observed estimate of the historical variance coefficient is significantly different from zero even when the market price of variance risk is forced to be zero.

Panels E-G report the results of simulations in which the market price of variance risk is assumed to be $\lambda = -7.6$ under various assumptions about the errors in the futures and options prices. In these cases the mean of the simulated β coefficients drop to 0.76 or 0.77 as compared to the observed value of 0.68. The one-sided empirical *p*-value for the observed value is about 20% which indicates that there is a 20% chance of observing a value as small as 0.68 even if the investors are predicting future variance optimally. Furthermore, the one-sided empirical *p*-values for the *F*-statistic increase to at least 0.90 regardless of the assumptions on the futures and options errors which indicates that there is a 90% chance of observing a value as large as the actual 4.94 even if investors are predicting future variance optimally.

The Monte Carlo simulations presented in this section of the paper construct the distributions of the various test statistics when researchers estimate equations (1) and (2) assuming the option pricing model (3)-(7) when option market investors, in fact, subscribe to model (11)-(15) as estimated by Pan (1999). Regardless of the various assumptions that are made on the errors in the futures and options data, a comparison of the simulated distributions of the $\alpha, \beta, \gamma,$ and F statistics to the values estimated from the real SPX data indicates that neither investor rationality nor market efficiency is rejected by investor forecasts of future variance from option prices in this market.

A future version of this paper will include Monte Carlo simulations of the empirical tests presented in Section IV in which investors first estimate model (11)-(15) from past futures and options data which contains error and then use the estimates to forecast future variance under model (11)-(15).

VI. Conclusion

A large literature studying the ability of option prices to forecast the future variance of the underlying asset concludes that such forecasts are systematically biased, and the bias tends to be most severe for options written on equity indexes. The prevailing interpretation of these findings is that option investors may be forming unbiased forecasts of the future variance of underlying assets but that these unbiased forecasts fail to get impounded into option prices because of either (1) the difficulty in carrying out the necessary arbitrage strategies that would force the prices to their proper level, or (2) the availability of lucrative alternative strategies for market makers which profit from the large bid-ask spreads in the option markets. A recent study bolsters this interpretation by arguing that orthogonal error in the forecasts of future variance are a substantial source of the standard bias. This interpretation has significant consequences for the entire range of research into option pricing, because it implies that non-continuous trading, bid-ask spreads, and other market imperfections substantially influence option prices. This implication is important because incorporating these types of market imperfections into option pricing models is much more difficult than, for example, altering the dynamics of the underlying asset.

The present paper studies the variance forecasting ability of SPX option prices against the backdrop of the standard interpretation of the findings in the variance forecasting literature. The paper finds first of all that approximately one third of the usual bias can be eliminated by using high frequency futures data rather than daily closing data to construct measures of realized variance. Second of all, it finds that roughly another third of the bias disappears when forecasts of future variance are extracted from option prices via an option pricing model that – unlike the standard model – permits a non-zero market price of variance risk and a non-zero correlation between innovations to the level and variance of the SPX index. Once bias is removed in these two ways, the remaining bias is no longer significant. Finally, a Monte Carlo study that accounts for errors in the futures and option prices that are used to construct the measures of realized and forecast variance indicates that the regression coefficients obtained in the standard forecasting regressions are not unlikely to occur if option prices are set by investors who optimally forecast future variance and subscribe to a standard option pricing model that allows a non-zero market price of variance risk and a non-zero correlation between innovations to the level and the variance of the SPX index.

Appendix

This appendix derives equation (16) in the text. Given the instantaneous variance at time t, $Var_{Instantaneous}(t)$, the problem is to find the expected average variance over the time interval t to $t+T$ under the assumption that the instantaneous variance follows the stochastic differential equation (12). Let $\bar{V}$ be the average instantaneous variance over the time interval t to $t+T$

$$\bar{V} = \frac{1}{T}\int_t^{t+T} V_\tau d\tau.$$

Then the expected average instantaneous variance at time t over the interval t to $t+T$ is given by

$$E_t\left[\bar{V}\right] = \frac{1}{T}E_t\left[\int_t^{t+T} V_\tau d\tau\right]$$

or

$$E_t\left[\bar{V}\right] = \frac{1}{T}\left[\int_t^{t+T} E_t\left[V_\tau\right] d\tau\right]$$

where the last equation follows from an application of Lebesques's Dominated Convergence Theorem. Now equation (19) in Cox, Ingersoll, and Ross (1985) states that for $\tau > t$

$$E_t\left[V_\tau\right] = V_t e^{-k(\tau-t)} + \theta\left(1-e^{-k(\tau-t)}\right).$$

Substituting into the previous equation gives

$$E_t\left[\bar{V}\right]=\frac{1}{T}\left[\int_t^{t+T} V_t e^{-k(\tau-t)}+\theta\left(1-e^{-k(\tau-t)}\right)d\tau\right].$$

Integrating and simplifying yields

$$E_t\left[\bar{V}\right]=\theta+\left[\frac{1-e^{-kT}}{kT}\right]\left[V_{Instantaneous}(t)-\theta\right]$$

which is equation (16) in the text.

References

Aï t-Sahalia, Y., and A. W. Lo, 1998, Nonparametric estimation of state-price densities implicit in financial asset prices, *Journal of Finance* 53, 499-547.

Bakshi, G., C. Cao, and Z. Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003-2049.

Beckers, S., 1981, Standard deviations implied in option prices as predictors of future stock price variability, *Journal of Banking and Finance* 5, 363-381.

Bevan, B., S.-H. Poon, and S. J. Taylor, 1999, Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency index returns, working paper, Lancaster University.

Black, F., and M. Scholes, 1972, The valuation of option contracts in a test of market efficiency, *Journal of Finance* 27, 399-417.

Canina, L., and S. Figlewski, 1993, The informational content of implied volatility, *Review of Financial Studies* 6, 659-681.

Chiras, D. P., and S. Manaster, 1978, The information content of option prices and a test of market efficiency, *Journal of Financial Economics* 6, 213-234.

Christensen, B. J., and N. R. Prabhala, 1998, The relation between implied and realized volatility, *Journal of Financial Economics* 50, 125-150.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385-407.

Day, T., and C. Lewis, 1992, The information content of option prices and a test of market efficiency, *Journal of Econometrics* 52, 267-287.

Day, T., and C. Lewis, 1993, Forecasting futures market volatility, *Journal of Derivatives* 1, 33-50.

Dumas, B., J. Fleming, and R. E. Whaley, 1998, Implied volatility functions: Empirical tests, *Journal of Finance* 53, 2059-2106.

Feinstein, S., 1989, The Black-Scholes formula is nearly linear in for at-the-money options; therefore implied volatilities from at-the-money options are virtually unbiased, working paper, Federal Reserve Bank of Atlanta.

Figlewski, S., 1997, Forecasting volatility, *Financial Markets, Institutions & Instruments* 6, 1-88.

Fleming, J., 1998, The quality of market volatility forecasts implied by S&P 100 index option prices, *Journal of Empirical Finance* 5, 317-345.

Hansen, L. P., 1982, Large sample properties of generalized method of moment estimators, *Econometrica* 50, 1029-1054.

Heston, S. L., 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327-344.

Hull, J., 1993, *Options, Futures, and Other Derivative Securities* (Prentice Hall, Englewood Cliffs, NJ).

Jackwerth, J. C., and M. Rubinstein, 1996, Recovering Probability Distributions from Option Prices, *Journal of Finance* 51, 1611-1631.

Johnston, J., 1984, *Econometric Methods, 3rd Edition* (McGraw-Hill, Inc., New York).

Jorion, P., 1995, Predicting volatility in the foreign exchange market, *Journal of Finance* 50, 507-528.

Lamoureux, D., and W. Lastrapes, 1993, Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities, *Review of Financial Studies* 6, 293-326.

Latané, H. A., and R. J. Rendleman, 1976, Standard deviations of stock price ratios implied in option prices, *Journal of Finance* 31, 369-381.

LeRoy, S. F., and R. D. Porter, 1981, Stock price volatility: Tests based on implied variance bounds, *Econometrica* 49, 555-574.

Pan, J., 1999, "Integrated" time-series analysis of spot and option prices, working paper, Graduate School of Business, Stanford University.

Poteshman, A., 1999, Does investor misreaction to new information increase in the quantity of previous similar information? Evidence from the options market, working paper, University of Illinois at Urbana-Champaign.

Shiller, R. J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* 71, 421-436.

Table I
A Selection of Studies that use Option Prices to Predict Future Variance

$$f\left(V_{Realized}\left(t\right)\right)=\alpha+\beta f\left(V_{Implied}\left(t\right)\right)+\varepsilon\left(t\right)$$

$$f\left(V_{Realized}\left(t\right)\right)=\alpha+\beta f\left(V_{Implied}\left(t\right)\right)+\gamma f\left(V_{Historical}\left(t\right)\right)+\varepsilon\left(t\right)$$

$V_{Realized}\left(t\right)$ is a measure of the ex-post realized variance of the underlying over the period *t* to $t+T$, $V_{Implied}\left(t\right)$ is the variance that is forecast over the period *t* to $t+T$ from an option price observed at time t that expires at $t+T$, $V_{Historical}\left(t\right)$ is the variance forecast over the period *t* to $t+T$ from some subset of the information available to the market at time t – usually from the history of past returns on the underlying asset, and $\varepsilon\left(t\right)$ is a zero mean forecast error that is uncorrelated with the forecasting variables. *t*-statistics are shown in parentheses. The *t*-statistics reported for $\hat{\alpha}$ and $\hat{\gamma}$ are for the hypothesis that the coefficient is equal to zero. The *t*-statistics reported for $\hat{\beta}$ are for the hypothesis that the coefficient is equal to one. The *t*-statistics are sometimes derived from reported standard errors or reported t-statistics on a different hypothesis. Some of the reported values are averages over different data sets examined in the paper. Fleming (1998) runs orthogonality (i.e. market efficiency) regressions. They are not reported here because they do not fit conveniently into the format of the table. $I\left(\bullet\right)$ is the identity function, i.e. $I\left(x\right)=x$.

Study	Underlying Asset	$f\left(\bullet\right)$	Observations	Forecast Horizon	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	R^2
Canina and Figlewski (1993)	OEX	$\sqrt{\bullet}$	Daily Calls, 3/83-3/87	7-127 Cal. Days	0.14 (11.3)	0.02 (19.6)		0.00
					0.08 (2.91)	-0.06 (12.47)	0.49 (2.21)	0.17
Day and Lewis (1992)	OEX	$I\left(\bullet\right)$	Weekly Calls, 3/83-12/89	Nearby > 7 Days	0.0005 (0.00)	0.72 (0.14)		0.03
					-0.0001 (-0.00)	0.60 (0.39)	0.30 (0.71)	0.03
Fleming (1998)	OEX	$\sqrt{\bullet}$	Daily Calls & Puts, 10/85-4/92	Nearby > 15 Days	-0.0191 (-3.63)	0.57 (9.30)		0.03
					NA	NA	NA	NA
Christensen and Prabhala (1998)	OEX	$\ln\left(\sqrt{\bullet}\right)$	Monthly Calls, 11/83-5/95	One Month	-0.56 (-3.47)	0.76 (2.99)		0.39
					-0.49 (-3.01)	0.56 (3.76)	0.23 (2.38)	0.41
Lamoureux and Lastrapes (1993)	Ind. Stocks	$I\left(\bullet\right)$	Daily Calls, 4/82-3/84	90-180 Cal. Days	NA	NA		NA
					1627.258 (4.20)	0.67 (0.90)	-1.75 (-3.13)	0.29

Table I – Continued

Study	Underlying Asset	$f(\bullet)$	Observations	Forecast Horizon	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	R^2
Jorion (1995)	Foreign Ex. Fut.	$\sqrt{\bullet}$	Daily Calls & Puts, 1/85 or 7/86 –2/92	3-100 Cal. Days	0.35	0.52		0.13
					(2.7)	(2.9)		
					0.33	0.63	-0.09	0.14
					(2.65)	(1.99)	(-0.89)	
Day and Lewis (1993)	Crude Oil Fut.	$I(\bullet)$	Daily Closing Prices, 11/86-3/91	2 and 4 Months	0.03	0.88		0.72
					(1.0)	(0.8)		
					0.004	0.97	-0.01	0.61
					(1.14)	(0.19)	(-0.01)	

Table II
Descriptive Statistics for CBOE Traded S&P 500 Index Option Calls and Puts, June 1, 1993 through August 29, 1997

Descriptive statistics for daily observations on CBOE traded S&P 500 index options for the period June 1, 1993 through August 28, 1997 that meet four conditions: (1) The time-to-expiration is greater than or equal to 6 calendar days and less than or equal to 7/12 year; (2) The bid price for the option is greater than or equal to $ 3/8, and the bid-ask spread is less than or equal to $1; (3) The Black-Scholes implied volatility (BSIVol) is greater than zero and less than or equal to 0.7; and (4) The option is out-of-the-money which is determined by comparing the strike price of the option to the CRSP closing value for the S&P 500 index on the trade date. The price information for each option on each trade date is the last bid-ask quote prior to 3:00 PM CST. The call and put prices are defined as the bid-ask midpoint. The interest rate, *r*, is the one month LIBOR rate on the day that the option is observed. The moneyness of an option is defined as its strike price divided by its futures price, $Moneyness \equiv X/F$.

					Percentiles					
Variable	Obs.	Mean	S.D.	Min	1%	10%	50%	90%	99%	Max
Panel A: Calls										
Call Price ($)	56717	6.935	7.046	0.406	0.438	0.875	4.750	15.625	34.50	71.25
Spread ($)	56717	0.411	0.229	0.063	0.063	0.125	0.375	0.750	1.000	1.000
BSIVol (%)	56717	0.147	0.043	0.023	0.085	0.100	0.140	0.201	0.296	0.696
T (Cal. Days)	56717	90	55	7	9	24	77	176	209	212
X (Ind. Pnts.)	56717	503	178	245	265	305	460	785	970	995
r (%)	56717	0.059	0.018	0.029	0.030	0.032	0.057	0.084	0.099	0.103
Moneyness	56717	1.043	0.045	0.940	0.988	1.002	1.033	1.091	1.217	1.467
Panel B: Puts										
Put Price ($)	97575	5.071	4.944	0.406	0.438	0.750	3.563	11.375	23.125	55.500
Spread ($)	97575	0.347	0.200	0.000	0.063	0.125	0.250	0.625	1.000	1.000
BSIVol (%)	97575	0.199	0.058	0.080	0.110	0.137	0.189	0.271	0.397	0.700
T (Cal. Days)	97575	89	56	7	8	24	76	176	210	212
X (Ind. Pnts.)	97575	471	163	175	215	275	430	720	880	960
r (%)	97575	0.056	0.017	0.029	0.030	0.032	0.055	0.083	0.099	0.103
Moneyness	97575	0.914	0.065	0.527	0.700	0.826	0.929	0.982	0.996	1.005

Table III

$f(\bullet) = I(\bullet)$ Regressions Varying Data Used to Construct Realized Variance

$$Var_{Realized}(t) = \alpha + \beta Var_{Implied}(t) + \varepsilon(t)$$

OLS point estimates with *t*-statistics computed from OLS standard errors (in parentheses) for $\alpha = 0$ and $1-\beta = 0$. $Var_{Realized}(t)$ is the realized variance of the SPX returns computed under the restriction that the mean return is equal to zero. The SPX returns are derived either from the daily CRSP SPX closing levels, the 3:00 PM SPX levels inferred from SPX futures prices, or five minute interval SPX index levels inferred from SPX futures prices. $Var_{Implied}(t)$ is the Black-Scholes implied variance of the closest to ATM call. When computing the Black-Scholes implied variance, the quantity $Se^{-\delta T}$ which is computed from futures prices is used as the value of the SPX index, and it is then assumed that the SPX index pays no dividends. The *F*-statistic tests the hypothesis $\alpha = 0$ and $\beta = 1$. The *p*-value gives the probability that the *F*-statistic is as large as it is if the hypothesis that $\alpha = 0$ and $\beta = 1$ is true. There is one observation for each Wednesday in the data period that follows an option expiration date, and the time interval covered by each observation extends from that Wednesday to the next option expiration date.

Realized Variance Constructed From	α	β	Adj R^2	*F*-statistic	*p*-value
Panel A: June 1993 – August 1997					
SPX Close	0.0020	0.57	0.35	23.14	0.00
	(0.91)	(3.99)			
3:00 PM Futures	0.0018	0.68	0.34	8.11	0.00
	(0.68)	(2.48)			
Five Minute Futures	0.0013	0.76	0.47	5.94	0.00
	(0.54)	(2.09)			
Panel B: June 1988 – August 1997					
SPX Close	0.0022	0.60	0.35	40.60	0.00
	(1.17)	(5.24)			
3:00 PM Futures	0.0028	0.68	0.23	8.51	0.00
	(0.98)	(2.76)			
Five Minute Futures	0.0014	0.74	0.42	14.98	0.00
	(0.71)	(3.18)			

Table IV

$f(\bullet) = I(\bullet)$ Regressions Varying Options Used to Construct Implied Variance

$$Var_{Realized}(t) = \alpha + \beta Var_{Implied}(t) + \varepsilon(t)$$

OLS point estimates with *t*-statistics computed from OLS standard errors (in parentheses) for $\alpha = 0$ and $1 - \beta = 0$. $Var_{Realized}(t)$ is the realized variance of the SPX returns derived from five minute interval SPX index levels inferred from SPX futures prices. The realized variance is computed under the restriction that the mean return is equal to zero. $Var_{Implied}(t)$ is the Black-Scholes implied variance from either the closest to ATM call, the closest to ATM put, the two closest to ATM calls, the two closest to ATM puts, or the two closest to ATM calls and the two closest to ATM puts. The Black-Scholes implied variance is computed by using the quantity $Se^{-\delta T}$, inferred from futures prices, as the value of the SPX index level and then assuming that the SPX index pays no dividends. When the Black-Scholes implied variance is computed from more than one option, it is defined as the real number that minimizes the sum of the squared pricing errors of the options (i.e., it is *not* the average of the individual Black-Scholes implied variances). The *F*-statistic tests the hypothesis $\alpha = 0$ and $\beta = 1$. The *p*-value gives the probability that the *F*-statistic is as large as it is if the hypothesis that $\alpha = 0$ and $\beta = 1$ is true. There is one observation for each Wednesday in the data period that follows an option expiration date, and the time interval covered by each observation extends from that Wednesday to the next option expiration date.

Options Used to Construct Implied Variance	α	β	Adj R^2	*F*-statistic	*p*-value
Panel A: June 1993 – August 1997					
ATM Call	0.0013	0.76	0.47	5.94	0.00
	(0.54)	(2.09)			
ATM Put	0.0017	0.76	0.45	4.38	0.02
	(0.71)	(2.00)			
Two ATM Calls	0.0013	0.76	0.46	6.10	0.00
	(0.55)	(2.12)			
Two ATM Puts	0.0014	0.78	0.46	3.84	0.03
	(0.62)	(1.83)			
2 ATM Calls and 2 ATM Puts	0.0013	0.78	0.47	4.82	0.01
	(0.55)	(1.94)			
Panel B: June 1988 – August 1997					
ATM Call	0.0014	0.74	0.42	14.98	0.00
	(0.71)	(3.18)			
ATM Put	0.0028	0.66	0.39	22.16	0.00
	(1.45)	(4.47)			
Two ATM Calls	0.0016	0.73	0.40	14.94	0.00
	(0.80)	(3.25)			
Two ATM Puts	0.0031	0.65	0.38	22.30	0.00
	(1.58)	(4.60)			
2 ATM Calls and 2 ATM Puts	0.0022	0.69	0.40	17.82	0.00
	(1.11)	(3.80)			

Table V

Regression of Realized Variance on Predicted Variance for Various Specifications of $f(\bullet)$

$$f\left(Var_{Realized}(t)\right) = \alpha + \beta f\left(Var_{Implied}(t)\right) + \varepsilon(t)$$

OLS point estimates with *t*-statistics computed from OLS standard errors (in parentheses) for $\alpha = 0$ and $1 - \beta = 0$. $Var_{Realized}(t)$ is the realized variance of the SPX returns derived from five minute interval SPX index levels inferred from SPX futures prices. The realized variance is computed under the restriction that the mean return is equal to zero. $Var_{Implied}(t)$ is the Black-Scholes implied variance from the two closest to ATM calls and the two closest to ATM puts. The Black-Scholes implied variance is computed by using the quantity $Se^{-\delta T}$, inferred from futures prices, as the value of the SPX index level and then assuming that the SPX index pays no dividends. The Black-Scholes implied variance is defined as the real number that minimizes the sum of the squared pricing errors of the four options (i.e., it is *not* the average of the individual Black-Scholes implied variances). The *F*-statistic tests the hypothesis $\alpha = 0$ and $\beta = 1$. The *p*-value gives the probability that the *F*-statistic is as large as it is if the hypothesis that $\alpha = 0$ and $\beta = 1$ is true. There is one observation for each Wednesday in the data period that follows an option expiration date, and the time interval covered by each observation extends from that Wednesday to the next option expiration date. $I(\bullet)$ is the identity function.

$f(\bullet)$	α	β	Adj R^2	*F*-statistic	*p*-value
		Panel A: June 1993 – August 1997			
$I(\bullet)$	0.0013 (0.55)	0.78 (1.94)	0.47	4.82	0.01
$\sqrt{\bullet}$	0.0019 (0.12)	0.88 (0.98)	0.52	5.96	0.00
$\ln(\bullet)$	-0.1726 (-0.32)	1.02 (-0.18)	0.55	8.96	0.00
$\ln\left(\sqrt{\bullet}\right)$	-0.0863 (-0.32)	1.02 (-0.18)	0.55	8.96	0.00
		Panel B: June 1988 – August 1997			
$I(\bullet)$	0.0022 (1.11)	0.69 (3.80)	0.40	17.82	0.00
$\sqrt{\bullet}$	0.0106 (0.93)	0.80 (2.59)	0.48	23.09	0.00
$\ln(\bullet)$	-0.5470 (-1.71)	0.94 (0.75)	0.55	30.44	0.00
$\ln\left(\sqrt{\bullet}\right)$	-0.2735 (-1.71)	0.94 (0.75)	0.55	30.44	0.00

Table VI

$f(\bullet) = I(\bullet)$ Regressions for Historical Variance Prediction

$$Var_{Realized}(t) = \alpha + \beta Var_{Historical}(t) + \varepsilon(t)$$

OLS point estimates with *t*-statistics computed from OLS standard errors (in parentheses) for $\alpha = 0$ and $1 - \beta = 0$. $Var_{Realized}(t)$ is the realized variance of the SPX returns derived from five minute interval SPX index levels inferred from SPX futures prices. The realized variance is computed under the restriction that the mean return is equal to zero. $Var_{Historical}(t)$ is the historical variance computed from five minute observations on the SPX level derived from futures prices also under the restriction that the mean return is equal to zero. The historical variance is computed either from 1 month, 2 months, 3 months, 6 months, 1 year, or 2 years of SPX levels leading up to the observation time. The *F*-statistic tests the hypothesis $\alpha = 0$ and $\beta = 1$. The *p*-value gives the probability that the *F*-statistic is as large as it is if the hypothesis that $\alpha = 0$ and $\beta = 1$ is true. There is one observation for each Wednesday in the data period that follows an option expiration date, and the time interval covered by each observation extends from that Wednesday to the next option expiration date.

Past Data for Historical Prediction	α	β	Adj R^2	*F*-statistic	*p*-value
Panel A: June 1993 – August 1997					
1 Month	0.0045	0.75	0.41	2.39	0.10
	(2.13)	(2.02)			
2 Months	0.0035	0.84	0.40	1.19	0.31
	(1.49)	(1.10)			
3 Months	0.0028	0.90	0.38	0.94	0.40
	(1.14)	(0.64)			
6 Months	0.0006	1.10	0.42	1.61	0.21
	(0.24)	(-0.57)			
1 Year	-0.0031	1.50	0.45	5.63	0.01
	(-1.02)	(-2.14)			
2 Years	-0.0114	2.37	0.42	11.83	0.00
	(-2.53)	(-3.51)			
Panel B: June 1988 – August 1997					
1 Month	0.0074	0.57	0.28	12.21	0.00
	(4.12)	(4.99)			
2 Months	0.0052	0.69	0.32	5.17	0.01
	(2.76)	(3.25)			
3 Months	0.0041	0.76	0.34	2.89	0.06
	(2.12)	(2.43)			
6 Months	0.0045	0.73	0.28	3.01	0.05
	(2.10)	(2.47)			
1 Year	0.0161	0.04	0.01	597.22	0.00
	(11.53)	(34.20)			
2 Years	0.0163	0.02	0.00	631.86	0.00
	(10.49)	(33.28)			

Table VII

$f(\bullet) = I(\bullet)$ Regression with Option Implied and Historical Predictor Variables

$$Var_{Realized}(t) = \alpha + \beta Var_{Implied}(t) + \gamma Var_{Historical}(t) + \varepsilon(t)$$

OLS point estimates with *t*-statistics computed from OLS standard errors (in parentheses) for $\alpha = 0,\ 1-\beta = 0,$ and $\gamma = 0.$ $Var_{Realized}(t)$ is the realized variance of the SPX returns derived from five minute interval SPX index levels inferred from SPX futures prices. The realized variance is computed under the restriction that the mean return is equal to zero. $Var_{Implied}(t)$ is the Black-Scholes implied variance from the two closest to ATM calls and the two closest to ATM puts. The Black-Scholes implied variance is computed by using the quantity $Se^{-\delta T}$, inferred from futures prices, as the value of the SPX index level and then assuming that the SPX index pays no dividends. The Black-Scholes implied variance is defined as the real number that minimizes the sum of the squared pricing errors of the four options (i.e., it is *not* the average of the individual Black-Scholes implied variances). $Var_{Historical}(t)$ is the historical variance computed from five minute observations on the SPX level derived from futures prices also under the restriction that the mean return is equal to zero. The historical variance is computed either from 1 month, 2 months, 3 months, or 6 months of SPX levels leading up to the observation time. The *F*-statistic tests the hypothesis $\alpha = 0,\ \beta = 1,$ and $\gamma = 0.$ The *p*-value gives the probability that the *F*-statistic is as large as it is if the hypothesis that $\alpha = 0,\ \beta = 1,$ and $\gamma = 0$ is true. There is one observation for each Wednesday in the data period that follows an option expiration date, and the time interval covered by each observation extends from that Wednesday to the next option expiration date.

Past Data for Historical Prediction, i.e. $Var_{Historical}(t)$	α	β	γ	Adj R^2	*F*-stat	*p*-value
Panel A: June 1993 – August 1997						
1 Month	0.0015	0.64	0.15	0.46	5.02	0.01
	(0.63)	(1.37)	(0.55)			
2 Months	0.0011	0.60	0.25	0.47	5.39	0.01
	(0.45)	(1.84)	(0.97)			
3 Months	0.0011	0.68	0.13	0.46	4.94	0.01
	(0.46)	(1.32)	(0.42)			
6 Months	0.0003	0.57	0.36	0.47	5.43	0.01
	(0.13)	(1.80)	(1.00)			
Panel B: June 1988 – August 1997						
1 Month	0.0019	0.82	-0.14	0.40	18.30	0.00
	(0.92)	(1.02)	(-0.84)			
2 Months	0.0020	0.55	0.19	0.40	18.71	0.00
	(1.02)	(3.08)	(1.20)			
3 Months	0.0016	0.50	0.28	0.41	19.78	0.00
	(0.80)	(3.62)	(1.72)			
6 Months	0.0014	0.59	0.18	0.40	18.76	0.00
	(0.68)	(3.41)	(1.19)			

Table VIII

$f(\bullet) = I(\bullet)$ Regression for Heston Model and Various Estimation Methods, June 1993 through August 1997

$$Var_{Realized}(t) = \alpha + \beta Var_{Implied}(t) + \varepsilon(t)$$

OLS point estimates with *t*-statistics computed from OLS standard errors (in parentheses) for $\alpha = 0$ and $1 - \beta = 0$. $Var_{Realized}(t)$ is the realized variance of the SPX returns derived from five minute interval SPX index levels inferred from SPX futures prices. The realized variance is computed under the restriction that the mean return is equal to zero. $Var_{Implied}(t)$ is the Heston model implied variance from the two closest to ATM calls and the two closest to ATM puts. Estimates of the Heston model are obtained either from Pan (1999) which would not have been implementable for SPX market investors or using an implementable method. The implementable method obtains estimates of the parameters of the variance process $(\text{i.e., of } k, \theta, \text{ and } \eta)$ from a maximum likelihood estimate of the time-series of monthly variances obtained from CRSP SPX index closing level from July 1962 through the month before an observation time. Estimates of ρ and λ are obtained by minimizing the squared pricing errors of all out-of-the-money SPX option 3:00 PM bid-ask midpoints that meet some mild exclusionary criteria described in the text given $k, \theta,$ and η. Once the parameters of the Heston model are determined through either method, the Heston implied instantaneous variance is computed by using the quantity $Se^{-\delta T}$, inferred from futures prices, as the value of the SPX index level and then assuming that the SPX index pays no dividends. The Heston implied instantaneous variance is defined as the real number that minimizes the sum of the squared pricing errors of the four options under the Heston model (i.e., it is *not* the average of the individual Heston implied instantaneous variances). Once the Heston implied instantaneous variance has been obtained, the expected average variance over the life of the option is obtained from equation (16) in the text and assigned to $Var_{Implied}(t)$. The *F*-statistic tests the hypothesis $\alpha = 0$ and $\beta = 1$. The *p*-value gives the probability that the *F*-statistic is as large as it is if the hypothesis that $\alpha = 0$ and $\beta = 1$ is true. There is one observation for each Wednesday in the data period that follows an option expiration date, and the time interval covered by each observation extends from that Wednesday to the next option expiration date.

Estimation of Heston Model	α	β	Adj R^2	*F*-statistic	*p*-value
Pan (1999)	0.0009 (0.37)	0.97 (0.19)	0.46	0.11	0.89
Implementable	-0.0001 (-0.03)	0.88 (0.90)	0.46	2.25	0.12

Table IX

$f(\bullet) = I(\bullet)$ Regression with Heston Model Option Implied and Historical Predictor Variables, June 1993 through August 1997

$$Var_{Realized}(t) = \alpha + \beta Var_{Implied}(t) + \gamma Var_{Historical}(t) + \varepsilon(t)$$

OLS point estimates with *t*-statistics computed from OLS standard errors (in parentheses) for $\alpha = 0,\ 1-\beta = 0,$ and $\gamma = 0.$ $Var_{Realized}(t)$ is the realized variance of the SPX returns derived from five minute interval SPX index levels inferred from SPX futures prices. The realized variance is computed under the restriction that the mean return is equal to zero. $Var_{Implied}(t)$ is the Heston model implied variance from the two closest to ATM calls and the two closest to ATM puts. Estimates of the Heston model are obtained either from Pan (1999) which would not have been implementable for SPX market investors or using an implementable method. The implementable method obtains estimates of the parameters of the variance process $(\text{i.e., of } k, \theta, \text{ and } \eta)$ from a maximum likelihood estimate of the time-series of monthly variances obtained from CRSP SPX index closing level from July 1962 through the month before an observation time. Estimates of ρ and λ are obtained by minimizing the squared pricing errors of all out-of-the-money SPX option 3:00 PM bid-ask midpoints that meet some mild exclusionary criteria described in the text given $k, \theta,$ and $\eta.$ Once the parameters of the Heston model are determined through either method, the Heston implied instantaneous variance is computed by using the quantity $Se^{-\delta T},$ inferred from futures prices, as the value of the SPX index level and then assuming that the SPX index pays no dividends. The Heston implied instantaneous variance is defined as the real number that minimizes the sum of the squared pricing errors of the four options under the Heston model (i.e., it is *not* the average of the individual Heston implied instantaneous variances). Once the Heston implied instantaneous variance has been obtained, the expected average variance over the life of the option is obtained from equation (16) in the text and assigned to $Var_{Implied}(t).$ $Var_{Historical}(t)$ is the historical variance computed from five minute observations on the SPX level derived from futures prices also under the restriction that the mean return is equal to zero. The historical variance is computed either from 1 month, 2 months, 3 months, or 6 months of SPX levels leading up to the observation time. The *F*-statistic tests the hypothesis $\alpha = 0,\ \beta = 1,$ and $\gamma = 0.$ The *p*-value gives the probability that the *F*-statistic is as large as it is if the hypothesis that $\alpha = 0,\ \beta = 1,$ and $\gamma = 0$ is true. There is one observation for each Wednesday in the data period that follows an option expiration date, and the time interval covered by each observation extends from that Wednesday to the next option expiration date.

Past Data for Historical Prediction, i.e. $Var_{Historical}(t)$	α	β	γ	Adj R^2	*F*-stat	*p*-value
Panel A: Heston Model Pan (1999) Estimation						
1 Month	0.0007	1.05	-0.08	0.45	0.18	0.84
	(0.29)	(-0.19)	(-0.34)			
2 Month	0.0009	0.87	0.13	0.45	0.23	0.80
	(0.36)	(0.49)	(0.48)			
3 Month	0.0009	0.98	-0.01	0.45	0.11	0.89
	(0.37)	(0.09)	(-0.03)			
6 Month	0.0002	0.80	0.29	0.46	0.38	0.68
	(0.08)	(0.71)	(0.71)			
Panel B: Heston Model Implementable Estimation						
1 Month	-0.0004	0.98	-0.11	0.45	2.36	0.10
	(-0.16)	(0.09)	(-0.44)			

Table IX – Continued

Past Data for Historical Prediction, i.e. $Var_{Historical}(t)$	α	β	γ	Adj R^2	F Stat	p-Value
2 Month	0.0000 (0.00)	0.79 (0.90)	0.13 (-0.36)	0.45	2.38	0.10
3 Month	-0.0001 (-0.03)	0.87 (0.49)	0.01 (0.03)	0.45	2.25	0.12
6 Month	-0.0007 (-0.25)	0.71 (1.17)	0.33 (0.82)	0.45	2.62	0.08

Table X

Simulation of $f(\bullet) = I(\bullet)$ Forecasting Regression Under the Null, June 1993 through August 1997

$$Var_{Realized}(t) = \alpha + \beta Var_{Implied}(t) + \varepsilon(t)$$

These simulations assume that the true (unobserved) dynamics of the underlying SPX index level and SPX variance are described by equations (11)-(15). The drift of the index return is set to $\mu(S_t, V_t, t) = 0.12,$ and the parameters of the variance process and the correlation between the Weiner processes are taken from Pan (1999) and have the values $k = 7.1,$ $\theta = 0.0137,$ $\eta = 0.32,$ and $\rho = -0.53.$ For each experiment, 1000 sample paths are simulated. Each sample path begins at an SPX level of 210 and a variance level of 0.0137. The sample paths are obtained by simulating equations (11)-(13) from the initial values for 238,869 steps using a bivariate Euler scheme where each step corresponds to five minutes of trading time. This procedure generates sample paths that start with an underlying value equal to the level of the SPX index on January 2, 1986 and a variance level equal to the assumed long-run mean of the variance process and that cover at five minute trading time intervals the period until September 18, 1997 which is the expiration date for the option observed at the final observation date in August 1997. The interest rate is set to $r = 0.05$ and the dividend rate is set to $\delta = 0.025.$ Error was introduced into both the SPX index levels from which the realized variances were computed and also the option prices from which the implied variances were computed. Error was introduced into the SPX index levels as follows. For each of the 238,869 steps on each simulated path, spot-futures parity was used to convert the SPX level into a futures price for each of the three delivery times that were actually used at the corresponding time during the tests on the real data. A bandwidth of either $0.00, $0.10, or $0.20 was then centered on each of the spot-futures parity derived futures prices, and a price for each of the three futures was drawn uniformly from its band. Error was introduced into the option prices as follows. On each observation date (i.e. the Wednesday after an option expiration) the ATM strike price level was determined from the current (true) level of the SPX index, the time to expiration of the options, $r,$ and $\delta.$ The strike prices of the two closest to ATM calls and two closest to ATM puts were then set equal to the integer multiple of five less than and greater than the ATM strike price level. Next, the true price of these options was set to their values under the Heston model given the assumed parameters values. A band with width of either zero, the bid-ask spread of the corresponding call or put in the real data observed on the same trade date (i.e. the call or put whose strike price was also just below or just above the ATM strike price), or twice that bid-ask spread was then centered on each of the options prices, and a price for each of the four options was drawn uniformly from its band. Once the futures and options prices with error have been simulated, the $Var_{Realized}(t)$ and $Var_{Implied}(t)$ variables are constructed exactly as they were for the real data. The regression equation is estimated for each sample path using the sample path's simulated data for each of the observation dates from June 1993 through August 1997.

	Fractiles of Statistics												
Statistic	0.01	0.05	0.10	0.50	0.90	0.95	0.99	Mean	Std. Dev.	Obs. Stat.	Empirical Percentile	Emp. 1-sided *p*-value	Emp. 2-sided *p*-value
Panel A: Futures Spread = 0, Options Spread = 0, $\lambda = 0$													
α	-0.0020	-0.0013	-0.0008	0.0008	0.0025	0.0029	0.0037	0.0008	0.0013	0.0013	0.66	0.34	0.39
β	0.69	0.77	0.82	0.97	1.11	1.14	1.21	0.97	0.11	0.78	0.06	0.06	0.07
R^2	0.42	0.51	0.55	0.70	0.80	0.83	0.87	0.69	0.10	0.47	0.03	0.03	
F-stat	0.01	0.08	0.16	1.09	3.36	4.57	7.63	1.52	1.56	4.82	0.96	0.05	

Table X – Continued

Statistic	Fractiles of Statistics 0.01	0.05	0.10	0.50	0.90	0.95	0.99	Mean	Std. Dev.	Obs. Stat.	Empirical Percentile	Emp. 1-sided *p*-value	Emp. 2-sided *p*-value
Panel B: Futures Spread = 0, Options Spread = 1*x* actual, $\lambda = 0$													
α	-0.0022	-0.0013	-0.0008	0.0008	0.0026	0.0032	0.0045	0.0008	0.0014	0.0013	0.67	0.33	0.39
β	0.66	0.76	0.81	0.97	1.11	1.15	1.21	0.96	0.12	0.78	0.06	0.06	0.07
R^2	0.42	0.50	0.56	0.69	0.80	0.82	0.87	0.68	0.10	0.47	0.02	0.02	
F-stat	0.02	0.08	0.14	1.02	3.59	5.11	7.78	1.55	1.78	4.82	0.94	0.06	
Panel C: Futures Spread = 0.10, Options Spread = 0, $\lambda = 0$													
α	-0.0020	-0.0012	-0.0007	0.0009	0.0027	0.0033	0.0041	0.0010	0.0013	0.0013	0.61	0.39	0.43
β	0.72	0.79	0.83	0.98	1.10	1.15	1.23	0.97	0.11	0.78	0.05	0.05	0.06
R^2	0.45	0.53	0.56	0.70	0.80	0.83	0.87	0.69	0.10	0.47	0.02	0.02	
F-stat	0.02	0.10	0.21	1.24	3.86	4.98	8.76	1.74	1.77	4.82	0.95	0.06	
Panel D: Futures Spread = 0.10, Options Spread = 1*x* actual, $\lambda = 0$													
α	-0.0016	-0.0010	-0.0007	0.0009	0.0027	0.0033	0.0042	0.0010	0.0013	0.0013	0.60	0.40	0.42
β	0.70	0.78	0.82	0.97	1.11	1.14	1.22	0.97	0.11	0.78	0.05	0.05	0.06
R^2	0.40	0.51	0.55	0.70	0.80	0.83	0.87	0.69	0.10	0.47	0.03	0.03	
F-stat	0.02	0.10	0.19	1.27	3.81	4.50	7.79	1.73	1.74	4.82	0.95	0.05	
Panel E: Futures Spread = 0, Options Spread = 0, $\lambda = -7.6$													
α	-0.0019	-0.0009	-0.0006	0.0009	0.0027	0.0032	0.0040	0.0010	0.0013	0.0013	0.61	0.39	0.42
β	0.53	0.61	0.64	0.76	0.86	0.90	0.94	0.76	0.09	0.78	0.59	0.59	0.59
R^2	0.44	0.52	0.56	0.69	0.80	0.82	0.87	0.69	0.09	0.47	0.02	0.02	
F-stat	4.21	6.83	8.57	16.85	31.10	38.25	49.85	18.76	9.47	4.82	0.01	0.99	
Panel F: Futures Spread = 0.10, Options Spread = 1*x* actual, $\lambda = -7.6$													
α	-0.0019	-0.0010	-0.0004	0.0011	0.0029	0.0035	0.0046	0.0011	0.0013	0.0013	0.57	0.44	0.47
β	0.55	0.60	0.64	0.76	0.87	0.90	0.97	0.76	0.09	0.78	0.57	0.57	0.57
R^2	0.42	0.50	0.56	0.70	0.80	0.83	0.88	0.69	0.10	0.47	0.03	0.03	
F-stat	3.28	6.28	7.44	15.65	29.20	33.79	44.96	17.19	8.75	4.82	0.02	0.98	
Panel G: Futures Spread = 0.20, Options Spread = 2*x* actual, $\lambda = -7.6$													
α	-0.0016	-0.0006	-0.0001	0.0018	0.0038	0.0046	0.0055	0.0018	0.0016	0.0013	0.37	0.63	0.65
β	0.54	0.61	0.64	0.76	0.87	0.90	0.96	0.76	0.09	0.78	0.59	0.59	0.59
R^2	0.43	0.50	0.55	0.69	0.80	0.83	0.88	0.68	0.10	0.47	0.04	0.04	
F-stat	1.50	3.30	5.02	11.62	24.43	28.60	35.44	13.17	7.56	4.82	0.09	0.91	

Table XI

Simulation of $f(\bullet) = I(\bullet)$ Regression with Option Implied and Historical Predictor Variables

$$Var_{Realized}(t) = \alpha + \beta Var_{Implied}(t) + \gamma Var_{Historical}(t) + \varepsilon(t)$$

These simulations assume that the true (unobserved) dynamics of the underlying SPX index level and SPX variance are described by equations (11)-(15). The drift of the index return is set to $\mu(S_t, V_t, t) = 0.12,$ and the parameters of the variance process and the correlation between the Weiner processes are taken from Pan (1999) and have the values $k = 7.1,$ $\theta = 0.0137,$ $\eta = 0.32,$ and $\rho = -0.53.$ For each experiment, 1000 sample paths are simulated. Each sample path begins at an SPX level of 210 and a variance level of 0.0137. The sample paths are obtained by simulating equations (11)-(13) from the initial values for 238,869 steps using a bivariate Euler scheme where each step corresponds to five minutes of trading time. This procedure generates sample paths that start with an underlying value equal to the level of the SPX index on January 2, 1986 and a variance level equal to the assumed long-run mean of the variance process and that cover at five minute trading time intervals the period until September 18, 1997 which is the expiration date for the option observed at the final observation date in August 1997. The interest rate is set to $r = 0.05$ and the dividend rate is set to $\delta = 0.025.$ Error was introduced into both the SPX index levels from which the realized variances were computed and also the option prices from which the implied variances were computed. Error was introduced into the SPX index levels as follows. For each of the 238,869 steps on each simulated path, spot-futures parity was used to convert the SPX level into a futures price for each of the three delivery times that were actually used at the corresponding time during the tests on the real data. A bandwidth of either $0.00, $0.10, or $0.20 was then centered on each of the spot-futures parity derived futures prices, and a price for each of the three futures was drawn uniformly from its band. Error was introduced into the option prices as follows. On each observation date (i.e. the Wednesday after an option expiration) the ATM strike price level was determined from the current (true) level of the SPX index, the time to expiration of the options, $r,$ and $\delta.$ The strike prices of the two closest to ATM calls and two closest to ATM puts were then set equal to the integer multiple of five less than and greater than the ATM strike price level. Next, the true price of these options was set to their values under the Heston model given the assumed parameters values. A band with width of either zero, the bid-ask spread of the corresponding call or put in the real data observed on the same trade date (i.e. the call or put whose strike price was also just below or just above the ATM strike price), or twice that bid-ask spread was then centered on each of the options prices, and a price for each of the four options was drawn uniformly from its band. Once the futures and options prices with error have been simulated, the $Var_{Realized}(t),$ $Var_{Implied}(t),$ and $Var_{Historical}(t)$ variables are constructed exactly as they were for the real data. The regression equation is estimated for each sample path using the sample path's simulated data for each of the observation dates from June 1993 through August 1997.

Statistic	Fractiles of Statistics 0.01	0.05	0.10	0.50	0.90	0.95	0.99	Mean	Std. Dev.	Obs. Stat.	Empirical Percentile	Empirical 1-sided *p*-value	Empirical 2-sided *p*-value
Panel A: Futures Spread = 0, Options Spread = 0, $\lambda = 0$													
α	-0.0021	-0.0012	-0.0007	0.0011	0.0033	0.0039	0.0053	0.0012	0.0016	0.0011	0.50	0.50	0.56
β	0.68	0.76	0.81	0.99	1.16	1.21	1.28	0.99	0.13	0.68	0.01	0.01	0.01
γ	-0.33	-0.23	-0.19	-0.05	0.11	0.15	0.25	-0.05	0.12	0.13	0.93	0.07	0.32
R^2	0.42	0.51	0.55	0.70	0.81	0.83	0.87	0.69	0.10	0.46	0.02	0.02	
F-stat	0.08	0.24	0.41	1.71	4.46	5.79	8.84	2.19	1.85	4.94	0.93	0.07	

Table XI – Continued

Statistic	Fractiles of Statistics 0.01	0.05	0.10	0.50	0.90	0.95	0.99	Mean	Std. Dev.	Obs. Stat.	Empirical Percentile	Empirical 1-sided *p*-value	Empirical 2-sided *p*-value
Panel B: Futures Spread = 0, Options Spread = 1*x* actual, $\lambda = 0$													
α	-0.0022	-0.0013	-0.0007	0.0011	0.0032	0.0040	0.0060	0.0012	0.0017	0.0011	0.51	0.49	0.56
β	0.62	0.74	0.81	0.98	1.15	1.20	1.29	0.98	0.14	0.68	0.02	0.02	0.02
γ	-0.31	-0.24	-0.20	-0.05	0.11	0.18	0.29	-0.04	0.12	0.13	0.92	0.08	0.31
R^2	0.41	0.51	0.55	0.69	0.80	0.82	0.87	0.69	0.10	0.46	0.02	0.02	
F-stat	0.08	0.26	0.40	1.67	4.71	5.99	8.73	2.20	1.97	4.94	0.91	0.09	
Panel C: Futures Spread = 0.10, Options Spread = 0, $\lambda = 0$													
α	-0.0019	-0.0012	-0.0006	0.0012	0.0036	0.0043	0.0060	0.0013	0.0017	0.0011	0.49	0.51	0.57
β	0.69	0.77	0.83	0.98	1.15	1.20	1.28	0.98	0.13	0.68	0.01	0.01	0.01
γ	-0.30	-0.22	-0.19	-0.04	0.12	0.16	0.25	-0.04	0.12	0.13	0.92	0.08	0.32
R^2	0.44	0.53	0.56	0.70	0.81	0.83	0.87	0.69	0.10	0.46	0.02	0.02	
F-stat	0.07	0.31	0.52	1.89	4.94	6.32	9.35	2.42	2.05	4.94	0.90	0.10	
Panel D: Futures Spread = 0.10, Options Spread = 1*x* actual, $\lambda = 0$													
α	-0.0019	-0.0010	-0.0006	0.0012	0.0034	0.0044	0.0059	0.0013	0.0017	0.0011	0.48	0.52	0.56
β	0.69	0.77	0.81	0.97	1.15	1.20	1.28	0.98	0.13	0.68	0.01	0.01	0.01
γ	-0.30	-0.24	-0.19	-0.04	0.12	0.16	0.25	-0.04	0.12	0.13	0.92	0.08	0.31
R^2	0.40	0.50	0.55	0.70	0.80	0.84	0.88	0.69	0.10	0.46	0.02	0.02	
F-stat	0.11	0.28	0.47	1.97	4.67	5.96	8.46	2.41	2.01	4.94	0.92	0.08	
Panel E: Futures Spread = 0, Options Spread = 0, $\lambda = -7.6$													
α	-0.0019	-0.0010	-0.0004	0.0012	0.0035	0.0040	0.0055	0.0014	0.0016	0.0011	0.47	0.53	0.57
β	0.53	0.60	0.64	0.77	0.90	0.94	1.00	0.77	0.10	0.68	0.18	0.18	0.18
γ	-0.29	-0.22	-0.19	-0.05	0.10	0.14	0.22	-0.05	0.11	0.13	0.94	0.06	0.29
R^2	0.43	0.51	0.56	0.69	0.80	0.83	0.87	0.69	0.10	0.46	0.02	0.02	
F-stat	4.53	7.30	9.20	17.87	32.40	39.18	50.50	19.78	9.69	4.94	0.01	0.99	
Panel F: Futures Spread = 0.10, Options Spread = 1*x* actual, $\lambda = -7.6$													
α	-0.0021	-0.0009	-0.0004	0.0013	0.0036	0.0044	0.0059	0.0015	0.0016	0.0011	0.42	0.58	0.62
β	0.53	0.60	0.63	0.78	0.90	0.95	1.02	0.77	0.11	0.68	0.20	0.20	0.20
γ	-0.36	-0.23	-0.19	-0.04	0.11	0.16	0.25	-0.04	0.12	0.13	0.92	0.08	0.30
R^2	0.41	0.50	0.55	0.70	0.81	0.84	0.88	0.69	0.10	0.46	0.02	0.02	
F-stat	3.78	6.82	8.24	16.64	30.46	35.13	46.34	18.26	9.11	4.94	0.02	0.98	

Table XI – Continued

Statistic	Fractiles of Statistics 0.01	0.05	0.10	0.50	0.90	0.95	0.99	Mean	Std. Dev.	Obs. Stat.	Empirical Percentile	Empirical 1-sided *p*-value	Empirical 2-sided *p*-value
Panel G: Futures Spread = 0.20, Options Spread = 2*x* actual, $\lambda = -7.6$													
α	-0.0016	-0.0006	0.0000	0.0019	0.0045	0.0052	0.0073	0.0021	0.0018	0.0011	0.32	0.68	0.70
β	0.50	0.59	0.63	0.76	0.90	0.95	1.03	0.76	0.11	0.68	0.22	0.22	0.22
γ	-0.30	-0.22	-0.18	-0.03	0.15	0.19	0.28	-0.02	0.13	0.13	0.88	0.12	0.32
R^2	0.43	0.50	0.55	0.69	0.80	0.84	0.88	0.68	0.10	0.46	0.03	0.03	
F-stat	2.24	4.06	5.74	12.52	25.14	30.05	38.34	14.11	7.77	4.94	0.07	0.93	