



WORKING PAPER SERIES

Forecasting Foreign Exchange Volatility: Why Is Implied
Volatility Biased and Inefficient?
And Does It Matter?

Christopher J. Neely

Working Paper 2002-017D
<http://research.stlouisfed.org/wp/2002/2002-017.pdf>

August 2002
Revised March 2004

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
411 Locust Street
St. Louis, MO 63102

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors. Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Federal Reserve Bank of St. Louis Working Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

Photo courtesy of The Gateway Arch, St. Louis, MO. www.gatewayarch.com

Forecasting Foreign Exchange Volatility: Why Is Implied Volatility Biased and Inefficient?
And Does It Matter?

Christopher J. Neely*

First draft: August 20, 2002

This draft: March 25, 2004

Abstract: Research has consistently found that implied volatility is a conditionally biased predictor of realized volatility across asset markets. This paper evaluates explanations for this bias in the market for options on foreign exchange futures. No solution considered—including a model of priced volatility risk—explains the conditional bias found in implied volatility. Further, while implied volatility fails to subsume econometric forecasts in encompassing regressions, these forecasts do not significantly improve delta-hedging performance. Thus this paper deepens the implied volatility puzzle by rejecting popular explanations for forecast bias while demonstrating that statistical measures of bias and informational inefficiency should be treated with circumspection.

Keywords: exchange rate, option, implied volatility, GARCH, long-memory, ARIMA, high-frequency

JEL subject numbers: F31, G15

* Research Officer, Research Department
Federal Reserve Bank of St. Louis
P.O. Box 442 St. Louis, MO 63166
(314) 444-8568, (314) 444-8731 (fax)
neely@stls.frb.org

The author thanks David Bates, Toby Daglish, Jon Faust, Joshua Rosenberg, Charles Whiteman, Steve Weinberg, Paul Weller, Jonathan Wright, and other participants at the following presentations for helpful comments on earlier drafts: the Federal Reserve System Committee on International Economics Fall 2002; Quantitative Finance 2002; Rutgers; the New York Fed; the University of Iowa and Academia Sinica's Conference on Analysis of High-Frequency Financial Data and Market Microstructure. The author also thanks Gurdip Bakshi for correspondence on the assumptions of his option pricing model. Rob Dittmar provided computer code for the computation of stochastic volatility options pricing formulas. Charles Hokayem provided excellent research assistance. The views expressed are those of the author and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System. Any errors are my own.

Forecasting Foreign Exchange Volatility: Why Is Implied Volatility Biased and Inefficient? And Does It Matter?

Abstract: Research has consistently found that implied volatility is a conditionally biased predictor of realized volatility across asset markets. This paper evaluates explanations for this bias in the market for options on foreign exchange futures. No solution considered—including a model of priced volatility risk—explains the conditional bias found in implied volatility. Further, while implied volatility fails to subsume econometric forecasts in encompassing regressions, these forecasts do not significantly improve delta-hedging performance. Thus this paper deepens the implied volatility puzzle by rejecting popular explanations for forecast bias while demonstrating that statistical measures of bias and informational inefficiency should be treated with circumspection.

Options prices depend on the volatility of the underlying asset price. If the correct mapping between the option's price and realized variance is known, then the implied variance (IV) generated by inverting an options pricing formula should be the best possible forecast of future volatility. If IV were not the best forecast, it might be possible to earn excess returns by trading on better predictions. With this motivation, many authors have investigated whether IV is an unbiased and informationally efficient predictor of realized variance (RV) in a variety of markets. Almost all such research has found that IV is conditionally biased, but conclusions on informational efficiency have been mixed.

Of course, tests of the predictive properties of IV are implicitly joint tests of market efficiency and the testing procedures, including the correct option pricing model. Several deficiencies in testing procedures have been suggested for the apparent predictive bias in IV.¹ Engle and Rosenberg (2000) blame sample selection bias for the predictive bias in S&P 500 index options. Christensen, Hansen, and Prabhala (2001) argue that the traditional estimation with telescoping (overlapping) observations is inappropriate, producing inconsistent parameter estimates as the forecast horizon increases. Poteshman (2000) finds that measuring realized volatility with high-frequency data eliminates half the predictive bias in IV and that incorporating a model that prices volatility risk removes much of the rest. Along the same lines, Chernov (2002) makes a case that permitting volatility risk to be priced resolves the conditional bias puzzle in several markets. In short, the literature has attributed IV's conditional bias to testing procedures.

This paper evaluates these explanations for conditional bias and inefficiency in IV. In doing so, it extends the previous literature in several ways. The stochastic volatility (SV) options pricing model of Heston (1993) is used to derive an expectation of RV. High-frequency foreign exchange data precisely characterize volatility, and more sophisticated econometric models—some with long-memory—are used to forecast volatility. To alleviate distributional problems of estimates with telescoping samples, horizon-by-horizon forecasts are also used. To increase the power of such tests, this study uses a much longer span of data than previous research and pools across exchange rates. Finally, delta hedging exercises supplement the statistical criteria to assess the economic value of alternative volatility forecasts.

The evidence confirms that IV is a biased forecast of realized foreign exchange volatility. No

hypothesis considered here plausibly explains the bias and inefficiency found in options on foreign exchange futures. There is no evidence of sample selection bias. Estimating RV with intraday data does not reduce the conditional bias. Fixed horizon estimation does not eliminate the bias. Finally, this paper finds little support for the hypothesis that a non-zero price of volatility risk generates the conditional bias.

Contrary to much research, IV is not informationally efficient; econometric forecasts are useful adjuncts to IV in out-of-sample forecasts. The statistical metric might not be very informative, however. Econometric forecasts do not usefully augment IV in an out-of-sample delta hedging exercise.

The next section describes the expected relation between realized and IV. Then the data and econometric methods are reviewed. The fifth section presents and discusses the statistical forecasting results. The economic value of the forecasts is then evaluated through a delta hedging exercise. A final section concludes.

2. Option prices and realized volatility

Implicit variance from the Black-Scholes formula

Almost all research on the properties of IV derives estimates of IV from a variant of the Black-Scholes (1972) formula, which counterfactually assumes constant volatility. The justification for using a constant-volatility model to predict SV was provided by Hull and White (1987). Assuming that volatility evolves independently of the underlying price and that no priced risk is associated with the option, Hull and White (1987) showed that the correct price of a European option would equal the expectation of the Black-Scholes (BS) formula, evaluating the variance argument at average variance until expiry:

$$(1) \quad C(S_t, V_t, t) = \int_t^T C^{BS}(\bar{V}) h(\bar{V} | \sigma_t^2) d\bar{V} = E[C^{BS}(\bar{V}_{t,T}) | V_t],$$

where the average volatility until expiry is denoted as: $\bar{V}_{t,T} = \frac{1}{T-t} \int_t^T V_\tau d\tau$.²

Bates (1996) takes a second-order Taylor series approximation to the BS price for an at-the-money option to approximate the relation between the BS IV and expected variance until expiry (Appendix A):

$$(2) \quad \hat{\sigma}_{BS}^2 \approx \left(1 - \frac{1}{8} \frac{\text{Var}(\bar{V}_{t,T})}{(E_t \bar{V}_{t,T})^2} \right)^2 E_t \bar{V}_{t,T}.$$

In other words, IV from the BS formula (σ_{BS}^2) will understate the expected variance of the asset until expiry ($E_t \bar{V}_{t,T}$). This bias will be very small, however.³

Equation (2) implies that the BS IV is approximately the conditional expectation of RV ($\bar{V}_{t,T}$). The testable implication of this unbiasedness hypothesis is that $\{\alpha, \beta_1\} = \{0, 1\}$ in the following model:

$$(3) \quad \sigma_{RV,t,T}^2 = \alpha + \beta_1 \sigma_{IV,t,T}^2 + \varepsilon_t,$$

where $\sigma_{RV,t,T}^2$ is the realized variance of the asset return from time t to T and $\sigma_{IV,t,T}^2$ is IV (usually the BS IV) at t for an option expiring at T .⁴ Note that the approximate equality of IV and expected variance in (2) assumes that volatility risk is unpriced; BS IV is more properly called risk-neutral IV.

A related hypothesis is that IV is an informationally efficient predictor of volatility. This issue is usually investigated with variants of the following regression:

$$(4) \quad \sigma_{RV,t,T}^2 = \alpha + \beta_1 \sigma_{IV,t,T}^2 + \beta_2 \sigma_{FV,t,T}^2 + \varepsilon_t,$$

where $\sigma_{RV,t,T}^2$ is the RV from t to the expiration of the option at T , $\sigma_{IV,t,T}^2$ is the IV from t to T , and $\sigma_{FV,t,T}^2$ is some alternative forecast of variance from t to T .⁵ The coefficient estimates ($\hat{\beta}_1, \hat{\beta}_2$) measure the incremental forecasting value of the IV and statistical forecast. Leaving aside the data snooping problem, if one rejects that $\beta_2 = 0$ for some $\sigma_{FV,t,T}^2$, then one rejects that IV is informationally efficient.

The Properties of implicit volatility

Across many asset classes and sample periods, researchers estimating versions of (3) have found that $\hat{\alpha}$ is positive and $\hat{\beta}_1$ is less than one (Canina and Figlewski (1993), Lamoureux and Lastrapes (1993), Jorion (1995), Fleming (1998), Christensen and Prabhala (1998), Szakmary, Ors, Kim, and Davidson (2003)). That is, IV is a significantly biased predictor of RV: A given change in IV is associated with a larger change in the RV.

Tests of informational efficiency provide more mixed results. Kroner, Kneafsey, and Claessens (1993) concluded that combining time series information with IV could produce better forecasts than either technique singly. Blair, Poon, and Taylor (2001) discover that historical volatility provides no incremental information to forecasts from VIX IVs. Li (2002) and Martens and Zein (2002) find that intraday data and long-memory models can improve on IV forecasts of RV in currency markets.

Several hypotheses have been put forward to explain the conditional bias: errors in IV estimation, sample selection bias, estimation with overlapping observations, and poor measurement of RV. Perhaps the most popular solution is that volatility risk is priced. This theory requires some explanation.

The Price of volatility risk

To illustrate the volatility risk problem, consider that there are two sources of uncertainty about the value of an option in a SV environment: the change in the price of the underlying asset and the change in its volatility. An option writer will have to take a position both in the underlying asset (delta hedging) and in another option (vega hedging) to hedge both sources of risk. If the investor only hedges with the underlying asset—not using another option too—then the return to the investor’s portfolio is not certain. It depends on changes in volatility. If such volatility fluctuations represent a systematic risk, then investors must be compensated for exposure to them. In this case, the Hull-White result (1) does not apply and the IV from the BS formula will not be the expectation of objective variance as in (2).

The idea that volatility risk might be priced has been discussed for some time: Hull and White (1987) and Heston (1993) consider it. Lamoureux and Lastrapes (1993) argued that a price of volatility risk was likely to be responsible for the bias in IVs options on individual stocks. But most empirical work has assumed that this volatility risk premium is zero, that volatility risk could be hedged or is not priced.

Is it reasonable to assume that the volatility risk premium is zero? There is no question that volatility is stochastic, options prices depend on volatility, and risk is ubiquitous in financial markets. If option writers hedge their exposure to volatility by buying other options, who will hold a net short position in options to offset the net long position desired by customers? For these reasons, it seems reasonable to attribute IV’s bias to a non-zero price of volatility risk.

On the other hand, there seems little reason to think that volatility risk itself should be priced. While the volatility of the market portfolio is a priced factor in the intertemporal CAPM (Merton (1973), Campbell (1993)), it is more difficult to see why volatility risk in foreign exchange and commodity markets should be priced. One must appeal to limits-of-arbitrage arguments (Shleifer and Vishny (1997)) to justify a non-zero price of currency volatility risk.

Recently, researchers have paid greater attention to the role of volatility risk in options and equity markets (Poteshman (2000), Bates (2002), Benzoni (2002), Chernov (2002), Pan (2002), Bollerslev and Zhou (2003), and Ang, Hodrick, Xing and Zhang (2003). Poteshman (2000), for example, directly estimated the price of risk function and instantaneous variance from options data, then constructed a measure of IV until expiry from the estimated volatility process to forecast SPX volatility over the same horizon. Benzoni (2002) finds evidence that variance risk is priced in the S&P 500 option market. Using different methods, Chernov (2002) also marshals evidence to support this price of volatility risk thesis. This paper extends past research by examining whether a non-zero price of volatility risk can explain the bias in IV for futures on foreign exchange.

3. The Data

Four kinds of data are used in this paper: daily settlement prices on futures on foreign exchange, daily observations on quarterly options on the foreign exchange futures, high-frequency (30-minute) returns on foreign exchange, and daily U.S. interest rates from the Bank for International Settlements. All data begin on January 2, 1987, and end on December 31, 1998.

The daily futures and options-on-futures on foreign exchange data are Chicago Mercantile Exchange (CME) data from quarterly futures contracts on the German mark (DEM), Japanese yen (JPY), Swiss franc (CHF), and British pound (GBP) against the United States dollar (USD). These futures contracts expire in March, June, September, and December. To construct a series of the most liquid contracts, the futures and options contract data are spliced in the usual way at the beginning of each expiration month. That is, on each day prior to a delivery month, the settlement price (collected at 2:00 p.m. central time) and daily range (high minus low price) for the nearest-to-delivery futures contract are extracted. At the

beginning of each delivery month, the next-to-nearest contract is substituted to avoid illiquidity around delivery. For example, the March contract data are used for all trade dates between the first day of December and the last day of February.

Olsen and Associates provided 5-minute returns on the DEM, JPY, CHF, and GBP spot foreign exchange rates. To construct a daily measure of volatility from intraday data, this paper sums the squared 30-minute returns over each day, from 2:30 p.m. central time to 2:00 p.m. central time.⁶ The intraday (high-frequency) volatility measure for day t can be written as:

$$(5) \quad \sigma_{RV,t}^2 = \sum_{i=1}^{48} r_{i,t}^2 \quad .$$

where $r_{i,t}$ is the log return over the i th half-hourly period on day t .

Andersen and Bollerslev (1998) argue that high-frequency measures provide better estimates of the unobservable volatility process than do the standard deviations of daily returns. Specifically, they assert that squared daily innovations are unbiased but noisy estimates of the true variance process. As one samples the returns process at increasing frequency, estimates of underlying volatility become progressively more efficient, allowing volatility to be treated as observed data at very high frequencies. Poteshman (2000) extends this line of research to show that such a high-frequency measure eliminates ½ the bias in predictions from SPX options.

Rather than considering intraday volatility estimates as a superior substitute for daily volatility, however, one might consider it as a complement. The appropriate volatility horizon might depend on the application; perhaps one should examine the performance of IV for both intraday and daily volatility data.

4. Econometric methodology

Constructing realized variance

Two measures of volatility until expiry are used. The annualized futures measure of variance until expiry is the following:

$$(6) \quad \sigma_{RV,t,T}^2 = \frac{251}{T} \sum_{i=1}^T r_{t+i}^2 ,$$

where r_t is the log futures price change from $t-1$ to t . The high-frequency variance measure is constructed analogously.

Constructing implied variance

The benchmark measure of IV used in this paper comes from the Heston (1993) SV pricing model, under the assumption that volatility risk is unpriced.⁷ The SV model posits that the futures price and volatility evolve as follows:

$$(7) \quad dF = \mu F dt + \sqrt{V} F d\varpi_s ,$$

$$(8) \quad dV = (\theta_v - \kappa_v V) dt + \sigma_v \sqrt{V} d\varpi_v ,$$

where F is the futures price at t ; V is the instantaneous variance of F 's diffusion process, $d\varpi_s$ and $d\varpi_v$ are standard Brownian motion with correlation ρ ; and κ_v , θ_v / κ_v , and σ_v are the adjustment speed, long-run mean, and variation coefficient of the diffusion volatility.

The SV model describes options prices as functions of ρ , κ_v , θ_v , σ_v , as well as asset price (F), strike price (X), interest rates (i), time to expiry ($T-t$), and instantaneous variance (V). Values for ρ , κ_v , θ_v , and σ_v were obtained with joint time series estimation techniques as in Sarwar and Krehbiel (2000). Table 1 shows that the estimated parameter vectors appear to be stable across the four exchange rates.

Conditional on the ρ , κ_v , θ_v , σ_v , F , X , i , and $T-t$, instantaneous variance ($V(t)$) for each business day is chosen to minimize the unweighted sum of the squared percentage differences between the prices implied by the SV pricing formula and the settlement prices for the two nearest-to-the-money call options and two nearest-to-the-money put options for the appropriate futures contract.⁸ Instantaneous variance is

$$(9) \quad V(t) = \arg \min_{\sigma_{i,t}} \sum_{i=1}^4 \left((SV_i(V(t)) - Pr_{i,t}) / Pr_{i,t} \right)^2$$

where $Pr_{i,t}$ is the observed settlement premium (price) of the i th option on day t and $SV_i(*)$ is the appropriate call or put formula as a function of the IV and the parameters of the stochastic processes described in (7) and (8).⁹

Before being used in the minimization of (6), the data were checked to make sure that they obeyed the inequality restrictions implied by the no-arbitrage conditions on American options prices: $C \geq F - X$ and $P \geq X - F$. Options prices that violated these relations were discarded. In addition, the observation was discarded if there was not at least one call and one put price. For a very few cases, the quasi-maximum likelihood (QML) estimation failed to converge and a bisecting grid search was used to find IVs instead. The bisecting search estimates appeared consistent with IVs found through QML estimation.

Expected variance until expiry can be calculated by applying Ito's lemma to $e^{at}V_t$ and using the variance process (4), as shown by Poteshman (2000) and Chernov (2002).

$$(10) \quad E_t(\bar{V}_{t,T}) = \frac{1}{T-t} \int_t^T E_t(V_u) du = \frac{\theta_v}{\kappa_v} + \left(\frac{\theta_v}{\kappa_v} - V_t \right) \frac{1}{(T-t)\kappa_v} \left[e^{-\kappa_v(T-t)} - 1 \right],$$

The expected variance until expiry in (10) is the IV used to predict realized variance until expiry.

Why should one choose the two puts and two calls nearest to the money to estimate IV? Researchers have varied the number of options, the type of options, and the weighting procedure in the estimation process. Nevertheless, it has been common to rely heavily on at-the-money options for three reasons: 1) At-the-money options prices are most sensitive to changes in IV, meaning that changes in IV should be reflected in those options. 2) At-the-money options are usually the most heavily traded, resulting in fewer pricing errors due to illiquidity. 3) Research has suggested that IV from at-the-money options provides the best estimates of future realized volatility (e.g., Beckers (1981)). Therefore, choosing the two nearest calls and two nearest puts for estimating IV each day seems to be a reasonable procedure. Bates (1996) reports that using at-the-money options has become increasingly popular. Bates (2002) updates the excellent review in Bates (1996).

Alternative forecasts

In tests of informational efficiency, four types of models are used to provide alternative forecasts of RV: autoregressive integrated moving average (ARIMA) models, long-memory ARIMA (LM-ARIMA) models, generalized autoregressive conditional heteroskedastic (GARCH) models, and ordinary least squares (OLS) models with several independent variables.¹⁰ The Bayesian Information Criterion (BIC)

chose the specific structure (e.g., the AR and MA orders of the ARIMA model) of each of the four classes of models during an in-sample period, 1987 through 1991 (Schwarz (1978)). The in-sample structure and in-sample coefficient estimates were then fixed and used to forecast RV until expiry out-of-sample, 1992-1998. Appendix B describes the forecasting methods in detail.

Summary statistics

Table 2 displays the summary statistics in percent, per annum, for the DEM one-step standard deviation and its forecasts in the left-hand panel and the analogous statistics for root mean variance until option expiration in the right-hand panel. Although the forecasting exercises are done with variance measures, Table 2 presents standard deviation measures because they are easier to interpret. For brevity, only the DEM statistics are shown; other exchange rates are similar. The top panel measures daily volatility with daily sums of log 30-minute changes; the bottom panel uses daily squared returns.

Mean RV is 9.90 percent per annum by the high-frequency measure and 7.95 percent per annum by the daily futures price measure. The mean one-step-ahead forecast volatility is a bit higher than actual volatility, but the forecast means are based only on in-sample data. Actual volatility over the in-sample period (not shown) is closer to the forecast means. As one might expect, the forecasts are less variable than the RV but much more highly autocorrelated. First-order autocorrelations for the one-step forecasts range from 0.44 to 0.99.

Comparing the top panel with the bottom panel shows that the high-frequency volatility measure (column labeled $\sigma_{RV,t}$) is also much more highly autocorrelated than the measure constructed from daily futures prices. The former has first-order autocorrelation of 0.46 to the latter's figure of 0.08. This is consistent with the idea that daily futures volatility is a noisier estimate of true RV, which high-frequency RV approximates well.

Figure 1 illustrates the behavior of both measures of realized root mean variance until option expiry ($\sigma_{RV,t,T}$) and IV ($\sigma_{IV,t,T}$) for the DEM data. All three series appear to be mean reverting. Although the high-frequency behavior cannot be seen clearly in the figure, IV appears to track both RV series well.

5. Testing for bias and inefficiency using overlapping observations

Is implicit variance an unbiased forecast of realized variance?

The first hypothesis to be investigated is that—if IV is the market’s prediction of future volatility and expectations are rational—IV should be an unbiased estimator of future volatility. In other words, one should find that $\{\alpha, \beta_1\} = \{0, 1\}$ in the following model:

$$(3) \quad \sigma_{RV,t,T}^2 = \alpha + \beta_1 \sigma_{IV,t,T}^2 + \varepsilon_t,$$

This paper initially follows most previous research in estimating (3) with OLS and telescoping samples.¹¹ For overlapping horizons, the residuals in (3) will be autocorrelated and, while OLS estimates are still consistent, the autocorrelation must be dealt with in constructing standard errors. Such data sets are described as “telescoping” because correlation between adjacent errors declines linearly and then jumps up at the point at which contracts are spliced. To construct correct measures of parameter uncertainty, this paper follows Jorion (1995) in using the following covariance estimator:

$$(11) \quad \hat{\Sigma} = (X'X)^{-1} \hat{\Omega} (X'X)^{-1},$$

where $\hat{\Omega} = \sum_{t=1}^T \hat{\varepsilon}_t^2 X_t' X_t + \sum_{s=1}^T \sum_{t=s}^T I(s,t) \hat{\varepsilon}_t \hat{\varepsilon}_s (X_t' X_s + X_s' X_t)$, X is the T by K matrix of regressors, X_t

is the t th row of X , $\hat{\varepsilon}_t$ is the residual at time t , and $I(s,t)$ is an indicator variable that takes the value 1 if the forecast from period s overlaps with the forecast from period t in equation (3).

Table 3 presents the results of such estimation using both intraday and daily volatility measures, over the whole sample, 1987 through 1998. The estimates of the constant, $\hat{\alpha}_1$, are always positive and the $\hat{\beta}_1$ coefficient is also positive, but also much less than the hypothesized value of one, if IV is unbiased. For the high-frequency data, for example, the $\hat{\beta}_1$ coefficients range from 0.51 for the DEM to 0.86 for the JPY.¹² The daily futures data in the lower panel provide a similar range for the $\hat{\beta}_1$ coefficients. The fact that the $\hat{\beta}_1$ coefficients are less than one indicates that IV is an overly volatile predictor of subsequent RV. And the bias looks statistically significant; the penultimate column of Table 3 shows that joint Wald

tests—calculated with a robust covariance matrix—reject that $\{\alpha, \beta_1\}$ equals $\{0,1\}$ for three of the four series using each type of volatility data. The tests fail to reject for the JPY in each case.¹³

Consistent with the idea that high-frequency volatility is a less noisy measure of the unobserved underlying volatility in the market, the R^2 for the intraday data (top panel) are larger in each case than those from the daily futures price data in the bottom panel.

Is implicit variance an informationally efficient forecast of realized variance?

The second hypothesis to be tested is that IV subsumes all publicly available information. To test this proposition, one can forecast volatility until expiry— using the ARIMA, LM-ARIMA, GARCH, OLS models—to see if such predictions add information to IV, as in (4):

$$(4) \quad \sigma_{RV,t,T}^2 = \alpha + \beta_1 \sigma_{IV,t,T}^2 + \beta_2 \sigma_{FV,t,T}^2 + \varepsilon_t,^{14}$$

A statistically significant estimate of β_2 rejects informational efficiency (Fair and Shiller (1990)).

Table 4 presents the results of the OLS regression with telescoping samples and robust standard errors (equation (6)). The estimation was restricted to the out-of-sample period (1992 - 1998) to permit a genuinely ex ante exercise for the econometric forecasts. From left to right, the four panels of Table 4 display the results from forecasts with an ARIMA model, a long-memory ARIMA model, a GARCH(1,1) model, and the OLS model described in Appendix B. The top subpanels use the daily sums of 30-minute squared returns to measure variance while the bottom panel uses the daily futures price variance.

With the high-frequency measure of variance (top panel Table 4), the coefficients on the forecasts ($\hat{\beta}_2$) are positive in 14 of 16 cases and statistically significant in 8 of 16 cases. The futures price measure of volatility provides similar evidence against informational efficiency with positive coefficients on the forecasts ($\hat{\beta}_2$) in 12 of 16 cases, and statistically significant in 11 of 16 cases. All the forecasts do equally well. The positive and statistically significant values found for $\hat{\beta}_2$ contradict the hypothesis that IV subsumes all useful information about future volatility.

Comparing performance across exchange rates, the econometric forecasts do the worst in predicting JPY volatility, for which not a single $\hat{\beta}_2$ coefficient is statistically significant. The JPY was also an

unusual case in the bias tests in Table 3, having large $\hat{\beta}_1$ coefficients that are insignificantly different from one. One might conjecture that frequent, substantial intervention by the Bank of Japan provides IV some advantage over econometric forecasts. The commitment of the Japanese authorities to exchange market management might tie down expectations about volatility. And market participants would account for the Bank of Japan's intervention in option pricing, but the econometric rules could not do likewise.

At the same time, the R^2 s for the high-frequency regressions (top panel of Table 4) are again greater than the corresponding R^2 s for the futures-volatility regressions (bottom panel) for three of the four exchange rates, excluding the DEM.

6. Why is implicit variance biased and informationally inefficient?

Explanations for the apparent bias and inefficiency of IV forecasts fall into two categories: 1) failure of the EMH; 2) or failure of the testing procedures. Economists are understandably reluctant to attribute the puzzles to failure of the EMH, as it seems unlikely that traders would leave money on the table, so attention has focused on the testing procedures, including the possibility that a non-zero price of volatility risk could produce bias and inefficiency.

This section considers whether some known failures of the testing procedures could plausibly generate the bias and inefficiency. Such failures fall into several categories: 1) peso/finance minister problems; 2) measurement error in IV; 3) sample selection bias; 4) use of overlapping samples; or 5) a non-zero price of volatility risk.

Peso problems

Peso or finance-minister problems might lead to estimates that appear to be biased in-sample because of unusual sampling variation. That is, agents might have rationally priced options while taking into account low probability events that were not observed in the sample. Conversely, other low probability events might have been observed too often in the sample. For example, the market might have rationally priced in a low probability of periods of very high volatility that never occurred. If so, IV would appear unconditionally and conditionally biased, producing overly volatile predictions.

Given the ubiquity of the result that IV is a biased predictor of RV across assets and across samples, it seems very unlikely that sample-specific variation is to blame, as discussed by Poteshman (2000). Further, the only way to correct for such problems is through longer spans of data; the 12-year data set used here is already very long by the standards of the options literature. Sampling variation is a very unlikely explanation for the findings of bias and inefficiency.

Measurement error

An obvious candidate explanation for the apparent bias and inefficiency of IV is that IV is measured with error. It is well known that error in the independent variable creates attenuation bias; the estimated coefficient is inconsistent, smaller in absolute value than the true coefficient. Errors-in-variables could also explain IV's failure to subsume other forecasts. Specifically, if both IV and econometric forecasts constitute "noisy" predictions of future volatility, then an optimal predictor will put some weight on each. Christensen and Prabhala (1998) illustrate this point in more detail.

There are two types of measurement error: specification error from using the wrong options model and idiosyncratic error from microstructure effects like asynchronous prices and bid-ask spreads in both the options and the underlying futures. Such spreads could produce a wedge between the theoretically correct options price and the observed settlement price. Bates (2000) decomposes the two types of errors for options on S&P 500 futures contracts and concludes that IV is not very sensitive to the choice of two-factor pricing models. Similarly Bates (1996) concludes that while there is some error in options pricing, at-the-money options provide robust, relatively precise estimates of IV.

Such conclusions also apply to the data set used in this paper. The IV estimates are very similar across three different options pricing models. Table 5 illustrates the similar summary statistics of the IVs from the three option pricing models: the SV pricing model of Heston (1993), the Barone-Adesi and Whaley (1987) early exercise correction to the Black (1976) model, and the Black (1976) model. The IVs from the three models are extremely highly correlated and have very similar summary statistics. This similarity suggests why the results in this paper are robust to the choice of (risk-neutral) option pricing model.

How much error is there in IV estimation? One measure of microstructure error in IV estimation is the maximal difference between the four IVs implied by the two closest calls and two closest puts (six possible pairs of IVs) each day. Results from other exchange rates are omitted for brevity. These differences are small. The median difference is about 22 basis points and the 90th percentile is about 190 basis points.¹⁵

Because these options have slightly different degrees of moneyness, they might imply different volatility (Hull (2002)). To remove variation caused by different degrees of moneyness, one can examine the absolute difference between IVs from put-call pairs of options with exactly the same strike price on the same day.¹⁶ In the absence of bid-ask spreads, transactions costs, or early exercise, these differences should be exactly zero. Indeed, they are very small. The median difference for the DEM, for example, is only 4.3 basis points and the 90th percentile is 12 basis points. These statistics indicate that transactions costs are probably within the 30-basis-point bid-ask spreads (in vols) considered by Jorion (1995) in options on foreign exchange.¹⁷ Experiments conducted in simulations—not reported for brevity—indicate that error of this size has almost no effect on the estimates of bias and efficiency.

Sample selection bias

Engle and Rosenberg (2000) suggest that sample selection bias is responsible for bias in S&P 500 index options. That is, if one cannot observe IV or RV during periods of extremely high RV—perhaps because liquidity dries up because of uncertainty—then there will be sample selection bias in the regression of RV on IV. If there was selection bias, then one might expect that volatility until expiry would be systematically higher or lower on days with missing IV than on other days. To investigate this issue, the difference between mean realized variance until expiry on days with IV missing and other days was calculated. Table 7 shows no pattern in the differences: half are positive, half are negative. Some are statistically significant, some are not. This exercise shows no evidence of selection bias in these data.

Overlapping samples

Regressions in this literature are usually conducted on data with overlapping forecasts, which might produce very poor small-sample estimates (Christensen, Hansen, and Prabhala (2001)). There are at least

two ways to investigate the consequences of such overlapping forecasts. First, one can simulate the distribution of the test statistics under the null hypothesis of unbiased forecasts. This method has the advantage of greater power in pooling all horizons together. Second, one can independently estimate the predictive equation (3) for each forecast horizon. This fixed horizon method is computationally simpler and does not require one to assume that the regression has the same coefficient vector at each horizon. This paper confronts the overlapping observations problem in both ways.

What effects do autocorrelation and errors-in-variables have on the IV coefficient? Both autocorrelation and measurement error in the dependent variable will tend to bias the coefficient toward zero (Mankiw and Shapiro (1986), Stambaugh (1986)). To investigate these effects, this paper follows papers such as Mark (1995), Jorion (1995), Kilian (1999) and Berkowitz and Giorgianni (2001) in judging the significance of the parameters by using a plausible data generating process to simulate the distribution of the parameter estimates under the null.

Both GARCH and log-ARIMA models are used to simulate and predict RV and IV until expiry. Because the daily variance data are truncated at zero, highly skewed, and kurtotic, the ARIMA model was estimated and simulated on a modified logarithmic transformation of the these data. The ARIMA forecasts were then transformed with a Taylor series expansion to produce approximately conditionally unbiased forecasts of RV until expiry. Appendix C describes these transformations in detail.

The simulation procedure for the GARCH/ARIMA models were as follows:

1. For each data set, estimate the GARCH/ARIMA model with the whole sample, saving the estimated coefficients and residuals.
2. Construct 1000 simulated log variance samples by drawing errors from the bootstrap distribution.
3. For each of the 1000 samples, construct RV as the annualized sum until expiry of the squared returns. The sample sizes will be the same as those in Table 3. Construct IV as the optimal multiperiod forecast of RV over the appropriate horizon.
4. Regress simulated RV-until-expiry on simulated IV and save the coefficients and test statistics.
5. Examine whether the coefficients and test statistics from the real data are consistent with the

distribution of the coefficients and test statistics from the simulated data.

The simulated data were checked to ensure that the summary statistics—especially the autocorrelations—of the simulated data were reasonably close to the analogous statistics in the real data and the simulated IV was an approximately unbiased predictor of simulated RV.

Table 8 displays the results of the Monte Carlo experiment simulating the regression of RV on IV using a GARCH-t generating process in the upper panel and a log ARIMA process in the lower panel. The first four columns show statistics on the distribution of the estimates of α ; columns five to eight show statistics on the distribution of the estimates of β ; and the final four columns display the percentage of rejections from the simulated Wald statistics and the simulated R^2 s.

The simulated GARCH model generates considerable bias in the estimates of β , the 5th percentiles of the distributions range from 0.46 to 0.62. And the Wald statistics reject the null from 19 to 67 percent of the time. Even with this substantial bias, the $\hat{\beta}_1$ estimates from the real data are in the left-hand tail of the simulated distributions, except for the JPY. For example, 94 percent of the simulated $\hat{\beta}_1$ s are greater than the $\hat{\beta}_1$ estimated from real DEM intraday data. Although the GARCH model provides a plausible answer in a few cases, it cannot consistently fully replicate the bias found in the IV data.

The lower half of Table 8 shows that data produced under the null of unbiasedness from the log-ARIMA model also produces substantial bias in the estimates of β . The 5th percentiles of the $\hat{\beta}_1$ distributions range from 0.45 to 0.74. The median estimates of $\hat{\beta}_1$ are much larger, however, ranging from 0.84 to 0.95 and the Wald statistics reject from 15 to 24 percent of the time. The simulated R^2 distributions appear to be consistent with most, but not all, of the R^2 s from the actual data (see Table 3).

It appears that persistent-regressor bias in the log-ARIMA data generating process cannot explain the whole conditional bias observed in IV, except for the JPY volatility measures and the daily GBP volatility measure. It would be overreaching, however, to rule out the possibility that some generating process with greater persistence in the independent variable produced the apparently biased and informationally inefficient coefficients that we observe in the data. For example, the fractionally cointegrated relation

between implied and RV found by Bandi and Perron (2003) might produce the necessary persistence.

Is IV Unbiased in Horizon-by-Horizon Estimation? The second method of correcting problems with overlapping samples is to use nonoverlapping samples—fixed forecast horizons—as Christensen, Hansen, and Prabhala (2001) advocate.¹⁸ To examine how IV might vary as a predictor of RV across forecast horizons, one can reestimate equation (3) separately for each forecast horizon ($k = T - t$),

$$(12) \quad \sigma_{RV,t,T}^2 = \alpha_k + \beta_{k,1} \sigma_{IV,t,T}^2 + \varepsilon_t.$$

where α_k and $\beta_{k,1}$ denote the coefficients for a horizon of k days until expiry. Under the null hypothesis that IV is an unbiased predictor of RV, the parameters of (12) are the same across exchange rates. And pooling the data across exchange rates might produce more precise estimates.

One can estimate a system of equations by maximum likelihood, imposing the same coefficient vector on all four exchange rates. The system can be written as follows:

$$(13) \quad \begin{bmatrix} \sigma_{RV,DEM,t,T}^2 \\ \sigma_{RV,JPY,t,T}^2 \\ \sigma_{RV,CHF,t,T}^2 \\ \sigma_{RV,GBP,t,T}^2 \end{bmatrix} = \begin{bmatrix} 1 & \sigma_{IV,DEM,t,T}^2 \\ 1 & \sigma_{IV,JPY,t,T}^2 \\ 1 & \sigma_{IV,CHF,t,T}^2 \\ 1 & \sigma_{IV,GBP,t,T}^2 \end{bmatrix} \begin{bmatrix} \alpha_k \\ \beta_{k,1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t,T}^{DEM} \\ \varepsilon_{t,T}^{JPY} \\ \varepsilon_{t,T}^{CHF} \\ \varepsilon_{t,T}^{GBP} \end{bmatrix}, \quad \text{where} \quad \begin{bmatrix} \varepsilon_{t,T}^{DEM} \\ \varepsilon_{t,T}^{JPY} \\ \varepsilon_{t,T}^{CHF} \\ \varepsilon_{t,T}^{GBP} \end{bmatrix} \sim N(0, \Omega_k).$$

Cases (horizons) with fewer than 20 observations per exchange rate were not estimated, leaving a minimum forecast horizon of five business days and a maximum horizon of 67 business days. There were 28 to 48 observations per exchange rate for each forecast horizon. Most horizons had close to 48 observations per exchange rate.

Figure 2 shows the series of pooled values for $\hat{\alpha}_k$, $\hat{\beta}_{k,1}$, and the p-values for the likelihood ratio (LR) test that $\{\alpha_k, \beta_{k,1}\} = \{0, 1\}$. As in the overlapping horizon results in Table 3, the $\hat{\alpha}_k$ are positive and the $\hat{\beta}_{k,1}$ are less than one. The pooled $\hat{\beta}_{k,1}$ for the futures data (solid line) are marginally greater than those for the high-frequency volatility measure (dashed line). Consistent with this, the LR tests on the futures volatility are often unable to reject the null that $\{\alpha_k, \beta_{k,1}\} = \{0, 1\}$, while the intraday data tests always reject.

Comparing results in Figure 2 with those in Table 3, one sees that the horizon-by-horizon $\hat{\beta}_{k,1}$ are a bit greater than the average $\hat{\beta}_1$ in the overlapping results. The mean intraday and futures $\hat{\beta}_{k,1}$ s in Figure 2 are 0.71 and 0.75, respectively. Pooling the exchange rates, rather than horizon-by-horizon estimation, drives most of this increase in $\hat{\beta}_{k,1}$ from the overlapping results. The means of the unpooled, horizon-by-horizon $\hat{\beta}_{k,1}$ —results omitted for brevity—are just slightly larger than those in the overlapping results. And the tests of the unpooled, horizon-by-horizon tests often fail to reject the unbiasedness null with intraday data. The rejections with the pooling procedure illustrate its greater power; a lack of such power might explain why horizon-by-horizon tests do not reject unbiasedness.

Figure 3 shows the degree of first-order autocorrelation in IV and both volatility measures until expiry, by forecast horizon. As one might expect, the autocorrelation in IV and volatility until expiry tends to increase with the forecast horizon. But it is modest at all horizons. The horizon-by-horizon results in Figure 2 probably suffer little endogenous regressor bias. First-order autocorrelation for IV ranges from 0.1 to 0.5, depending on the horizon. Full results are available in the working paper version of this paper.

Is pooling across exchange rates wise? Under the null hypothesis of unbiasedness, $\{\alpha_k, \beta_{k,1}\} = \{0, 1\}$, one would expect to see evidence in favor of pooling. Therefore testing the pooling restrictions provides another test of unbiasedness; homogeneity is a necessary but insufficient condition for IV to be an unbiased predictor of RV. To formally investigate whether pooling the data over exchange rates is reasonable, one can estimate (13) with and without the constraint that $\{\alpha_k, \beta_{k,1}\}$ are common across exchange rates and perform a likelihood ratio (LR) test of the restriction. Figure 4 illustrates the p-values for such a test for the intraday (high-frequency) volatility measure (dashed lines) and futures price volatility (solid lines) as a function of forecast horizon, k . The LR test rejects pooling in each case for the intraday data and for almost all horizons less than 12 or more than 31 days for the daily futures data. This confirms the consistent LR rejections of the $(\{\alpha_k, \beta_{k,1}\} = \{0, 1\})$ hypothesis for the intraday data. For the futures data, however, the LR pooling tests tell us something that was not apparent from the coefficient

tests: One can reject the (joint) unbiasedness hypothesis for horizons greater than 40 business days.

However, as is well known, classical statistical procedures such as the LR test fix the probability of Type I error (falsely rejecting the true model) but do not consider the probability of Type II error (failing to reject the null under the alternative). To provide a more balanced assessment of the wisdom of pooling at each horizon, one might use a goodness-of-fit measure, such as the BIC. If the BIC for the constrained model is greater than that of the unconstrained model at a given horizon, the BIC favors pooling the data at that horizon. Figure 5 shows the BIC for the constrained model less the BIC for the unconstrained model at each horizon. Positive values favor pooling. The data favor pooling the futures data at almost all horizons but only favor pooling the intraday data at horizons less than 30 days. The fact that the data do not favor pooling intraday data at longer horizons is further evidence of the failure of the unbiasedness hypothesis at longer horizons.

Is IV informationally efficient in fixed horizon tests? The potentially poor properties of overlapping data sets are also worrisome in tests of informational efficiency. There are, at most, 28 out-of-sample observations for each exchange rate with horizon-by-horizon estimation; one must be concerned about the power of such tests. And the coefficient vector in each prediction equation (4) is the same—{0,1,0}—under the joint null that IV is unbiased and subsumes other information. Therefore one can again pool coefficients across exchange rates and estimate a system for each horizon as follows:

$$(14) \quad \begin{bmatrix} \sigma_{R,DEM,t,T}^2 \\ \sigma_{R,JPY,t,T}^2 \\ \sigma_{R,CHF,t,T}^2 \\ \sigma_{R,GBP,t,T}^2 \end{bmatrix} = \begin{bmatrix} 1 & \sigma_{IV,DEM,t,T}^2 & \sigma_{F,DEM,t,T}^2 \\ 1 & \sigma_{IV,JPY,t,T}^2 & \sigma_{F,JPY,t,T}^2 \\ 1 & \sigma_{IV,CHF,t,T}^2 & \sigma_{F,CHF,t,T}^2 \\ 1 & \sigma_{IV,GBP,t,T}^2 & \sigma_{F,GBP,t,T}^2 \end{bmatrix} \begin{bmatrix} \alpha_k \\ \beta_{k,1} \\ \beta_{k,2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t,T}^{DEM} \\ \varepsilon_{t,T}^{JPY} \\ \varepsilon_{t,T}^{CHF} \\ \varepsilon_{t,T}^{GBP} \end{bmatrix}, \quad \text{where} \quad \begin{bmatrix} \varepsilon_{t,T}^{DEM} \\ \varepsilon_{t,T}^{JPY} \\ \varepsilon_{t,T}^{CHF} \\ \varepsilon_{t,T}^{GBP} \end{bmatrix} \sim N(0, \Omega_k).$$

The forecasting model structure and coefficients were fixed by a search over the in-sample (1987-1991) period and only out-of-sample forecasts (1992-1998) were used to estimate (14). Deleting forecast horizons with fewer than 20 observations per exchange rate left forecast horizons from 6 to 66 business days. Each forecast horizon had between 20 and 28 observations per exchange rate during the out-of-sample period. The BIC favors pooling both types of data at all but the longest horizons. Results are

omitted for brevity.

The top row of Figure 6 displays the series of coefficients from maximum likelihood estimation of the pooled model described by (14) on high-frequency (intraday) volatility data. The pooled $\hat{\beta}_{k,2}$ coefficients are always positive, ranging from 0.1 to 1.5. The bottom row of Figure 6 shows the LR test p-values for the hypothesis that the pooled $\beta_{k,2}$ coefficients are equal to zero. The low p-values reject the hypothesis that the pooled $\beta_{k,2}$ coefficients equal zero for all the forecasts for almost all horizons. In other words, the pooled, out-of-sample forecasts from the ARIMA, LM-ARIMA, and OLS models improve IV estimates of high-frequency RV. This confirms the overlapping results in Table 4.

The futures price measure of volatility provides similar evidence against the joint hypothesis of correct testing procedures and efficiency. Only for horizons less than 15 days do the data consistently fail to reject the null. The overall picture is not consistent with the null.

The only evidence that IV subsumes other forecasts is at short horizons. The relative value of IV at such horizons might be due to forward-looking news that is not embedded in the time series of volatility. Such information might include exchange rate intervention, macroeconomic news, and/or financial crises such as the Russian default. These factors might be more forecastable by traders at short horizons. Or, the failure to reject informational efficiency at short horizons might stem from a power problem due to smaller samples—there are fewer observations per exchange rate.

A Non-zero price of volatility risk

Recall that the approximate equality of IV with the conditional expectation of RV—equation (2)—requires risk-neutral valuation. In other words, the risks associated with a position in an option either can be hedged or are not systematic. This is not necessarily the case. Recent research has considered the possibility that volatility risk is responsible for the bias found in BS IVs (e.g., Poteshman (2000), Benzoni (2002) and Chernov (2002)).

Chernov (2002) derives an expression for expected realized variance until expiry in terms of the implied risk-neutral expected variance until expiry (see Appendix D for details). That is, one estimates

the following prediction equation to see if the coefficient, β_1 , is equal to one:

$$(15) \quad \bar{V}_{t,T} = \alpha_{t,T}(\theta_v^Q, \kappa_v^Q, \theta_v^M, \kappa_v^M) + \beta_1 \sigma_{IV,t,T}^2 + \gamma_{t,T}(\theta_v^Q, \kappa_v^Q, \theta_v^M, \kappa_v^M) \mathcal{N}_t + \varepsilon_t,$$

where the coefficients, $\alpha_{t,T}$, and $\gamma_{t,T}$, are functions of time to expiry and the parameters of the risk-neutral (θ_v^Q, κ_v^Q) and objective SV processes (θ_v^M, κ_v^M) . If volatility risk is not priced, the risk-neutral and objective SV processes are the same and the coefficient vector $\{\alpha_{t,T}, \beta_1, \gamma_{t,T}\}$ in (15) equals $\{0, 1, 0\}$. In this case, the conventional prediction equation (3) is appropriate.

Citing equation (15), Chernov (2002) argues that the conventional regression of average RV on the risk-neutral IV is misspecified because instantaneous variance—which is heavily correlated with the risk-neutral IV—is omitted. This omission biases the coefficient (β_1) on the risk-neutral IV downwards. Thus the price-of-volatility-risk hypothesis blames the bias of risk-neutral IV on a correlated omitted variable, the volatility risk premium.

There are potentially at least 2 ways to estimate equation (15). The first method is to pool over horizons, estimating the three hyperparameters $(\theta_v^Q, \kappa_v^Q, \beta_1)$ and using the time to expiry ($T-t$) and the parameters from the time series estimation of the process— (κ^M, θ^M) in Table 1—to help construct the constants and coefficient terms. This imposes the restrictions across horizons and requires that one estimate only 3 free parameters, then test whether the data reject that $\beta_{k,1}$ is equal to one.

$$(16) \quad \sigma_{R,t,T}^2 = \alpha_k(T-t, \kappa^Q, \theta^Q) + \beta_{k,1} \sigma_{IV,t,T}^2 + \delta_{k,1}(T-t, \kappa^Q, \theta^Q) \mathcal{N}_{t,T} + \varepsilon_{k,t}$$

This method is potentially powerful, getting precise estimates, but imposes fairly strong restrictions on the functional form—ruling out jumps in volatility, for example.

The second method of estimating the price-of-volatility-risk model is to pool over exchange rates and estimate the parameter vector $\{\alpha_{t,T}, \beta_1, \gamma_{t,T}\}$ for each horizon with no cross-horizon restrictions. One can test whether the data reject the model's implication that risk-neutral IV is unbiased after the inclusion of the volatility risk premium. That is, one tests if $\beta_1 = 1$.

Either estimation method requires estimates of instantaneous variance as well as risk-neutral IV.

Instantaneous variance is taken to be the daily volatility estimate from intraday data. Because this measure and—to a lesser extent—the IV regressors are estimated with error, one should use an instrumental variables procedure to estimate (16) (Chernov (2002)). The BIC selects instruments from lagged values of the estimated instantaneous variance, IV and forecasts of realized variance.

Table 9 presents the results of Generalized Method of Moments (GMM) (Hansen (1982)) estimation of (16), accounting for the overlapping observations in the standard error (equation (11)). T-statistics for $\hat{\beta}_1$ reject the null that IV is an unbiased forecast of RV for all cases except the JPY. (Recall that the JPY IV has been an unusually good performer in all tests.)

Using overlapping data and only three hyperparameters imposes fairly strong restrictions on the model, however. And the parameter estimates are highly correlated and estimated imprecisely. The price-of-volatility-risk model might do better if $\{\alpha_k, \beta_{k,1}, \gamma_k\}$ were estimated separately for each horizon. To increase the power of such a test, one can again pool across exchange rates. This is similar to Chernov’s procedure, except that he estimated only one horizon (one-month) and did not pool across assets. As a result, while he could not reject unbiasedness, the estimates were not very precise.

Figure 8 shows the results of estimating (16) by GMM with the instrumental variables procedure. The model rejects the null that $\hat{\beta}_{k,1}$ equals one for almost all horizons. Permitting a price of volatility risk as in equation (16) does not solve the puzzle of IV’s apparent bias for foreign exchange futures.

To examine the robustness of the results in Figure 8, the figure was reconstructed under alternative assumptions, including a range-based estimate of instantaneous variance, a moving average based estimate of instantaneous variance, and OLS estimation of the model. None of the alternative estimation methods—results omitted for brevity—support the hypothesis that $\hat{\beta}_{k,1}$ equals one.

6. Economic implications of inefficiency: tracking error

The issues of bias and efficiency are statistically interesting, but the frequent rejection of those nulls does not tell us about economical significance. To judge whether the econometric forecasts can usefully augment the IV in delta hedging, one can compare the tracking error from delta hedging the option using

1) only the IV of the option or 2) an ex ante function of IV and variance from econometric forecasts.¹⁹

The exercise seeks to determine whether one-step-ahead ex ante econometric forecasts can improve delta hedging over the delta hedging of a benchmark model using only the implied instantaneous variance.

To calculate the tracking error, assume that a hypothetical agent delta hedges a call option on a futures contract as follows:

1. In the first period of a contract, the agent sells a call option, puts the proceeds into bonds, and takes a position of $\Delta(1)$ units in the futures contract. The value of this portfolio is initially zero because the long bond position offsets the short call.

$$(17) \quad V(0) = V_B(0) + V_C(0) = C_{BS}(F^{(0)}/X, 1) - C_{BS}(F^{(0)}/X, 1) = 0$$

2. On subsequent business days, the agent's bond holdings are augmented by interest on the bond position and gains (losses) on the previous futures position. The agent also adjusts the futures position according to the current delta, $\Delta(t)$. The portfolio's value on day t is

$$(18) \quad V(t) = V(t-1) + \left(e^{r \frac{n}{365}} - 1 \right) V_B(t-1) + \Delta F(t) \Delta(F^{(t)}/X, t-1) - \Delta C(F^{(t)}/X, t),$$

where $V_B(t-1)$ is the value of the bonds on t-1, there were n calendar days since the last business day, $\Delta F(t)$ is the change in the value of the futures price from t-1 to t, $\Delta(t-1)$ is the futures position at t-1 and $\Delta C(t)$ is the change in the call price from t-1 to t.

3. On the last day of the contract, the agent closes the futures position. The contract tracking error is the difference between the value of bond holdings and the option's value.

The aim of the exercise is to determine whether one-step-ahead ex ante econometric forecasts can improve delta hedging over the delta hedging of a benchmark model using only the implied instantaneous variance. The forecast-augmented model picks the relative weight of instantaneous volatility (λ) versus the econometric forecast ($1-\lambda$) and a constant to construct the instantaneous variance estimate for the delta function to minimize tracking error in the in-sample period (1987-1991). The augmented model's instantaneous variance is given by the following:

$$(19) \quad V^{Aug}(t) = \lambda V^{SV}(t) + (1-\lambda) V^F(t) + k,$$

where $V^{SV}(t)$ is the instantaneous variance from the SV model, $V^F(t)$ is the one-step-ahead forecast volatility from one of the four statistical forecasting models and k is a constant. The benchmark model constrains $\lambda = 1$ and chooses k to minimize in-sample delta hedging error. Both models generally choose a negative k to make the delta larger and more volatile to hedge better at the daily frequency.

To implement the delta hedging experiment, at the beginning of each contract period the call with the longest time continuously near-the-money was chosen as the call to hedge. When that call left the money, the position was closed out, the tracking error calculated, and a new call chosen to hedge. All positions were closed at the end of splicing periods.

Table 10 shows the percentage improvement in the out-of-sample tracking error using the four econometric forecasts. The tracking error is the sum of the absolute tracking errors for each option contract followed. Standard errors are calculated with the Newey-West procedure. The econometric forecasts reduce tracking errors over the IV benchmark in 14 of 16 intraday cases and 9 of 16 daily volatility cases. The JPY IV again returned the best performance versus the augmented model. Recall that the JPY also had the best IV according to the statistical measures. The OLS forecasts consistently improve on the IV benchmark, but the improvement is only statistically significant in three of eight cases.

Econometric forecasts do not consistently reduce delta-hedging tracking error.²⁰ This conclusion is consistent with Kim and Kim (2003) who examine a different trading strategy and conclude that the options market in foreign exchange is efficient. But it sheds new light on the usual rejections of bias and recent rejections of informational efficiency (Li (2002) and Martens and Zein (2002)). It also highlights the need to evaluate forecasts with the most relevant criteria.

7. Conclusions

This paper has examined explanations for IV's bias and inefficiency as a forecast of RV. No explanation found useful in other contexts is able to account for the bias and inefficiency of options on foreign exchange futures. Contrary to results in Poteshman (2000), high-frequency measures of volatility do not reduce the apparent bias. Horizon-by-horizon estimation, advocated by Christensen, Hansen, and Prabhala (2001), does not eliminate the bias when one pools across exchange rates to increase power.

Autocorrelation in IV could plausibly explain much of the bias, however. There is no evidence of sample selection bias distorting the results (Engle and Rosenberg (2000)). Finally, permitting a price-of-volatility risk, as recommended by Chernov (2002), does not eliminate the bias.

Contrary to most previous studies, this research finds that out-of-sample forecasts from ARIMA, LM-ARIMA, and OLS models of RV are helpful adjuncts to IV. Specifically, IV fails to subsume time series forecasts of RV using either overlapping or fixed horizon estimation.

Despite the success of econometric forecasts by statistical criteria, augmenting IV with econometric forecasts does not consistently reduce tracking error in out-of-sample delta hedging exercises. In most cases, tracking error declined when IV was augmented with out-of-sample econometric forecasts, but the improvement was not usually large or statistically significant. The strong statistical evidence of bias and inefficiency appears to lack much economic significance.

The prominent exception to the results that IV is biased and inefficient is that of the JPY. IV for JPY futures appears to be unbiased and efficient. The Ministry of Finance's exchange rate management might tie down expectations of volatility and create this phenomenon.

IV's bias as a predictor of RV is still a puzzle, but it does not matter economically.

Appendix A: Black-Scholes implicit variance and expected variance until expiry

If volatility evolves independently of the underlying price and all risk associated with the option can be hedged away, Hull and White (1987) showed that the correct price of a European option would be equal to the expectation of the BS price, evaluating the variance argument at average variance until expiry:

$$(A.1) \quad C(S_t, V_t, t) = \int C^{BS}(\bar{V}) h(\bar{V} | \sigma_t^2) d\bar{V} = E[C^{BS}(\bar{V}) | V_t],$$

where the average volatility over a period from t to T is denoted as: $\bar{V} = \frac{1}{T-t} \int_t^T V_\tau d\tau$.

Bates (1996) refined this relation to discern the relation between the volatility recovered by inverting the BS formula and the expected variance of the asset price until expiry. For at-the-money (ATM) options, the BS formula for futures reduces to $C^{BS} = e^{-rT} F \left[2N\left(\frac{1}{2}\sigma\sqrt{T}\right) - 1 \right]$. This can be approximated with a second-order Taylor expansion of $N(*)$ around zero, which yields:

$C^{BS} = e^{-rT} F \sigma \sqrt{T/(2\pi)}$. Another second-order Taylor expansion of that approximation around the expected value of variance until expiry produces an approximate expression for the BS IV in terms of the expected variance until expiry:

$$(A.2) \quad \hat{\sigma}_{BS}^2 \approx \left(1 - \frac{1}{8} \frac{Var(\bar{V}_{t,T})}{(E_t \bar{V}_{t,T})^2} \right)^2 E_t \bar{V}_{t,T}.$$

This approximation implies that IVs from the BS formula will underestimate the expected variance of the asset until expiry. This bias will be very small, however.

Appendix B: Forecasting methods

Four types of forecasting models are used to examine the informational efficiency of IV:

ARIMA, LM-ARIMA, GARCH, and an OLS model. The ARIMA class of models was chosen because of its ability to model a variety of economic and financial time series (Box and Jenkins (1976)). The ARIMA $(p,0,q)$ model for the daily realized variance can be written as follows:

$$(B.1) \quad r_t^2 = \gamma_0 + \sum_{i=1}^p \phi_i (r_{t-i}^2 - \gamma_0) + e_t - \sum_{i=2}^q \theta_i e_{t-i}, \text{ or } (1 - \phi(L))(r_t^2 - \gamma_0) = (1 - \theta(L))e_t.$$

The maximum lag length permitted was 5.

Andersen, Bollerslev, Diebold, and Labys (2001) pioneered the use of the LM-ARIMA structure in modeling conditional variance. They recommend it as a parsimonious model that fits the exchange rate variance process well. In addition, because it is a long-memory model, it can generate non-trivial variance forecasts for many periods. This is very useful in forecasting RV at the horizons needed for quarterly options (up to 72 business days). The LM-ARIMA model for the daily realized variance measure may be written as follows:

$$(B.2) \quad (1 - \phi(L))(1 - L)^d (r_t^2 - \gamma_0) = (1 - \theta(L))e_t,$$

where r_t^2 is again the realized variance at time t and d is the fractional differencing parameter.²¹

This paper follows Andersen, Bollerslev, Diebold, and Labys (2001) and Andersen, Bollerslev, Diebold, and Ebens (2001) in fitting the LM-ARIMA model in a two-step process. The first step is to perform the Geweke-Porter-Hudak (1983) (GPH) regression:

$$(B.3) \quad \log[I(\varpi_j)] = \beta_0 + \beta_1 \log(\varpi_j) + u_j,$$

where $I(\varpi_j)$ is the sample periodogram $\left(I(\varpi_j) = (2\pi m)^{-1} \left| \sum_{t=1}^n (h_t - \gamma_0) e^{it\varpi_j} \right|^2 \right)$ of the spectrum at the j th

Fourier frequency, $\varpi_j = 2\pi j/T$, and $j = 1, 2, 3, \dots, m$. The parameter m is chosen to equal $[T^{1/2}]$. The

fractional differencing parameter, $d = -\beta_1/2$, is asymptotically normal with a standard error of $\pi(24m)^{-1/2}$.

The second step in the LM-ARIMA estimation is to fit an ARIMA model to the residuals from the

fractional differencing operation implied by the GPH regression. Constructing the forecast of the LM-ARIMA model reverses this two-stage process.

The third forecasting model is the ubiquitous GARCH(1,1) benchmark (Bollerslev (1986)). The GARCH model was chosen because of its ability to fit the conditional heteroskedasticity in a variety of daily financial time series, including exchange rates. The quasi-maximum likelihood version of this model may be written as follows:

$$(B.4) \quad r_t \sim N(0, h_t), \quad h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1},$$

where h_t is the forecast conditional variance and the restrictions that $\omega > 0$, $\alpha \geq 0$, and $\beta \geq 0$ are sufficient to ensure that h_t is positive. If $\alpha + \beta < 1$, the variance process displays (geometric) mean reversion to the unconditional expectation of σ_t^2 , $\omega / (1 - \alpha - \beta)$.

The fourth forecasting model is an OLS regression that uses up to four variables to predict variance: up to five lags of the realized daily variance measure, up to five lags of absolute daily returns, up to five lags of the futures price range during the day, and a Friday/holiday indicator variable. The general regression can be written as follows:

$$(B.5) \quad h_t = \gamma_0 + \sum_{i=1}^5 a_i r_{t-i}^2 + \sum_{i=1}^5 b_i |r_{t-i}| + \sum_{i=1}^5 c_i range_{t-i} + d \cdot Friday/holiday_t + e_t.$$

When the absolute return or futures price range was used in the regression, an auxiliary autoregression was employed to construct truly ex ante, multi-period forecasts.

Appendix C: The Simulations

This study uses an ARIMA model to model conditional variance to construct simulated RV and IV until expiry. To alleviate the problem that the distribution of daily variance is truncated at zero, skewed, and kurtotic, the daily variance data were transformed with a logarithmic transformation. That is, we defined a new daily variance measure,

$$(C.1) \quad R_t = \ln(r_t^2 + 0.0000001).$$

The additive constant was necessary because a few daily returns were equal to zero. The transformed data were then modeled as ARIMA data and 1000 simulated data samples were drawn. The ARIMA $(p,0,q)$ model for the daily realized variance can be written as follows:

$$(C.2) \quad R_t = \gamma_0 + \sum_{i=1}^p \phi_i (R_{t-i} - \gamma_0) + e_t - \sum_{i=2}^q \theta_i e_{t-i}, \text{ or } (1 - \phi(L))(R_t - \gamma_0) = (1 - \theta(L))e_t.$$

The simulated data $(\tilde{R}_t)$ are exponentiated to recover simulated daily variance

$$(\tilde{r}_t^2 = \exp(\tilde{R}_t) - 0.0000001), \text{ which was summed in the usual way to recover variance until expiry.}^{22}$$

While the conditional forecasts of $\tilde{R}_t$ are easy to recover from the ARIMA model, one must transform these forecasts to recover an approximately conditionally unbiased forecast of realized variance until

$$\text{expiry, } E_t \left[\frac{251}{T-t+1} \sum_{i=1}^T \tilde{r}_{t+i}^2 \right].$$

If $\tilde{r}_t$ were conditionally normal, the n-step-ahead expectation of $\tilde{r}_t^2$ could be recovered using the moment-generating function of the normal distribution, $E_t(\tilde{r}_{t+n}^2) = \exp\left(\hat{\tilde{R}}_{t+n|t} + \sigma_R^2/2\right)$, where $\hat{\tilde{R}}_{t+n|t}$ and σ_R^2 denote the predicted mean and n-step variance of $\tilde{R}_{t+n}$ from the estimated ARIMA model. However, $\tilde{r}_t$ is not normally distributed with bootstrapped errors and the above approximation is very poor. To improve the approximation, one can numerically compute the expectation of the n-step-ahead prediction of $\tilde{r}_{t+n|t}^2$ by bootstrapping from the empirical distribution of the estimated ARIMA errors:

$$(C.3) \quad E_t(\tilde{r}_{t+n}^2) = E \left[\exp \left(\hat{\tilde{R}}_{t+n|t} \right) \exp \left(\sum_{i=1}^n \kappa_{n-i+1} \tilde{e}_{t+i} \right) \right] = \exp \left(\hat{\tilde{R}}_{t+n|t} \right) k_i,$$

where κ_{n-i+1} is the coefficient on the $t+i$ th shock in an n -step-ahead ARIMA forecast—a function of the elements of ϕ and θ —and $k_i = E \left[\exp \left(\sum_{i=1}^n \kappa_{n-i+1} \tilde{e}_{t+i} \right) \right]$, and the additive constant is omitted to simplify notation.

Computing an approximately unbiased estimate of RV required taking the expectation of a second-order Taylor series expansion of that quantity.

$$(C.4) \quad \tilde{\sigma}_{RV,t,T}^2 = \frac{251}{T-t+1} \sum_{i=1}^T \tilde{r}_{t+i}^2$$

$$(C.5) \quad E_t(\tilde{\sigma}_{RV,t,T}^2) = \frac{251}{T-t+1} \left(\sum_{i=1}^{T-t} \mu_{t+i} - \frac{1}{8} \left(\sum_{i=1}^{T-t} \mu_{t+i} \right)^2 \sum_{i=1}^{T-t} \sum_{j=1}^{T-t} \Omega_{ij} \right),$$

where the n -step-ahead prediction of $\tilde{r}_{t+n|t}^2$ is $\mu_{t+n} = E \left[\exp \left(\hat{\tilde{R}}_{t+n|t} \right) \exp \left(\sum_{i=1}^n \kappa_{n-i+1} \tilde{e}_{t+i} \right) \right] = \exp \left(\hat{\tilde{R}}_{t+n|t} \right) k_i$

and Ω_{ij} is the covariance of the i and j step-ahead prediction errors:

$$\Omega_{ij} = \exp \left(\hat{\tilde{R}}_{t+i} \right) \exp \left(\hat{\tilde{R}}_{t+j} \right) E \left[\exp \left(\sum_{g=1}^i \kappa_{i-g+1} \tilde{e}_{t+g} + \sum_{g=1}^j \kappa_{j-g+1} \tilde{e}_{t+g} \right) - k_i k_j \right].$$

Like the k_i , the expectation in the Ω_{ij} function is unconditional and can be computed prior to the simulation by bootstrapping from the ARIMA residuals. Thus, the forecast correction is not as computationally intensive as one might think at first glance. These forecasts are approximately conditionally unbiased forecasts of variance until expiry in the absence of serial correlation or errors in variables.

Appendix D: Implicit variance and realized variance with a price of volatility risk

Direct estimation of the time varying price of volatility risk from options prices is potentially computationally difficult. Chernov (2002) develops a technique that is more tractable. Recall the Hull-White result that the BS IV is approximately equal to the expected value of volatility until expiry under the risk-neutral probability measure:

$$(D.1) \quad \sigma_{BS,t,T}^2 \approx E_t^Q(\bar{V}_{t,T}),$$

where the exponent Q denotes the expectation with respect to the risk-neutral probability measure.

Chernov then uses the definition of the Radon-Nikodym derivative ($\xi_{t+\tau,t}$) of the risk-neutral probability measure with respect to the objective probability measure and the definition of covariance to show that

$$(D.2) \quad \sigma_{t,T}^2 = E_t^Q(\bar{V}_{t,T}) = E_t^M(\xi_{t,T} \bar{V}_{t,T}) = E_t^M(\bar{V}_{t,T}) E_t^M(\xi_{t,T}) + Cov_t^M(\xi_{t,T} \bar{V}_{t,T}),$$

where the exponent M denotes the expectation with respect to the objective (market) probability measure.

Because the risk-neutral and objective probability measures are equivalent, the expectation of $\xi_{t,T}$ equals one and (D.2) implies the following:

$$(D.3) \quad E_t^M(\bar{V}_{t,T}) = \sigma_{t,T}^2 - Cov_t^M(\xi_{t,T} \bar{V}_{t,T}).$$

Chernov then observes that integrating the variance process (equation (8)) from t to T and taking expectations shows that the expected variance until expiry is linear in instantaneous variance under any probability measure:

$$(D.4) \quad E_t^M(\bar{V}_{t,T}) = A_t^M V_t + B_t^M \quad \text{or} \quad E_t^Q(\bar{V}_{t,T}) = A_t^Q V_t + B_t^Q,$$

where A_t^M (A_t^Q) and B_t^M (B_t^Q) are functions of the parameters of the objective (risk-neutral) variance processes.

$$(D.5) \quad A_{t,T}^Q = -\frac{1}{(T-t)\kappa^Q} [e^{-\kappa^Q(T-t)} - 1] \quad \text{and} \quad A_{t,T}^M = -\frac{1}{(T-t)\kappa^M} [e^{-\kappa^M(T-t)} - 1]$$

$$(D.6) \quad B_{t,T}^Q = \frac{\theta^Q}{\kappa^Q} [1 - A_t^Q] \quad \text{and} \quad B_{t,T}^M = \frac{\theta^M}{\kappa^M} [1 - A_t^M]$$

The linearity of variance combined with (D.3) implies that the covariance is linear too.

$$(D.7) \quad Cov_t^M(\xi_{t,T} \bar{V}_{t,T}) = E_t^Q(\bar{V}_{t,T}) - E_t^M(\bar{V}_{t,T}) = [A_{t,T}^Q - A_{t,T}^M]V_t + [B_{t,T}^Q - B_{t,T}^M]$$

Substituting the expression for the covariance back into (D.3) and using (D.1), one gets the expected variance to expiry as a linear function of the BS IV and instantaneous variance.²³

$$(D.8) \quad E_t^M(\bar{V}_{t,T}) = (B_{t,T}^M - B_{t,T}^Q) + \sigma_{BS,t,T}^2 + (A_{t,T}^M - A_{t,T}^Q)V_t = \alpha_{t,T} + \sigma_{BS,t,T}^2 + \gamma_{t,T}V_t$$

That is, the conditional expectation of average variance until expiry is a linear function of BS IV and instantaneous variance. Note that the coefficient on instantaneous variance $(A_{t,T}^M - A_{t,T}^Q)$ goes to zero as time to expiry goes to zero.

References

- Andersen, Torben G., and Tim Bollerslev. (1998). "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts." *International Economic Review* 39, 885-905.
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens, 2001, The distribution of realized stock return volatility, *Journal of Financial Economics* 61, 43-76.
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, January 2001, Modeling and forecasting realized volatility. Unpublished Manuscript, Northwestern University.
- Ang, Andrew, Robert J. Hodrick, Yuhang Xing, Xiaoyan Zhang. (2003). "The Cross-Section of Volatility and Expected Returns." Unpublished Manuscript, Columbia University.
- Bakshi, Gurdip, Charles Cao and Zhiwu Chen, 1997, Empirical Performance of Alternative Option Pricing Models, *The Journal of Finance*, Vol. 52, No. 5, December, 2003-2049.
- Bandi, Federico, and Benoit Perron. (2003). "Long Memory and the Relation Between Implied and Realized Volatility." Working Paper, University of Chicago and University of Montreal.
- Barone-Adesi, Giovanni, and Robert E. Whaley. (1987). "Efficient Analytic Approximation of American Option Values." *Journal of Finance* 42, 301-20.
- Bates, David S. (2000). "Post-'87 Crash Fears in the S&P 500 Futures Option Market." *Journal of Econometrics* 94, 181-238.
- Bates, David S. (2003). "Empirical Option Pricing: A Retrospection." *Journal of Econometrics* 116, 387-404.
- Bates, David S. (1996). "Testing Option Pricing Models." In G.S. Maddala and C. R. Rao (eds.), *Statistical Methods in Finance (Handbook of Statistics, v. 14)*. Amsterdam: Elsevier Publishing.
- Beckers, Stan. (1981). "Standard Deviations Implied in Options Prices as Predictors of Futures Stock Price Variability." *Journal of Banking & Finance* 5, 363-81.
- Benzoni, Luca. (2002). "Pricing Options Under Stochastic Volatility: An Empirical Investigation." Unpublished Manuscript, Carlson School of Management.
- Berkowitz, Jeremy and Giorgianni, Lorenzo, 2001, Long-Horizon Exchange Rate Predictability? *Review of Economics and Statistics*, 83(1), pp. 81-91.
- Blair, Bevan J., Ser-Huang Poon, and Stephen J. Taylor. (2001). "Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High-Frequency Index Returns." *Journal of Econometrics* 105, 5-26.
- Black, Fischer, and Myron Scholes. (1972). "The Valuation of Option Contracts and a Test of Market Efficiency." *Journal of Finance* 27, 399-417.
- Black, Fischer. (1976). "The Pricing of Commodity Contracts." *Journal of Financial Economics* 3, 167-79.
- Bollen, Nicolas, and Robert Whaley. (2003). "Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?" forthcoming in *Journal of Finance*.

- Bollerslev, Tim. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31, 307-27.
- Bollerslev, Tim, and Hao Zhou. (2003). "Volatility Puzzles: A Unified Framework for Gauging Return-Volatility Regressions." Finance and Economics Discussion Series 2003-40, Board of Governors of the Federal Reserve System.
- Box, G.E.P., and G.M. Jenkins. (1976). "Time Series Analysis: Forecasting and Control." Revised Edition. San Francisco, CA: Holden Day.
- Campbell, John Y. (1993). "Intertemporal Asset Pricing without Consumption Data." *American Economic Review* 83, 487-512
- Canina, Linda, and Stephen Figlewski. (1993). "The Informational Content of Implied Volatility." *Review of Financial Studies* 6, 659-81.
- Chernov, Mikhail. (2002). "On the Role of Volatility Risk Premia in Implied Volatilities Based Forecasting Regressions." Unpublished Manuscript, Columbia University.
- Chernov, Mikhail, and Eric Ghysels. (2000). "A Study towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purposes of Options Valuation." *Journal of Financial Economics* 56, 407-58.
- Christensen, B. J., and N.R. Prabhala. (1998). "The Relation Between Implied and Realized Volatility." *Journal of Financial Economics* 50, 125-50.
- Christensen, B. J., C.S. Hansen, and N.R. Prabhala. (2001). "The Telescoping Overlap Problem in Options Data." Unpublished Manuscript, School of Economics and Management, University of Aarhus.
- Davidson, Wallace D. III, Jin K. Kim, Evren Ors, and Andrew Szakmary. (2001). "Using Implied Volatility on Options to Measure the Relation between Asset Returns and Variability." *Journal of Banking & Finance* 25, 1245-69.
- Engle, Robert F., and Joshua Rosenberg. (2000). "Testing the Volatility Term Structure Using Option Hedging Criteria." *Journal of Derivatives* 8, 10-28.
- Fair, Ray C., and Robert J. Shiller. (1990). "Comparing Information in Forecasts from Econometric Models." *American Economic Review* 80, 375-89.
- Figlewski, Stephen, and T. Clifton Green. (1999). "Market Risk and Model Risk for a Financial Institution Writing Options." *Journal of Finance* 54, 1465-99.
- Fleming, Jeff. (1998). "The Quality of Market Volatility Forecasts Implied by S&P 100 Index Option Prices." *Journal of Empirical Finance* 5, 317-45.
- Garcia, René, Eric Ghysels and Eric Renault. (2003). "The Econometrics of Option Pricing." forthcoming in Y. Ait-Sahalia and L.P. Hansen (eds.), *The Handbook of Financial Econometrics*.
- Geweke, John, and Susan Porter-Hudak. (1983). "The Estimation and Application of Long Memory Time Series Models." *Journal of Time Series Analysis* 4, 221-38.
- Hansen, L. (1982). "Large Sample Properties of Generalized Method of Moment Models." *Econometrica* 50, 1029-53.

- Heston, Steven L. (1993). "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options." *Review of Financial Studies* 6, 327-43.
- Hull, John C. (2002). "Options, Futures, and Other Derivatives." Prentice Hall; 5th edition.
- Hull, John C., and Alan White. (1987). "The Pricing of Options on Assets with Stochastic Volatilities." *Journal of Finance* 42, 281-300.
- Jorion, Philippe. (1995). "Predicting Volatility in the Foreign Exchange Market." *Journal of Finance* 50, 507-28.
- Kilian, Lutz. 1999, Exchange Rates and Monetary Fundamentals: What Do We Learn From Long-Horizon Regressions? *Journal of Applied Econometrics*, 14(5), pp. 491-510.
- Kim, Minho, and Minchoul Kim. (2003). "Implied Volatility Dynamics in the Foreign Exchange Markets." *Journal of International Money and Finance* 22, 511-28.
- Kroner, Kenneth F., Kevin P. Kneafsey, and Stijn Claessens. (1993). "Forecasting Volatility in Commodity Markets." Working Paper 93-3, University of Arizona.
- Lamoureux, Christopher G., and William D. Lastrapes. (1993). "Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities." *Review of Financial Studies* 6, 293-326.
- Latane, Henry A., and Richard J. Rendleman, Jr., 1976, Standard deviations of stock price ratios implied in option prices, *Journal of Finance* 31, 369-81.
- Li, Kai. (2002). "Long-Memory versus Option-Implied Volatility Prediction." *Journal of Derivatives* 9, 9-25.
- Mankiw, Gregory N., and Matthew D. Shapiro. (1986). "Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models." *Economics Letters* 20, 139-45.
- Mark, Nelson C. (1995). "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability." *American Economic Review* 85, 201-18.
- Martens, Martin, and Jason Zein. (2002). "Predicting Financial Volatility: High-Frequency Time-Series Forecasts vis-à-vis Implied Volatility." forthcoming in the *Journal of Futures Markets*.
- Merton, Robert C. (1973). "An Intertemporal Capital Asset Pricing Model." *Econometrica* 41, 867-87.
- Neely, Christopher J., and Paul A. Weller. (2002). "Predicting Exchange Rate Volatility: Genetic Programming vs. GARCH and RiskMetrics™." *Federal Reserve Bank of St. Louis Review* 84, 43-54.
- Pong, Shiuyan, Mark B. Shackleton, Stephen J. Taylor, and Xinzhong Xu. (2003). "Forecasting Currency Volatility: a Comparison of Implied Volatilities and AR(FI)MA Models." forthcoming in *Journal of Banking & Finance*.
- Pan, Jun. (2002). "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study." *Journal of Financial Economics* 63, 3-50.

- Poon, Ser-Huang, and Clive Granger. (2003). "Forecasting Volatility in Financial Markets: A Review." *Journal of Economic Literature* 41, 478-539.
- Poteshman, Allen M. (2000). "Forecasting Future Volatility from Option Prices." Unpublished Manuscript, Department of Finance, University of Illinois at Urbana-Champaign.
- Romano, Marc, and Nizar Touzi. (1997). "Contingent Claims and Market Completeness in a Stochastic Volatility Model." *Mathematical Finance* 7, 399-410.
- Sarwar, Ghulam, and Timothy Krehbiel. (2000). "Empirical Performance of Alternative Pricing Models of Currency Options." *Journal of Futures Markets* 20, 265-91.
- Schwarz, G. (1978). "Estimating the Dimension of a Model." *The Annals of Statistics* 6, 461-64.
- Shleifer, Andrei, and Robert W. Vishny. (1997). "The Limits of Arbitrage." *Journal of Finance* 54, 35-55.
- Stambaugh, R. F. (1986). "Bias in Regressions With Lagged Stochastic Regressors." Working Paper, Center for Research in Security Prices, University of Chicago.
- Szakmary, Andrew, Evren Ors, Jin Kyoung Kim, and Wallace N. Davidson III. (2003). "The Predictive Power of Implied Volatility: Evidence from 35 Futures Markets." *Journal of Banking & Finance* 27, 2151-75.
- Taylor, Stephen J. (2000). "Consequences for Option Pricing of a Long Memory in Volatility." Unpublished Manuscript, Department of Accounting and Finance, Lancaster University.
- Whaley, Robert E. (2003). "Derivatives." In G. M. Constantinides, M. Harris, and R. Stulz (eds.), *Handbook of the Economics of Finance*. Elsevier Science B.V.
- Wright, Jonathan. (2000). "Log-Periodogram Estimation of Long Memory Volatility Dependencies with Conditionally Heavy Tailed Returns." International Finance Discussion Paper 685, Board of Governors of the Federal Reserve System.

Table 1: SV model parameters

	DEM	JPY	CHF	GBP
$\rho * 100$	0.684	1.671	-0.859	1.344
θ_v	0.024	0.030	0.030	0.025
κ_v	1.487	1.503	1.466	1.612
σ_v^2	0.312	0.348	0.345	0.315

Notes: The table shows parameter estimates for the asset and volatility processes described by (7) and (8). Parameters were estimated by full information maximum likelihood for each exchange rate. Sarwar and Krehbiel (2000) describe the relation between the time series representation and the parameters of the SV option pricing model.

Table 2: Summary statistic for DEM volatility data.

One-step-ahead annualized volatility and forecasts						Annualized volatility until expiry and forecasts thereof						
	Statistic	$\sigma_{RV,t}$	ARIMAF1	LMARMAF1	GARCHF1	OLSF1	$\sigma_{RV,t,T}$	ARIMATX	LMARMATX	GARCHTX	OLSTX	$\sigma_{IV,t,T}$
High-frequency measure	TotalObs	2977	2977	2977	2977	2977	2977	2977	2977	2977	2977	2977
	Nobs	2977	2972	2972	2977	2973	2955	2972	2972	2977	2972	2977
	mean	9.90	10.71	10.70	10.57	10.98	10.48	10.90	10.83	10.87	10.99	11.71
	stddev	4.68	2.27	2.39	2.79	2.64	2.51	1.09	1.63	1.13	0.78	2.08
	max	59.77	26.65	27.21	32.55	27.23	30.23	16.69	19.41	20.56	17.67	21.24
	min	0.07	5.93	6.55	6.27	3.16	4.75	8.25	7.37	7.42	6.27	6.27
	AC1	0.46	0.87	0.89	0.89	0.70	0.97	0.95	0.95	0.87	0.81	0.93
	AC2	0.38	0.80	0.82	0.79	0.66	0.95	0.90	0.92	0.75	0.69	0.90
	AC3	0.32	0.82	0.82	0.70	0.70	0.93	0.86	0.90	0.65	0.61	0.88
	AC4	0.34	0.78	0.78	0.64	0.60	0.91	0.81	0.88	0.58	0.52	0.85
	AC5	0.29	0.75	0.76	0.59	0.54	0.89	0.77	0.86	0.52	0.46	0.83
Futures price volatility	mean	7.95	11.19	11.03	10.63	11.51	10.42	11.44	11.18	10.70	11.63	11.71
	stddev	7.37	1.56	1.72	2.19	2.27	2.84	0.83	1.25	1.87	0.73	2.08
	max	56.83	19.45	19.89	18.97	32.68	22.43	17.68	20.19	18.69	17.46	21.24
	min	0.00	8.79	7.89	6.45	3.37	2.93	9.46	8.10	6.77	7.61	6.27
	AC1	0.08	0.79	0.83	0.99	0.44	0.97	0.95	0.95	0.99	0.84	0.93
	AC2	0.07	0.71	0.77	0.97	0.63	0.95	0.90	0.92	0.97	0.72	0.90
	AC3	0.06	0.81	0.85	0.96	0.56	0.92	0.86	0.90	0.96	0.63	0.88
	AC4	0.07	0.79	0.83	0.95	0.48	0.90	0.81	0.88	0.94	0.52	0.85
	AC5	0.09	0.75	0.80	0.93	0.41	0.88	0.76	0.86	0.93	0.46	0.83

Notes: The table displays summary statistics for the DEM volatility data. The left-hand panel denotes one-step-ahead measures of volatility. The column labeled $\sigma_{RV,t}$ denotes the annualized volatility of the exchange rate in percentage terms. The columns labeled ARIMAF1, LMARMAF1, GARCHF1, and OLSF1 denote the properties of the one-step-ahead forecasts of the volatility of the exchange rate. The right-hand panel denotes analogous statistics for RV until expiry, the forecasts of that quantity, and the IVs. Similarly, the labels ARIMATX, LMARMATX, GARCHTX, and OLSTX denote recursive forecasts until the expiry of the option. The top panel uses the annualized daily sums of 30-minute log returns (intraday data) to measure actual volatility, while the bottom panel uses the annualized squared log change in the daily futures price for that purpose. Nobs is total observations less observations used for lags. Mean and stddev denote the mean and standard deviation of the series, while AC1 through AC5 denote the first five autocorrelations of the series. Summary statistics are computed over the whole sample, 1987 through 1998.

Table 3: Results from regressing RV on IV, using overlapping samples

	XRATE	Obs	α	(s.e.)	β_1	(s.e.)	Wald p-value	R ²
High-frequency volatility	DEM	2955	0.44	(0.14)	0.51	(0.09)	0.000	0.19
	JPY	2960	0.06	(0.14)	0.86	(0.12)	0.136	0.42
	CHF	2967	0.42	(0.19)	0.65	(0.13)	0.016	0.27
	GBP	2958	0.25	(0.09)	0.61	(0.08)	0.000	0.44
Futures price volatility	DEM	2977	0.40	(0.14)	0.54	(0.10)	0.000	0.18
	JPY	2981	0.19	(0.14)	0.80	(0.11)	0.155	0.27
	CHF	2989	0.55	(0.23)	0.58	(0.15)	0.011	0.16
	GBP	2980	0.11	(0.11)	0.79	(0.09)	0.014	0.36

Notes: The table displays the coefficients and robust standard errors (equation (11)) from a regression of RV on IV (equation $\sigma_{RV,t,T}^2 = \alpha + \beta_1 \sigma_{IV,t,T}^2 + \varepsilon_t$), pooling over horizons from 8 to 75 business days. The table uses the full sample from 1987 through 1998.

Table 4: Results from predicting RV with IV and out-of-sample econometric forecasts

	Xrate	ARIMA forecast				LM- ARIMA forecast				GARCH forecast				OLS forecast			
		α	β_1	β_2	R^2	α	β_1	β_2	R^2	α	β_1	β_2	R^2	α	β_1	β_2	R^2
HF volatility	DEM	-0.02 (0.28)	0.33 (0.10)	0.60 (0.21)	0.23	0.27 (0.21)	0.27 (0.10)	0.43 (0.14)	0.23	0.11 (0.27)	0.42 (0.09)	0.39 (0.16)	0.22	0.02 (0.36)	0.46 (0.10)	0.40 (0.22)	0.21
	JPY	0.14 (0.14)	0.93 (0.19)	-0.07 (0.22)	0.44	0.09 (0.13)	0.90 (0.20)	0.00 (0.16)	0.44	0.11 (0.13)	0.92 (0.16)	-0.03 (0.09)	0.44	-0.01 (0.25)	0.88 (0.13)	0.12 (0.22)	0.44
	CHF	0.02 (0.25)	0.66 (0.19)	0.30 (0.26)	0.38	0.23 (0.18)	0.61 (0.21)	0.20 (0.14)	0.39	0.08 (0.22)	0.69 (0.17)	0.21 (0.19)	0.38	-0.12 (0.37)	0.71 (0.15)	0.34 (0.27)	0.37
	GBP	-0.10 (0.14)	0.54 (0.12)	0.42 (0.19)	0.58	0.06 (0.10)	0.41 (0.14)	0.42 (0.15)	0.59	0.00 (0.09)	0.55 (0.10)	0.33 (0.12)	0.59	-0.10 (0.20)	0.65 (0.10)	0.29 (0.22)	0.57
Futures volatility	DEM	-0.80 (0.50)	0.26 (0.08)	1.15 (0.43)	0.32	-0.16 (0.23)	0.20 (0.09)	0.77 (0.21)	0.31	0.12 (0.13)	0.00 (0.09)	0.85 (0.18)	0.37	-0.21 (0.31)	0.44 (0.10)	0.48 (0.20)	0.29
	JPY	0.26 (0.34)	0.86 (0.12)	-0.13 (0.26)	0.31	0.19 (0.63)	0.86 (0.16)	-0.07 (0.60)	0.31	-0.07 (0.27)	0.70 (0.21)	0.32 (0.35)	0.32	3.95 (1.65)	0.86 (0.11)	-3.08 (1.34)	0.31
	CHF	-0.44 (0.50)	0.59 (0.16)	0.61 (0.32)	0.28	-0.53 (0.47)	0.38 (0.24)	0.91 (0.43)	0.29	0.06 (0.22)	0.00 (0.19)	0.96 (0.21)	0.36	-1.24 (0.98)	0.64 (0.15)	1.05 (0.59)	0.27
	GBP	-0.51 (0.25)	0.68 (0.14)	0.53 (0.27)	0.52	-0.32 (0.16)	0.47 (0.18)	0.61 (0.23)	0.54	-0.09 (0.10)	0.52 (0.17)	0.39 (0.12)	0.53	2.29 (6.58)	0.87 (0.11)	-1.79 (5.00)	0.51

Notes: The table displays the results of regressing RV on IV and forecasts of RV, pooling over forecast horizons (equation (4)). The four panels display the results from forecasts with an ARIMA model, a long-memory ARIMA model, a GARCH(1,1) model, and an OLS model, respectively. The top subpanels use the daily sums of high-frequency (HF) RV, while the bottom panel uses the daily futures price variance. β_1 indicates the estimated coefficients on IV while β_2 indicates the estimated coefficients on the forecasts. Boldfaced β_2 estimates indicate a t-ratio of 1.64 or more. Models are estimated over the out-of-sample period, 1992 through 1998.

Table 5: Comparative statistics on IVs from three option pricing models

Statistic	SV	BAW	BS
	$\sigma_{\tau,T}^{IV}$	$\sigma_{\tau,T}^{IV}$	$\sigma_{\tau,T}^{IV}$
μ	11.71	11.17	11.19
σ	2.08	2.05	2.05
max	21.24	20.36	20.38
min	6.27	6.21	6.22
ρ_1	0.93	0.97	0.97
ρ_2	0.90	0.94	0.94
ρ_3	0.88	0.91	0.91
ρ_4	0.85	0.89	0.89
ρ_5	0.83	0.88	0.87
Corr w/ BS	0.97198	0.99998	
Corr w/ BAW	0.97145		

Notes: The table compares summary statistics from three options pricing models: the SV pricing model of Heston (1993), the Barone-Adesi and Whaley (1987) early exercise correction to the Black (1976) model and the Black (1976) model. The statistics shown are IV means (μ); standard deviation (σ); maximum, minimum, first five autocorrelations (ρ); and the contemporaneous correlations between the IV series from the three models.

Table 6: Measures of uncertainty in IV estimation

	Percentile of the distribution														
	obs	min	10%	25%	50%	75%	90%	max	μ	σ	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
max difference in 4 IVs	3000	0.002	0.080	0.129	0.223	0.475	1.928	11.046	0.74	1.47	0.06	0.07	0.03	0.01	0.00
diffs in strike-matched IVs	5336	0.000	0.007	0.019	0.043	0.078	0.121	2.192	0.06	0.07	0.10	0.15	0.17	0.11	0.07

Notes: The table shows two measures of the uncertainty associated with IV estimation for DEM. The first row shows the statistics for the maximal difference each day between the four IVs implied by the two closest calls and two closest puts (six possible pairs of IVs). The second row of each panel shows analogous results for strike-matched pairs of IVs. All percentiles are expressed in percentage terms, i.e., 0.5 is $\frac{1}{2}$ percent.

Table 7: Realized variance until expiry between days when IV is missing and other days

	OBS	IVOBS	HF difference	Futures difference
DEM	3009	2977	-0.086 (0.110)	-0.330 (0.086)
JPY	3009	2989	0.722 (0.342)	0.167 (0.223)
CHF	3008	2979	0.416 (0.179)	0.246 (0.184)
GBP	3007	2993	-0.260 (0.060)	-0.442 (0.090)

Notes: The table displays the differences between mean RV until expiry on days when IV is missing and other days. The fourth and fifth columns calculate the differences with intraday and daily futures data, respectively. Standard errors for the differences are in parentheses.

Table 8: Results of simulated errors-in-variables regression

GARCH model													
		Percentiles of α hat distribution				Percentiles of β_1 hat distribution				Wald test	R ² distribution		
		5th	50th	95th	% > est α	5th	50th	95th	% > est β_1	% rejections	5th	50th	95th
Intraday volatility	DEM	-0.26	0.24	0.51	18	0.51	0.77	1.12	94	56	0.06	0.13	0.21
	JPY	-0.13	0.35	0.65	87	0.46	0.70	1.03	25	67	0.08	0.15	0.24
	CHF	-0.29	0.28	0.62	31	0.53	0.79	1.09	79	57	0.08	0.17	0.27
	GBP	-0.19	0.13	0.30	21	0.62	0.87	1.13	96	36	0.16	0.26	0.38
Futures volatility	DEM	-0.16	0.15	0.40	10	0.55	0.86	1.10	96	24	0.11	0.26	0.43
	JPY	-0.29	0.25	0.54	58	0.47	0.80	1.17	50	35	0.05	0.15	0.27
	CHF	-0.27	0.19	0.52	8	0.51	0.86	1.12	92	19	0.08	0.23	0.39
	GBP	-0.09	0.18	0.39	68	0.53	0.82	1.06	56	30	0.11	0.26	0.44
ARIMA model													
		Percentiles of α hat distribution				Percentiles of β_1 hat distribution				Wald test	R ² distribution		
		5th	50th	95th	% > est α	5th	50th	95th	% > est β_1	% rejections	5th	50th	95th
Intraday volatility	DEM	-0.27	0.14	0.35	6	0.58	0.84	1.18	99	24	0.08	0.19	0.30
	JPY	-0.38	0.13	0.36	62	0.63	0.89	1.23	59	21	0.13	0.24	0.35
	CHF	-0.52	0.09	0.28	5	0.70	0.95	1.17	99	18	0.09	0.21	0.31
	GBP	-0.31	0.06	0.20	2	0.74	0.94	1.16	100	15	0.21	0.32	0.40
Futures volatility	DEM	-0.27	0.13	0.30	6	0.64	0.88	1.18	100	16	0.12	0.23	0.37
	JPY	-0.51	0.10	0.59	43	0.45	0.89	1.29	63	16	0.04	0.11	0.19
	CHF	-0.44	0.09	0.42	5	0.60	0.94	1.21	97	15	0.09	0.24	0.33
	GBP	-0.31	0.09	0.29	44	0.62	0.91	1.19	65	24	0.15	0.25	0.42

Notes: The table describes the results of simulating the predictive equation (3) $\sigma_{RV,t,T}^2 = \alpha + \beta_1 \sigma_{IV,t,T}^2 + \varepsilon_t$. The top panel shows the results from simulation with a GARCH model. The bottom panel does the same for results from a log ARIMA model. The left-hand panels shows statistics on the distribution of the estimates of α ; the right-hand panels shows statistics on the distribution of the estimates of β_1 . The column labeled “% > estimated α ” shows the percentage of simulated estimates of α that were greater than the estimated α . The column labeled “% > estimated β_1 ” similarly shows the percentage of simulated estimates of β_1 that were greater than the estimated β_1 .

Table 9: Price of Volatility Risk Model with Overlapping Observations

	θ_v^Q	κ_v^Q	β_1
DEM	-18.76 (7.14)	26.26 (5.50)	0.25 (0.11)
JPY	-0.78 (3.04)	3.67 (1.31)	0.78 (0.18)
CHF	-11.84 (5.75)	12.26 (2.55)	0.41 (0.17)
GBP	-8.05 (1.93)	11.71 (0.97)	0.40 (0.06)

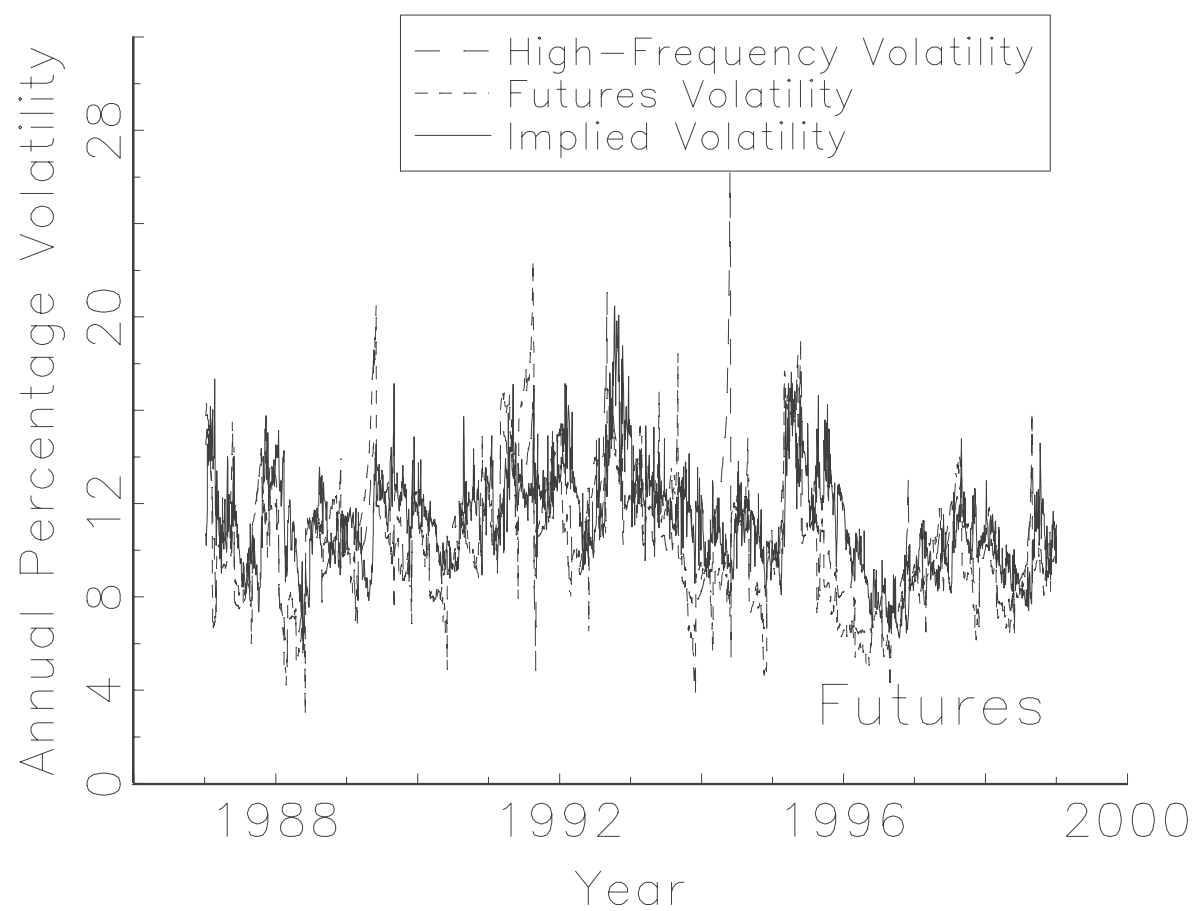
Notes: The table presents the results of GMM estimation of θ_v^Q , κ_v^Q , and β_1 in equation (16) as well as their standard errors, constructed accounting for the overlapping observations.

Table 10: Tracking error improvement from econometric forecasts

		ARIMA				LM-ARIMA				GARCH				OLS			
		TrError	Std Error	Weight	K	TrError	Std Error	Weight	K	TrError	Std Error	Weight	K	TrError	Std Error	Weight	K
HF	DEM	6.28	(5.80)	0.40	-0.09	5.28	(5.92)	0.40	-0.09	6.11	(5.89)	0.50	-0.09	11.07	(5.65)	0.60	-0.09
	JPY	1.39	(5.28)	0.50	-0.07	0.55	(5.28)	0.40	-0.07	0.37	(5.27)	0.60	-0.07	0.09	(5.28)	0.70	-0.07
	CHF	6.37	(4.26)	0.30	-0.08	3.77	(4.37)	0.30	-0.08	4.63	(4.27)	0.30	-0.08	5.52	(4.10)	0.30	-0.08
	GBP	1.08	(8.02)	0.70	-0.06	-1.04	(8.04)	0.60	-0.06	-0.29	(8.00)	0.70	-0.06	2.54	(7.90)	0.20	-0.05
Daily	DEM	0.03	(6.48)	0.90	-0.10	-0.41	(6.50)	0.90	-0.10	-0.63	(6.54)	0.90	-0.10	12.21	(5.62)	0.40	-0.09
	JPY	-1.76	(5.15)	0.00	-0.08	-1.23	(5.53)	0.40	-0.08	-1.76	(5.32)	0.00	-0.07	0.14	(5.21)	0.10	-0.08
	CHF	2.75	(4.17)	0.30	-0.09	0.21	(4.35)	0.10	-0.09	-0.75	(4.71)	0.90	-0.09	8.15	(4.26)	0.00	-0.09
	GBP	1.96	(8.03)	0.80	-0.06	0.56	(8.01)	0.90	-0.06	-1.26	(8.14)	0.70	-0.06	2.46	(7.99)	0.90	-0.06

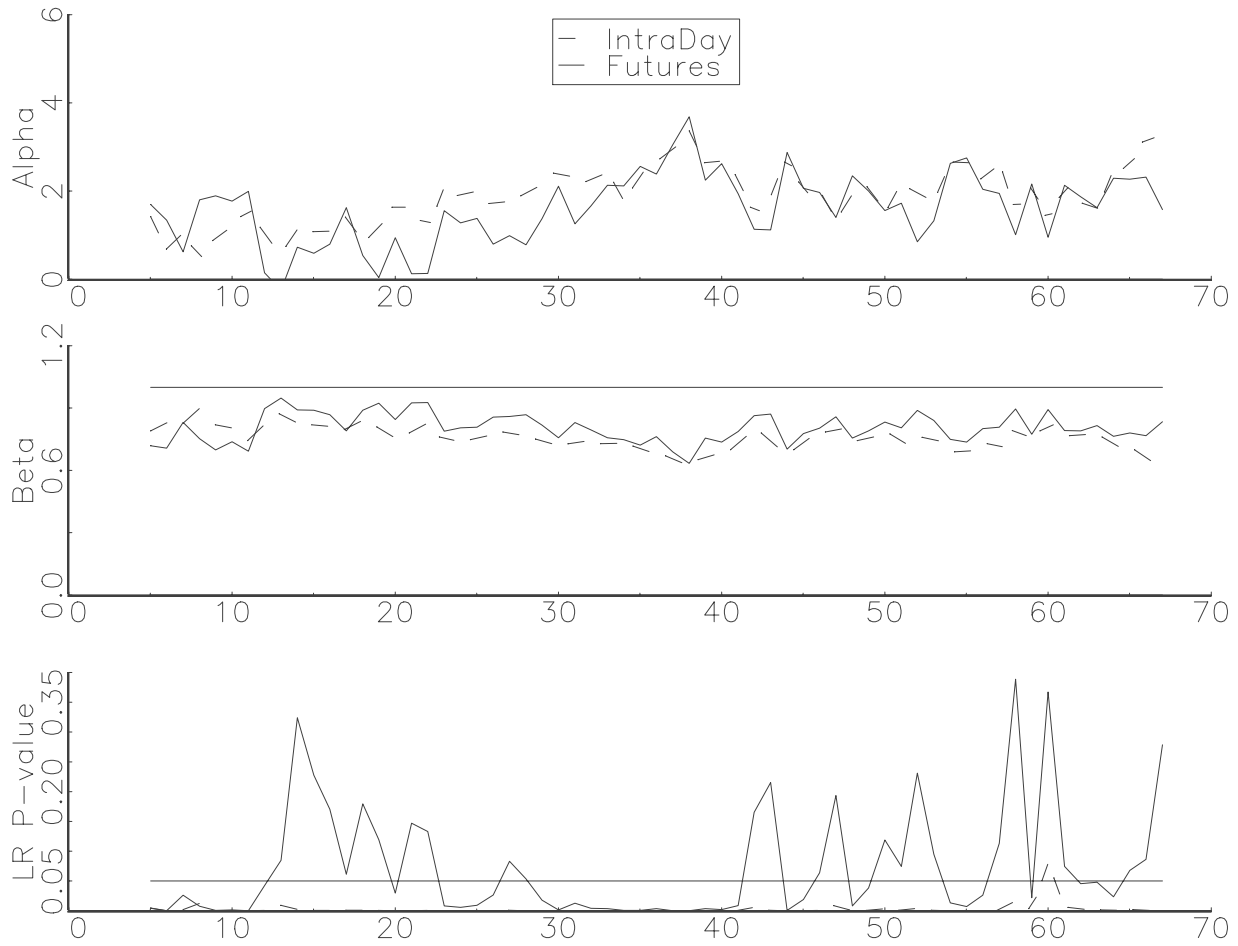
Notes: The table displays the improvement in augmenting the delta hedging rule with a statistical forecast, over the pure IV benchmark. The top panel forecasts volatility with intraday volatility, while the bottom panel uses daily futures prices for the same purpose. Within each subpanel, “TrError” shows the out-of-sample percentage improvement in tracking error with the statistical forecast. Boldfaced tracking error numbers are statistically significant. “Std Error” is the Newey-West standard error of the tracking error. “Weight” is the ex ante weight on IV, and “K” is the ex ante constant in the construction of the volatility to use in delta hedging.

Figure 1: Annual and percentage realized and IV for the DEM



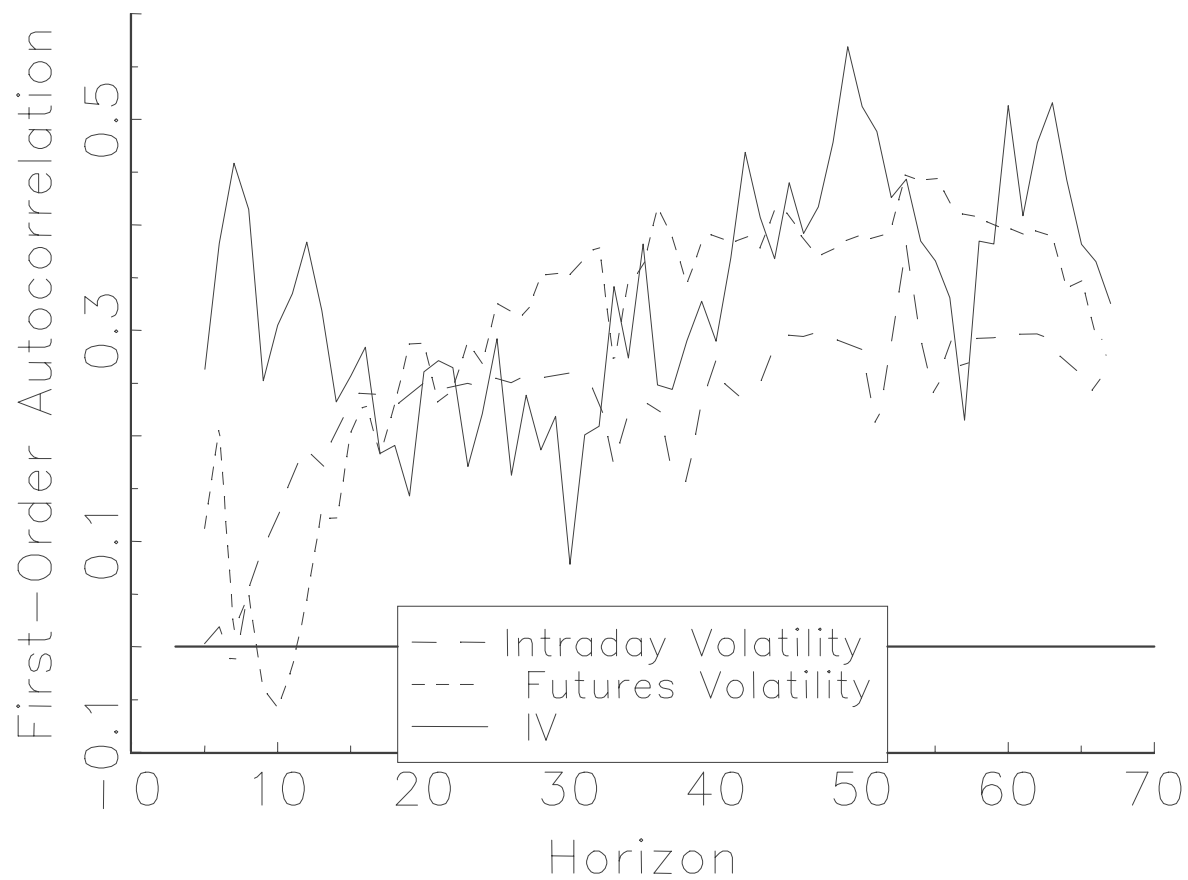
Notes: The figure displays IV and two measures of RV until expiry, a high-frequency measure and the futures volatility measure.

Figure 2: Estimated α_k and $\beta_{k,1}$ coefficients from predicting RV with IV



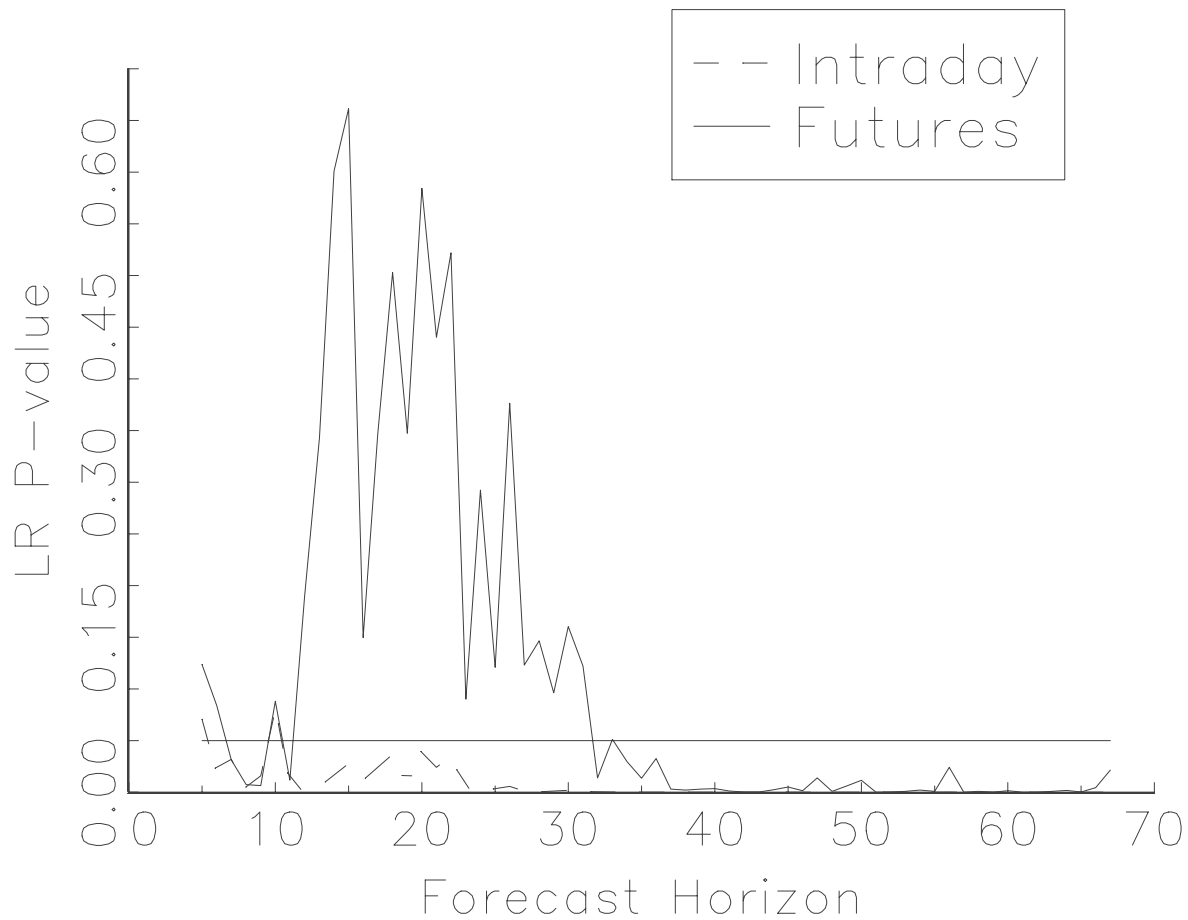
Notes: The figure displays the series of $\hat{\alpha}_k$ and $\hat{\beta}_{k,1}$ from a pooled (across exchange rates) regression of RV on IV, horizon-by-horizon, with the LR test p-values that $\{\alpha_k, \beta_{k,1}\} = \{0, 1\}$. P-values less than 0.05 reject the null hypothesis that IV is an unbiased predictor of RV. The intraday p-values are so close to zero that the dashed line is rarely visible. Figures for the intraday volatility measure are represented by dashed lines while those from futures price volatility are represented by solid lines. The models were estimated by maximum likelihood over the whole sample, 1987 through 1998.

Figure 3: First-order autocorrelation from IV and both measures of RV, by forecast horizon.



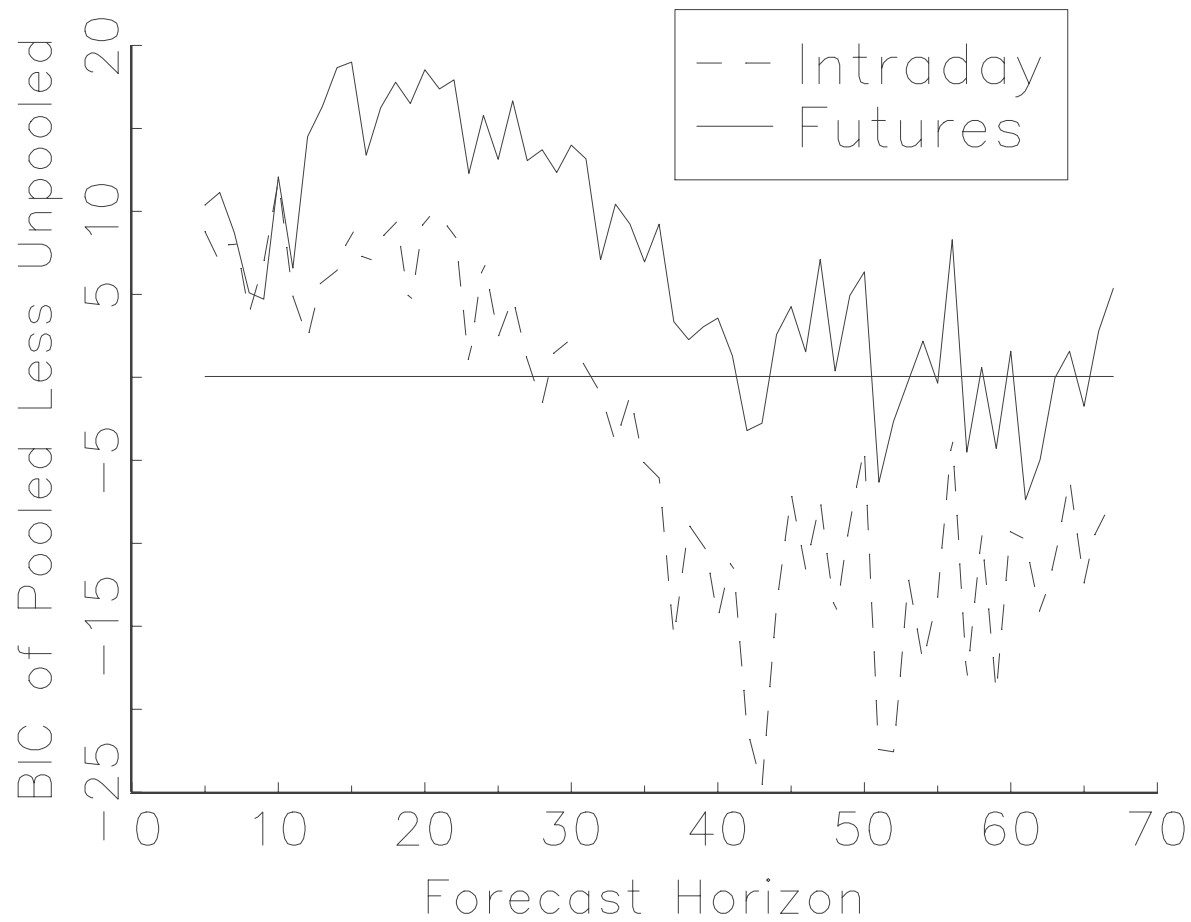
Notes: The figure displays the first-order autocorrelations, by horizons, for both volatility measures. The solid horizontal line denotes zero.

Figure 4: P-values from the likelihood ratio test that pooling coefficients across exchange rates is appropriate, as in equation (13).



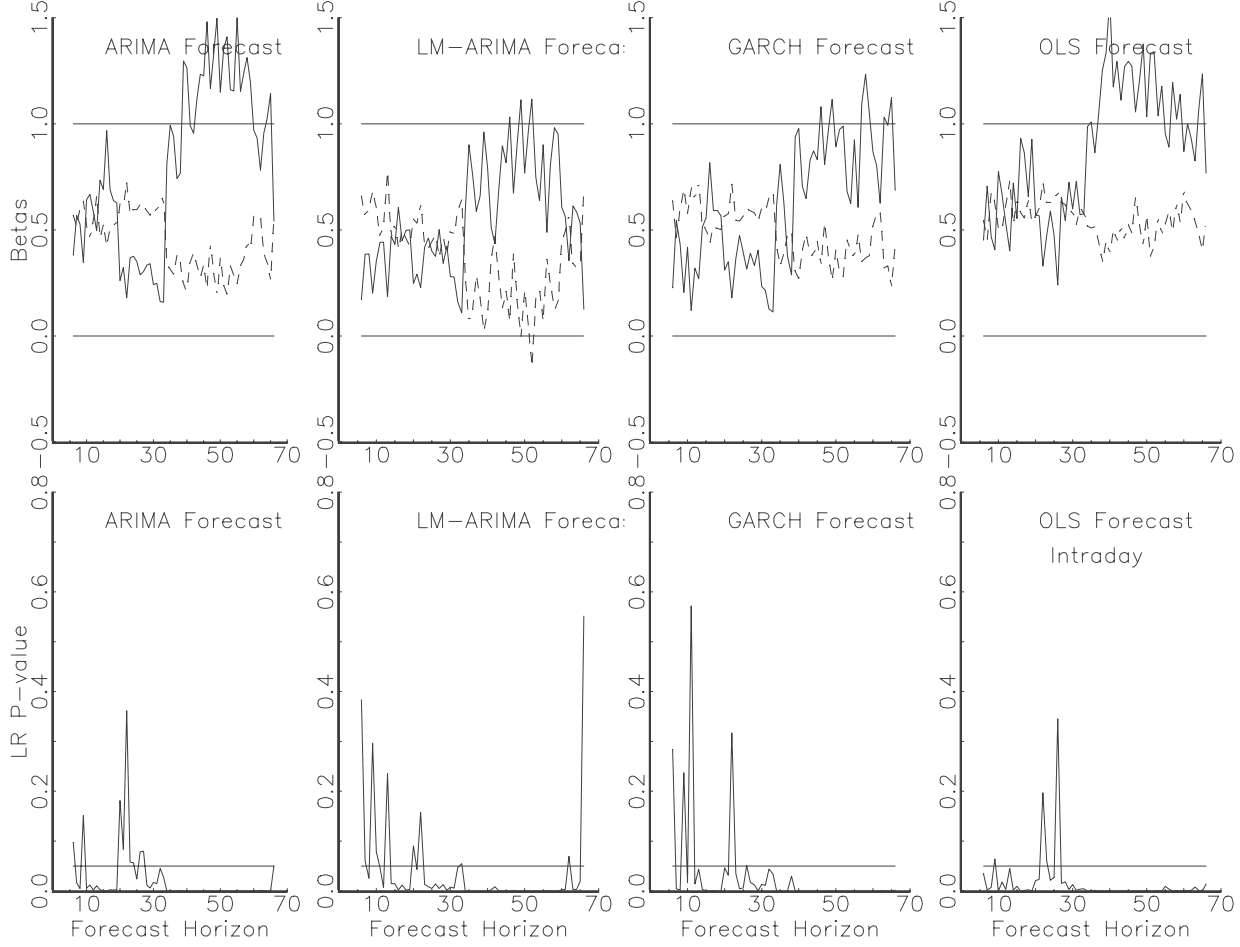
Notes: The figure displays p-values from the likelihood ratio test that assesses the wisdom of pooling coefficients across exchange rates in equation (13). Low p-values reject the null hypothesis that the pooling restriction is appropriate. P-values for the intraday (high-frequency) volatility measure are represented by dashed lines while those from futures price volatility are represented by solid lines. Some p-values are not visible because they are so close to zero.

Figure 5: Estimated differences in the Bayesian information criteria from pooled vs. unpooled models predicting RV with IV.



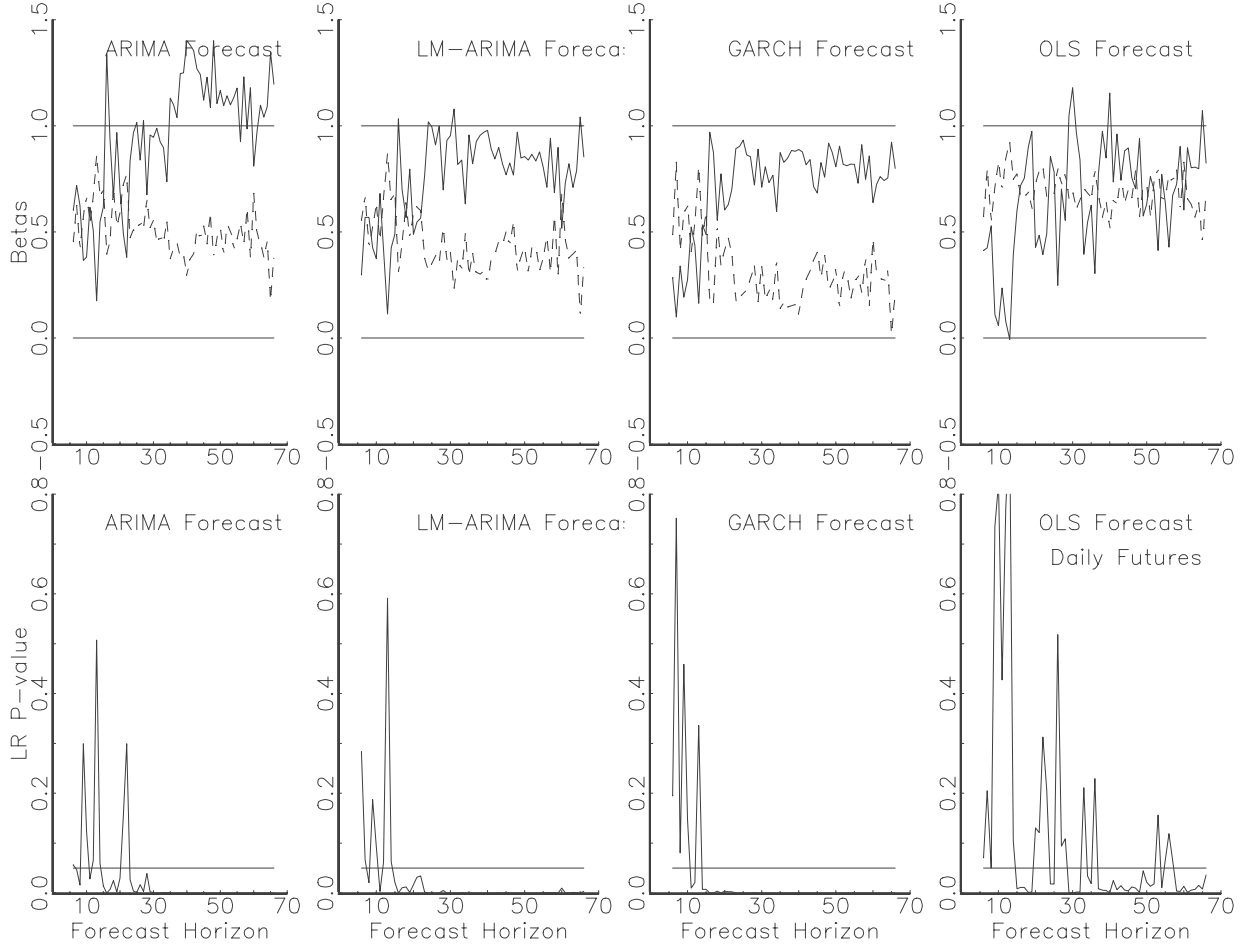
Notes: The figure displays the difference in the Bayesian information criteria for equation (13) under the pooling constraint that all exchange rates share a common $\{\alpha_k, \beta_k\}$ and the unpooled model. Positive values of the statistic indicate that the BIC favors pooling the data, negative values indicate the reverse. BIC differences for the intraday (high-frequency) volatility measure are represented by dashed lines while those from futures price volatility are represented by solid lines.

Figure 6: Estimated $\beta_{k,1}$, $\beta_{k,2}$, and p-values for the null that $\beta_{k,2}$ equals zero, from a model predicting high-frequency RV with IV and forecasts of high-frequency RV until expiry, pooling over exchange rates.



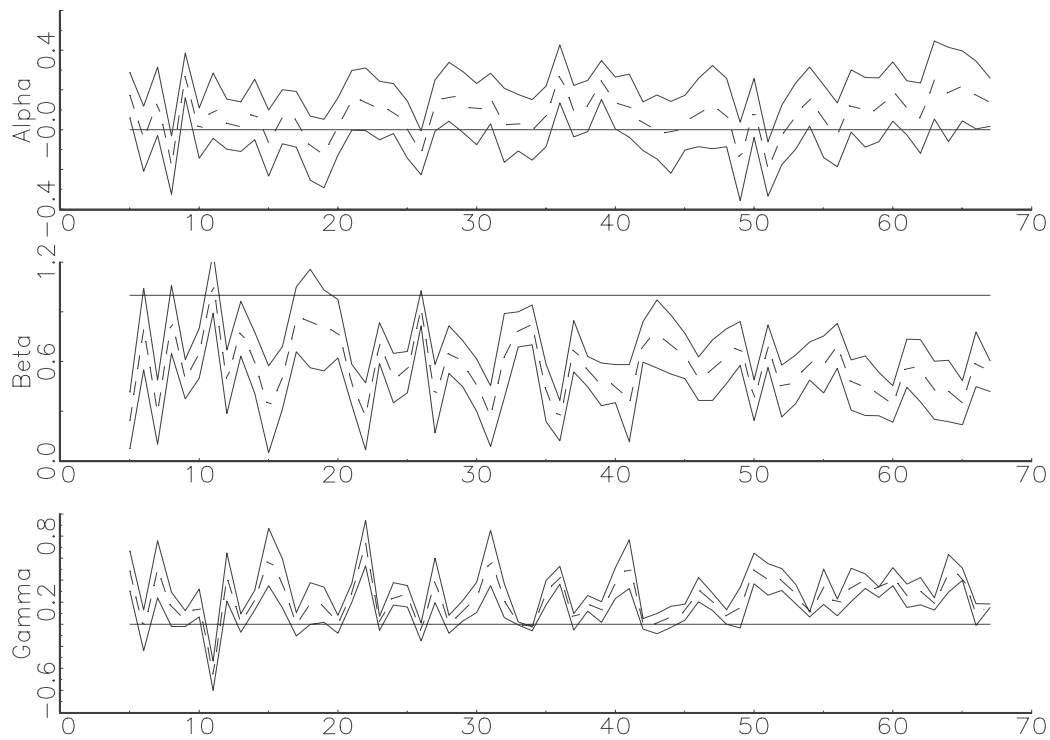
Notes: The top panels of the figure displays the series of $\hat{\beta}_{k,1}$ (dashed line) and $\hat{\beta}_{k,2}$ (solid line) from pooled (over exchange rates) models of RV predicted by IV and four forecasts of RV until expiry (equation (14)). Horizontal lines in the upper panels denote 0 and 1. The lower panels of the figure display the p-values from the LR tests that IV is informationally efficient (each $\beta_{k,2}$ is equal to zero). Horizontal lines in the lower panels denote 0.05. The four statistical forecast models are (from left to right in the panels) ARIMA, LM-ARIMA, GARCH, and OLS. Coefficients on IV ($\hat{\beta}_{k,1}$) are represented by dashed lines while those from the econometric forecasts ($\hat{\beta}_{k,2}$) are represented by solid lines. The out-of-sample forecast period was 1992-1998. High-frequency data were used to construct the volatility measure.

Figure 7: Estimated $\beta_{k,1}$, $\beta_{k,2}$, and p-values for the null that $\beta_{k,2}$ equals zero, from a model predicting daily futures RV with IV and forecasts of daily futures RV until expiry, pooling over exchange rates.



Notes: The top panels of the figure displays the series of $\hat{\beta}_{k,1}$ (dashed line) and $\hat{\beta}_{k,2}$ (solid line) from pooled (over exchange rates) models of RV predicted by IV and four forecasts of RV until expiry (equation (14)). Horizontal lines in the upper panels denote 0 and 1. The lower panels of the figure display the p-values from the LR tests that IV is informationally efficient (each $\beta_{k,2}$ is equal to zero). Horizontal lines in the lower panels denote 0.05. The four statistical forecast models are (from left to right in the panels) ARIMA, LM-ARIMA, GARCH, and OLS. Coefficients on IV ($\hat{\beta}_{k,1}$) are represented by dashed lines while those from the econometric forecasts ($\hat{\beta}_{k,2}$) are represented by solid lines. The out-of-sample forecast period was 1992-1998. Daily futures prices were used to construct the volatility measure.

Figure 8: Coefficients of the volatility risk model, by horizon



Notes: The panels of the figure display the series of $\hat{\alpha}_k$, $\hat{\beta}_{k,1}$ and $\hat{\gamma}_k$ and two-standard-error bands estimated by an instrumental GMM procedure following equation (16). The horizontal axis indexes the forecast horizon, k . Solid horizontal lines denote 0 or 1.

¹ There are other explanations that are beyond the scope of this paper. For example, Bandi and Perron (2003) find VIX and OEX implied volatility is fractionally cointegrated with realized volatility.

² The Hull and White (1987) result—that the option price is given by the BS formula evaluated at expected average variance until expiry—is not valid for the Heston (1993) model, which permits arbitrary correlation between parameters. Romano and Touzi (1997) show, however, that—with minor modifications—the HW result that the correct option price is the expected BS implicit variance is still correct. Because there is very low correlation between asset return and variance innovations in foreign exchange futures, the Romano and Touzi (1997) adjusted formula is extremely close to the ordinary BS formula for the relatively short-term options considered in this paper.

³ These biases can be avoided entirely, however, by using a stochastic volatility pricing model rather than the BS model, solving for implicit instantaneous variance and using the parameters of the instantaneous variance process to derive the desired expectation of variance until expiry.

⁴ Researchers also estimate (3) with the standard deviation of asset returns and the implicit standard deviation (ISD), rather than variances. The results from such estimations provide similar inference to those done with variances. Previous versions of this paper used ISDs and the results were very similar. The current version uses variances for consistency with Chernov's (2002) model.

⁵ One might wonder whether the statistical forecast of realized volatility should be transformed to be orthogonal to IV before being used in the predictive equation (3). This is not necessary. The t statistic on $\hat{\beta}_2$ will provide the same inference (asymptotically) as the appropriately constructed F test for the hypothesis that $\beta_2 = 0$. And the F test—which is based on the R^2 of the regression—will be invariant to orthogonalization of the regressors.

⁶ Anderson and Bollerslev (1998) measure integrated volatility with the sum of squared 5-minute returns. This paper uses 30-minute returns to avoid serial correlation at such high frequencies. All exercises were also done with 5-minute returns, however, and the very minor differences between 5- and 30-minute volatility are noted where appropriate.

⁷ Garcia, Ghysels, and Renault (2003) survey recent options pricing models. Whaley (2003) looks at the wider derivatives literature.

⁸ Using the full sample to estimate ρ , κ_v , θ_v , and σ_v potentially introduces a look-ahead bias into the IVs. This is not very worrisome in practice, however. IV estimates were not very sensitive to reasonable values of the parameters, which

were fairly stable across exchange rates and over time. The insensitivity of IV to model and parameter estimates is consistent with Bates (2000). An alternative procedure would be to derive all parameters from options prices, each day, but this method is computationally very difficult in many cases and impossible for some. Although Chernov and Ghysels (2000) find that relying only on options data in pricing and hedging the S&P 500 index contract is best, experiments indicate that it is unlikely to make much difference.

⁹ Latane and Rendleman (1976) pioneered the concept of recovering market expectations of volatility from options prices. Bakshi, Cao, and Chen (1997) and Sarwar and Krehbiel (2000) apply versions of the stochastic volatility (SV) option pricing model to hedging and pricing problems. Bakshi, Cao, and Chen (1997) use a more elaborate version of the SV model with stochastic jumps and interest rates.

¹⁰ Poon and Granger (2003) and Figlewski (1997) review the literature on forecasting volatility in financial markets. Neely and Weller (2002) apply genetic programming and other techniques to forecast foreign exchange volatility. Pong, Shackleton, Taylor, and Xu (2003) compare the forecasting ability of ARFIMA models with IV by mean-squared error and R^2 metrics.

¹¹ Christensen and Prabhala (1998) also estimate versions of (3) and (4) with feasible generalized least squares (FGLS) for one short subperiod but find it does not help IVs bias or efficiency. See Table 6 in that paper.

¹² The $\hat{\beta}_1$ coefficients using 5-minute data to measure realized volatility are slightly smaller (0 to 0.07) than those in Table 3 that use 30-minute data. Results with 5-minute data are omitted for brevity.

¹³ Davidson, Kim, Ors, and Szakmary (2003) study IV in 37 markets and emphasize that findings from one are not necessarily generalizable to all.

¹⁴ In independent work, Li (2002) and Martens and Zein (2002) use intraday data and long-memory models to forecast RV. Both papers conclude that such models can improve on forecasts from IV alone.

¹⁵ Full results are omitted for brevity but are available in the working paper version of this paper.

¹⁶ Experiments with very similar options on gold futures—not reported for brevity—suggest that correcting IV estimates for the volatility smile made very little difference in the bias or informational efficiency of IV. One must conclude that error generated by the volatility smile is not important for the issues of bias and informational efficiency.

¹⁷ Jorion concluded that only implausibly large errors could generate the apparent bias in the coefficients. A previous version of this paper (Neely's) argued incorrectly that there is large measurement error in IV estimation. This conclusion was due to the author's error.

¹⁸ There is a modest amount of overlap at horizons greater than 50 business days.

¹⁹ Bollen and Whaley (2003) also examine vega hedging for S&P 500 index options. This exercise is not pursued here because IV is relatively successful in the simpler delta hedging environment.

²⁰ Green and Figlewski (1999) examine the risks inherent in pricing and hedging options and the extent to which using a higher than expected IV can compensate for these risks.

²¹ Taylor (2000) examines the consequences of long memory in volatility for the term structure of IV. Wright (2000) argues that estimating long memory in squared returns—as opposed to log squared or absolute returns—is likely to bias the estimate of the long memory parameter, d , downward.

²² In future expressions, the additive constant is omitted to simplify notation.

²³ If the variance process has a jump component, one still obtains variance until expiry as a linear function of instantaneous variance and the BS variance, but the coefficients are then functions of more hyperparameters.