

The information content of realized volatility forecasts*

Torben G. Andersen[†] Per Frederiksen[‡] Arne D. Staal[§]
Northwestern University *Nordea Bank* *Lehman Brothers*

(preliminary and incomplete)

Abstract

We examine the relative information content of monthly volatility forecasts derived from option prices and time-series forecasts of realized volatility. If volatility risk is priced the return variance process is different under the physical and risk-neutral pricing measures, which has important implications for the interpretation of implied volatility as an expectation of future volatility. We use recent developments in volatility measurement and modeling to capture and forecast the variance process. Our results suggest that forecasts of future realized volatility based on past realized volatility are unbiased, more efficient than VIX model-free implied volatility, and roughly as efficient as Black and Scholes implied volatility. Forecasts of future volatility based on historical realized volatility contain incremental information relative to both model-free and Black-Scholes implied volatility. This finding can be explained in the context of priced variance risk. Given the finding of different information contained in individual volatility forecasts, we show that conditional volatility forecasts can be improved upon by combining individual measures. We also show that predictive regressions of realized volatility on option implied volatility are likely to be misspecified due to the presence of long memory in such measures and the variance risk premium.

*We would like to thank M. Ø. Nielsen for sharing option price data with us and providing helpful comments.

[†]Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208, and NBER; phone: +1-847-467-1285; e-mail: t-andersen@kellogg.northwestern.edu

[‡]Equity Trading & Derivatives, Nordea Markets, Strandgade 3, 0900 Copenhagen C, Denmark; phone: +45-3333-6683; e-mail: per.frederiksen@nordea.com

[§]Lehman Brothers, 745 7th Ave., New York, NY 10019; e-mail: arne.staal@lehman.com

1 Introduction

Return volatility is at the center of many theories within financial economics, be it asset and derivatives pricing or risk management, so it is hardly surprising that great effort has been made to determine reliable, if not optimal, procedures for forecasting future volatility. Likewise, the practical import of volatility for financial performance has spurred product innovation, leading to a rapid increase in organized trading of financial derivatives written directly on volatility variables, such as variance swaps and futures and options written on volatility indices as well as an over-the-counter market in variance and volatility swaps on individual assets. In short, the financial industry views volatility as a distinct asset class endowed with separate risk factors and novel opportunities for both strategic trading and hedging. Obviously, the latter developments also have generated a surge in the demand for practical volatility forecast procedures.

There are two main academic paradigms for formal volatility forecasting. The first involves building direct time series models for return volatility using past asset prices, possibly in combination with other relevant historical information, to capture the dynamic evolution of the process. Given estimates of model parameters and the current volatility state variable, expected future volatility can then be computed. This approach encompasses the GARCH and stochastic volatility (SV) model class along with variants of each. The second paradigm backs out market expectations of future volatility from observed derivatives prices. In particular, in an efficient market the option price incorporates all available relevant information, including past returns. Hence, implied volatility is often argued, a priori, to provide an optimal forecast of future volatility and subsume any forecast generated by time series models based solely on historical prices. Numerous studies, cast in a predictive regression setting, have tested whether implied volatility is an unbiased and information efficient forecast of future realized volatility. Unbiasedness is typically assessed through the associated regression coefficients while a forecast, loosely, is defined to be information efficient if it is not subsumed by other forecast variables. Although the conclusions are somewhat contradictory, and often emphasize the need to correct for various methodological issues, most authors contend that option implied volatilities provide biased, but efficient, forecasts of future volatility, see, e.g., Jiang and Tian (2005) for a recent example.

One potential explanation for the observed bias in option implied volatility vis-a-vis future realized volatility is the presence of a volatility risk premium. When invested, say, in an equity portfolio, one faces at least two sources of uncertainty regarding the future returns. One is, of course, the inherent random nature of the return, as typically captured by the return variance, while a second is uncertainty about the return variance itself. Since, indisputably, returns are risky and the evidence for stochastic volatility is conclusive, see, e.g., Andersen, Bollerslev, Diebold & Ebens (2001)), it is natural to expect the risk associated with an uncertain return variance to be priced, at least at the aggregate market level, as argued by Lamoureux and Lastrapes (1993), Poteshman (2000) and Chernov (2006). In fact, when examining variance swap contracts, Carr & Wu (2004) and Bondarenko (2007) find large, negative and statistically significant unconditional variance risk premia for stock indices, as the option implied volatility, on average, exceeds the ex-post realized volatility of the underlying asset by a substantial amount. Hence, an exposure to equity index volatility is associated with negative expected returns. They also find that these premia cannot be rationalized by the CAPM or other commonly used risk factors, suggesting that the market variance risk premium involves a distinct set of risk exposures. Moreover, the variance risk premia appear to fluctuate over time, typically increasing with the level of expected volatility itself and showing sensitivity to the business cycle. Qualitatively consistent with this finding is the evidence that a volatility factor is priced in the cross-section of stock returns, see, e.g., Ang, Hodrick, Xing and Zhang (2006).

A second potential explanation for the observed bias, and the varied empirical findings more generally, is that the underlying OLS predictive regression framework may be fundamentally flawed. The realized return volatility is a highly persistent, yet seemingly stationary, process whose dynamic dependence structure is well captured by a long memory or fractionally integrated process, denoted $I(d)$, $0 < d < 1/2$. But if the realized volatility is fractionally integrated, then the implied option volatility process will also tend to display long range persistence. This invalidates the standard OLS regression set-up which operates with short memory, or $I(0)$, random variables only. Instead, the predictive regressions must be viewed as fractionally cointegrated relations with an associated statistical inference theory that have some features in common with integrated, or $I(1)$, variables within cointe-

grating regressions. In the fractional cointegration setting it may indeed not be feasible to estimate the short-run relationship between the realized volatility and the underlying option market expectations of actual volatility but, arguably, the unbiasedness hypothesis may be tested as a long-run equilibrium condition. This point has been made forcefully by Christensen and Nielsen (2006) and Bandi & Perron (2006).

A third potential explanation for the observed biases is that the option implied volatilities typically are model-dependent and the underlying model, even ignoring the issue of risk premia, may be misspecified. The use of at-the-money (ATM) Black-Scholes option implied volatilities (IV) is widespread because they are tractable and they are approximately linear functions of the underlying true volatilities under some popular stochastic volatility models. Nonetheless, it is problematic to rely on a misspecified model for formal testing. One possible response to this problem is to exploit the newly developed theory for model-free implied volatility (MFIV) as exemplified by (Jiang & Tian (2005*b*)). The underlying methodology, based on the theory for variance swap contracts and with pricing obtained from the full cross-section of options prices, has been adopted by the Chicago Board of Exchange (CBOE) in constructing the monthly so-called VIX, or model-free implied volatility, for the S&P500 equity index, building on the original work of Carr and Madan (1998) and Britten-Jones & Neuberger (2000). While the methodology is widely applicable, the publically observed VIX index has become a standard measure of market volatility with documented correlation with the trading conditions in a variety of global markets, leading to the moniker of an "investor fear gauge." Since the VIX is theoretically superior and more robust it may perform better than the alternative implied volatility measures. At the same time, the theory rationalizing the MFIV clearly identifies the measure with the expected quadratic return variation under the risk-neutral (Q) measure as opposed to the expectation under the actual or physical (P) measure. In contrast, it is the quadratic return variation under the actual measure which constitutes the relevant volatility measure for the underlying asset. Hence, the model-free implied volatility forecasts do not predict the future (physical) realized asset volatility, but rather this quantity in conjunction with the associated volatility risk premium, as also emphasized above.

A fourth potential explanation for the wide variation in reported findings is the different empirical measures adopted for realized and historical volatility. The earlier studies computed realized volatility over a multi-day horizon as the subsequent cumulative squared daily returns over the corresponding period. However, a recent literature documents that use of high-frequency data leads to dramatic improvements in measuring and modeling volatility. The simplest approach is to follow, e.g., Andersen & Bollerslev (1998), and exploit the summation of high-frequency intraday squared returns as model-free volatility measures. In effect, the associated measurement errors are quite small and uncorrelated across days so, apart from a minor measurement error, we may treat realized volatility as observed. This improved measurement of the regressand in the predictive regression framework leads to more efficient inference. Equally important, the relatively small measurement error renders direct time series modeling of the realized volatility process practical. Such models provide significant improvements over GARCH or stochastic volatility forecasts based on daily data (Andersen, Bollerslev, Diebold & Labys (2003)). As such, they constitute a natural benchmark for gauging the quality of implied volatility forecasts because they utilize the information in past asset prices quite effectively. In contrast, the so-called historical volatility, typically constructed from inefficient measures of lagged realized volatility, are outright poor volatility forecast candidates and hence not proper benchmarks for forecast performance. In sum, there is potential for significant improvements in both measurement and selection of time series based forecast variables for realized volatility, leading to modifications in both the regressors and the time series regressand of the usual predictive regressions.

In this paper we explore the relative informativeness of implied volatility and time series based volatility forecasts, taking into account the set of issues discussed above. We study the volatility of monthly return series derived from the S&P 500 futures market and the associated SPX options. This allows us to minimize the impact of other features that may cloud the empirical results. First, these options are European style and are heavily traded. The high liquidity is important as the pricing of less liquid options may reflect liquidity premia and a host of microstructure issues that may induce significant measurement errors. The model-free implied volatilities are readily available given the existence of the CBOE VIX index. We are careful in constructing realized volatility measures

that can be compared to implied volatility by matching the number of trading days until expiration and including the overnight and weekend returns. Moreover, the use of high-frequency futures data minimizes the nonsynchronous trading problem and thus largely avoids the spurious return autocorrelation and noise that is induced into the corresponding high-frequency cash equity-index series. Overall, these procedures alleviate the impact of errors-in-variables and overlapping data concerns, see, e.g., Christensen & Prabhala (1998)), that have been stressed in the preceding literature.

Within this fairly controlled setting, we find supportive evidence for the presence of both risk premia and long memory in the implied volatility series. Hence, regular predictive regressions involving a single implied volatility regressor may indeed be statistically suspect. However, we also find that the corresponding regression involving time series forecasts are well specified and, perhaps even more constructively, we present evidence that the associated encompassing regressions, having both implied volatility and time series forecasts as regressors, appear sound. Intuitively, this occurs because the realized volatility forecast, almost by construction, is integrated of the same order as the actual realized volatility series so we obtain a "balanced" regression with an error term that is purged of long memory features. In the absence of the realized volatility forecast regressor, the combination of a volatility risk premium and the long memory persistence in the system is absorbed in the regression residual, rendering standard inference techniques invalid. Nonetheless, as indicated, we are able to assess the short run unbiasedness and forecast efficiency issue within the usual predictive regression framework by expanding the set of regressors in a suitable fashion and paying close attention to related residual diagnostic tests. Hence, we are to a large extent able to circumvent the fundamental inference problem associated with the short-run predictive regressions identified by Bandi and Perron (2006). Moreover, we also expand upon their study of longer run unbiasedness of the implied volatility measures by casting them within a corresponding encompassing regression system that also includes time series forecast regressors.

We also deviate from prior studies by using a carefully selected realized volatility model estimated from intraday data to generate the time series forecasts. In fact, we are seemingly the first to compare the model-free VIX with such time series forecasts. Perhaps as a reflection of these features, our

empirical findings turn out to be radically different from those obtained by existing studies. We find the time series based forecasts to be unbiased and fully competitive with the (biased) BS implied volatilities, while the VIX forecasts perform worst. Specifically, we show that forecasts of future volatility based on the past high-frequency return data provide incremental information relative to implied volatilities, and vice versa. Our interpretation is that implied volatilities carry information beyond a pure expectation of future volatility due to the divergence between the risk-neutral pricing measure and the physical measure in the presence of variance, and most likely also jump, risk premia. Hence, the wider information set exploited by option traders is partially offset by the differences across the two probability measures. We also confirm that one may further improve upon the performance of any of these forecasts by combining them through a natural weighting scheme. Finally, we speculate that the disappointing forecast precision of the VIX index may be tied to the way in which the measure is extracted from observed option prices by the CBOE rather than reflecting an inherently poorer procedure. However, given the widespread popularity of the index it is of direct interest to establish its forecast power vis-a-vis natural competitors.

The existing literature comparing high-frequency based time-series forecasts with implied volatilities is limited, but the topic is addressed by, e.g., Pong, Shackleton, Taylor & Xu (2003), Li (2002), and Martens & Zein (2004). Our study differs substantially from these in scope, implementation, and conclusion. For example, only Martens & Zein (2004) consider equity volatility and none of them deal with the confounding issues arising from the joint existence of risk premia and long range (co-)dependence. In summary, we implement the forecast evaluation procedures in a methodologically distinct manner, inspired by inherent properties of the underlying volatility series, and obtain new empirical results regarding the relative efficacy of time series and implied volatility forecasts. Given the relative advantages and limitations of implied and time series approaches, it should really not be surprising that we find it useful, if not optimal, to combine the forecasts.

The rest of the paper proceeds as follows. Section 2 discusses the different measures of expected volatility employed in this paper. In section 3 the data is described. Section 4 contains a discussion of the econometric framework. Section 5 gives the empirical results and an interpretation. Conclusions

and suggestions for future research are in the final section.

2 Expectations of future volatility and the variance risk premium

In this section we define the underlying return volatility concept, we review the recently developed high-frequency intraday data based measurement, time series modeling and forecasting of return volatility, and we formally introduce two distinct notions of implied volatility.

2.1 Realized volatility and long memory modeling

In the standard arbitrage-free asset pricing framework, the log-price of a financial asset follows a continuous-time semi-martingale process with stochastic volatility and, possibly, jumps. In particular, we assume the univariate log-price process, $p(t) = \ln(S(t))$, follows a general jump diffusion,

$$dp(t) = \mu(t)dt + \sigma(t)dw(t) + \kappa(t)dq(t), \quad t \geq 0.$$

The drift process $\mu(t)$ is continuous and locally bounded, the spot volatility process $\sigma(t) > 0$ is cadlag, $w(t)$ is a standard Brownian motion, and for the counting process, $q(t)$, we have $dq(t) = 1$ corresponding to a jump at time t and $dq(t) = 0$ otherwise. Finally, when a jump occurs, $\kappa(t)$ denotes the corresponding jump size. Within this setting, a special case may be used to highlight the relevant volatility concept. Over short horizons, the semi-martingale property implies that the mean return is negligible relative to the price variation associated with the jump and diffusive terms. Hence, without loss generality, we may ignore the drift, $\mu(t) = 0$. If, in addition, we assume that the volatility process is independent of the Brownian motion then, in the absence of jumps,

$$p(t) \mid \sigma^{2*}(t) \sim N(0, \sigma^{2*}(t)),$$

where $\sigma^{2*}(t) = \int_0^t \sigma^2(s)ds$ denotes the integrated variance.

A related quantity is the quadratic variation of the process $p(t)$, which is defined for any semi-martingale (see Protter (1990)) by $[p](t) = p^2(t) - 2 \int_0^t p^2(s) dp(s)$. If returns are obtained from a given sampling scheme satisfying, $0 = s_0 < s_1 < \dots < s_m = t$ and $\max_j |s_j - s_{j-1}| \rightarrow 0$ as $M \rightarrow \infty$, we have

$$[p](t) = p \lim \sum_{j=1}^M (p(s_j) - p(s_{j-1}))^2.$$

Under very general conditions allowing, e.g., for spot volatility to jumps, the quadratic variation process for the jump diffusion model equals the sum of the integrated variance and the squared jumps through time t ,

$$[p](t) = \sigma^{2*}(t) + \sum_{s=0}^{q(t)} \kappa^2(s).$$

Realized variance, defined as the sum of intraday squared returns, provides consistent estimates of the quadratic variation for continuous semi-martingales (Andersen, Bollerslev, Diebold & Labys (2001, 2003), Andersen et al. (2003)). Letting time be measured in daily units and denoting the intraday returns during day t obtained by sampling M times over the day,

$$r_{t,i} = p_{t-1+\frac{i}{M}} - p_{t-1+\frac{i-1}{M}}, \quad i=1, \dots, M,$$

we may then define the realized variance for day t as,

$$RV_t = \sum_{i=1}^M r_{t,i}^2,$$

where the precision of the estimate is related to the choice of M . Theoretically, choosing a higher M improves the precision of the RV measure. Practically, we need a compromise between the improvement in precision obtained from using more observations and the increased sensitivity to market microstructure noise (bid-ask bounce, measurement error, nonsynchronous prices, etc.). Several studies have considered the preferred choice of M (see Bandi & Russell (2003), Zhang, Mykland & At-Sahalia (2003)). We adopt a conservative approach and follow the earlier literature in using 5-minute returns. This is in line with the recommendations from the recent literature regarding the appropriate choice of a sparse sampling frequency, even if alternative procedures may be seen as preferable.

The RV measure is consistent for the increment to the quadratic variation process as $M \rightarrow \infty$,

$$\lim_{M \rightarrow \infty} RV_t = \int_{t-1}^t \sigma^2(u) du + \sum_{s=q(t-1)}^{s=q(t)} \kappa^2(s).$$

The consistency result only requires that the maximum distance between any two consecutive observation within the exogenously chosen sampling scheme approaches zero in the limit. Barndorff-Nielsen & Shephard (2002) show that, in the absence of jumps and leverage effects, RV converges to $\sigma^{2*}(t) = \int_0^t \sigma^2(s) ds$ at rate \sqrt{M} and satisfies a (mixed) Gaussian asymptotic distribution theory. Moreover, they have subsequently relaxed the absence of leverage condition. The bottom line is that realized volatility approximates ex-post realizations of quadratic well, but there are invariably some discretization error and microstructure frictions that induce a measurement error in the computed realized return variation measures

Realized volatility is an observable series and may thus be modeled directly. Recent work concludes that log-transformed realized volatility exhibits long-memory features (the correlogram dies out more slowly than exponentially). To model these properties and provide associated volatility forecasts, Andersen et al. (2003) adopt the class of autoregressive fractionally integrated moving average (ARFIMA) processes, introduced into econometrics by Granger & Joyeux (1980) and Hosking (1981). In particular, the d 'th difference of each series is a stationary and invertible ARMA process where d may be any real number such that $-1/2 < d < 1/2$ to ensure stationarity and invertibility. More precisely, σ_t is an ARFIMA(p, d, q) process if

$$\alpha(L)(1-L)^d(\sigma_t - \mu) = \beta(L)v_t,$$

where $\alpha(z) = 1 - \alpha_1 z - \dots - \alpha_p z^p$ and $\beta(z) = 1 + \beta_1 z + \dots + \beta_q z^q$ are polynomials of order p and q , respectively, in the lag operator L ($L\sigma_t = \sigma_{t-1}$) with roots strictly outside the unit circle, v_t is $iid(0, \sigma_v^2)$, and $(1-L)^d$ is defined by a binomial expansion involving the gamma function, $\Gamma(\cdot)$,

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j.$$

The parameter d determines the memory of the process. If $d > 0$ the process is said to possess long memory as the autocorrelations die out at a slow hyperbolic rate and thus fail to be absolutely summable, in contrast to the much faster exponential rate in the weak dependence case ($d = 0$). That is, the covariance function for volatility at lags j , for $j \rightarrow \infty$, are of the order j^{2d-1} , in contrast to the usual order of ρ^j , for $|\rho| < 1$, within the usual short memory GARCH style volatility models.

Given an estimated ARFIMA model, forecasting is carried out by extrapolating the estimated model. Deo, Hurvich & Lu (2006), Andersen et al. (2003), and Martens, van Dijk & de Pooter (2004) show that forecasting log realized volatility based on a simple ARFIMA(1,d,0) model is a very good competitor to other time-series methods of forecasting realized volatility¹.

Modeling volatility as a long-memory process is also theoretically consistent with an option pricing framework. Comte & Renault (1998) derive the relationships between realized, implied, and spot volatility in a long memory framework. They show that if the unobservable spot volatility in a stochastic volatility model displays long-range dependence then so do implied and realized volatility.

2.2 Implied volatility

The most commonly adopted notion of implied volatility is obtained from the renowned Black and Scholes option pricing formula. It provides the arbitrage-free price of a European option given the parameters of the hypothesized model for asset returns. A European call option with T periods to expiration and strike price K is given by

$$\begin{aligned} c(S, K, T, r, \sigma) &= s \Phi(\delta) - \exp(-rT)K\Phi(\delta - \sigma\sqrt{T}), \\ \delta &= \frac{\ln(\frac{S}{K}) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \end{aligned}$$

where S is the price of the underlying asset, r is the risk free interest rate, Φ denoted the c.d.f. of the standard normal distribution, and σ denotes the volatility of the underlying asset. The parameters in

¹In the empirical analysis we have also used the HAR-RV model to predict future realized volatility. These forecasts are slightly less efficient in the predictive regressions (results are available on request).

the model, including the volatility σ , are assumed to be constant over the life of the option. Given a market price of the option and observed values for all parameters except σ , we can invert the formula to find the value of σ implied by the market price, assuming the model is valid². Feinstein (1989) shows that the Black Scholes option pricing formula is nearly linear in volatility for at-the-money (ATM) options. We denote the implied volatility measures derived accordingly BS implied volatility. In principle, this is strictly model dependent and rests on the counterfactual assumption of constant volatility. However, it may be shown, in the absence of volatility risk premia, that the BS implied volatility, σ , approximately equals the expected volatility (more precisely, the square root of the integrated variance) until option expiration, even if volatility is time-varying. Comte & Renault (1998) demonstrate that the BS implied volatility of a short-term option remains a good proxy for the expected realized volatility of the underlying price process in the case of long-range dependent volatility. Of course, in the case of nonzero volatility and/or jump risk premia the difference between implied volatility and expected future return volatility is governed by these premia and a non-trivial bias may be expected. Nonetheless, the basic linear relationship between the volatility expectations and the BS implied volatility measures serves to motivate the emphasis on linear regression tests in the literature.

2.3 Model-free implied volatility

Carr & Madan (1998) and Britten-Jones & Neuberger (2000) independently show that the risk-neutral expected sum of squared returns on a financial asset over any interval $[t, T]$ can be obtained from the set of market prices at time t for European options expiring on time T . The derivation assumes a full set of call options with a continuum of strike prices and a continuum of maturities is traded on the asset. Under regularity assumptions, they find that

$$E_t^Q \left[\int_t^T \left(\frac{dS_t}{dt} \right)^2 \right] = \frac{2}{T-t} e^{r(T-t)} \left[\int_0^{F_t} \frac{P_t(T, K)}{K^2} dK + \int_{F_t}^{\infty} \frac{C_t(T, K)}{K^2} dK \right], \quad (1)$$

²More precisely, given the observed call option price C , we need to solve the nonlinear equation $C = c(S, K, T, r, \sigma)$ numerically for the implied value of σ .

where S is the price of the underlying asset, K is the strike price of an option, C_t and P_t denote the current call and put option prices, F_t is the time- t forward price, and E_t^Q is the expectations operator w.r.t. the risk neutral distribution. Jiang and Tian (2005a) provide a simplified derivation of the result and generalize it by showing that it applies approximately for processes with jumps as well. The variance dynamics can be general and it may, for example, contain jump components.

The most interesting aspect of this identity is that it is based on, at least in principle, observed option prices and remains valid under very general assumptions on the underlying stochastic process so that is effectively is a model-free measure. The continuum of option prices needed for estimation of the risk-neutral expected integrated variance is not available in the financial markets, but accurate approximations can be obtained through a discretization procedure. The VIX index, provided by the Chicago Board of Options Exchange (CBOE), is defined as the square-root of a measure calculated on the basis of this approach. We denote this a model-free implied volatility measure. Jiang & Tian (2005a) identify problems in the way the CBOE calculates the VIX as it leads to systematic biases. Instead, they provide a no-arbitrage smoothing method to calculate the expected risk-neutral variance that avoids the problem of inducing systematic errors through the approximation technique. Bondarenko (2007) further generalizes the conditions underlying this procedure for measurement of the expected future return variation and robustifies the implementation significantly.

It is important to realize that the model-free implied volatility implicit in this identity is an expectation with respect to the risk-neutral distribution. If volatility risk is priced, the expected volatility under the risk-neutral measure and the expected volatility under the physical measure differ by a quantity reflecting the associated risk premium. We explore this issue in more depth in the following subsection.

2.4 The variance risk premium

The concept of a variance risk premium is particularly transparent in the context of a derivative instrument that is contingent on volatility itself. A variance swap is a contract with zero net market

value at entry, and a payoff to the long side (buying volatility) equal to the difference between realized volatility and the fixed variance swap rate set upon contract initiation. The payoff to the long side is,

$$L(RV_{t,T} - SW_{t,T})$$

for a contract entered into at time t with maturity date T , swap rate $SW_{t,T}$, a notional amount L . Since the contract is worthless at the time of entry, its value in the absence of arbitrage opportunities is equal to the expected quadratic return variation under the risk-neutral measure,

$$SW_{t,T} = E_t^Q[QV_{t,T}].$$

Importantly, even in the absence of an active swap variance market, the results of Carr & Madan (1998) and Britten-Jones & Neuberger (2000) provide us with a theory to approximate the risk-neutral expected return variation from observable option prices as in equation (1). The model-free implied volatility in the previous section is merely a relabeling of the variance swap rate which result from pricing the variance swap contract.

The model-free implied volatility or variance swap rate is an expectation with respect to the risk-neutral distribution. If volatility risk is priced, the expected volatility under the risk-neutral and the physical measure differ by a quantity reflecting the associated risk premium. The concept of a variance risk premium becomes explicit when pricing a variance swap contract in the standard stochastic discount factor framework (see Cochrane (2001)). Given a normalized pricing kernel,³ $m_{t,T}$, we have in the absence of arbitrage that

$$SW_{t,T} = E_t^P [m_{t,T}RV_{t,T}],$$

These equivalent pricing equations imply that the expected difference under the physical measure between realized volatility and the swap rate (model-free implied volatility) reflects a variance risk

³Normalized such that $E_t^P [m_{t,T}] = 1$.

premium,, $-cov_t^p(m_{t,T}, RV_{t,T})$, arising from the covariation between the normalized pricing kernel $m_{t,T}$ and return variance $RV_{t,T}$.

$$E_t^P [RV_{t,T} - SW_{t,T}] = -cov_t^p(m_{t,T}, RV_{t,T}).$$

Hence, the conditional expectation $E_t^P [RV_{t,T} - SW_{t,T}]$ equals the conditional variance risk premium in effect for the period $[t, T]$.

In many applications our goal is to forecast volatility over a certain horizon. From above,

$$E_t^P [RV_{t,T}] = SW_{t,T} - cov_t^p(m_{t,T}, RV_{t,T}).$$

Hence, the realized volatility over $[t, T]$ can be decomposed into its conditional expectation and a conditionally mean zero random shock $\epsilon_{t,T}$,

$$RV_{t,T} = SW_{t,T} - cov_t^p(m_{t,T}, RV_{t,T}) + \epsilon_{t,T},$$

which has important implications for regressions of realized volatilities on model-free implied volatilities if variance risk is priced. Since the variance risk premium is not observed (and therefore cannot be explicitly incorporated in the regressions) it will invariably show up in the regression error. It seems likely that the variance risk premium is a function of expected volatility which suggests that it will inherit some of the dynamic properties of the volatility process.

Again, the direct linkage of expected return variation and the implied volatility measure motivates the exploration of linear relationships between realized and (model-free) implied volatility as exemplified by Jiang and Tian (2005) among others. At the same time, the specific implications of a volatility risk premium within the regression setting identified above point towards potential complications that we also explore in the subsequent empirical work.

3 The econometric approach

The arguments in the previous sections suggest a linear relationship between realized volatility and both BS and model-free implied volatility. We employ several univariate and multivariate encompassing regressions to analyze the forecast efficiency and information content of different volatility expectation and forecast measures.

3.1 Predictive and encompassing regressions

Studies on the implied-realized volatility relation have generally emerged from the perception that implied volatility from option prices is an information efficient predictor of the future realized volatility of the underlying asset. A particularly simple and intuitive approach for testing this conjecture is to run Mincer & Zarnowitz (1969) predictive regressions in which realized volatility is regressed on the corresponding implied, and possibly alternative, volatility forecasts, see, e.g., Christensen & Prabhala (1998), Poteshman (2000), Chernov (2001) and Jiang & Tian (2005*b*). This methodology is still by far the dominant approach within the literature addressing the efficiency and bias issue of implied volatility forecasts.

We follow suit and consider univariate and encompassing regressions of the type

$$y_{t+1} = \alpha + \beta'x_t + \varepsilon_{t+1},$$

where y_{t+1} is a measure of realized volatility over period $t + 1$, say one month, α and $\beta = (\beta_i)_{i=1,\dots,I}$ are, respectively, a scalar and an $I \times 1$ vector of regression coefficients and $x_t = \{x_{i,t}\}_{i=1,\dots,I}$ is a corresponding $I \times 1$ vector of candidate volatility forecasts, with the latter consisting of one or more of the following: a time series based (ARFIMA) volatility forecast, a BS implied volatility measure, and a model-free implied volatility forecast. Hence, the setting accommodates traditional univariate predictive regressions as well as extended encompassing regressions. The generic notation for the regressor and regressands emphasizes the fact that the exposition applies across alternative transformations

of volatility measures and forecasts, such as basic variances, standard deviations or volatilities and log-volatilities, which each may have desirable empirical or theoretical properties. The null hypothesis that a specific volatility forecast, $x_{i,t}$, exploits all available information rationally and provides an unbiased and efficient forecast may then be explicated as follows,

$$H_0 : [x_{i,t}] = E_t^P[y_{t+1}] \quad \text{for all } t \quad \text{and} \quad \alpha = 0, \beta_i = 1, \text{ and } \beta_j = 0 \text{ for all } j \neq i .$$

The population parameter constraints of a zero intercept and a unitary slope is an immediate consequence of unbiasedness, while the additional constraints of zero slope coefficients on the alternative forecasts reflect the assumed efficiency of the candidate forecast. Since the other forecasts are assumed to exploit strictly less of the available information they should be dominated by the fully efficient forecast.

A number of factors may complicate inference, as discussed extensively in the earlier literature, including standard measurement error in the regressors, problems associated with the use of overlapping forecast horizons, imprecise measures of the realized volatility and general market frictions and non-synchronous price observations. However, these issues are well understood and careful data gathering and construction in conjunction with appropriately robust inference can serve to alleviate most pressing concerns along these dimensions. In fact, at the cost of eliminating observations, overlapping forecasts can be avoided and instrumental variable techniques can be used to avoid the measurement error problem if this is deemed a serious concern. Moreover, focusing on a liquid market reduces the market friction problems. Hence, even in the face of these complications, one may test for unbiasedness and efficiency of the implied forecast measures through standard (robust) Wald tests, subject only to the usual stationarity and regularity conditions associated with least squares inference. In addition, for diagnostic purposes, one can test for the presence of serial correlation and long memory in the regression residuals. We discuss the significance of the latter features in some detail in the following subsection.

3.2 Volatility Risk Premia

More critical for the interpretation of predictive regressions is the existence of, potentially time-varying, risk premia in the implied volatility measures. Such premia may serve not only to compensate for uncertainty future volatility but also for the possibility of jumps in the underlying asset price. The evidence for such variance and jump risk premia is accumulating rapidly in the studies exploring the pricing of variance swap contracts, see, e.g., Carr and Wu (2004), Wu (2005), Chernov (2006), Bondarenko (2007) and Todorov (2007). Specifically, the consensus is that the risk premium is substantial, usually negative and varies positively in (absolute) size with the expected level of volatility.

Since the variance swap contracts provide a concrete illustration of features relevant to our subsequent empirical analysis we explore this setting in some additional detail. First, we define the variance risk premium, π_t , as the expected gain on a long position in a variance swap contract of nominal size one dollar, as measured by the payoff over the swap rate. As shown in Section 2.4, this amounts to,

$$\pi_t = E_t^P [RV_{t,T} - SW_{t,T}] = E_t^P [RV_{t,T}] - E_t^Q [RV_{t,T}] = -cov_t^p(m_{t,T}, RV_{t,T}).$$

The covariance term on the right is invariably found to be positive for equity-index variance swap contracts, rendering the risk premium negative. Intuitively, payoffs in poor states-of-the-world are more highly valued than in rich states-of-the-world and, since realized volatility is strongly negatively correlated with the return on the equity market through the so-called leverage effect, being long a variance contract provides a hedge against negative aggregate shocks to the wealth portfolio. Consequently, in equilibrium, investors willingly pay a hedge premium.

For notational convenience, we adapt this relation into our current generic setting, and thus define the risk premium for realized volatility, y_{t+1} , and an associated implied volatility measure, x_t , as,

$$\pi_t = E_t^P [y_{t+1} - x_t] = E_t^P [y_{t+1}] - E_t^Q [y_{t+1}] = y_{t+1}^* - x_t,$$

where we explicitly adopt the null hypothesis that the forecast, in the absence of a risk premium, or $\pi_t = 0$, is an unbiased estimate of the future volatility, and we denote the conditionally expected realized volatility of the underlying asset, at time t for month $t + 1$, by y_{t+1}^* . Obviously, a non-zero (negative) risk premium induces a, potentially systematic, bias in the implied volatility vis-a-vis the future realized volatility. Although this is widely acknowledged, the literature has not developed a consistent response to the issue, as most authors simply amend the empirical framework through ad hoc assumptions regarding the nature of the risk premium. It is instructive to explore the implications within a few stylized settings.

A simple scenario is obtained if one postulates the existence of a constant (negative) premium, apart from a random idiosyncratic component, so that we have $\pi_t = -\pi_0 - u_t^x$ where u_t^x is a white noise, i.e., serially uncorrelated with $E[u_t^x] = 0$, and $E[(u_t^x)^2] = \sigma_x^2$, and $\pi_0 > 0$. The implied volatility series will then be of the form, $x_t^a = y_{t+1}^* - \pi_t = y_{t+1}^* + \pi_0 + u_t^x$. The implied volatility is thus upward biased relative to the expected volatility and subject to random noise akin to a classical measurement error in the regressor. The latter may reflect inefficiencies and random pricing errors in the derivatives market. Christensen and Prahbala (1998) argue that it is critical to accommodate this type of idiosyncratic error in the predictive regression set-up. It is straightforward to derive the implications of this specification. The basic null hypothesis for the case of no risk premium is,

$$y_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1} = \alpha + \beta y_{t+1}^* + \varepsilon_{t+1}, \text{ and thus } \alpha = 0 \text{ and } \beta = 1.$$

However, under the alternative representation for the risk premium, we have,

$$y_{t+1} = (\alpha - \beta \pi_0) + \beta x_t^a + (\varepsilon_{t+1} - \beta u_t^x).$$

The random error in the regressor will bias the regression slope coefficient downward while the regression intercept will be subject to partially offsetting effects as the measurement error induces an upward bias but the (negative) risk premium pulls in the opposite direction. Specifically, letting $Var[y_t^*] = \sigma_{y^*}^2$, one may readily show that, asymptotically, the regression coefficients become

$$\hat{\beta} = 1 / (1 + \gamma) \quad \text{and} \quad \hat{\alpha} = -\pi_0 + [\gamma / (1 + \gamma)] E[x_t^a],$$

where $\gamma = \sigma_x^2 / \sigma_{y^*}^2$ may be interpreted as a "noise-to-signal ratio" for the regressor. If the average risk premium is zero ($\pi_0 = 0$), the intercept will be positive. In contrast, if there is no idiosyncratic component in the risk premium ($\gamma = 0$), the intercept will be negative while the slope coefficient, more importantly, will be estimated consistently. However, in the general case OLS will be inconsistent. Of course, the remedy for measurement error is to do inference through an instrumental variable estimator. Since implied volatility typically is highly persistent, it is natural to use the lagged implied volatility as an instrument for the current implied volatility. In the current set-up this provides valid inference and Christensen and Prabhala (1998) report that the evidence of bias in implied volatility forecast is eliminated through this procedure. That is, they obtain a slope coefficient close to unity as well as an insignificant intercept, suggesting that the unconditional risk premium is negligible, or $\pi_0 \approx 0$. Nonetheless, various authors deem this explanation incomplete. Poteshman (2000) argue strongly that the "measurement" error in implied volatility is too small to account for the large bias in the original regression and this is furthermore supported by the simulation results in Jorion (1995). Moreover, the above results imply a negligible variance risk premium which is strikingly at odds with extensive recent evidence.

An alternative illustrative scenario, more in line with the extant empirical evidence, is for the risk premium to obey the following affine relationship with expected volatility, where we ignore the small idiosyncratic risk premium component, i.e. $\gamma = 0$, and assume

$$\pi_t = -\pi_0 - \pi_1 y_{t+1}^*, \quad \pi_0, \pi_1 > 0.$$

The implied volatility measure now takes the form $x_t^a = \pi_0 + (1 + \pi_1) y_{t+1}^*$, so that the (negative) risk premium embodied in the implied volatility measure increases with the level of expected volatility. One may readily derive the following relationship between the population values of the univariate predictive regression coefficients and the risk premia coefficients,

$$\alpha = -\pi_0 / (1 + \pi_1) < 0 \quad \text{and} \quad 0 < \beta = 1 / (1 + \pi_1) < 1.$$

Again, the implied volatility forecast is biased, resulting in a negative intercept and a slope coefficient below unity. Notice furthermore that these biased forecasts arise naturally from a rational asset

pricing framework and are unrelated to inefficiencies in the derivatives markets. However, one may still exploit the linear relationship to provide unbiased forecasts of future volatility. Consequently, this is a scenario consistent with the contention in the literature that implied volatilities provide efficient, yet biased forecasts.

Such direct parameterization of the volatility risk premium is tantamount to the imposition of a specific option pricing model. The linear version above is consistent with the popular and tractable affine class of models, see. e.g., Poteshman (2000), Bollerslev & Zhou (2006) and Chernov (2007). In particular, the latter notes that if the expected volatility is an affine function of current volatility, as for one-factor affine stochastic volatility models, then predictive regressions enhanced with an estimate of the current volatility state variable may serve as the basis for consistent tests of the unbiasedness hypothesis.

Although the above formulation of the risk premium produces results that are qualitatively consistent with the empirical evidence, the specific representation is overwhelmingly rejected by the recent literature. One may, of course, continue to search for better tractable option pricing models as a basis for the implied volatility forecast. However, this renders the methodology model dependent, while a main attraction of the BS and model-free implied volatility measures is that they stem from model-free procedures which may be implemented for any given option market without prior specification search and inference. Hence, we continue to focus on the standard implied volatility measures in the sequel. The empirical question is whether these measures, or simple transformation thereof, provide approximately unbiased forecasts for future volatility and whether the associated linear forecasts are efficient compared to natural benchmarks constructed from time series of underlying historical returns.

In order to set the stage for a more general analysis, we finally consider a less restrictive specification of the volatility risk premium, allowing for both classical measurement error and a nonlinear dependence of the risk premium on expected volatility. We thus assume that the volatility risk premium is an unknown function of expected volatility and all other information available at time t , denoted Φ_t , so that we have $\pi_t = -\pi_0^* - u(y_{t+1}^*, \Phi_t)$ with the stipulation that the premium is a stationary process with a finite second moment and $E [u(y_{t+1}^*, \Phi_t)] = 0$. For simplicity, we henceforth suppress

the dependence on the information set and simply write $u_t^* = u(y_{t+1}^*, \Phi_t)$, and we let $Var [u_t^*] = \sigma_*^2$, so it follows from above that $E [\pi_t] = -\pi_0^*$, where $\pi_0^* \geq 0$ will ensure a non-positive unconditional variance risk premium. Moreover, we expect u_t^* to be an increasing function of expected volatility, albeit not constrained to be linear which is attained for the special case, $u_t^* = \pi_1 y_{t+1}^*$. Hence, the implied volatility is now $x_t^a = y_{t+1}^* - \pi_t = y_{t+1}^* + \pi_0^* + u_t^* + u_t^x$, with the corresponding "noise-to-signal ratio" becoming $\gamma = \sigma_x^2 / Var(y_{t+1}^* + u_t^*)$. The predictive regression takes the following form,

$$y_{t+1} = (\alpha - \beta \pi_0^*) + \beta x_t^a + (\varepsilon_{t+1} - \beta u_t^* - \beta u_t^x).$$

Direct computations reveal that the population value of the regression slope becomes,

$$\beta = \beta \left(1 - \frac{Cov(y_{t+1}^* + u_t^*, u_t^*)}{Var(y_{t+1}^* + u_t^*)} \right) \left(\frac{1}{1+\gamma} \right) = \beta \left(\frac{\sigma_{y^*}^2 + Cov(y_{t+1}^*, u_t^*)}{\sigma_{y^*}^2 + \sigma_*^2 + 2 Cov(y_{t+1}^*, u_t^*)} \right) \left(\frac{1}{1+\gamma} \right).$$

It is readily checked that both parentheses are positive and less than unity so the regression coefficient is again downward biased relative to the benchmark of no measurement error and risk premium. Letting $u_t^* = \pi_1 (y_{t+1}^* - E [y_{t+1}^*])$, we are back in the affine risk premium setting, and we have $\beta = [1 / (1 + \pi_1)] [1 / (1 + \gamma)]$, which, of course, constitutes a direct generalization of the two previous scenarios, reflected a classical measurement error and an affine risk premium, respectively.

From an econometric perspective it follows that we should expect a downward bias in the slope coefficient. And this bias does not stem exclusively from a measurement error problem which may be alleviated through instrumental variable estimation. Given that the risk premium is latent, or unobserved, we cannot readily purge the (negative) premium from the implied volatility measure, so it will also contribute to a slope coefficient below unity. Since the risk premium component is positively correlated with expected volatility and this is a highly persistent process, the lagged implied volatility will also tend to be correlated with the current risk premium, so the lagged volatility measure does not constitute a valid instrument. The current approach in the literature is to accept a rational basis for a coefficient value less than unity and largely ignore the classical measurement error, i.e., $\gamma \approx 0$. As argued above, this may be a reasonable first order approximation in this context. Nonetheless, absent an exact affine functional form for the risk premium, the error term will still contain a component of the risk premium and the implied volatility regressor will be correlated with the error term, implying

inconsistent estimates of the relationship between the realized and implied volatility. Moreover, it is not evident how to obtain effective instruments for this general case. The response in the literature has been largely pragmatic. It is argued that the affine specification may serve as an acceptable approximation. Thus, the predictive and encompassing regressions continue to be meaningful as efficiency tests and the linear relation between implied and realized volatility may still serve as a basis for volatility forecasting. At the same time, volatility forecast residuals are invariably conditionally heteroskedastic and time-variation in the risk premium will introduce serial correlation in the error term as well so inference is typically conducted through (HAC) standard errors which are robust to such features.

We now turn to another issue largely neglected in the literature that is potentially even more troublesome for our ability to conduct meaningful inference regarding the questions of interests within these predictive volatility regressions.

3.3 Long memory in volatility

An extensive literature has documented strong empirical evidence of long range dependence in return volatility, see, e.g., Ding, Granger & Engle (1993), Baillie, Bollerslev & Mikkelsen (1996), Comte & Renault (1998), Ray & Tsay (2000), Andersen, Bollerslev, Diebold & Ebens (2001), Andersen, Bollerslev, Diebold & Labys (2001) Andersen et al. (2003), Wright (2002), Hurvich & Ray (2003) and Arteche (2004). From a theoretical perspective, Comte & Renault (1998) establish a link between the long memory in integrated, realized and implied volatility. Hence, one would generally expect the series to possess the identical degree of fractional integration, so they may all be covariance stationary $I(d)$ processes with $0 < d < 1/2$. This is also in line with the more limited empirical evidence regarding long memory features in implied volatility, as established in some of the above studies. Since the risk premium is linked to the level of volatility, it is almost inevitable that these features also will be present in premium component of the implied volatility measure. If the volatility process is driven strictly by one factor and the functional relationship is affine, such that we have, within the general volatility risk premium specification above, that $u_t^* = \pi_1 (y_{t+1}^* - E [y_{t+1}^*])$, then the risk premium component is

entirely absorbed in the implied volatility measure and the linear predictive regression is well specified. Consequently, the regression will consistently estimate the relationship between realized and implied volatility, even if the slope coefficient is below unity.

However, the above represents an empirically implausible knife-edge scenario. The evidence strongly suggests a multi-factor diffusive representation for volatility, supplemented by a jump factor. The presence of jump risk premia in the implied volatility measures is also stressed in several empirical studies, including Pan (2002) and Todorov (2007). The jump factor is often found to be decidedly less persistent than the diffusive volatility factors and the associated premia appear economically large but correspondingly less long lived than the other dynamic components of the premium. As such, the validity of a one-factor affine model for the volatility risk premium seems, at best, dubious. The implication is that the error term in the general predictive regression, say, $e_{t+1} = \varepsilon_{t+1} - \beta u_t^* - \beta u_t^x$, will be nonlinearly related to the implied volatility regressor through the term, u_t^* , and likely display rich dynamics of a less persistent nature than associated with the degree of fractional integration of the volatility measures. The presence of a classical measurement error in the implied measure will simply reinforce this feature.

This line of reasoning has inspired Bandi & Perron (2004) and Christensen & Nielsen (2005) to provide a fundamental reinterpretation of the conventional predictive regressions. The regressions involve two fractionally integrated series of the identical order, so it is natural to invoke a fractional cointegration framework where the relation between implied and realized volatility is viewed as a long-run equilibrium. From this perspective, y_{t+1} and x_t are both $I(d)$ processes with a cointegrating vector of $(1, -\beta)$ with $0 < \beta < 1$, and the regression takes the form,

$$y_{t+1} = \alpha + \beta x_t + e_{t+1},$$

where e_{t+1} is $I(d_e)$, with $0 \leq d_e < d < 1/2$, and x_t is correlated with e_{t+1} . The correlation between the regressor and error term generally results in inconsistent inference regarding the true underlying long run relationship between realized and implied volatility. These issues render direct interpretation of the empirical findings based on the predictive regression set-up difficult and the diverse findings

may in part be attributable to misspecification within the conventional framework.

One solution to this problem is to apply the frequency-domain (or narrow-band) least squares (FDLS) estimator of Robinson (1994), provided that the regressor and regressand are, indeed, linked by a (fractional) cointegrating relation. Intuitively, this works by only running the regression, in the frequency domain, for low frequencies where the coherence among the fractionally integrated variables dominates the confounding effects of the less persistent features inherent in the error term. The conclusion is that consistent inference is feasible but it will be less efficient as the procedure only exploits information obtained at the lower frequency so the effective sample size is reduced. This is a relatively important drawback as the highest non-overlapping sampling frequency typically is monthly, and reliable data series for implied volatility only are available, for a few series, from around 1990, thus allowing for less than 200 observations. Below, we describe both methods for conducting inference regarding the long run relationship through the frequency domain based estimation techniques and for conducting inference regarding the short run relationship through the usual encompassing regressions utilizing all the available data.

3.4 Estimating the Degree of Fractional Integration

Our first step in assessing the nature of the predictive volatility regressions is to explore the evidence for a common degree of integration across the realized volatility and the associated volatility forecast or implied volatility series. The short run dynamics of the various series may be quite different and not readily characterized in a tractable parametric framework. As such, a natural choice is to apply the so-called narrow-band semiparametric estimators which, asymptotically, are valid estimators of the degree of long-range dependence and robust to the specification of short memory and noise components. Although this is, in principle, straightforward, the recent literature has documented a variety of practical problems with the performance of the standard estimation procedures for the long memory parameter, d , within moderately sized samples. In response, a number of useful refinements have been developed over the last few years and we exploit some of these insights in our inference.

The traditional approaches exploit the intuition that the short memory, $I(0)$, processes will be dominated by long memory components at the lowest frequencies, so estimation is performed in a "narrow band" of frequencies just above zero. The log-periodogram regression method of Geweke and Porter-Hudak (1983) and the local Whittle or Gaussian semiparametric estimator associated with Künsch (1987) and Robinson (1995a) fall in this class. Unfortunately, the estimates obtained from either procedure are subject to serious finite sample biases in many cases. There are two main effects, pulling in opposite directions. First, the presence of a noise component, with no or little serial correlation, in the regressor will tend to induce a negative bias in the long memory estimates. Although this effects disappears asymptotically, it is a potential concern for standard realized volatility measures constructed, say, as the cumulative sum of daily squared returns. The squared return comprise both the latent volatility of interest and the corresponding squared return innovation, with the latter typically modelled as an i.i.d. mean zero series with fat-tails. In this context, the squared innovations are pure noise but cannot be separated from the (fractionally integrated) volatility process. As such, they weaken the measured serial correlation and induces the downward bias. Wright (2002) concludes from a large simulation study that the bias is sufficiently strong that researchers should avoid using squared returns in the semiparametric estimation of long memory volatility dependencies. Instead, emphasis should be put on the estimation in the transformed logarithmic or absolute return series which generally retain the underlying long memory dependence. Confirmatory simulation evidence is provided by, e.g., Hurvich and Ray (2003). Of course, one way to alleviate the problem is to use intraday squared returns as the basis for construction of the realized volatility measure, as this dramatically reduces the idiosyncratic noise component in the measure, see, e.g., Andersen and Bollerslev (1997), Andersen (2000), and Bollerslev and Wright (2000). Nonetheless, the problem remains, in part because monthly forecasts cover extended non-trading hours such as overnight periods, weekends and Holidays. Although these periods displays much lower per hour volatility than the trading hours, the close-to-open period from one trading day to the next is a significant source of overall return variation while the high-frequency strategy only provides a remedy for the within trading hours component. Likewise, the existence of components in the volatility process with purely transient effects such as, potentially, jumps and news arrivals will have a similar impact on the finite sample inference.

Consequently, the fact that the observed realized volatility series is distinct from the ideal volatility process which we seek to draw inference about, remains a concern. Technically, such series are denoted perturbed fractional processes, and we briefly describe our adaptation of one procedure designed for estimation in this scenario below.

The second factor inducing a potentially serious finite-sample bias in estimation of the degree of fractional integration is the presence of persistent short run components, along with the long memory process, in the volatility series. In this case, the short memory dynamics will tend to inflate the serial correlation, even at low frequencies, and thus induce an upward bias in the estimates, see, e.g., Andrews and Guggenberger (2003) and Andrews and Sun (2004). Both Hurvich and Ray (2003), Hurvich, Moulines and Soulier (2005), and Bandi and Perron (2007) provide simulation evidence that this also can be a concern within the current setting. However, the former find the first factor, inducing a downward bias, more prominent and empirically relevant. In fact, one may, cautiously, conjecture that an initial transformation of the realized volatility series, constructed from high-frequency data, will suffice to reduce the magnitude of the first effect to the point where the two offsetting effects largely balance each other out. Nonetheless, it seems proper to conduct inference in a manner that serves to accommodate either of these features.

In light of the discussion above, we rely on the so-called Local Polynomial Whittle (LPW) estimator of Frederiksen (2006) as well as the standard Local Whittle (LW) estimator for inference regarding the memory parameter, d . The LPW estimator directly extends the LW approach by approximating the impact of the short memory components on the local Whittle likelihood at the lowest frequencies by an even polynomial rather than a constant. The specification is related to, but different from, those of Hurvich and Ray (2003), Sun and Phillips (2003), and Hurvich, Moulines and Soulier (2005). The advantages of the LPW estimator is that, under appropriate regularity conditions, a faster rate of convergence to the true long memory parameter is obtained independently of the distributional characteristics of the perturbing noise processes and the specific nature of the short memory components. The cost is an additional inflation of the asymptotic variance. However, most importantly, these types of estimators have been found to reduce the finite sample bias in estimation by a signifi-

cant amount, justifying the added complication and asymptotic variance. For brevity, we relegate the details regarding the LPW procedure to the appendix.

3.5 Fractionally cointegrated predictive and encompassing regressions

We now return to the question of inference within predictive and encompassing regressions in the presence of fractional cointegration between the involved volatility measures and forecasts. This is a relevant scenario given the evidence of long memory in all the volatility series and the existence of a theoretical rationale for the degree of integration to be identical across the series. Hence, we again consider the general predictive and encompassing regression setting introduced in Section 3.1 where $y_{t+1} = \alpha + \beta'x_t + e_{t+1}$, but we now stipulate that e_{t+1} is $I(d_e)$, with $0 \leq d_e < d < 1/2$, while y_{t+1} and all volatility forecast series in x_t are $I(d)$ series which individually may or may not be correlated with e_{t+1} . This is a formal representation of fractional cointegration, where there exists a linear combination of two integrated series which produces a (residual) series with a lower degree of integration.

We first consider the univariate predictive regression setting, where x_t is a scalar. If this forecast is an implied volatility measure, be it Black Scholes or model-free, we have previously established that OLS regressions likely are inconsistent in the presence of a significant volatility risk premium. Importantly, the same argument is not true for time series based predictors which, by construction, represent forecasts under the actual probability measure and thus embody no risk premium. As a consequence, we expect the residual error to have short memory in that case, i.e., be an $I(0)$ process. We are then mostly left with a potential concern regarding the presence of a type of the classical measurement error in the volatility forecast. However, if the forecast model is well specified we would expect only a relatively small degree of misspecification of the expected volatility forecasts relative to the time series variation in the level of expected volatility over time, implying a low value for the noise-to-signal ratio of the regressor, γ . As such, the correlation between regressor and error term may not be a major concern and OLS based inference on the regression coefficients would indeed be valid as long as the common long memory parameter remains below one half so that all series are covariance stationary.

In light of these observations, it is of interest to explore the standard regression based tests for short run unbiasedness of the time series based volatility forecasts, and we do so in the empirical section below. Moreover, to provide some sense of whether there are indications of long range dependent serial correlation in the residuals of predictive regressions using implied volatility predictors, we report results for these as well. In either case, as a diagnostic, we provide estimates for the degree of long memory in the time series of regression residuals. If the risk premium induces the type of complications discussed above, it should manifest itself in distinctly different persistence properties of the residual series, depending on whether the forecast variable is based on a time series model or observed option prices.

On the other hand, if there is long range dependence in the residuals and we suspect correlation between the implied volatility regressor and error term, we are squarely back to the Christensen and Nielsen (2006) and Bandi and Perron (2007) reinterpretation of the predictive regressions. They view the setting as a long term relationship and exploit the fact that the spectral densities of the more persistent variables will dominate the spectral density of the error term at frequencies near zero. Hence, one runs a narrow band regression in the frequency domain, using a set of Fourier frequencies, m , which grows more slowly than the overall number of observations, n . Formally, the frequency-domain least squares (FDLS) estimator of β in (??) becomes,

$$\hat{\beta}_m = \hat{F}_{xx}^{-1}(1, m) \hat{F}_{xy}(1, m)$$

and

$$\alpha_m = \bar{y} - \hat{\beta}_m^t \bar{x},$$

where \hat{F}_{ab} is an estimate of the spectral density matrix, and \bar{y} and \bar{x} denote sample averages. The estimate of the requisite spectral density matrix is obtained through standard procedures and we defer the details to an appendix. Typically, the bandwidth is chosen so that $m = \lfloor n^\lambda \rfloor$, where λ is on the order of 0.3 to 0.8. Contrary to OLS the FDLS allows the regressor and error terms to share the same short- and medium-run dynamics, and even allows long memory in the relation between the two. This makes the FDLS widely applicable and even in nonstationary fractional cointegration it may

entail faster rates of convergence than OLS although OLS is consistent (see Robinson & Marinucci (2001), Robinson & Marinucci (2003)).

Under fully general circumstances, the asymptotic distribution for the parameter estimates is unknown, but Christensen & Nielsen (2005) have established asymptotic normality under specific parameter restrictions for the univariate regressor case. In contrast, Bandi & Perron (2004) avoid such conditions by applying subsampling to derive inference. Here we apply the bootstrap methodology of Frederiksen, Nielsen & Shimotsu (2006) to obtain the confidence intervals of the regression parameters. The advantage is that no block length or subsample size has to be chosen and that it is easy to implement by just resampling from the normalized discrete Fourier transform of the frequency-domain least squares (FDLS) residuals, see the appendix for details.

The FDLS approach may readily be applied for the encompassing regressions involving multiple volatility forecast regressors. The issue of whether the standard (time domain) regressions are well specified in this case is largely empirical issue. It is possible that the inclusion of a well specified time series forecast of realized volatility among the regressors will serve to absorb the dominant long memory features of the realized volatility regressand and thus purge the error term for significant long range dependencies, even in the presence of additional implied volatility regressors. In that scenario, standard inference based on a multiple regression approach should work well in practice. As for the univariate predictive regression scenario, it is pertinent to provide auxiliary estimates of the long memory parameter of the regression residual series to gauge the plausibility of the underlying assumptions. As we indicate in the empirical work below, this scenario indeed appears empirically relevant and thus points towards a strategy for circumventing the inference problems highlighted by Christensen and Nielsen (2006) and Bandi and Perron (2007). It may hence, after all, be sensible to test for short run unbiasedness and efficiency of specific forecast variables within the standard encompassing regression set-up. Given the obvious advantages of directly addressing the short-run forecast questions raised in the literature and allowing for utilization of data across all frequencies, this strategy has some attractive properties relative to the narrow band FDLS approach.

4 The data

Our empirical analysis of the information content of volatility forecasts is based on monthly implied volatilities and realized volatilities for the S&P 500 from January 1990 to December 2002. The starting point coincides with the availability of the VIX index. Our sampling period starts after the October 1987 stock market crash. Casual analysis confirms that the inclusion of this earlier data leads to fragile results in terms of the statistical inference. There is also substantial evidence that the crash led to a structural break in the time-series properties of option derived volatility measures. Other studies have used a variety of different sample periods, but none appear to exploit as long a monthly post-crash series as we do.

Following prior research we seek to ensure that our option data have minimal measurement error. The Black-Scholes implied volatilities for at-the-money call options are computed on the basis of option prices from the Berkeley Options Data Base, supplemented by hand collected option prices from the Wall Street Journal. By convention these options expire on the Friday immediately before the third Saturday of each month and we calculate implied volatilities for options with one month to expiration. The dividend yield is taken into account as described in Hull (1997), where dividend yield data is obtained from Datastream. The time series of monthly implied volatilities are non-overlapping, i.e. every observation is based on its own unique window. This is important since previous work has shown that the use of overlapping data leads to problems in the statistical analysis. In particular, the use of overlapping data in the implied volatility series is thought to favor time-series forecasts (Christensen & Prabhala (1998), Fleming (1998)). For the risk-neutral model-free implied volatilities we use the VIX index provided by the Chicago Board of Options Exchange (CBOE). The VIX implied volatilities, computed by the CBOE based on the approach of Britten-Jones & Neuberger (2000), are available from January 1990 onwards.

Our realized volatilities are based on high-frequency futures prices traded on the S&P 500 index over the period from the first trading day of January 1988 to the last trading day of December 2002. The data are from the Chicago Mercantile Exchange (CME). Realized volatilities over a month are

calculated by summing the daily realized volatilities of the trading days within a particular month. In turn, these daily realized volatilities are based on the summation of five-minute squared intraday returns and the squared overnight return. The treatment of the overnight return is somewhat of an open question in the literature, but seems without consequences in the current application. Our results are qualitatively unchanged when considering other methods of incorporating the overnight return, see, e.g., Hansen & Lunde (2005) for a comprehensive discussion. We scale monthly realized volatilities to make them comparable with the implied volatilities. Let n be the particular number of trading days in the window under consideration, then the annualized realized volatility in month t is given by

$$RV_t^{monthly} = \sqrt{\frac{252}{n} \sum_{i=1}^n RV_i^{daily}}.$$

We are careful to match the exact period covered by each of the implied volatilities in the construction of our realized volatility series and time-series forecasts. Forecasts corresponding to the monthly periods covered by the implied volatilities are constructed by forecasting the realized variance for all trading days in the period via an ARFIMA model, as described in Section 2.1, and summing. Importantly, these forecasts are out-of-sample with the AR and MA parameters determined by means of the BIC information criterion using only prior observed data. In line with Andersen, Bollerslev, Diebold & Ebens (2001) we find the logarithmic daily realized volatilities to be approximately normally distributed. Consequently, we follow Andersen et al. (2003) in constructing time series estimates for the log realized volatility and confirm that these produce well calibrated forecasts relative to alternative models estimated directly on the realized volatility. When converting a forecast of logarithmic volatility back to volatility, measured as the standard deviation, the lognormal assumption on daily realized volatility allows for a simple correction for the effect of Jensen’s inequality on the forecast.

In sum, our basic data consists of monthly time-series of realized volatility, Black-Scholes implied volatility, and VIX model-free implied volatility stated in standard deviation form. In addition, we have corresponding ARFIMA volatility and log volatility time series forecasts. All volatility measures are annualized (assuming 252 days per year) and reflect volatility over the identical monthly periods.

Each series contain 155 non-overlapping observations. Inspired by the recommendations of Wright (2002), we avoid making inference directly on the realized variance measures. Instead, we analyze the realized volatility measures in standard deviation form and complement these results with findings for the log transformed series, as also suggested by Christensen & Prabhala (1998) and Jiang & Tian (2005*b*). Of course, the two specifications are not mutually consistent, if one is correct the other is necessarily misspecified. However, the bias in the expected volatility measure induced by the nonlinear transformation is likely to be small, even if the distributional properties are altered significantly. The results of Patton (2006) also confirm that comparisons based on different transformation of volatility tend to produce similar findings as long as the realized volatility proxy is satisfactory, which he finds to be the case for sufficiently high-frequency intraday return based measures. Thus, while the theoretical relationship between the expected values of the alternative volatility measures are not altered much by the transformations, some representations are more conducive to reliable inference. For example, the approximate Gaussianity of the log volatility series alleviates the impact of outliers and the right-skew in the volatility series, rendering finite sample inference more robust.

The descriptive statistics for the three volatility measures and our time-series forecasts, provided in Table 1, illustrate the general comments above. Both the implied volatilities and the realized volatility display heavy tails and positive skewness, while the logarithmic transformation dramatically reduces the positive skewness and eliminates the excess kurtosis relative to a Gaussian benchmark. We also note that the means of both implied volatility measures are larger than the average realized volatility, possibly reflecting the presence of a risk premium.

5 The information content of realized volatility forecasts

In this section we examine the relationship between time-series forecasts of future realized volatility, Black-Scholes implied volatility, model-free implied volatility, and realized volatility⁴. We discuss both the short- and long-run properties of the different forecasts of future realized volatility. We refrain from

⁴In additional regressions we included the implied volatility that is provided by the CBOE based on the old VIX methodology instead of the Black-Scholes implied volatility. Results are qualitatively the same and therefore not reported.

instrumental variable regressions as classical measurement error, as argued, is only a minor concern for the realized volatility forecasts while, at the same time, a time-varying risk premium correlated with the level of volatility in the implied volatility measures renders the usual choice of instrument, namely lagged volatility measures, unsuitable.

We are not the first to study model-free implied volatility as an expectation of future realized volatility. For example, Jiang & Tian (2005*b*) compare the information content of a model-free implied volatility with Black-Scholes implied volatility and a measure based on historical data. They conclude that model-free implied volatility completely subsumes the information in the other measures. Surprisingly, we come to the opposite conclusion. We find that the VIX computed by the CBOE performs poorly as a forecast of future realized volatility. Part of this might be explained by the fact that Jiang & Tian (2005*b*) construct their own model-free implied volatility measure over a shorter sample period, only partly overlapping with ours. Although they base the construction on the theoretical procedure of Britten-Jones & Neuberger (2000), their measure has different properties than the official VIX. And perhaps even more importantly, they make no attempt to construct efficient time series forecasts of volatility based on historical data.

5.1 Information efficiency in the short run

We first consider standard univariate predictive regressions. Table 2 displays results from OLS predictive regressions for volatility measured in standard deviation form as well as its logarithmic transformation in the first three lines of Panels A and B. The results are so similar in terms of statistical implications that we do not discuss them separately. In all univariate regressions we test the null hypothesis that the forecast under consideration is an unbiased estimate of future realized volatility, i.e. $\beta = 1$ and $\alpha = 0$. The reported p-values for the Wald statistic corresponding to this hypothesis suggest that only time-series forecasts can reasonably be assumed to provide unbiased estimates of future volatility⁵. On the other hand, across all regressions the slope coefficient is significantly differ-

⁵Standard errors of the parameter estimators in the short-run regressions are calculated following the heteroscedasticity and autocorrelation robust approach of Newey & West (1987).

ent from zero, verifying that all three forecast measures contain information about future volatility. Regression R^2 s are fairly high in general, and not surprisingly, somewhat larger for the log specification. Model-free implied volatility is associated with the lowest R^2 , while the ARFIMA forecasts have the highest R^2 , albeit not significantly different from the one associated with Black-Scholes implied volatility. The Durbin-Watson test statistics are very close to two for the regressions that involving the ARFIMA forecast as regressor, suggesting that the autocorrelation in the associated residuals is moderate. The p-values for the Diebold-Mariano test for equal forecast accuracy of two competing forecasts (with a squared loss function), suggest that our time-series forecasts are superior to model-free implied volatility forecasts. The table also reports the estimated long-memory parameter in the residuals and the p-value for the null hypothesis $d = 0$. It is also evident that tests based on a linear regression framework including either of the implied volatilities may be misspecified as the regression residuals display substantial long-range dependencies. Robinson (1994) shows that conventional OLS estimators are inconsistent in the stationary case when the errors are fractionally integrated. Interestingly, when our ARFIMA based forecasts serve as regressors the residuals loose all indications of long run dependence, seemingly implying that the OLS regressions are well-specified. If the implied volatility predictive regressions are misspecified, we can no longer examine the short-run unbiasedness of these forecasts in this univariate framework, but we can still test the hypothesis of long-run unbiasedness in a fractional cointegration framework, and we do so later on.

Consider next the results from the encompassing regressions in Table 2. These regressions can be interpreted as a horse race between volatility forecasts. Since we focus on the information content of the time series forecasts, we present regression results only for specifications that include the ARFIMA forecast among the regressors. Even though all these regressions include at least one of the implied volatility forecasts, it is striking that we now find no evidence of long memory or serial correlation in the residuals, in sharp contrast to the results for the predictive regressions involving the implied volatility forecasts alone. In other words, the inclusion of the time series forecasts have restored the empirical properties of the residuals which are compatible with valid inference in the OLS setting. We use the Wald statistic to test whether the coefficient on the ARFIMA forecasts is equal to one,

while all other coefficients are zero (including the intercept), i.e. $\beta^{RVF} = 1, \alpha = \beta^{BS} = \beta^{MF} = 0$. When regressing future realized volatility on our time series forecasts and BS implied volatility, the significant slope coefficients on each regressor imply that both carry substantial information about future volatility. In the standard deviation form regression the coefficients are statistically equal, while our forecasts do significantly better in the log specification. In both regressions we reject the null hypothesis of $\beta^{RVF} = 1, \alpha = \beta^{BS} = 0$. When regressing realized volatility on our time series forecasts and model-free implied volatility, the picture changes. In both the log and standard deviation form our forecasts do significantly better. While the Wald test certainly does not support the null hypothesis $\beta^{RVF} = 1, \alpha = \beta^{MF} = 0$, it is less strongly rejected than the analog hypothesis with Black-Scholes implied volatility as the competitor. These results imply that the time-series forecasts and implied volatility may have information in common, but neither fully subsumes the information content of the other.

It is somewhat surprising that the VIX implied volatility performs so relatively poorly in the encompassing regressions. Jiang & Tian (2005*b*) find that model-free implied volatility completely subsumes the information in historical volatility and BS volatility. As mentioned, this may stem from their use of a more sophisticated model-free implied volatility measure although this is hard to judge given their shorter sample period. In fact, Jiang & Tian (2005*a*) suggest that the VIX measure is systematically biased because of inappropriate approximation methods. However, in the presence of risk premia related to volatility and jumps, we expect model-free implied volatility to be a bit compromised as a pure forecast of future realized volatility. The model-free measure estimates the expected integrated variance under the risk neutral measure using information extracted from a large cross-section of option prices. This may well enhance the extent to which the risk premia are incorporated in the model-free implied volatility relative to BS implied volatility. For example, large premia associated with deep out-of-the-money put options are often interpreted as insurance against a large downward move, and an implied volatility derived from such options are more likely to reflect the jump risk premia than otherwise identical at-the-money options.

To summarize, unlike many previous studies our OLS results show that forecasts based on time-

series modeling of historical data contain at least as much information for short-run prediction of future realized volatility as implied volatilities. An important result is that these forecasts are unbiased while implied volatilities provide biased forecasts of future volatility over short horizons. The evidence suggests that the time-series forecasts explain the variation in future realized volatility better than Black-Scholes and model-free implied volatility. Of the two implied volatility measures the Black-Scholes implied volatility is clearly a more efficient forecast of future volatility. Model-free VIX implied volatility is neither an efficient nor unbiased forecast of future volatility. This measure performs poorly as a forecast of future realized volatility, presumably because it reflects information on risk premia for volatility and jumps to a larger extent than Black-Scholes implied volatility.

Since Black-Scholes implied volatility and time-series forecasts seem to incorporate different information, it is natural to consider a weighted forecast method.

5.2 Combined forecasts

The idea that combining different point forecasts of the same variable can improve upon the forecasting ability of any single forecast has gained widespread acceptance since the seminal paper of Granger & Bates (1969). While the theoretical rationale for this approach has not been fully developed, it continues to meet empirical success, as illustrated, e.g., in recent papers by Stock and Watson (1999, 2003, 2004); for additional discussion, see the recent survey by Timmerman (2006).

We follow Bates and Granger (1969) and consider the case of two competing point forecasts $fc_{1,t}$ and $fc_{2,t}$, of the quantity y_t . The forecast errors are

$$e_{i,t} = y_t - fc_{i,t}, \quad i = 1, 2$$

with corresponding forecast variances σ_1^2 , σ_2^2 and covariance σ_{12} . The combined forecast is the weighted average

$$fc_t = w \cdot fc_{1,t} + (1 - w) \cdot fc_{2,t}.$$

The variance of the forecast error of the combined forecast is minimized by setting

$$w = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}.$$

These weights can be found by means of a regression of y_t on the individual forecasts, in which the intercept is zero and the coefficients are restricted to sum to one. The forecast error variance of the combined forecast is no greater than the minimum of the two individual forecast error variances. Bates and Granger (1969) suggest that it might be beneficial to neglect the covariance term (so that the weights lie within $(0, 1)$) and use

$$w = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}},$$

which assigns the forecasting weights based on the precision of the individual forecasts.

We implement the combined forecast by estimating weights using a rolling window of two years of realized volatility and forecasts, the weights are then used to construct a combined forecast for the following month. Forecasts are evaluated by means of the mean squared forecasting error (*MSFE*) and the R^2 of the regression of the separate forecasts on realized volatility. In all cases modest, but seemingly significant, gains in terms of both *MSFE* and the R^2 are feasible using these simple weighting schemes (see Table 3). The p-values for the Diebold-Mariano test for equal forecast accuracy of the combined forecasts versus the individual component forecasts, suggest that the combined forecasts have significantly different predictive accuracy from the underlying forecasts at reasonable levels of significance.

This preliminary analysis suggests that a combined forecast variable, based on simple weighting schemes taking into account past forecast errors, is able to raise the predictive ability relative to the individual forecasts. This suggests the possibility of further gains by means of more sophisticated forecast combinations. The variability of the forecast weights suggests that the relative importance of time-series forecasts and implied volatilities in explaining future volatility is time-varying. In the presence of a dynamic variance risk premium this finding is not surprising. When this risk premium is relatively small the difference between the risk-neutral expected volatility (in the form of option implied

volatility) and the expected volatility under the physical measure are also relatively small. In this case it would not be surprising that option implied volatilities outperform time-series forecasts because of their nature as market determined expectations. When the variance risk premium is relatively large, the wedge between expectations under the different measures is larger and the time-series forecasts contain important information on the expected volatility under the physical measure.

5.3 Information efficiency in the long run

As argued earlier, it is likely that a time-varying volatility risk premium as well as measurement errors induce dependence between the regressor and the residuals and thus cloud the inference regarding the efficiency of option markets for predicting realized volatility. In particular, the risk premium would naturally induce a lower order of fractional integration in the errors than in the regressor, which is in line with our empirical evidence. To explore whether the longer run components of the implied volatility measures provide improved forecasts vis-a-vis the low frequency movements in the realized volatility, we turn to the narrow band least squares spectral methods to estimate the long-run coefficients consistently, even in the presence of stationary residuals that are correlated with the regressor.

Table 4 depicts the memory properties of the three predictors and the realized volatility of the SP500 index. The first part of the table shows the NLW estimates of the fractional integration parameter d for each series for the bandwidth parameter $m = \lceil n^{0.8} \rceil$, where $\lceil x \rceil$ denotes the largest integer smaller than or equal to x . This choice of bandwidth is motivated by the evidence in Frederiksen (2006). However, we first focus on the bottom of the table where standard augmented Dickey-Fuller (ADF) unit root test statistics are provided. It is evident that there is no unit root in the volatility series. Since it is of interest to test whether the series are stationary ($d < 0.5$), the middle part of the table depict the variance ratio test of Nielsen (2006b) and Nielsen (2006a) where we test the null of $d = 0.5$ against $d < 0.5$.

The long-memory estimates are all around 0.45 indicating in stationary volatility series, which is

in agreement with the ADF tests ⁶ and the empirical literature. However, from the standard errors and the variance ratio test,⁷ we cannot reject that nonstationarity is a possibility. Nonetheless, with the full set of point estimates all falling close to $d = 0.45$, which is consistent with existing evidence from the literature, it is reasonable to assume that we are within a stationary fractional cointegration setting in which the long-run unbiasedness hypothesis can be tested formally

Table 5 depicts the results of the univariate fractional cointegration regressions. The estimation of β is based on four different bandwidths, $m = \lceil n^\lambda \rceil$ for $\lambda = \{0.8, 0.6, 0.4, 0.2\}$. Following Robinson & Marinucci (2003), Bandi & Perron (2004) and Christensen & Nielsen (2005), we place considerable weight on the estimates based on low values of m as we thus focus on the (very) low frequencies that avoid contamination from less persistent residual dynamics arising from a volatility risk premium and even measurement errors. We find Black-Scholes implied volatility to be a biased forecast of the realized S&P 500 volatility since even the 99% confidence intervals do not contain $\hat{\alpha} = 0$ and $\hat{\beta} = 1$. In contrast, for model-free implied volatility the cointegration coefficient is very close to unity while the negative α 's and the dynamic dependence of the residuals ($\hat{d}_\epsilon > 0.1$) indicate a non-negligible impact from a risk premium, as is likely the case for Black-Scholes implied volatility. These results complement findings of Bandi & Perron (2004) and Christensen & Nielsen (2005) who lean towards the conclusion that the implied volatility measures provide long-run efficient forecasts. However, this interpretation needs to be judged against our novel finding, reported in the last panel of the table, that *ARFIMA* time series forecasts of realized volatility are near unbiased and, in contrast to the implied volatility regressions, have a (insignificantly different from) zero intercept, $\hat{\alpha}$, for all m . In retrospect, this result is perhaps to be expected as the long-memory feature of the *ARFIMA* time series forecasts are constructed to reflect the historical long-run properties of realized S&P 500 volatility. The mild evidence of an ARFIMA slope coefficient in excess of unity may potentially reflect the impact of a

⁶As an alternative Phillips & Perron (1988) tests were carried out with same conclusion.

⁷The middle part of the table depicts the variance ratio test of Nielsen (2006b) and Nielsen (2006a) for the null of $d = 0.5$ against $d < 0.5$. The version of the variance ratio test we employ here is

$$\rho = n^{0.2} \frac{\sum_{t=1}^n \sigma_{t,T}^2}{\sum_{t=1}^n (1-L)^{-0.1} \sigma_{t,T}^2} \rightarrow_d \frac{\int_0^1 W_{-0.5}(s)^2 ds}{\int_0^1 W_{-0.4}(s)^2 ds},$$

where W_d is the demeaned type II fractional standard Brownian motion of order d . The critical values are calculated from 20,000 Monte Carlo simulations.

leverage effect which tends to bias in the slope coefficient upward relative to unity, see, e.g., the related discussion in Bandi and Perron (2007). Nonetheless, the time series forecasts appear to provide better guidance for the long run behavior of realized volatility than any of the implied measures, thus casting some doubt on the relative efficacy of implied volatility measures as volatility forecasts.

Turning to the new multiple fractional cointegration estimates in Table 6 we find some very intriguing results, further collaborating the tentative interpretation provided above. For the three smallest bandwidths, we cannot reject the hypothesis, $\hat{\alpha} = 0$, $\hat{\beta}_{m,BSIV} = 0$ and $\hat{\beta}_{m,RV} = 1$ or $\hat{\alpha} = 0$, $\hat{\beta}_{m,MFIV} = 0$ and $\hat{\beta}_{m,RV} = 1$, suggesting that, in the long run, the realized volatility forecast is informationally efficient relative to both the Black-Scholes and model-free implied volatilities. Since the correlations (coherences) between the residuals of the encompassing regressions, and the residuals obtained from the univariate regressions involving the two regressors from the encompassing regressions, in general are small, we are confident in our results.⁸ We also ran regressions based on volatility expressed in terms of the variance and the associated log-transform, and the results are similar to those shown.

In summary, the results add a new dimension to the discussion on volatility forecasting and option market efficiency. Our analysis does find evidence consistent with a stable long-run equilibrium relation exists between the S&P 500 index volatility and the implied volatilities, but the impact of a variance risk premium seems to complicate the inference regarding option market efficiency in terms of long horizon volatility forecasts. More importantly, our results also reveal that carefully constructed time series predictions of future realized volatility are, at a minimum, equally informative. Moreover, the information embedded in our ARFIMA forecasts contribute significant beyond what may be inferred solely from option-implied volatility, thus indicating that the latter does not subsume the available time series information and cannot, in practice, be seen as fully efficient volatility forecasts.

⁸This result is supported by Monte Carlo studies where we used the empirical estimates for simulation.

6 Conclusion

We examine the properties of time-series forecasts and implied volatilities as expectations of future realized volatility of the *S&P* 500 index. In doing so we contribute to the literature by employing the most recent insights into volatility measurement and modeling and explicitly incorporating the possible role of a variance risk premium.

Using volatility measures based on high frequency data, we find that forecasts based on long memory time-series models of historical realized volatility provide unbiased estimates of future realized volatility. Option implied volatilities appear to reflect transient risk premium dynamics as well as rational volatility forecasts which render them less reliable predictors of future realized volatility than one may perhaps expect given the broad information set processed by market participants. In particular, model-free implied volatility, as given by the VIX index, is less relevant as a forecast for future volatility than at-the money Black-Scholes implied volatility or time-series forecasts. In encompassing predictive regressions, the contribution of model-free implied volatility in explaining future realized volatility is negligible. A likely explanation is that model-free implied volatility is highly sensitive to the volatility risk premia.

The existence of a fractional cointegrating relationship between future realized volatility and the forecasts suggests that predictive regressions have to be considered with care. Although the evidence suggests that the use of implied volatility as a regressor can lead to a break down of statistical inference in the conventional regression setup, a long memory analysis of the relationship between time-series forecasts, implied and future realized volatility confirms that all the forecast measures considered are unbiased in the long run. It is possible that this reflects long-run stationary properties of the variance risk premium process.

Since the results in this paper suggest that implied volatility and time-series forecasts contain independent information about future volatility, a promising and important avenue for future research lies in combining time-series forecasts with implied volatility to use the information in both jointly. Since time-series forecast are based on historical price information only and provide forecasts of the

observed volatility process, while implied volatility can be interpreted as expectations under the risk-neutral measure that presumably incorporate the whole of the investors information set, it seems natural to use both when predicting future volatility. An initial analysis confirms this intuition.

A Bayesian approach to weighting volatility forecasts might be worth investigating. In such a framework the weights might be adjusted through Bayesian updating based on past forecasting performance, possibly including multiple models in a model-averaging framework. It could also coherently introduce beliefs of the decision maker on the relative importance of historical measures as a function of the variance risk premium. Another open question is the relationship between time-series forecasts, implied volatility, and future realized volatility for individual stocks.

References

- Andersen, T. (2000), ‘Some reflections on analysis of high-frequency data’, *Journal of Business and Economic Statistics* **18**, 146–153.
- Andersen, T. G. & Bollerslev, T. (1998), ‘Answering the skeptics: Yes, standard volatility models do provide accurate forecasts’, *International Economic Review* **39**(4), 885–905.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), ‘The distribution of realized stock return volatility’, *Journal of Financial Economics* **61**, 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), ‘The distribution of realized exchange rate volatility’, *Journal of the American Statistical Association* **96**(453), 42–55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), ‘Modelling and forecasting realized volatility’, *Econometrica* **71**, 579–625.
- Arteche, J. (2004), ‘Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models’, *Journal of Econometrics* **119**, 131–154.
- Baillie, R. T., Bollerslev, T. & Mikkelsen, H. O. (1996), ‘Fractionally integrated generalized autoregressive conditional heteroscedasticity’, *Journal of Econometrics* **74**(1), 3–30.
- Bandi, F. M. & Perron, B. (2004), ‘Long memory and the relation between implied and realized volatility’, *Manuscript, Universit de Montral* .
- Bandi, F. M. & Russell, J. R. (2003), ‘Microstructure noise, realized volatility, and optimal sampling’, *Working paper* (Graduate School of Business, University of Chicago).
- Bandi, F. & Perron, B. (2006), ‘Long memory and the relation between implied and realized volatility’, *Journal of Financial Econometrics* **4**, 636–670.
- Barndorff-Nielsen, O. E. & Shephard, N. (2002), ‘Estimating quadratic variation using realized variance’, *Journal of Applied Econometrics* **17**(5), 457–477.

- Bollerslev, T. & Zhou, H. (2006), ‘Volatility puzzles: a simple framework for gauging return-volatility regressions’, *Journal of Econometrics* **131**, 123–150.
- Bondarenko, O. (2007), ‘Variance trading and the price of variance risk’, *Working Paper, University of Illinois at Chicago* .
- Britten-Jones, M. & Neuberger, A. (2000), ‘Option prices, implied price processes, and stochastic volatility’, *Journal of Finance* **55**, 839–866.
- Carr, P. & Madan, D. B. (1998), Towards a theory of volatility trading, *in* R. Jarrow, ed., ‘Volatility’, Risk Publications, pp. 417–427.
- Carr, P. & Wu, L. (2004), ‘Variance risk premia’, *Working paper* .
- Chernov, M. (2001), ‘Implied volatilities as forecasts of future volatility, the market risk premia, and returns variability’, *Mimeo* .
- Christensen, B. J. & Nielsen, M. Ø. (2005), ‘Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting’, *Forthcoming, Journal of Econometrics* .
- Christensen, B. J. & Prabhala, N. R. (1998), ‘The relation between implied and realized volatility’, *Journal of Financial Economics* **50**(2), 125–150.
- Cochrane, J. (2001), ‘Asset pricing’, *Princeton University Press* .
- Comte, F. & Renault, E. (1998), ‘Long memory in continuous-time stochastic volatility models’, *Mathematical Finance* **8**, 291–323.
- Deo, R. S., Hurvich, C. M. & Lu, Y. (2006), ‘Forecasting realized volatility using a long memory stochastic volatility model: Estimation, prediction and seasonal adjustment’, *Journal of Econometrics* **131**(1-2), 29–58.
- Ding, Z., Granger, C. W. J. & Engle, R. F. (1993), ‘A long memory property of stock returns and a new model’, *Journal of Empirical Finance* **1**(1), 83–106.

- Feinstein, S. (1989), ‘Forecasting stock market volatility using options on index futures’, *Economic Review, Federal Reserve Bank of Atlanta* **97**, 115–154.
- Fleming, J. (1998), ‘The quality of market volatility forecasts implied by s&p 100 index option prices’, *Journal of Empirical Finance* **5**(4), 317–345.
- Frederiksen, P. H., Nielsen, M. . & Shimotsu, K. (2006), ‘Narrow-band bootstrap inference in fractional cointegration’, *Working Paper, Cornell University* .
- Granger, C. W. J. & Bates, J. (1969), ‘The combination of forecasts’, *Operations Research Quarterly* **20**, 451–468.
- Granger, C. W. J. & Joyeux, R. (1980), ‘An introduction to long-memory time series models and fractional differencing’, *Journal of Time Series Analysis* **1**(1), 15–29.
- Hansen, P. R. & Lunde, A. (2005), ‘A realized variance for the whole day based on intermittent high-frequency data’, *Journal of Financial Econometrics* **3**(4), 525–554.
- Hosking, J. R. M. (1981), ‘Fractional differencing’, *Biometrika* **68**(1), 165–176.
- Hull, J. C. (1997), *Options, Futures, and Other Derivatives*, third edn, Prentice-Hall, Englewood Cliffs, NJ.
- Hurvich, C. M. & Ray, B. K. (2003), ‘The local whittle estimator of long-memory stochastic volatility’, *Journal of Financial Econometrics* **1**(3), 445–470.
- Jiang, G. J. & Tian, Y. S. (2005a), ‘Gauging the ’investor fear gauge’: implementation problems in the cboe’s new volatility index and a simple solution’, *Working paper* .
- Jiang, G. J. & Tian, Y. S. (2005b), ‘Model-free implied volatility and it’s information content’, *Review of Financial Studies* **18**, 1305–1342.
- Li, K. (2002), ‘Long-memory versus option-implied volatility predictions’, *Journal of Derivatives* **9**(3), 9–25.

- Martens, M., van Dijk, D. & de Pooter, M. (2004), ‘Modeling and forecasting s&p 500 volatility: Long memory, structural breaks and nonlinearity’, *Working Paper, Erasmus University Rotterdam* .
- Martens, M. & Zein, J. (2004), ‘Predicting financial volatility: High-frequency time-series forecasts vis--vis implied volatility’, *The journal of futures markets* **11**, 1005–1028.
- Mincer, J. & Zarnowitz, V. (1969), The evaluation of economic forecasts, *in* J. Mincer, ed., ‘Economic Forecasts and Expectations’, NBER, New York.
- Newey, W. & West, K. (1987), ‘A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix’, *Econometrica* **55**, 703–708.
- Nielsen, M. . (2006a), ‘Nonparametric cointegration analysis of fractional systems with unknown integration orders’, *Working Paper, Cornell University* .
- Nielsen, M. . (2006b), ‘A tuning parameter free nearly optimal test of the autoregressive unit root hypothesis’, *Working Paper, Cornell University* .
- Patton, A. (2006), ‘Volatility forecast comparison using imperfect volatility proxies’, *Working Paper, London School of Economics* .
- Phillips, P. C. B. & Perron, P. (1988), ‘Testing for a unit root in time series regression’, *Biometrika* **75**, 335–346.
- Pong, S.-Y., Shackleton, M. B., Taylor, S. J. & Xu, X. (2003), ‘Forecasting currency volatility: A comparison of implied volatilities and ar(fi)ma models’, *Lancaster University Working Paper* .
- Poteshman, A. M. (2000), ‘Forecasting future volatility from option prices’, *Mimeo* .
- Protter, P. (1990), ‘Stochastic integration and differential equations: A new approach’, *Springer-Verlag, New York* .
- Ray, B. K. & Tsay, R. (2000), ‘Long-range dependence in daily stock volatility’, *Journal of Business and Economic Statistics* **18**, 254–262.

- Robinson, P. M. (1994), ‘Semiparametric analysis of long-memory time series’, *The Annals of Statistics* **22**(1), 515–539.
- Robinson, P. M. & Marinucci, D. (2001), ‘Narrow-band analysis of nonstationary processes’, *Annals of Statistics* **29**, 947–986.
- Robinson, P. M. & Marinucci, D. (2003), Semiparametric frequency domain analysis of fractional cointegration, in P. M. Robinson, ed., ‘Time Series with Long Memory’, Oxford University Press.
- Sun, Y. & Phillips, P. C. B. (2003), ‘Nonlinear log-periodogram regression for perturbed fractional processes’, *Journal of Econometrics* **115**, 355–389.
- Timmerman, A. (2006), ‘Forecast combinations’, *Handbook of Economic Forecasting*, Elliot G., Granger, C.W.J., And Timmerman A. (editors) **4**, 135–196.
- Wright, J. H. (2002), ‘Log periodogram estimation of long memory volatility dependencies with conditionally heavy tailed returns’, *Econometric Reviews* **21**, 397–417.
- Zhang, L., Mykland, P. A. & At-Sahalia, Y. (2003), ‘A tale of two time scales: Determining integrated volatility with noisy high frequency data’, *Working Paper* (w10111, NBER).

Appendix - The Local Polynomial Whittle Estimator

If we assume that the volatility series can be described by the class of autoregressive fractionally integrated moving average (ARFIMA) processes outlined in section 2.1 plus some added noise, the spectral density can be written as

$$\begin{aligned}
 f_{\sigma}(\lambda) &= f_{\sigma^*}(\lambda) + f_u(\lambda) \\
 &= \frac{\sigma_v^2}{2\pi} |1 - e^{i\lambda}|^{-2d} \frac{|\beta(e^{i\lambda})|^2}{|\alpha(e^{i\lambda})|^2} + \frac{\sigma_u^2}{2\pi} \\
 &\approx g\lambda^{-2d} \exp(\theta\lambda^{2d}) \text{ as } \lambda \rightarrow 0^+,
 \end{aligned} \tag{2}$$

where $f_{\sigma^*}(\lambda)$ and $f_u(\lambda)$ are the spectral densities of the true volatility process and the measurement error, respectively, $\alpha(\cdot)$ and $\beta(\cdot)$ are the autoregressive and moving average polynomials, respectively, and the symbol " \sim " means that the ratio of the left and the right hand sides tends to one in the limit.

The NLW estimator is based on the approximation in (2) and is defined as the minimizers of

$$R(d, \theta) = \ln \left(\frac{1}{m} \sum_{j=1}^m I(\lambda_j) \lambda_j^{2d} \exp[-\theta\lambda_j^{2d}] \right) + \frac{1}{m} \sum_{j=1}^m [-2d \ln \lambda_j + \theta\lambda_j^{2d}], \tag{3}$$

where $m = m(n)$ is a bandwidth number which tends to infinity as $n \rightarrow \infty$, but at a slower rate than n , $\lambda_j = 2\pi j/n$ are the Fourier frequencies and $I(\lambda) = \frac{1}{2\pi n} |\sum_{t=1}^n \sigma_t e^{it\lambda}|^2$ is the periodogram of the predictor under consideration. Note that the estimator is invariant to a possible non-zero mean since $j = 0$ is left out of the minimization and that it enjoys robustness to short-memory dynamics since it only uses information from the periodogram ordinates in the vicinity of the origin. Compared to the GSP estimator the measurement error is directly modelled in (3) by the term $\theta\lambda_j^{2d}$ where θ is restricted to be positive⁹. Note that by setting $\theta = 0$, we obtain the GSP estimator.

⁹As noted by Sun & Phillips (2003) this restriction tends to bias the estimate of the long-memory parameter positively if the series is a pure fractional process (the true θ is zero).

Appendix - The Spectral Density Estimator

If we define the Discrete Fourier Transform (DFT) of the observed vector $\{a_t, t = 1, \dots, n\}$

$$w_a(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n a_t e^{it\lambda}$$

and the vector b_t similarly, the cross periodogram matrix between a_t and b_t is

$$I_{ab}(\lambda) = w_a(\lambda) w_b^*(\lambda) = I_{ab}^c(\lambda) + iI_{ab}^q(\lambda),$$

where the asterisk is transposed complex conjugation and c, q indicate the co- and quadrature periodograms, respectively. Then

$$\widehat{F}_{ab}(k, l) = \frac{2\pi}{n} \sum_{j=k}^l I_{ab}^c(\lambda_j), \quad 1 \leq k \leq l \leq n-1, \quad (4)$$

for the Fourier frequencies $\lambda_j = 2\pi j/n$. Note that the FDLS estimator enjoys robustness to short-memory dynamics if one only uses periodogram ordinates in the vicinity of the origin and that $\widehat{\beta}_{n-1}$ is the OLS estimate of β with allowance for a non-zero mean in $\epsilon_{t,T}$. If z_t is a vector ARFIMA process then $\widehat{\beta}_m$ is in the narrow-band FDLS class for $m = o(n^{0.8})$.

Appendix - The Bootstrap Procedure

1. Calculate the FDLS estimator

$$\widehat{\beta} = \left(\sum_{j=1}^m I_{xx}(\lambda_j) \right)^{-1} \sum_{j=1}^m I_{xy}(\lambda_j)$$

where $I_{xx}(\lambda) = w_x(\lambda)w_x'(-\lambda)$ and $I_{xy}(\lambda) = w_x(\lambda)w_y'(-\lambda)$ are the periodogram of x_t and the cross-periodogram of x_t and y_t , respectively, $w_a(\lambda) = (2\pi T)^{-1/2} \sum_{t=1}^T a_t e^{it\lambda}$ denotes the discrete Fourier transform of a_t and m is the chosen bandwidth used in the FDLS estimate.

Step 1 suggest that $\hat{\beta}$ can be regarded as the FDLS estimator in the model

$$w_y(\lambda_j) = \beta w_x(\lambda_j) + w_u(\lambda_j), \quad j = 1, \dots, m,$$

where $w_x(\lambda_j)$ and $w_u(\lambda_j)$ are the regressor and error term, respectively. Since $w_u(\lambda_j)$ might possibly be hetereskedastic it is interchanged with $v_u(\lambda_j) = w_u(\lambda_j)/|w_u(\lambda_j)|$, which can be regarded as a zero mean asymptotically independent homoskedastic random variable, to give

$$w_y(\lambda_j) = \beta w_x(\lambda_j) + |w_u(\lambda_j)| v_u(\lambda_j), \quad j = 1, \dots, m.$$

2. Obtain the residuals

$$\hat{u}_t = y_t - \hat{\beta}x_t, \quad t = 1, \dots, T.$$

3. Compute the discrete Fourier transform of the residuals \hat{u}_t , denoted $w_{\hat{u}}(\lambda_j)$, and let $v_{\hat{u}}(\lambda_j) = w_{\hat{u}}(\lambda_j)/|w_{\hat{u}}(\lambda_j)|$.

4. Draw independent bootstrap residuals η_j^* , $j = 1, \dots, m$, from the empirical distribution function of

$$\tilde{v}_{\hat{u}}(\lambda_j) = \hat{\sigma}_v^{-1}(v_{\hat{u}}(\lambda_j) - \bar{v}_{\hat{u}}), \quad j = 1, \dots, m,$$

where $\bar{v}_{\hat{u}} = m^{-1} \sum_{j=1}^m v_{\hat{u}}(\lambda_j)$ and $\hat{\sigma}_v^2 = m^{-1} \sum_{j=1}^m |v_{\hat{u}}(\lambda_j) - \bar{v}_{\hat{u}}|^2$. That is, for all $j = 1, \dots, m$,

$$\Pr\{\eta_j^* = \tilde{v}_{\hat{u}}(\lambda_j)\} = m^{-1}.$$

5. Obtain the bootstrap FDLS

$$w_y^*(\lambda_j) = \hat{\beta}w_x(\lambda_j) + |w_{\hat{u}}(\lambda_j)|\eta_j^*, \quad j = 1, \dots, m.$$

6. Compute the bootstrap estimator of $\hat{\beta}$ as

$$\hat{\beta}^* = \left(\sum_{j=1}^m I_{xx}(\lambda_j) \right)^{-1} \sum_{j=1}^m Re(I_{xy^*}(\lambda_j)),$$

where $I_{xy^*}(\lambda_j) = w_x(\lambda_j)w_{y^*}'(-\lambda_j)$ and $Re(a)$ denotes the real part of the complex number a .

7. Redo step 4 – 6 e.g. 9,999 times, sort the 9,999 bootstrap estimates and select the appropriate values for the chosen confidence level.

Table 1: Descriptive statistics. The table reports descriptive statistics for the realized and implied volatility series, as well as the ARFIMA forecasts for the S&P 500 index from January 1990 to December 2002.

	Mean	Std. dev.	Skewness	Kurtosis
Panel A: σ_t				
Realized Volatility	0.1558	0.0742	1.5511	6.1259
ARFIMA forecast	0.1527	0.0607	1.2147	4.5807
B&S Implied Volatility	0.1662	0.0567	0.8276	3.2068
VIX	0.1954	0.0652	0.9702	3.8708
Panel B: $\log \sigma_t$				
Realized Volatility	-1.9554	0.4321	0.3179	2.7055
ARFIMA forecast	-1.9490	0.3693	0.3282	2.5438
B&S Implied Volatility	-1.8491	0.3295	0.2004	2.2230
VIX	-1.6842	0.3184	0.2546	2.3785

Table 2: Univariate and encompassing regressions of 1-month volatilities. The table reports the results of the predictive regressions of realized volatility on an ARFIMA volatility forecast and both model-free and implied volatility for the S&P 500 index from January 1990 to December 2002. The numbers in parentheses are Newey-West (1987) standard errors of the estimated parameters.

α	β^{RVF}	β^{BS}	β^{MF}	$adj.R^2$	Durbin-Watson	Wald-test	d	$H_0 : d = 0$
Panel A: σ_t								
0.0021 (0.0105)	1.0059 (0.0784)			0.6760	2.049	0.542	-0.020	0.764
-0.0226 (0.0163)		1.0736 (0.1141)		0.6728	1.715	0.000	0.188	0.004
-0.0230 (0.0162)			0.9155 (0.0981)	0.6172	1.473	0.000	0.275	0.000
-0.0225 (0.0130)	0.5487 (0.1413)	0.5685 (0.1714)		0.7251	1.970	0.008	0.039	0.558
-0.0160 (0.0128)	0.6275 (0.1918)		0.3939 (0.1721)	0.7002	1.947	0.130	0.057	0.392
-0.0242 (0.0132)	0.5132 (0.1889)	0.5205 (0.1697)	0.0772 (0.1728)	0.7257	1.949	0.000	0.053	0.426
Panel B: $\log \sigma_t$								
0.0244 (0.0871)	1.0158 (0.0434)			0.7533	2.117	0.859	0.009	0.892
0.1061 (0.1371)		1.1149 (0.0715)		0.7226	1.674	0.000	0.188	0.004
-0.0419 (0.0.1321)			1.1366 (0.0730)	0.7014	1.544	0.000	0.243	0.000
0.1780 (0.0960)	0.6239 (0.0989)	0.4966 (0.1099)		0.7846	2.002	0.000	0.051	0.443
0.0801 (0.0911)	0.7251 (0.1295)		0.3695 (0.1364)	0.7658	2.061	0.030	0.046	0.652
0.1787 (0.0968)	0.6146 (0.1297)	0.4842 (0.1195)	0.0243 (0.1487)	0.7847	1.999	0.000	0.053	0.426

Table 3: Combined forecasts. The table reports the mean squared forecasting error and R^2 for individual forecasts and forecasts that combine the ARFIMA volatility forecast and Black and Scholes implied volatility. The combined forecasts are constructed using either weights based on estimated individual forecast precisions (fc_{simple}) or restricted least squares weights (fc_{rols}).

	$MSFE$		R^2	
Panel A: σ_t				
fc_{simple}	0.016	0.017	0.729	0.715
fc_{rols}	0.018	0.018	0.717	0.701
ARFIMA forecast	0.021	0.021	0.660	0.660
B&S Implied Volatility	0.021		0.682	
VIX		0.033		0.671
Panel B: $\log \sigma_t$				
fc_{simple}	0.045	0.046	0.809	0.789
fc_{rols}	0.047	0.047	0.777	0.769
ARFIMA forecast	0.051	0.051	0.757	0.757
B&S Implied Volatility	0.066		0.739	
VIX		0.105		0.729

Table 4: Memory Properties of the Volatility Series

	Realized Volatility SP500 Index	Realized Volatility Forecast	Black-Scholes Implied Volatility	Model-Free Implied Volatility
NLW (bandwidth) ^a $\hat{d}(m = \lfloor n^{0.8} \rfloor)$	0.4680 (0.1382)	0.4481 (0.1414)	0.4371 (0.1433)	0.4523 (0.1407)
VRT ^b ρ	1.7681 (0.9469)	1.7805 (0.9259)	1.7482 (0.9610)	1.7466 (0.9642)
ADF ^c t	-6.2012**	-4.9773**	-3.6832*	-5.4715**

^aNLW are nonlinear local Whittle estimates of the fractional integration order. The numbers in parenthesis are asymptotic standard errors using $\sqrt{m}(\hat{d} - d) \rightarrow_d N\left(0, \frac{(1+2d)^2}{16d^2}\right)$.

^bVRT are the variance ratio tests of the null of $d = 0.5$. The critical value at the 5% significance level is 2.339. p-values are in parenthesis.

^cADF are Augmented Dickey-Fuller tests of the null of a unit root where only significant arguments are included. Two asterisk indicate significance at the 1% level (critical value including both constant and trend: -4.02) and one asterisk indicate significance at the 5% level (critical value including both constant and trend: -3.44).

Table 5: The Univariate Fractional Cointegration Analysis of the Volatility Series

<i>The Black-Scholes Implied Volatility Analysis</i>					
Bandwidth	$\hat{\alpha}_{m,BSIV}$	$\hat{\beta}_{m,BSIV}$	Confidence Interval		\hat{d}_ϵ
			95%	99%	
m_1	-0.033**	1.138	[1.025,1.256]	[0.991,1.290]	0.067 (0.540)
m_2	-0.051**	1.243	[1.141,1.347]	[1.105,1.382]	0.053 (0.698)
m_3	-0.045**	1.208	[1.136,1.280]	[1.114,1.304]	0.058 (0.642)
m_4	-0.045**	1.206	[1.137,1.276]	[1.113,1.296]	0.058 (0.640)
<i>The Model-Free Implied Volatility Analysis</i>					
Bandwidth	$\hat{\alpha}_{m,MFIV}$	$\hat{\beta}_{m,MFIV}$	Confidence Interval		\hat{d}_ϵ
			95%	99%	
m_1	-0.030*	0.949	[0.819,1.079]	[0.775,1.134]	0.123 (0.339)
m_2	-0.042**	1.015	[0.902,1.132]	[0.870,1.171]	0.114 (0.361)
m_3	-0.042**	1.014	[0.917,1.112]	[0.883,1.142]	0.114 (0.360)
m_4	-0.046**	1.034	[0.946,1.121]	[0.916,1.150]	0.112 (0.365)
<i>The Realized Volatility Forecast Analysis</i>					
Bandwidth	$\hat{\alpha}_{m,RV}$	$\hat{\beta}_{m,RV}$	Confidence Interval		\hat{d}_ϵ
			95%	99%	
m_1	-0.006	1.056	[0.948,1.173]	[0.914,1.212]	-0.028 (0.067)
m_2	-0.018	1.140	[1.029,1.248]	[0.998,1.293]	-0.072 (0.067)
m_3	-0.018	1.136	[1.020,1.230]	[0.990,1.262]	-0.073 (0.067)
m_4	-0.015	1.119	[1.026,1.211]	[0.986,1.244]	-0.072 (0.067)

The table shows semiparametric estimates of the univariate stationary fractional cointegration for the bandwidths $m_k = \lceil n^{\lambda_k} \rceil$, $\lambda_k = \{0.8, 0.6, 0.4, 0.2\}$ with 95% and 99% bootstrap confidence intervals. Regarding the intercept, two asterisk indicate that the 99% confidence interval does not contain zero and one asterisk indicates that the 95% confidence interval does not contain zero. The fractional integration orders of the residuals \hat{d}_ϵ are based on the NLW for the implied volatilities and the GSP for the realized volatility forecast with $m_\epsilon = \lceil n^{0.8} \rceil$.

Table 6: The Multiple Fractional Cointegration Analysis of the Volatility Series

Bandwidth	$\hat{\alpha}_m$	$\hat{\beta}_{m,BSIV}$	$\hat{\beta}_{m,RV}$	\hat{d}_ϵ	$\hat{\alpha}_m$	$\hat{\beta}_{m,MFIV}$	$\hat{\beta}_{m,RV}$	\hat{d}_ϵ
m_1	-0.028*	0.553 [0.307,0.801]	0.602 [0.366,0.839]	0.007 (0.067)	-0.021*	0.336 [0.119,0.554]	0.731 [0.492,0.968]	0.000 (0.067)
m_2	-0.022	0.109 [-0.299,0.517]	1.047 [0.663,1.429]	-0.056 (0.067)	-0.018	0.002 [-0.206,0.218]	1.138 [0.901,1.373]	-0.072 (0.067)
m_3	-0.014	-0.102 [-0.363,0.142]	1.242 [0.977,1.476]	-0.084 (0.067)	-0.017	-0.012 [-0.181,0.147]	1.147 [0.956,1.338]	-0.073 (0.067)
m_4	-0.017	0.059 [-0.061,0.179]	1.067 [0.918,1.215]	-0.064 (0.067)	-0.015	-0.005 [-0.070,0.059]	1.124 [0.998,1.201]	-0.072 (0.067)

The table shows semiparametric estimates of the multivariate stationary fractional cointegration for the bandwidths $m_k = \lceil n^{\lambda_k} \rceil$, $\lambda_k = \{0.8, 0.6, 0.4, 0.2\}$. The numbers in square brackets are 95% bootstrap confidence intervals. Regarding the intercept, one asterisk indicates that the 95% confidence interval does not contain zero. The fractional integration orders of the residuals are based on the GSP estimator with $m_\epsilon = \lceil n^{0.8} \rceil$.