Implied Volatilities as Forecasts of Future Volatility, Time-Varying Risk Premia, and Returns Variability[1]

Mikhail Chernov[2]

First Draft: July 2000

This Revision: October 11, 2001

[1] I would like to thank Andrew Ang, Federico Bandi, Luca Benzoni, Rob Engle, Eric Ghysels, Larry Glosten, Bob Hodrick, Gur Huberman, Mike Johannes, Peter Lakner, Scott Mixon, David Robinson, Tano Santos as well as seminar participants at Columbia, Minnesota, Montreal and NYU for comments. I am grateful to the Center for International Business Education (CIBE) at Columbia for financial support.

[2] Division of Finance and Economics, Columbia Business School, 413 Uris Hall, 3022 Broadway, New York, NY 10027, tel: 212-854-9009, fax: 212-316-9180, e-mail: mc1365@columbia.edu

Implied Volatilities as Forecasts of Future Volatility, Time-Varying Risk Premia, and Returns Variability

Abstract

The unbiasedness tests of implied volatility as a forecast of future realized volatility have found implied volatility to be a biased predictor. We explain this puzzle by recognizing that option prices contain a market risk premium not only on the asset itself, but also on its volatility. Hull and White (1987) show using a stochastic volatility model that a call option price can be represented as an expected value of the Black-Scholes formula evaluated at the average integrated volatility. If we allow volatility risk to be priced, this expectation should be taken under the risk-neutral probability measure, and can be decomposed into the expectation with respect to the physical measure and the risk-premium term. This term is just a linear function of the unobservable spot volatility. The decomposition explains the bias documented in the empirical literature and shows that the realized and historical volatility, which are used in the tests, are in fact the estimates of the unobserved quadratic variation and spot volatility of the stock-return generating process. Therefore, the use of these estimates generates the error-in-the-variables problem. We generalize the above results from a stochastic volatility model to a model with multiple volatility and jump factors. We provide an empirical illustration based on two US equity indices and three foreign currency rates. We find, that when we take into an account the risk-premium and use efficient methods to estimate volatility, the unbiasedness hypothesis can not be rejected, and the point estimate of the loading on the implied volatility in the traditional regression is equal to 1.

JEL classification:

Key Words: Implied Volatility, Realized Volatility, Historical Volatility, Spot Volatility, Quadratic Variation, Jump-Diffusion Processes, Market Prices of Risk, Error-in-the-Variables Problem

1 Introduction

Option prices continue to attract widespread interest because of their sensitivity to the distribution of returns over a specified time to maturity. This feature allows for a possibility that options contain more information about future volatility than time series of the underlying asset. Indeed, the stochastic volatility (SV) model of Hull and White (1987) implies that the Black and Scholes (1973) implied volatility is an unbiased forecast of subsequent realized volatility. However, the overwhelming majority of empirical work finds that implied volatilities are generally higher than realized ones.[1] Attempts to understand this puzzle are based on critiques of the econometric methodology. The improvements these critiques propose are not robust to different types of financial assets or different sample time periods.

In this paper we suggest an alternative explanation to the unbiasedness puzzle. We show that accounting for the volatility risk premium goes a long way towards explaining the previous empirical results. First, we explain how the disparity between objective and risk-neutral probability measures leads to the disparity between the realized and implied volatilities. Then, we illustrate these ideas empirically using a comprehensive dataset containing several financial assets previously considered in the literature and spanning the time frames of all previous studies.

Generally, prior tests concentrate on the following regression:

$$RV_{t,\tau} = a + b \cdot \sigma^2_{t,\tau} + c \cdot HV_{t-\tau,\tau} + \epsilon_{t+\tau} \tag{1.1}$$

where $\sigma^2_{t,\tau}$ is the Black-Scholes volatility implied at time t from an option contract with time to maturity τ, RV is the realized volatility over the period t to $t+\tau$, and HV is the historical volatility over the preceding period $t-\tau$ to t. Three hypotheses are typically tested:

Informativeness The loading on the implied volatility, b, should be significantly different from zero.

Unbiasedness The slope b should be equal to 1 and intercept a to 0.

Informational efficiency No other variables (e.g. historical volatility) are significant, i.e. $c = 0$

[1] We will refer to this evidence as the unbiasedness puzzle.

Most of the studies find evidence in favor of informativeness and against unbiasedness. The evidence of informational efficiency is mixed.[2]

We show that taking into account the volatility risk premium can explain the unbiasedness puzzle. Our starting point is the work of Hull and White (1987). Their option pricing formula implies that, in the framework of an SV model, a call option price can be represented as an expected value of the Black-Scholes formula evaluated at the average integrated volatility. If, however, we allow volatility to have a market risk premium, this expected value should be taken under the risk-neutral probability measure. We decompose the expectation of the integrated volatility with respect the risk-neutral measure into the expectation with respect to the objective measure and the risk-premium term. The decomposition immediately explains the bias documented in the empirical literature.

Our decomposition also shows that RV and HV, which are typically used in regression (1.1) as the realized and historical volatility, respectively, are in fact the estimates of the unobserved quadratic variation and spot volatility of the stock-return generating process. Therefore, their appearance in the regression is tied to an error-in-the-variables problem. It turns out that these measurement errors have a much more dramatic impact on the results than errors in the measurement of implied volatilities accounted for in prior studies. We show this using data free of implied volatility measurement problems.

We use daily data on U.S. Equity Indices (S&P 100 and Nasdaq 100) and foreign exchange (British pound, Japanese yen, and Swiss franc). The respective implied volatility series are constructed in such a way that every day they correspond to an at-the-money (ATM) contract which has exactly one month to maturity. We find, that once the risk-premium is taken into account, the point estimate of the slope b is equal to 1.

The paper is organized as follows. The next section outlines the theoretical justification for the regression (1.1) based on Hull and White (1987). We also discuss the recent literature, which attempts to explain the rejection of the unbiasedness puzzle. In section 3 we show how volatility risk premium changes the relationship between the future and Black-Scholes volatilities. We derive the particular relationship for the most popular SV models. Section 4 further develops the results for more general processes describing the dynamics of the stock-return generating process. These

[2]Refer to Bates (1996) and Poteshman (2000) for reviews of this subject and additional references.

results allow us to show that RV and HV in the traditional regression (1.1) can be viewed as estimates of the latent volatility. These observations imply the error-in-the-variables problem, which is investigated in the empirical example in the section 5. The last section concludes.

2 Justification of the Unbiasedness Hypothesis and Related Literature

We would like to justify the regression (1.1), which is routinely tested in the literature. This discussion will allow us to formalize the issues we are facing, to establish links to the recent research as well as to introduce notation used in the remainder of the paper. We start out by the formalism, which is based on the work of Hull and White (1987). Namely, they show in the framework of a stochastic volatility (SV) model without the leverage effect, that a call option price can be represented as an expected value of the Black-Scholes formula evaluated at the average integrated volatility:

$$C^{HW}(S_t, V_t, r, K, \tau) = E_t^Q \left\{ C^{BS}(S_t, \bar{V}_{t,\tau}, r, K, \tau) \right\} \tag{2.1}$$

where S_t and V_t are current underlying asset value and its spot variance respectively, r is the risk-free interest rate, τ is the option's time to maturity, K is the strike price, $\bar{V}_{t,\tau}$ is the average integrated volatility:

$$\bar{V}_{t,\tau} = \frac{1}{\tau} \int_t^{t+\tau} V_s ds \tag{2.2}$$

and Q indicates that the expectation is taken with respect to the risk-neutral probability measure. In addition to this, the Black-Scholes formula is known to be nearly linear in volatility for the at-the-money (ATM) close to maturity options.[3] Hence for the options with $S_t/K \approx 1$, we conclude from (2.1), that

$$E_t^Q \left(\bar{V}_{t,\tau} \right) = \sigma_{t,\tau}^2 \tag{2.3}$$

where $\sigma_{t,\tau}$ is the Black-Scholes volatility at time t, implied from the contract which matures in τ periods. See Fleming (1998) and Jones (2000) for further details. Jones also argues, that this approximation is still valid if one introduces the leverage effect into the SV model.

[3] To be more precise, the Black-Scholes formula is linear in volatility when the strike, K, is delta-neutral, i.e. when the call's Black-Scholes delta using the corresponding implied volatility is equal to the negative of the put's Black-Scholes delta using the same volatility. Details are available in the appendix A.

The last relationship forms a theoretical basis for the described above regressions. The unbiasedness hypothesis is based on the Hull and White (1987) assumption that the volatility risk can be diversified away and, hence, its risk premium is zero. Therefore Q in (2.3) can be dropped, i.e. we can take expectations with respect to the physical (true) probability measure. This immediately implies the regression:

$$\bar{V}_{t,\tau} = a + b \cdot \sigma^2_{t,\tau} + \epsilon_{t+\tau} \tag{2.4}$$

This is essentially the view taken in Lamoureux and Lastrapes (1993) in their evaluation of option markets' informational efficiency. An alternative approach, taken in Canina and Figlewski (1993), is to consider realized volatility

$$RV_{t,\tau} \equiv \frac{1}{\tau}\sum_{i=1}^{\tau} r^2_{t+i}(1) \tag{2.5}$$

$$r_t(n) = \log(S_t/S_{t-1/n}) \tag{2.6}$$

as a direct object of interest and evaluate whether the implied volatility is its rational forecast.

These considerations form a theoretical basis for the regression (1.1) and the hypotheses outlined in the introduction. Several studies address the unbiasedness hypothesis, most notably Canina and Figlewski (1993), Day and Lewis (1992), Jorion (1995), and Lamoureux and Lastrapes (1993). They find the value of $\hat{b}$ to be between 0.02 and 0.72 using various datasets, time periods, and observation frequencies (refer to Poteshman (2000) for a detailed comparison of the data and results in these studies). Moreover, Lamoureux and Lastrapes (1993) find negative and significant $\hat{c}$, while Day and Lewis (1992) find evidence in favor of positive $\hat{c}$. The overall conclusion from these studies is that implied volatility is a biased and perhaps inefficient estimate of the future volatility.

More recent studies attempt to explain these results by improving the econometric methodology. Namely, Fleming (1998) observes that daily options data have a telescoping time to maturity. Hence, the standard Hansen and Hodrick (1980) correction for autocorrelation in the errors does not work. He develops a procedure which takes this problem into an account and finds that implied volatilities are informationally efficient albeit biased. Christensen and Prabhala (1998) take this criticism one step further and construct a data set with monthly observations of volatilities implied from one-month-to-maturity contracts. Such nonoverlapping observations remove many econometric issues. Furthermore, they observe that the implied volatility is measured with error due to the American feature of the option contract, dividends, and the equity-options markets nonsynchronicity bias. After they correct for this problem in the error-in-the-variables (EIV) framework the unbiasedness and informational efficiency of the implied volatility are restored.

Poteshman (2000) critiques these studies. One of his points is that some of the studies use log and square-root transformations of the explanatory and dependent variables in (1.1). Such transformations will bias the relevant parameters because of the Schwartz inequality. Poteshman (2000) demonstrates that the log transformation biases $\hat{b}$ upwards the most. He agrees with Christensen and Prabhala (1998) that σ is measured imprecisely, so he suggests to use the SPX contracts (which do not have the American feature), several contracts simultaneously as in VIX, and intraday prices to avoid the nonsynchronicity bias. He also argues that one has to rely on a more realistic model, e.g. SV, to imply volatilities. This suggestion, however, violates the model-free spirit of the volatility forecasting. In addition, it is hard to settle on one particular more realistic model. Different models would clearly lead to different values of implied volatilities.

We would like to argue in this paper that while all the issues raised in the literature are critical, they are of second order importance compared to the omitted volatility risk-premium and errors in the measurement associated with the use of the realized volatility RV and historical volatility HV. We elaborate on these issues in the following sections.

3 Market Price of Volatility Risk

In this section we allow for non-diversifiable assets volatility risk. In other words, we assume that investors require a premium for bearing this risk. This assumption can be defended on empirical grounds. For instance, Lamoureux and Lastrapes (1993) conclude: "... equilibrium models of option pricing that do not assume investor indifference to volatility risk appear necessary to reconcile the theory and data. The data suggest that the market premium on variance risk is time varying; it is a decreasing function of the level of the stock's variance". More recent empirical studies by Benzoni (1999), Chernov (2001), Chernov and Ghysels (2000), Jones (2000), and Pan (2000) explicitly model the volatility risk premium as a linear function of volatility and find it to be significant based on the information contained in the options prices and respective underlying assets' returns. These findings are fully consistent with the conclusion of Lamoureux and Lastrapes (1993) and justify our approach in this paper.

The presence of a required volatility risk premium translates into the requirement to explicitly take into an account the risk-neutral probability measure Q in (2.3):

$$\begin{aligned}\sigma_{t,\tau}^2 &= E_t^Q\left(\bar{V}_{t,\tau}\right) = E_t\left(\xi_{t,\tau}\bar{V}_{t,\tau}\right) \\ &= E_t\left(\xi_{t,\tau}\right)\cdot E_t\left(\bar{V}_{t,\tau}\right) + \mathrm{Cov}_t\left(\xi_{t,\tau},\bar{V}_{t,\tau}\right) = E_t\left(\bar{V}_{t,\tau}\right) + \mathrm{Cov}_t\left(\xi_{t,\tau},\bar{V}_{t,\tau}\right) \end{aligned}\tag{3.1}$$

where $\xi_{t,\tau}$ is the Radon-Nikodym derivative of the risk-neutral probability measure Q with respect to the objective probability measure P.[4] The last formula immediately yields:

$$E_t\left(\bar{V}_{t,\tau}\right) = \sigma_{t,\tau}^2 - \mathrm{Cov}_t\left(\xi_{t,\tau},\bar{V}_{t,\tau}\right)\tag{3.2}$$

We can see that the covariance term, which represents the volatility risk premium can be non-negligible depending on the specification of $\xi_{t,\tau}$ and the values of relevant parameters. Therefore, we have shown that non-zero volatility premium can be responsible for the unbiasedness puzzle. Indeed, the state-dependent negative volatility risk-premium documented in the literature implies positivity of the covariance between the stochastic discount factor and the averaged integrated volatility in (3.2).[5] It means that the expected future volatility is less than the implied volatility. If one omits the covariance term from the analysis the implied volatility's slope will decrease to preserve this relationship.

To make the discussion more concrete and to see how this decomposition can reconcile the empirical evidence, we have to make an additional step and assume the underlying asset dynamics. We consider a fairly generic specification of an SV model:

$$\frac{dS_t}{S_t} = \delta_t dt + \sqrt{V_t}dW_{St}\tag{3.3}$$

$$dV_t = \left(\theta - \kappa V_t\right)dt + \sigma(V_t)dW_{Vt}\tag{3.4}$$

The Brownian motions W_S and W_V are assumed to be uncorrelated (no leverage effect) for the Hull-White formula (2.1) to hold. However, as we discussed in the introduction, the non-zero correlation

[4] The mathematical concept of the Radon-Nikodym derivative has an economic interpretation because it is equal to the stochastic discount factor scaled by the value of the money market account.

[5] According to the Girsanov theorem,

$$\xi_{t,\tau} = \exp\left(-\int_t^{t+\tau}\lambda_s dW_s - \int_t^{t+\tau}\lambda_s'\lambda_s ds\right)$$

where λ_t is a vector of market prices of risk. In empirical literature, the volatility market risk is often modeled as $\lambda\sqrt{V_t}$ and λ is found to be negative. This explains positive covariance term.

will not affect the approximation in (2.3) by much. We do not specify the functional form of the stock price drift δ or of the volatility of volatility $\sigma(\cdot)$.[6] The volatility is assumed to be reverting to the long-run mean θ/κ with the speed κ. This is a standard assumption in all empirically relevant SV models and, hence, is not very restrictive.

The assumed form of the volatility drift allows us to use a well-known result

$$E_t^M(\bar{V}_{t,\tau}) = A_\tau^M V_t + B_\tau^M \tag{3.5}$$

where M denotes a probability measure. The details regarding this formula are provided in the appendix B. Linearity of the expected integrated volatility in the spot volatility yields a particularly simple expression for the volatility risk term. Combining (3.1) with the above result we obtain:

$$\mathrm{Cov}_t\left(\xi_{t,\tau}, \bar{V}_{t,\tau}\right) = E_t^Q\left(\bar{V}_{t,\tau}\right) - E_t\left(\bar{V}_{t,\tau}\right) = \left[A_\tau^Q - A_\tau^P\right] V_t + \left[B_\tau^Q - B_\tau^P\right] \tag{3.6}$$

Now we can substitute this expression into (3.2):

$$E_t\left(\bar{V}_{t,\tau}\right) = \sigma_{t,\tau}^2 - \left[A_\tau^Q - A_\tau^P\right] V_t - \left[B_\tau^Q - B_\tau^P\right] = c_\tau^B + \sigma_{t,\tau}^2 + c_\tau^A V_t \tag{3.7}$$

with obvious notations for constants c^A and c^B, which depend only on the option time-to-maturity τ and the model parameters. This expression allows us to contrast our theoretical results and typical empirical findings in the literature.

This decomposition implies the following regression equation:

$$\hat{\bar{V}}_{t,\tau} + \epsilon_{t+\tau}^1 = a + b \cdot \sigma_{t,\tau}^2 + c \cdot \left(\hat{V}_t + \epsilon_t^2\right) + \epsilon_{t+\tau}^3 \tag{3.8}$$

where the error terms ϵ^1 and ϵ^2 are an explicit recognition of the fact that both realized and spot volatility are not observable and, hence, are measured with an error. We consider σ to be a transformation of the respective option price. Therefore, it is observable without any error if we abstract from the microstructure effects.

The fact that the spot volatility is not observable also complicates the estimation because it introduces the error-in-the-variables (EIV) problem. The EIV problem in the explanatory variable V leads to the attenuation effect, i.e. the coefficient c will be biased towards zero. The same EIV

[6] We assume that δ and $\sigma(\cdot)$ satisfy the regularity conditions necessary for the existence of the solution of (3.3)-(3.4).

problem also leads to the bias in b. Moreover, since V is correlated with σ, omitting it from the regression will bias b. Namely, if we implement (2.4) instead of (3.7)

$$E\left(\hat{b}\right) = b + \frac{\sum_t \sigma_{t,\tau}^2 V_t}{\sum_t \sigma_{t,\tau}^4} \cdot c \tag{3.9}$$

Apart from the mathematical formulation of the bias size, the expression in (3.9) has an economic interpretation as well. We can follow Fama's (1984) interpretation of the expectation hypothesis puzzle and think of $\hat{b}$ obtained from the regression (2.4) with an omitted variable as a coefficient which incorporates the volatility risk-premium. It means that since the true b should be equal to 1, we can compute the risk premium by estimating b via (2.4) and by subtracting 1 afterwards.

The dependence of the coefficients in (3.7) on time horizon may also explain some of the results. We can use the Taylor approximations from (B.9) and (B.10) in the appendix B to obtain from (3.7):

$$E_t\left(\bar{V}_{t,\tau}\right) = \sigma_{t,\tau}^2 + o(1) \tag{3.10}$$

when τ approaches 0. This result is not surprising, because, as Ledoit and Santa-Clara (1999) show, as time to maturity of an ATM option approaches zero, its implied volatility approaches the spot volatility. The same can be said about the integrated volatility. This is consistent with the expression in (3.10). If, however, the option contract horizon is very long, e.g. $\tau = 129$ business days as in Lamoureux and Lastrapes (1993), both the implied and spot volatilities may have a substantial presence in our representation, and, therefore, may explain the unbiasedness rejection.

In summary, our representation provides intuitive explanation of the empirical results on unbiasedness testing previously reported in the literature. Namely, accounting for non-zero volatility risk premium introduces additional terms into the usual realized-implied volatilities relationship. Omitting this terms in empirical studies leads to reported biases. We will perform an empirical exercise based on the representation in (3.7) to confirm the theoretical findings.

4 Returns Variability

We established in the previous section the important role played by the market risk premium. Omitting the component related to this premium leads to downward biases of the slope coefficient, which is reported in many previous studies. The decomposition into the Black-Scholes implied volatility and the volatility risk premium is quite constructive. We can use the relationship in (3.7) and the respective regression equation in (3.8) to implement the appropriate tests of the unbiasedness hypothesis.

Moreover, since we have to work with estimates of spot volatility we can use the representation in the original regression (1.1) to come up with the required estimates. Indeed, if we use the estimate:

$$\hat{V}_t = HV_{t-\tau,\tau} = \frac{1}{\tau}\sum_{i=0}^{\tau-1} r_{t-i}^2(1) \tag{4.1}$$

$$\hat{\bar{V}}_{t,\tau} = RV_{t,\tau} = \frac{1}{\tau}\sum_{i=1}^{\tau} r_{t+i}^2(1) \tag{4.2}$$

then our theory will be completely consistent with the typical implementation. Moreover, there is an appealing theoretical argument in Foster and Nelson (1996) which supports this choice of the estimators. They belong to the class of backward-looking flat-weight rolling regressions, which were first considered by Fama and Macbeth (1973). Foster and Nelson (1996) show for a large class of data-generating processes, that these estimators are asymptotically (in a sense of n in the definition of $r_t(n)$ in (2.6) approaching infinity) unbiased. However, if the process is more complicated than the one-factor SV model considered in the previous section, then the estimator converges to the quadratic variation of the process.

The quadratic variation of returns $r_t(n)$ is defined as:

$$QV_{t,\tau} \equiv [r,r]_t^{\tau} = \operatorname*{plim}_{n\to\infty} \sum_{i=1}^{n\tau} r_{t+\frac{i}{n}}^2(n) \tag{4.3}$$

As we can see, this definition is operational: adding up squared returns observed within a particular interval will give a consistent estimate of the quadratic variation over this period.[7] [8] As we show in the appendix C, if the true data generating process has multiple volatility factors and/or jumps, the averaged quadratic variation defined as:

$$\overline{QV}_{t,\tau} = \frac{1}{\tau} QV_{t,\tau} \tag{4.4}$$

[7]To be precise, the right-hand expression in (4.3) is the property, not the definition of the quadratic variation. Please refer to Protter (1990) for details.

[8]This idea is used in Andersen, Bollerslev, Diebold, and Labys (2000) to estimate daily quadratic variation based on five-minute returns.

substitutes the averaged integrated volatility in many relationships derived above, most notably (2.3) and (3.2).

As an example consider a model with a stochastic volatility factors and jumps to both returns and volatility (SVIJ):

$$\begin{aligned} \frac{dS_t}{S_t} &= \delta_t dt + \sqrt{V_t} dW_{St} + (J_t dN_t - \lambda\mu dt) && (4.5) \\ dV_t &= (\theta - \kappa V_t)\, dt + \sigma(V_t) dW_{Vt} + J_{Vt} dN_{Vt} && (4.6) \end{aligned}$$

where N (N_V) is the Poisson (counting) process with intensity λ (λ_V) and jump size J (J_V) is distributed normally with mean μ and variance σ^2 (exponentially with mean μ_V). This is a realistic model of equity indices returns (see Eraker, Johannes, and Polson, 2000).

In this case, the averaged quadratic variation is equal to:

$$\overline{QV}_{t,\tau} = \bar{V}_{t,\tau} + \frac{1}{\tau} \sum_{t<s\leq t+\tau} J_s^2 \Delta N_s = \bar{V}^c_{t,\tau} + \frac{1}{\tau} \sum_{t<s\leq t+\tau} J_{Vs} \Delta N_{Vs} + \frac{1}{\tau} \sum_{t<s\leq t+\tau} J_s^2 \Delta N_s \qquad (4.7)$$

where V^c denotes the continuous part of volatility V in (4.6). Then we can show (see appendix C) that our decomposition into the implied volatility and risk-premium components in (3.2) will translate into:

$$E_t\left(\overline{QV}_{t,\tau}\right) = c^J + c^B_\tau + \sigma^2_{t,\tau} + c^A_\tau V_t \qquad (4.8)$$

The first conclusion that we can make is that this representation does not differ qualitatively from the univariate one in (3.7). All the considerations we had then apply here as well. The real difference comes from the empirical implementation. We will concentrate on these issues in the next section. We make just a few observations to give a flavor of the possible problems.

As we already mentioned, the left-hand side of all these relationships is estimated with the sum of squared log-returns. Since it converges to the quadratic variation of the process, we know that no matter which model we consider an estimate of the quadratic variation is obtained. Hence, this approach is essentially non-parametric. However, the right-hand side of the representation features jumping spot volatility. The Foster and Nelson (1996) results on nonparametric estimation of the spot volatility do not apply in this case. It means, that accurate test of unbiasedness is impossible without some parametric assumptions about the volatility.

Moreover, the implied volatility will reflect a possibility that in a population the underlying asset and the spot volatility will jump. However, a particular sample may not contain any extreme movements and, therefore, the estimated quadratic variation in (4.7) will be smaller than the implied volatility, thus exacerbating the bias in the regression (1.1).[9] We address this issue by considering the largest sample possible in our empirical work. There is still no guarantee, that the chosen sample will reflect the population properties.

5 Empirical Illustration

5.1 Data and econometric strategy

We illustrate the theoretical results of the paper by implementing the forecasting regressions for several assets prominent in the literature: volatilities implied from U.S. equity indices and foreign exchange (FX) rates.

VIX and VXN are CBOE Market Volatility indices. They are constructed from 8 near-the-money, nearby, and second nearby puts and calls written on S&P 100 (OEX) and Nasdaq 100 (NDX) respectively. The implied volatility takes into an account the American feature of the option contract, as well as discrete cash dividends, bid-ask spread and other market frictions, such as nonsynchronicity bias. As a result the level of volatility index is equal to the implied volatility from a synthetic option contract, which always has $\tau = 1$ month to maturity and is exactly at-the-money (ATM). Whaley (2000) provides additional details. Mixon (2001) points out that the two indices constructed by CBOE do not have an immediate interpretation as either volatility per trading day or volatility per calendar day. We rescale the data to obtain volatility per calendar day.

We also use implied volatilities for US dollar exchange rate with British pound (GBP), Japanese yen (JPY), and Swiss franc (CHF).[10] The implied volatilities are marked directly by the traders, who see it in the market. Specifically, they correspond to Black-Scholes implied volatilities for ATM options, where "ATM" means delta-neutral strike.[11] Therefore, we use a slightly different notion of ATM for our equity and FX data. Discussion in the appendix A shows that for delta-neutral strike our main relationship (2.3) is more accurate. Also, FX data better conforms to the Hull-White assumption of zero correlation between the returns and spot volatility. Every day the FX implied

[9] Penttinen (2001) makes a similar observation in the framework of a regime-switching model of volatility.

[10] The data were provided by a major New York investment bank.

[11] Appendix A provides more details about the delta-neutral strike.

volatilities correspond to a one-month option contract and are reported on a volatility per calendar day basis.

It is important to note that this set of implied volatilities does not have the problems mentioned in Christensen and Prabhala (1998), Fleming (1998), and Poteshman (2000): errors in computing implied volatilities, telescoping time to maturity, and varying moneyness. This allows us to uncover the contribution of the volatility risk premium and problems associated with the unobservability of the spot volatility to the biases in the forecasting regression.

We use series which start at different points in time and end on the same date, July 27, 2001. Since the data frequency is daily, we have a substantial amount of observations even for the relatively small NDX sample. Figure 1 plots the time series of the underlying and implied volatilities. Table 1 reports the sample information. Both the underlying and volatility series suggest the presence of jumps in the data generating process. The measures of skewness and kurtosis support this observation.[12]

The presence of the 1987 market crash in the OEX series leads to very large negative skewness in the underlying and very large positive skewness in the implied volatility. Kurtosis is very large for both series as well. NDX has very different properties: it has a larger variance of returns, however, overall, it conforms to normality much more, even when compared with the FX series. It means that extreme movements in NDX are much more symmetric: it can jump not only down, but up as well. The spikes in the FX series can be associated with well-known crises such as the British pound devaluation, and the Asian crisis.

Because of the presence of extreme moves in the data, the estimator RV captures not only the continuous part of volatility in (4.7), but jumps as well. Also, though the results of Foster and Nelson (1996) apply to a broad set of data generating processes, they do not apply to discontinuous processes. Hence, most likely, the estimator HV for the spot volatility will be biased. Moreover, if there is a second volatility factor present, then HV is not capable of disentangling of V_1 and V_2 in (C.9). In this case the regression (1.1) cannot be saved: if we implement it, we should expect the estimated loading on the unobserved volatility to be biased despite the attempts to take into an account the EIV problems.

Our only hope is that even if the jump or second volatility are present in the data generating

[12] Changes in implied volatility are measured as $(\sigma_{t+1,\tau} - \sigma_{t,\tau})/\sigma_{t,\tau}$.

process they are small enough to introduce only a negligible bias.[13] This assumption immediately implies the decomposition (3.7), which corresponds to the standard SV model. Note that the estimator RV is consistent in this case as well, so generalization of the model affects only the properties of the right-hand side of our regression. Because of all these considerations we will proceed as if the data generating process is a standard SV model (or can be accurately approximated buy it).

As a result we adopt the following strategy. We want to estimate the regression:

$$\widehat{\overline{QV}}_{t,\tau} = a + b \cdot \sigma^2_{t,\tau} + c \cdot \hat{V}_t + \epsilon_{t+\tau} \tag{5.1}$$

where the noise term ϵ is correlated with the explanatory variable $\hat{V}$. V is estimated based on the rolling regression results of Foster and Nelson (1996):

$$\hat{V}_t = \frac{1}{\tau} \sum_{i=0}^{\tau-1} r^2_{t-i}(1) \equiv HV_{t-\tau,\tau} \tag{5.2}$$

We estimate $\overline{QV}$ based on its definition in (4.3):

$$\widehat{\overline{QV}}_{t,\tau} = \frac{1}{\tau} \sum_{i=1}^{\tau} r^2_{t+i}(1) \equiv RV_{t,\tau} \tag{5.3}$$

Since we have to estimate both the dependent and independent variables, we have to correct for the EIV problem by incorporating instrumental variables Z_t. We do this in the Hansen (1982) GMM framework with the Hansen and Hodrick (1980) error correction procedure. The just-identified set of moment conditions is:

$$E\left[\left(\widehat{\overline{QV}}_{t,\tau} - a - b \cdot \sigma^2_{t,\tau} - c \cdot \hat{V}_t\right) \otimes Z_t\right] = 0 \tag{5.4}$$

Another potential issue with our approach is the reliability of the GMM standard errors. Since we use daily data, there is a high degree of implied volatility autocorrelation. Asymptotically, the Hansen-Hodrick procedure accounts for it. However, in small samples these standard errors may be downward biased as explained in Hodrick (1992). First, we have very large datasets, so it is very likely that asymptotic standard errors would be appropriate here. Second, as will be seen in the subsequent discussion, even if the standard errors underestimated the true ones, it will not be critical for our conclusions.

[13] The FX data conforms to these requirements the most.

We also report simple ordinary least squares (OLS) estimates of the regression (5.1) to compare our results with the basic approach. We provide two sets of estimates. One with the coefficient c restricted to be equal to zero, i.e. assuming the informational efficiency of the implied volatility, and the other one with the unconstrained c.

5.2 Empirical results

Table 2 reports the estimates of (5.1). The first two columns contain the OLS estimates with c constrained to be equal to zero and unconstrained respectively. The third column of Table 1 reports GMM estimates.

We observe a typical picture for the OLS estimates: both the unbiasedness and informational efficiency hypotheses are rejected in most cases. The exceptions are NDX and GBP. In the case of NDX, the implied volatility loading is roughly equal to one only when state-varying part of the volatility risk premium is assumed to be equal to zero. However, the second column of the table shows that this assumption is not supported by the data: the historical volatility coefficient is significantly different from zero. The picture is different for GBP: it is clear that c, coefficient related to the random part of the volatility risk premium, is not significantly different from zero. This result combined with our decomposition (3.2) implies that GBP has a constant volatility risk premium.

For the GMM results, we observe that the implied volatility slope coefficient estimate $\hat{b}$ is almost exactly equal to 1, as predicted by (3.2). The slope coefficient $\hat{c}$ for the spot volatility, which reflects the volatility risk premium, is negative which is consistent with the combination of the decomposition (3.2) and extant empirical results mentioned in section 3.

The only exception is NDX. The value of $\hat{c}$ is positive indicating a potentially positive covariance term in our decomposition (3.2). While we cannot make definitive statements without estimating the structural model of NDX dynamics, this result can be interpreted as the effect of the volatility jump risk premium. Indeed, Eraker, Johannes, and Polson (2000) find that a stochastic volatility process for NDX has very different properties than that of S&P: the jumps in NDX are much more frequent and larger in magnitude. Schwert (2001) provides additional highlights of the VXN jumpiness by comparing VXN and VIX and RV's computed based on NDX and OEX.

The GMM estimates are very imprecise. It is not surprising because unobservability of QV and

V adds a lot of noise: recall the error terms ϵ^1 and ϵ^2 in the regression (3.8). Their variability will naturally affect the regression standard errors.

Results of Barndorff-Nielsen and Shephard (2001) give us insight into this problem. Using their asymptotic theory, we can say that ϵ^1 – the estimation error of quadratic variation – is normally distributed with variance $\frac{2}{3}\sum_{i=1}^{\tau} r_{t+i}^4(1)$.[14] We use this result in finite sample as an indication of what kind of error could be introduced.

Figure 2 plots the ratio of the stochastic standard deviation of ϵ^1 to the standard deviation of the residuals from the OLS regression (5.1) with unconstrained c. Therefore, we can evaluate the incremental size of the additional error as a percentage of the usual regression error if the quadratic variation were observed. It is clear from the plot that the error size tracks the volatility pretty well. It means that at the times of the market turmoil, the estimate of quadratic variation is the least precise. The median additional error ranges from 17% for OEX to 39% for CHF. Moreover, it can be as high as 400%. This explains the large standard errors we see in the table 2: realized volatility, RV, is a very noisy estimate of the quadratic variation.

One way to improve the efficiency of estimates is to use high frequency data in the spirit of Andersen, Bollerslev, Diebold, and Labys (2000). This is something that has not been previously considered in previous tests. This is an appealing direction to pursue, however a trade-offs between higher data frequency and efficiency of estimates are not completely understood. Hence, more work needs to be done before we could implement these improvements.

6 Conclusion

This paper is concerned with the analysis of the reasons for the failures of the tests of the implied volatility as an unbiased forecast of the future realized volatility. Typically, the extant literature argues that these failures are related to the measurement errors in the implied volatility and to the imperfect econometric procedures. We show that even if we remove these problems, we are still bound to reject the unbiasedness hypothesis for two reasons.

One reason, theoretical, is that previous studies did not take into an account the volatility risk premium required by economic agents. Once we take it into an account, we find that the

[14]Foster and Nelson (1996) obtain asymptotics for ϵ^2. Their asymptotic variance, however, relies on the knowledge of structural parameters of the underlying model and, therefore, can not be implemented in our case.

implied volatility has to be discounted by this risk premium, which is related to the unobserved spot volatility. This brings us to the second, econometric, reason of the unbiasedness hypothesis testing failures. We show how the measures of historical and future realized volatility, which are typically used in these tests, are related to the estimators of the latent spot volatility and averaged integrated quadratic variation of the asset-return generating process. The need to estimate these unobservable variables introduces the error-in-the-variables problem, which affects the results of the unbiasedness hypothesis testing and was not recognized before in the literature. This problem can be remedied in the instrumental variables estimation framework.

We illustrate the above ideas on the dataset of synthetically constructed implied volatilities on the S&P 100 and Nasdaq 100 indices and directly observed implied volatilities on foreign currency exchange rates. The appealing feature of these data, is that by construction they do not have the measurement problems outlined in the previous work. To avoid the peso problem related to absence of extreme asset movements (jumps) in particular samples, we select the longest sample period as compared to previous studies. Thus we can concentrate only on the deficiencies pointed out in this paper. We show that if we take into an account the volatility risk premium and the estimation error, the loading on the implied volatility in the forecasting regression is equal to 1.

However, the estimates are very imprecise. The reason is that we have to estimate the unobservable quadratic variation and spot volatility of the series by realized volatility and historical volatility respectively. These estimates introduce a large stochastic error. We were able to compute approximate standard errors for the difference between the realized volatility and quadratic variation. They can be as high as 400% relative to the standard deviation of the error term in the regular OLS.

We, therefore, conclude, that if we do not impose particular assumptions about the data generating process, we will not be able to reliably verify or reject the unbiasedness hypothesis. If we decide to impose some assumptions, we might as well use a realistic model of asset returns to forecast the future volatility based on modern filtering methods. Some progress along these lines is reported in, for instance, Chernov and Ghysels (2000), and Eraker, Johannes, and Polson (2000).

A Linearity of the Black-Scholes formula

This appendix describes the conditions under which the Black-Scholes formula is linear in volatility. Consider the second derivative of the Black-Scholes formula with respect to variance, $\Sigma = \sigma^2_{t,\tau}$:

$$\frac{\partial^2 C^{BS}(S_t, \Sigma, r, K, \tau)}{\partial \Sigma^2} = \frac{S_t \sqrt{\tau}}{4\Sigma^{3/2}} n(d_1)\left(d_1 d_2 - 1\right) \tag{A.1}$$

with

$$d_{1,2} = \frac{\log S_t/K + (r \pm \Sigma/2)\tau}{\sqrt{\Sigma\tau}} \tag{A.2}$$

and $n(\cdot)$ – probability density of the standard normal distribution. The strike K, at which the second derivative in (A.1) is equal to zero, ensures linearity of the formula in variance Σ. In other words, K solves

$$d_1 d_2 - 1 = 0 \tag{A.3}$$

Substituting in the expressions from (A.2), we obtain:

$$K_- = S_t \exp\left(r\tau - \frac{1}{2}\Sigma\tau\right) \tag{A.4}$$

$$K_+ = S_t \exp\left(r\tau + \frac{1}{2}\Sigma\tau\right) \tag{A.5}$$

as the solution. It is easy to verify that K_+ is actually delta-neutral, i.e. when the call's Black-Scholes delta using the corresponding implied volatility

$$\frac{\partial C^{BS}(S_t, \Sigma, r, K, \tau)}{\partial S_t} = N(d_1) \tag{A.6}$$

is equal to the negative of the put's Black-Scholes delta using the same volatility

$$-\frac{\partial P^{BS}(S_t, \Sigma, r, K, \tau)}{\partial S_t} = N(-d_1) \tag{A.7}$$

Note that for small maturities, τ, both strikes are approximately equal to the current stock price.

B Expected value of the average integrated volatility

We are interested in computing $E^M_t(\bar{V}_{t,\tau})$ where V_t follows the jump-diffusion process

$$dV_t = \left(\theta^M - \kappa^M V_t\right) dt + \sigma(V_t) dW^M_{Vt} + J^M_{Vt} dN^M_{Vt} \tag{B.1}$$

under some probability measure M. The diffusion part corresponds to the standard generic SV specification in (3.4). The jump part has two components: the jump size J_V^M, which is assumed to be exponentially distributed with mean μ_V^M and the counting process N_V^M with intensity λ_V^M.[15]

Following the standard approach, we first apply Itô's lemma to $\exp(\kappa t)V_t$:

$$d\left(e^{\kappa t}V_t\right) = \kappa e^{\kappa t}V_t dt + e^{\kappa t}dV_t = \theta e^{\kappa t}dt + e^{\kappa t}\sigma(V_t)dW_{Vt} + e^{\kappa t}J_{Vt}dN_{Vt} \tag{B.2}$$

Now we can compute V_s for $s \geq t$:

$$V_s = e^{-\kappa(s-t)}V_t + \theta e^{-\kappa s}\int_t^s e^{\kappa u}du + e^{-\kappa s}\int_t^s e^{\kappa u}\sigma(V_u)dW_{Vu} + e^{-\kappa s}\int_t^s e^{\kappa u}J_{Vu}dN_{Vu} \tag{B.3}$$

Therefore,

$$\begin{aligned} E_t\left(V_s\right) &= e^{-\kappa(s-t)}V_t + \theta e^{-\kappa s}\int_t^s e^{\kappa u}du + e^{-\kappa s}\int_t^s e^{\kappa u}E_t\left(J_{Vu}dN_{Vu}\right) \\ &= e^{-\kappa(s-t)}V_t + \frac{\theta}{\kappa}\left(1 - e^{-\kappa(s-t)}\right) + \frac{\lambda_V\mu_V}{\kappa}\left(1 - e^{-\kappa(s-t)}\right) \end{aligned} \tag{B.4}$$

Now, using Fubini's formula, we can compute the expected value of integrated volatility:

$$E_t(\bar{V}_{t,\tau}) = \frac{1}{\tau}\int_t^{t+\tau} E_t\left(V_s\right)ds \tag{B.5}$$

Then (B.4) implies:

$$E_t^M(\bar{V}_{t,\tau}) = A_\tau^M V_t + B_\tau^M \tag{B.6}$$

where

$$A_\tau^M = -\frac{1}{\kappa^M\tau}\left(e^{-\kappa^M\tau} - 1\right) \tag{B.7}$$

$$B_\tau^M = \frac{\theta^M + \lambda_V^M\mu_V^M}{\kappa^M}(1 - A_\tau^M) \tag{B.8}$$

and M emphasizes a particular probability measure. In particular a blank or P will denote the physical measure and an asterisk or Q the risk-neutral one.

This result does not depend on the functional form of $\sigma(\cdot)$, therefore the results are identical for the Heston and CEV models.[16] In the case of the Hull-White model, $\theta = 0$ and, hence $B_{t,\tau} = 0$.

[15] In the course of derivation we drop the superscript M to avoid cumbersome notation.

[16] Obviously, in these cases the jump part would be absent: $\lambda_V^M = \mu_V^M = 0$.

The volatility is known to be quite persistent, which manifests itself in empirical studies by the values of $\kappa\tau$ close to zero when τ stands for 1 day. Therefore we can apply the Taylor formula:

$$A_\tau^M = 1 + o(1) \tag{B.9}$$

$$B_\tau^M = o(1) \tag{B.10}$$

C Extensions to multiple SV factors and jumps

We discuss extensions into multiple factors in anticipation of the empirical work, which tests the unbiasedness hypothesis. Namely, the realized volatility is estimated as a sum of squared log-returns over τ periods. The theoretical motivation for such an approach is that such a sum converges to the quadratic variation of the stochastic process governing the dynamics of log-return. For instance, if log-returns follow the process in (3.3)-(3.4), then its quadratic variation is equal to $\tau\bar{V}_{t,\tau}$. If, however, the log-return process involves multiple SV factors and/or jumps, then quadratic variation will be equal to the sum of integrated volatilities and jump terms. Andersen, Bollerslev, Diebold and Labys (2000) provide several examples along these lines. Our concern is that, while the sum of squared returns can still be interpreted as integrated volatility, the right-hand terms in the expression (3.7) will require modification because they are based on particular assumptions about the returns model.

C.1 Models with two volatility factors

In this section we expand the model in (3.3)-(3.4) by allowing for a second SV factor:

$$\frac{dS_t}{S_t} = \delta_t dt + \sqrt{V_{1t} + V_{2t}} dW_{St} \tag{C.1}$$

$$dV_{it} = (\theta_i - \kappa_i V_{it})\, dt + \sigma_i(V_{it}) dW_{it}, \; i = 1, 2 \tag{C.2}$$

We can of course let i vary from 1 to N, but such a generality will not give us any additional insights. Therefore, we limit ourselves to a model with two stochastic volatility factors. In order to retain the Hull-White result in (2.1) we have to require that the V_i's are independent of S. We note parenthetically, that for the ATM near maturity options required for (2.3), the presence of such dependence may not affect the results much. Also, for the simplicity of exposition, we assume that the volatility factors are not correlated.

When a model specification involves multiple SV factors we have to distinguish between the individual average integrated volatility

$$\bar{V}_{i,t,\tau} = \frac{1}{\tau}\int_t^{t+\tau} V_{is}ds \tag{C.3}$$

and quadratic variation of the log-return process $r_t^\tau = \log(S_{t+\tau}/S_t)$ over the interval $[t, t+\tau]$:

$$QV_{t,\tau} = [r,r]_t^\tau = \int_t^{t+\tau} \left(V_{1s} + V_{2s}\right) ds = \tau \left(\bar{V}_{1,t,\tau} + \bar{V}_{2,t,\tau}\right) \tag{C.4}$$

We also introduce the notion of averaged quadratic variation to simplify some of the notations. It is simply quadratic variation divided by the length of the respective time interval:

$$\overline{QV}_{t,\tau} = \frac{1}{\tau} QV_{t,\tau} \tag{C.5}$$

Under the above assumptions we can follow the derivation of Hull and White (1997) and show that, conditional on $\overline{QV}_{t,\tau}$, r_t^τ has normal distribution with variance $QV_{t,\tau}$. Therefore, the analogue of the Hull and White option pricing formula (2.1) would be:

$$C^{HW}(S_t, V_{1t}, V_{2t}, r, K, \tau) = E_t^Q\left\{C^{BS}(S_t, \overline{QV}_{t,\tau}, r, K, \tau)\right\} \tag{C.6}$$

which implies

$$E_t^Q(\overline{QV}_{t,\tau}) = \sigma_{t,\tau}^2 \tag{C.7}$$

Our decomposition into the expectation under the physical measure and the volatility risk premium component in (3.2) still holds in this case. We just have to substitute $\bar{V}_{t,\tau}$ by $\overline{QV}_{t,\tau}$, which is not surprising because these two objects coincide in the univariate case. The decomposition in (3.7) will of course change because it relies on particulars of the assumed stochastic process of returns. By analogy with (3.6) and (3.7), one can show that:

$$\begin{aligned} \mathrm{Cov}_t\left(\xi_{t,\tau}, \overline{QV}_{t,\tau}\right) &= E_t^Q\left(\overline{QV}_{t,\tau}\right) - E_t\left(\overline{QV}_{t,\tau}\right) = E_t^Q\left(\bar{V}_{1,t,\tau} + \bar{V}_{2,t,\tau}\right) - E_t^Q\left(\bar{V}_{1,t,\tau} + \bar{V}_{2,t,\tau}\right) \\ &= -c_{1\tau}^A V_{1t} - c_{1\tau}^B - c_{2\tau}^A V_{2t} - c_{2\tau}^B \end{aligned} \tag{C.8}$$

and, therefore,

$$E_t\left(\overline{QV}_{t,\tau}\right) = c_\tau^B + \sigma_{t,\tau}^2 + c_{1\tau}^A V_{1t} + c_{2\tau}^A V_{2t} \tag{C.9}$$

where $c_\tau^B = c_{1\tau}^B + c_{2\tau}^B$.

C.2 Models with jumps

C.2.1 The Merton jump diffusion model

We start our discussion of jump diffusion with the Merton (1976) constant volatility case. This example will allow us to gain intuition about issues specific to the jump component. We make slight modifications to the Merton model and assume that asset prices dynamics can be described by the following process:

$$\frac{dS_t}{S_t} = \delta_t dt + \sqrt{V} dW_{St} + (J_t dN_t - \lambda \mu dt) \tag{C.10}$$

where N is the Poisson (counting) process with intensity λ and jump size J is distributed normally with mean μ and variance σ^2. Also, contrary to Merton, we assume that the jump risk is systematic, i.e. it is priced in the economy. In particular, we assume that intensity of the jump process is equal to λ^* and mean jump to μ^* under the risk-neutral measure Q.

Based on the above assumptions and Merton (1976) we can immediately write down the call option price:

$$\begin{aligned} C^M(S_t, V, r, K, \tau) &= \sum_{n=0}^{\infty} \frac{e^{-\lambda^* \tau}(\lambda^* \tau)^n}{n!} C^{BS}\left(S_t, V + \frac{n\sigma^2}{\tau}, r - \lambda^* \mu_J^* + \frac{n \log(1+\mu_J^*)}{\tau}, K, \tau\right) \\ &= E_t^Q\left\{C^{BS}\left(S_t, V + \frac{n\sigma^2}{\tau}, r - \lambda^* \mu_J^* + \frac{n \log(1+\mu_J^*)}{\tau}, K, \tau\right)\right\} \end{aligned} \tag{C.11}$$

where expectation is taken with respect the distribution of the random variable n, which represents number of jumps over the interval τ. The risk-free interest rate does not enter this formula the same way volatility does. Therefore, we can not produce a relationship similar to the ones in (2.3) and (C.7). If, however, μ^* is equal to zero such a relationship follows immediately:

$$E_t^Q\left(V + \frac{n\sigma^2}{\tau}\right) = \sigma_{t,\tau}^2 \tag{C.12}$$

The first question is how reasonable the assumption $\mu^* = 0$? The evidence in the extant empirical research, such as Andersen, Benzoni, and Lund (2000), Bakshi, Cao, and Chen (2000), Eraker, Johannes, and Polson (2000), and Pan (2000), is not conclusive because these studies have to impose various restrictive assumptions to identify parameters. Thus, we will maintain this hypothesis with its potential shortcoming in mind.

Now, returning to the relationship in (C.12), we can notice that it is fairly simple, because we can immediately compute the expectation in the left-hand side:

$$E_t^Q\left(V + \frac{n\sigma^2}{\tau}\right) = V + \frac{\sigma^2}{\tau}E_t^Q(n) = V + \lambda^*\sigma^2 \tag{C.13}$$

It is, therefore, clear that empirically the relationship (C.12) does not give us much because we have a constant in the left-hand side and time-varying variable in the right-hand one. This result is not surprising since it is just one manifestation of the rejection of the Merton model. The data calls for at least one stochastic volatility factor. We will explore an SVJ model later in this section. What we would like to see in the Merton model is whether the jump component lends itself to the same quadratic variation interpretation we observed with SV models. Indeed, the quadratic variation of the process in (C.10) is equal to:

$$QV_{t,\tau} = V\tau + \sum_{t<s\leq t+\tau} J_s^2 \Delta N_s \tag{C.14}$$

Then,

$$E_t^Q(\overline{QV}_{t,\tau}) = V + \frac{1}{\tau}E_t^Q\left(\int_{\mathbf{R}\backslash 0} x^2 N_\tau(dx)\right) = V + \int_{-\infty}^{\infty} x^2\lambda^* \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2} = V + \lambda^*\sigma^2 \tag{C.15}$$

For details and reference see, for instance, Chernov (2000b). The last relationship combined with (C.12) yields the same relationship as in (C.7). Hence, jump components yield the same initial relationship as stochastic volatility models given the assumption of zero mean jump under the risk-neutral measure.

C.2.2 The SVIJ model

We also have to explore the functional form our decomposition (3.2) takes under the jump processes. We address this issue using more empirically plausible SVIJ model (see, e.g. Eraker, Johannes, and Polson, 2000). Namely, we assume the price process to follow:

$$\frac{dS_t}{S_t} = \delta_t dt + \sqrt{V_t}dW_{St} + (J_t dN_t - \lambda\mu dt) \tag{C.16}$$

$$dV_t = (\theta - \kappa V_t)\,dt + \sigma(V_t)dW_{Vt} + J_{Vt}dN_{Vt} \tag{C.17}$$

Refer to the descriptions of (B.1) and (C.10) for details.

First of all, we have to verify the relationship (C.7) for this process. It is not difficult given all the previous discussions. Indeed, combining the derivation steps in Hull and White (1987) with (C.4) and (C.15) we obtain:

$$C^{HW-M}(S_t, V_t, r, K, \tau) = E_t^Q \left\{ C^{BS}(S_t, \overline{QV}_{t,\tau}, r, K, \tau) \right\} \tag{C.18}$$

One just has to remember that the expectation is taken with respect to the distribution of $\overline{QV}_{t,\tau}$, which involves distributions of average integrated volatility and number of jumps. This formula immediately gives expression identical to (C.7).

Now, as in all the models before, we would like to establish the functional form of our decomposition (3.2) (with $\bar{V}$ substituted by $\overline{QV}$). This is also easy to do because:

$$\overline{QV}_{t,\tau} = \bar{V}_{t,\tau} + \frac{1}{\tau} \sum_{t<s\leq t+\tau} J_s^2 \Delta N_s \tag{C.19}$$

and, by analogy with (C.15):

$$E_t^M \left(\overline{QV}_{t,\tau} \right) = A_\tau^M V_t + B_\tau^M + \lambda^M \sigma^2 \tag{C.20}$$

Therefore, similarly to (3.6) and (3.7), one can show that:

$$E_t \left(\overline{QV}_{t,\tau} \right) = c^J + c_\tau^B + \sigma_{t,\tau}^2 + c_\tau^A V_t \tag{C.21}$$

where c^J is the constant involving jump parameters:

$$c^J = (\lambda - \lambda^*)\, \sigma^2 \tag{C.22}$$

It does not depend on the option's time to maturity. Hence if the true data generating process follows a process involving jumps, the corresponding regression will have a non-zero intercept even in the extreme case of $\tau = 0$ as long as the jump risk is systematic.

References

[1] Andersen, T., L. Benzoni, and J. Lund (2000): Estimating Jump-Diffusions for Equity Returns, *Working paper*, Northwestern University

[2] Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2000): Distribution of the Exchange Rate Volatility, *NBER working paper*

[3] Barndorff-Nielsen, O. and N. Shephard (2001): How accurate is the asymptotic approximation to the distribution of relaised volatility?, *Working paper*, University of Aarhus

[4] Bates, D. (1996): Testing Option Pricing Models, in G.S. Maddala and C.R. Rao (eds.) *Handbook of Statistics, Vol. 14*, Elsevier Science B.V.

[5] Benzoni, L. (1998): Pricing Options under Stochastic Volatility: An Econometric Analysis, *Working Paper*, Northwestern University

[6] Black, F. and M.S. Scholes (1973): The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637-659.

[7] Canina, L. and S. Figlewski (1993): The informational content of implied volatility, *Review of Financial Studies*, **6**, 659-681.

[8] Chernov, M. (2001): Empirical Reverese Engineering of the Pricing Kernel, *Working paper*, Columbia Business School

[9] Chernov, M. and E. Ghysels (2000): A Study towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation, *Journal of Financial Economics*, **56**, 407-458

[10] Christensen, B. and N. Prabhala (1998): The Relation Between Implied and Realized Volatility, *Journal of Financial Economics*, **50**, 125-150

[11] Day, T. and C. Lewis (1992): Stock Market Volatility And the Information Content of Stock Index Options, *Journal of Econometrics*, **52**, 267-287

[12] Eraker, B., M. Johannes, and N. Polson (2000): Return Dynamics with Jumps to Stochastic Volatility and Returns, *Discussion Paper*, Graduate School of Business, University of Chicago

[13] Fama, E. (1984): The Information in the Term Structure, *Journal of Financial Economics*, **13**, 509-528

[14] Fama, E. and J. Macbeth (1973): Risk, Return, and Equilibrium: Empirical Tests, *Journal of Political Economy*, , 607-636

[15] Fleming, J. (1998): The Quality of Market Volatility Forecasts Implied by S&P 100 Index Option Prices, *Journal of Empirical Finance*, **5**, 317-345

[16] Foster, D. and D. Nelson (1996): Continuous Record Asymptotics for Rolling Sample Estimators, *Econometrica*, **64**, 139-174

[17] Hansen, L. (1982): Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, **50**, 1029-1054.

[18] Hansen, L. and R. Hodrick (1980): Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis, *Journal of Political Economy*, **88**, 829-853

[19] Heston, S.L. (1993): A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies*, **6**, 327-343

[20] Hodrick, R. (1992): Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies,* **5**, 357-386

[21] Hull, J. and A. White (1987): The Pricing of Options on Assets with Stochastic Volatility, *Journal of Finance*, **42**, 281-300

[22] Jones, C. (2000): The Dynamics of Stochastic Volatility, *Working paper*, University of Rochester

[23] Jorion, P. (1995): Predicting volatility in the foreign exchange market, *Journal of Finance*, **50**, 507-528

[24] Lamoureux, C. and W. Lastrapes (1993): Forecasting Stock-return Variance: Toward an understanding of stochastic implied volatilities, *Review of Financial Studies*, **6**, 293-326

[25] Ledoit, O. and Santa-Clara, Pedro (1999): Relative Pricing of Options with Stochastic Volatility, Working Paper, UCLA.

[26] Merton, R. (1976): Option Pricing when the Underlying Stock Returns are Discontinuous, *Journal of Financial Economics*, **3**, 124-144

[27] Mixon, S. (2001): A Warning about Calendar Conventions and Implied Volatility, *Working paper,* UBS Warburg

[28] Pan, J. (2000): "Integrated" Time-Series Analysis of Spot and Option Prices, *Working paper*, MIT Sloan

[29] Penttinen, A. (2001): The sensitivity of implied volatility to expectations of jumps in volatility: An explanation for the illusory bias in implied volatility as a forecast, *Working paper*, New York University

[30] Poteshman, A. (2000): Forecasting Future Variance from Option Prices, *Working paper*, University of Illinois at Urbana-Champaign

[31] Protter, P. (1990): *Stochastic Integration and Differential Equations*, Springer-Verlag

[32] Schwert, G.W. (2001): Stock Volatility in the New Millenium: How wacky is Nasdaq? *Working paper*, No. 8436, NBER

[33] Whaley, R. (2000): The Investor Fear Gauge, *Working paper*, Duke University

Table 1 : The Sample Statistics

We report the sample period and sample statistics for the five time series used in the paper. The statistics are computed for the logarithmic changes in the underlying asset and for the simple changes in the respective implied volatility.

Series	Dates	Sample size	Changes in underlying				Changes in volatility			
			Mean	Std. dev.	Skewness	Kurtosis	Mean	Std. dev.	Skewness	Kurtosis
OEX	01/02/86 – 07/27/01	3906	0.0483	1.1447	-2.5739	56.2392	0.0023	0.0789	16.7222	644.9706
NDX	01/03/95 – 07/27/01	1634	0.0989	2.3809	0.0966	6.6949	0.0016	0.0453	0.4560	5.4240
GBP	10/05/84 – 07/27/01	4054	-0.0041	0.6629	-0.1056	8.7206	0.0007	0.0422	1.8588	15.9908
JPY	10/05/84 – 07/27/01	4060	-0.0171	0.7204	-0.5461	7.4323	0.0014	0.0552	1.8679	15.1592
CHF	10/05/84 – 07/27/01	4060	-0.0089	0.7637	-0.2262	5.9534	0.0007	0.0419	1.6147	15.8544

Table 2 : Estimation of the Volatility Forecasting Regression

We report the estimation results for the regression:

$$E\left[\left(\widehat{QV}_{t,\tau} - a - b \cdot \sigma^2_{t,\tau} - c \cdot \hat{V}_t\right) \otimes Z_t\right] = 0$$

The first column reports the standard OLS estimation with $c = 0$. The second column reports the OLS results with uncontrained c. In the last column the regression is estimated via GMM using the Hansen-Hodrick standard errors. Instrumental variables Z_t correct for the error-in-the-variables (EIV) problem.

Parameters	OLS–$c=0$	OLS	GMM
OEX			$Z_t = \left(1, \hat{V}_{t-24}, \sigma^2_{t-46,22}\right)$
$\hat{a}$	0.6117	0.6177	0.1035
s.e.	(0.0457)	(0.0458)	(0.1839)
$\hat{b}$	0.3258	0.2823	0.9907
s.e.	(0.0267)	(0.0369)	(1.0813)
$\hat{c}$	0	0.0388	-0.1384
s.e.	–	(0.0227)	(1.2082)
NDX			$Z_t = \left(1, r^2_{t-16}(1), \hat{V}_{t-26}\right)$
$\hat{a}$	-0.7400	-0.5591	-0.0909
s.e.	(0.0485)	(0.0455)	(0.5769)
$\hat{b}$	1.0598	0.7757	1.0381
s.e.	(0.0105)	(0.0190)	(0.3503)
$\hat{c}$	0	0.2059	0.1495
s.e.	–	(0.0119)	(0.1837)
GBP			$Z_t = \left(1, \hat{V}_{t-22}, \hat{V}_{t-44}\right)$
$\hat{a}$	0.0056	0.0055	0.0277
s.e.	(0.0076)	(0.0081)	(0.0472)
$\hat{b}$	0.9481	0.9496	1.1121
s.e.	(0.0197)	(0.0367)	(0.4127)
$\hat{c}$	0	-0.0012	-0.2359
s.e.	–	(0.0230)	(0.3545)
JPY			$Z_t = \left(1, r^2_{t-25}(1), \sigma^2_{t-46,22}\right)$
$\hat{a}$	0.1020	0.1137	0.1099
s.e.	(0.0096)	(0.0096)	(0.0919)
$\hat{b}$	0.7644	0.5523	1.0141
s.e.	(0.0227)	(0.0343)	(1.1154)
$\hat{c}$	0	0.1735	-0.2608
s.e.	–	(0.0212)	(0.8025)
CHF			$Z_t = \left(1, \sigma^2_{t-38,22}, \hat{V}_{t-42}\right)$
$\hat{a}$	0.0985	0.1233	0.2372
s.e.	(0.0116)	(0.0121)	(0.1497)
$\hat{b}$	0.8088	0.5849	1.0464
s.e.	(0.0264)	(0.0424)	(1.2423)
$\hat{c}$	0	0.1542	-0.5467
s.e.	–	(0.0230)	(1.0303)

Figure 1. The Data

On this figure we plot the data series used in the empirical example.

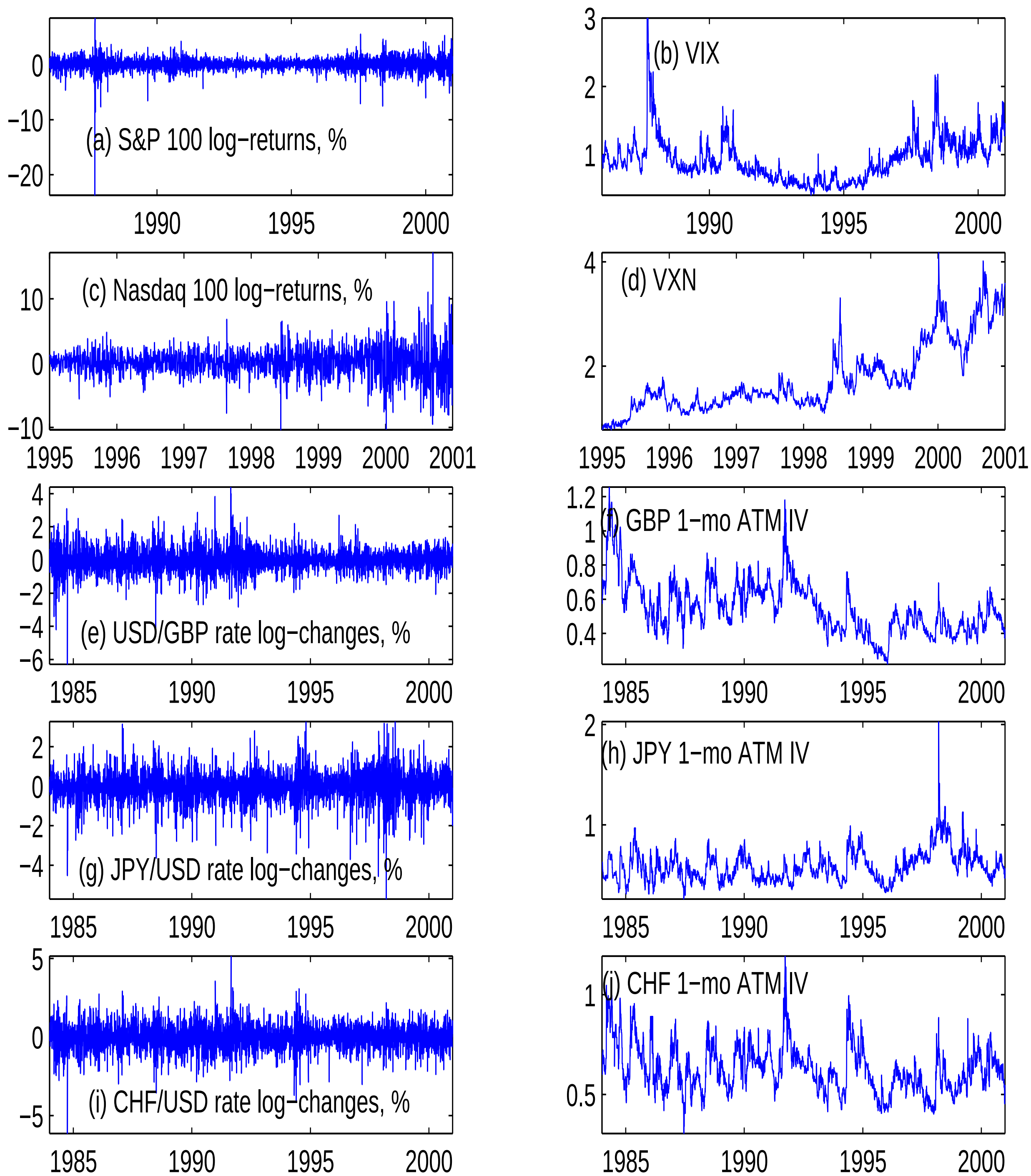

Figure 2. The precision of the quadratic variation estimate

This figure isllustrates the imprecision of the realized volatility, RV, as the estimate of the averaged quadratic variation $\overline{QV}$. The estimate error is denoted ϵ^1. We plot the ratio of the standard deviation of ϵ^1 to the standard deviation of the error term from the regular OLS regression. This allows us to see the perecentage increase in the standard errors of the volatility forecasting regression because of unobservability of $\overline{QV}$.