# Diagnostic Tests for Volatility Models*

Valentina Corradi[†]
*University of Warwick*

Walter Distaso[‡]
*Imperial College Business School*

December 2010

**Abstract**

This paper provides a nonparametric testing procedure for continuous time volatility models, under minimal assumptions. In particular, apart from standard regularity conditions, no assumptions are made on the functional forms of either the drift or the variance term. Our test is constructed by comparing two estimators of integrated volatility: one is a kernel estimator of the instantaneous variance, averaged over the sample realization; the other is a localized version of realized volatility. Under the hypothesis of the class of endogenous volatility model, the test statistic has a standard normal limiting distribution, while under the alternative hypothesis of stochastic volatility it diverges at an appropriate rate. The findings from a Monte Carlo study indicate that the suggested tests have good finite sample properties.

*Keywords and phrases:* Diffusions, local times, realized volatility.
*JEL classification:* G13, G12, C22.

*We are grateful to the Co-Editors, two anonymous referees, Karim Abadir, Carol Alexander, Federico Bandi, Andrea Buraschi, James Davidson, Mahmoud El-Gamal, Marcelo Fernandes, Marc Hallin, Javier Hidalgo, Oliver Linton, Rod McCrorie, Nour Meddahi, Antonio Mele, Joon Park, Hashem Pesaran, Peter C. B. Phillips, Neil Shephard, Viktor Todorov, Nizar Touzi, Nick Webber and seminar participants at the University of Exeter, Cass Business School, LSE, Cambridge University, University of Leicester, Rice University, Tanaka Business School, Texas A&M University, Università di Rimini, Università di Venezia, Cà Foscari, 2004 SIS Conference in Bari, 2005 Winter Meeting of the Econometric Society, 2007 Duke Conference on Forecasting in Macroeconomics and Finance,, 2007 European Meeting of the Econometric Society in Budapest, the 2007 Brazilian Time Series and Econometrics School in Gramado and at the Conference in honor of the 60th birthday of Peter C. B. Phillips in Singapore, 2008, for very helpful comments and suggestions. The authors gratefully acknowledge financial support from the ESRC, grant codes R000230006 and RES-062-23-0311.

[†]University of Warwick, Department of Economics, Coventry CV4 7AL, UK, `v.corradi@warwick.ac.uk`.

[‡]Imperial College, Business School, Finance Group, London, SW7 2AZ, UK, email: `w.distaso@imperial.ac.uk`.

# 1 Introduction

Continuous time diffusion models are nowadays an accepted paradigm in finance. Since the introduction of the simple homoskedastic diffusion:

$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t \tag{1}$$

in the 1970s through the work of Merton, they have been an extremely popular choice to characterize the behaviour of financial assets and have formed the basic infrastructure for pricing contingent claims.

A lot of work has been done in the past twenty years in order to make more realistic and empirically plausible assumptions on the data generating process. Particular attention has been put on the volatility coefficient $\sigma$, which is a key input in any derivative pricing formula and is of crucial importance in financial risk management.

The first important departure from the basic diffusion described in (1) has been the introduction of time varying volatility models; this has been accomplished by modeling volatility as a known measurable function of the underlying asset. Notable examples are the popular models by Cox, Ingersoll and Ross (1985) and Black, Derman and Toy (1990). In the paper, we refer to this class as either endogenous volatility or one factor models, since there is only one source of randomness driving the underlying variable.

A second strand of literature has treated volatility as an additional state variable (albeit latent), driven by random shocks which can be correlated with those driving the asset. These models, labeled stochastic volatility models, have been introduced by Hull and White (1987); further important contributions to this literature include Stein and Stein (1991) and Heston (1993). More recently, an interesting literature has used stochastic volatility models to study the variation of the yield curve and of the stock market as a function of multiple macroeconomic and unobservable factors. See, for example, the work by Ang and Piazzesi (2003), Piazzesi (2005), Bibkov and Chernov (2006), Buraschi and Jiltsov (2006) and Corradi, Distaso and Mele (2009).

Overall, the correct choice between the class of one factor models and of stochastic volatility models is important for the pricing of derivative securities. In fact, when there is only one source of randomness driving the traded asset, derivative securities can be priced by no arbitrage, via a pure replication argument, without the need to specify the behavior of the risk premium. However, in the presence of stochastic volatility, one needs to specify a risk premium, in order to recover a pricing function.

This paper provides a testing procedure which allows to discriminate between the *classes* of one factor and stochastic volatility models, under minimal assumptions. Apart from standard regularity conditions, no assumptions are made on the functional forms of either the drift or the diffusion term. Being able to choose between classes of models, our testing procedure is nonparametric in nature.

We derive our testing procedure by comparing two estimators of the spot volatility of the underlying asset. One is a kernel estimator of the instantaneous variance, the other is a localized version of realized volatility. Under the null hypothesis of a one factor model, both estimators are consistent. Under the alternative hypothesis, the kernel type estimator is not consistent, while realized volatility retains the consistency property. We show that the test statistic converges to a standard normal distribution under the null hypothesis and diverges under the alternative. The derived asymptotic theory is based on both the time interval between successive observations approaching zero, and the time span increasing. Because the limiting distribution of the statistic is standard normal, our test is very easy to implement.

Tests for the null hypothesis of one factor models have been already suggested in the financial literature. In fact, one factor models have three important implications for options: (i) monotonicity. Call (put) option prices are monotonically increasing (decreasing) in the price of the underlying asset; (ii) perfect correlation. As there is only one Brownian motion driving the behaviour of the asset price, option prices are perfectly correlated with the underlying asset prices; (iii) redundancy. Option payoffs can be perfectly replicated with the risk free asset and the underlying asset, and so are redundant securities. Bakshi, Cao and Chen (2000) have derived testable implications of the monotonicity and perfect correlation properties, while Buraschi and Jackwerth (2001) have suggested a test for redundancy.

However, there are important differences between the hypotheses tested in the two papers mentioned above and our procedure.

First, as Bergman, Grundy and Wiener (1996) pointed out, there are counterexamples to the monotonicity and redundancy properties. Second, in the papers cited above, the construction of the tests requires the use of option data. Therefore, one has to choose both the moneyness and the maturity of the options, and the outcome of the test may be rather sensitive to that. On the contrary, the test suggested here simply requires the availability of high frequency observations on the price of the underlying asset. We believe this is an important advantage of our procedure.

The rest of this paper is organized as follows. In Section 2, the testing procedure is outlined and the relevant limit theory is derived. Section 3 reports the findings from a Monte Carlo exercise, in order to assess the finite sample behavior of the proposed tests. Concluding remarks are given in Section 4. All the proofs are gathered in the Appendix.

In the paper, $\xrightarrow{d}$ denotes convergence in in distribution. We write $1_{\{\cdot\}}$ for the indicator function, $\lfloor \varpi \rfloor$ for the integer part of $\varpi$ and $\mathrm{N}(\cdot,\cdot)$ to indicate a normal distribution.

# 2 The testing procedure

We consider the following class of endogenous volatility models:

$$\begin{aligned} \mathrm{d}X_t &= \mu(X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \\ \sigma_t^2 &= \sigma^2(X_t) \end{aligned} \tag{2}$$

and the following class of stochastic volatility models:

$$\begin{aligned} \mathrm{d}X_t &= \mu_t\mathrm{d}t + \sigma_t \mathrm{d}W_t \\ \sigma_t^2 &= \sigma^2(V_t), \end{aligned}$$

where $V_t$ is a diffusion process not perfectly correlated with $X_t$, so that possible leverage effects are allowed.

We state the hypotheses of interest as:

$$\mathrm{H}_0 : \sigma_t^2 = \sigma^2(X_t) \in \mathcal{F}_t^X \quad \text{a.s.},$$

versus the alternative:

$$\mathrm{H}_A : \sigma_t^2 = \sigma^2(V_t) \notin \mathcal{F}_t^X - \text{measurable, a.s.},$$

where $\mathcal{F}_t^X = \sigma(X_s, s \leq t)$. Under the null hypothesis, the volatility process is a measurable function of the observable $X_t$. Under the alternative, the volatility process is a measurable function of $V_t$, hence is not $\mathcal{F}_t^X$-measurable. The derivation of the limiting distribution relies on the notion of local time for homogeneous univariate diffusions. In this sense, we require the drift to be $\mathcal{F}_t^X$-measurable. The test is constructed for discriminating between *classes* of models for financial volatility and, because the contribution of the drift term is asympotically negligible, has no power against the case of a drift depending on some latent factor, provided it satisfies the Lipschitz conditions stated in A1.

We have a sample of $n$ observations on the underlying process, i.e. we observe $X_{j\Delta_{n,T}}$, $j = 1, \ldots, n$, and $\Delta_{n,T} = T/n$. We assume that, as $n \to \infty$, the time span $T$ increases and the discrete interval $\Delta_{n,T}$ decreases, i.e. we consider both long-span and in-fill asymptotics. Define the following estimators:

$$\widehat{\sigma}^2_{n,T}(X_t) = \frac{\frac{1}{n\xi_{n,T}} \sum_{j=1}^{n-1} K\left(\frac{X_t - X_{j\Delta_{n,T}}}{\xi_{n,T}}\right) \Delta_{n,T}^{-1} \left(X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}}\right)^2}{\frac{1}{n\xi_{n,T}} \sum_{i=1}^{n-1} K\left(\frac{X_t - X_{i\Delta_{n,T}}}{\xi_{n,T}}\right)}, \tag{3}$$

and:

$$\widetilde{\sigma}^2_{n,T,t} = \sum_{j=0}^{\lfloor T^{-\gamma}\Delta_{n,T}^{-1}\rfloor - 1} T^{\gamma} \left(X_{t+(j+1)\Delta_{n,T}} - X_{t+j\Delta_{n,T}}\right)^2.$$

Here, $\widehat{\sigma}^2_{n,T}(X_t)$ is a nonparametric estimator of the volatility process evaluated at $X_t$. Florens-Zmirou (1993) has established consistency and the asymptotic distribution of a scaled version of (3) when the variance process follows (2), for the case of a uniform kernel function. $\widehat{\sigma}^2_{n,T}(X_t)$ has been used by Bandi and Phillips (2003, 2007), in the context of fully nonparametric and parametric estimation of diffusion processes, respectively. The estimator $\widetilde{\sigma}^2_{n,T,t}$ averages $\lfloor T^{-\gamma}\Delta_{n,T}^{-1}\rfloor - 1$ squared price differences in a local neighborhood of $t$, determined by the localizing factor $T^{-\gamma}$. It can be interpreted as a localized version of the usual realized volatility estimator.

Our test statistic is based on the difference:

$$Z_{n,T} = \sqrt{T/2}\left(\frac{2}{T}\sum_{t=1}^{T/2}\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2 I_t - \frac{2}{T}\sum_{t=T/2+1}^{T}\left(\widetilde{\sigma}^2_{n,T,t}\right)^2 I_t\right),$$

where $I_t = 1_{\{X_t \in B\}}$ and $B$ is a bounded set in $\mathbb{R}$.

The statistic is based on the scaled difference between the average of the squared kernel estimator and the average of the squared localized realized volatility. Under the null hypothesis, $\widehat{\sigma}^2_{n,T}(X_t) = \sigma^2(X_t) + o_p(1)$ and $\widetilde{\sigma}^2_{n,T,t} = \sigma^2(X_t) + o_p(1)$, where the $o_p(1)$ holds uniformly in $t$. Thus, we split the sample to avoid the degeneracy of the statistic. In principle we can split the sample into $\lfloor Tr\rfloor$ and $T - \lfloor Tr\rfloor$ observations, based on a generic fraction of the data $r$. However, because the limiting distribution does not depend on $r$, the most natural choice is to split the sample into halves. The reason why we are averaging only over $X_t \in B$, is that we need to control the denominator in the construction of $\widehat{\sigma}^2_{n,T}(X_t)$. Basically, this is the usual trimming device used when averaging nonparametric estimators of conditional moments. Under the alternative, $\widetilde{\sigma}^2_{n,T,t}$ is consistent for $\sigma^2_t$, while $\widehat{\sigma}^2_{n,T}(X_t)$ is not. However, while for any given $X_t$, we have a non vanishing bias, by averaging over the evaluation points the bias becomes negligible. This is why we base the statistic on the difference between $\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2$ and $\left(\widetilde{\sigma}^2_{n,T,t}\right)^2$, rather than on the difference between $\widehat{\sigma}^2_{n,T}(X_t)$ and $\widetilde{\sigma}^2_{n,T,t}$. Next, we define the scaling factor:

$$
\begin{aligned}
&\widehat{V}_{n,T} \qquad (4)\\
= \ & \frac{2}{T}\sum_{\tau=-l_T}^{l_T}\sum_{t=l_T+1}^{T/2-l_T}\omega_\tau\left(\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2 I_t-\frac{2}{T}\sum_{t=1}^{T/2}\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2 I_t\right)\\
&\times\left(\left(\widehat{\sigma}^2_{n,T}(X_{t-\tau})\right)^2 I_{t-\tau}-\frac{2}{T}\sum_{t=1}^{T/2}\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2 I_t\right)\\
&+\frac{2}{T}\sum_{\tau=-l_T}^{l_T}\sum_{t=T/2+l_T+1}^{T-T/2-l_T}\omega_\tau\left(\left(\widetilde{\sigma}^2_{n,T,t}\right)^2 I_t-\frac{2}{T}\sum_{t=T/2+1}^{T}\left(\widetilde{\sigma}^2_{n,T,t}\right)^2 I_t\right)\\
&\times\left(\left(\widetilde{\sigma}^2_{n,T,t-\tau}\right)^2 I_{t-\tau}-\frac{2}{T}\sum_{t=T/2+1}^{T}\left(\widetilde{\sigma}^2_{n,T,t}\right)^2 I_t\right)\\
&-\frac{4}{T}\sum_{t=T/2-l_T+1}^{T/2}\sum_{s=T/2+1}^{T/2+l_T}\omega_{t,s}\left(\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2 I_t-\frac{2}{T}\sum_{t=1}^{T/2}\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2 I_t\right)\\
&\times\left(\left(\widetilde{\sigma}^2_{n,T,s}\right)^2 I_s-\frac{2}{T}\sum_{t=T/2+1}^{T}\left(\widetilde{\sigma}^2_{n,T,t}\right)^2 I_t\right),
\end{aligned}
$$

where $\omega_\tau = 1-\tau/(l_T-1)$, $\omega_{t,s} = 1-(s-t)/(l_T-1)$. Note the variance estimator in (4) is composed of three pieces. The first two terms are HAC estimators of the variances of $\frac{1}{\sqrt{T/2}}\sum_{t=1}^{T/2}\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2 I_t$ and $\frac{1}{\sqrt{T/2}}\sum_{t=T/2+1}^{T}\left(\widetilde{\sigma}^2_{n,T,t}\right)^2 I_t$, respectively, while the third term is an estimator of the covariance between the two. Given the memory conditions in A2 and in A5, the covariance between the two terms is asymptotically negligible. Nevertheless, the inclusion of the covariance estimator may improve the finite sample performance of the statistic. Under the null, the first two terms converge to the same probability limit, as they are both consistent for the "true" variance. Under the alternative, the first term in (4) is no longer a consistent estimator of the true variance, but nevertheless it is bounded in probability, thus ensuring the proper divergence of the statistic.

Finally, the test statistic is just a scaled version of $Z_{n,T}$, namely:

$$\widehat{Z}_{n,T} = \widehat{V}_{n,T}^{-1/2} Z_{n,T}.$$

Before stating our main result, we need the following assumptions.

**Assumption A1:**

$(i)$ $\mu(.)$ and $\sigma(.)$ are time-homogeneous, at least twice continuously differentiable. and satisfy local Lipschitz and growth conditions. Thus, for every compact subset $B$ of the range of the process, there exist constants $C_1^B$ and $C_2^B$ so that, for all $x$ and $y$ in $B$:

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq C_1^B |x - y|,$$

and:

$$|\mu(x)| + |\sigma(x)| \leq C_2^B \{1 + |x|\}.$$

$(ii)$ $\sigma^2(.) > 0$ on $\mathfrak{D}$.

**Assumption A2:**

$(i)$ $T\xi_{n,T} \log(T^{-1}) \to \infty$, $(ii)$ $\sqrt{T}\xi_{n,T}^2 \to 0$, $(iii)$ $T^2 n^{-1} \log n \to 0$, $(iv)$ $n^{-1}\xi_{n,T}^{-2} T \log(n) \to 0$, $(v)$ $T^\gamma \Delta_{n,T} \to 0$, $(vi)$ $T^{1-\gamma} \log(T^\gamma) \to 0$, $(vii)$ $l_T = o(T^{1/2})$.

**Assumption A3:**

$(i)$ $X_t$ is strictly stationary, geometrically strong mixing, $(ii)$ the stationary density of $X_t$ (denoted by $f$) is ultimately decreasing (i.e. $\lim_{u\to\infty} |u| f(u) = 0$), $(iii)$ $\sup_{t\in[0,T]} |X_t|$ is measurable for each $T$, $(iv)$ $\mathrm{E}\big(\sup_{t\in[0,1]} |X_t|^a\big) < \infty$, for $a > 0$, $(v)$ $f(x) \geq \varepsilon > 0$, for all $x \in B$.

**Assumption A4:**

The kernel function $K(\cdot)$ is a second-order, symmetric and nonnegative function with bounded support, continuously differentiable in the interior of its support, with bounded first derivative, which satisfies $\int_{-\infty}^{\infty} K(u)\mathrm{d}u = 1$.

**Assumption A5:**

$(i)$ $V_t$ is a strictly stationary ergodic process, geometrically strong mixing, $(ii)$ $\mathrm{E}\left(\left(\sigma^2(V_t)\right)^{2k}\right) < \infty$ and $\mathrm{E}\left(\left(\sigma^{2\prime}(V_t)\right)^{2k}\right) < \infty$, with $k > 1/\delta$ and $\delta < 1/2$, $(iii)$ $n^{-1}\xi_{n,T}^{-2} T^{1+2\delta} \to 0$, $(iv)$ $T^{2(1+\delta)} n^{-1} \log n \to 0$, $(v)$ $T^{1+\delta-\gamma} \log(T^{\gamma+\delta}) \to 0$.

Assumptions 1 ensures the existence of a unique strong solution under both hypotheses (see e.g. Karatzas and Shreve, 1991, Chapter 5). These assumptions are standard in the literature (see Aït-Sahalia, 1996), because global Lipschitz and growth conditions fail to be met by a lot of models of interest in financial applications.

Assumption A2 states rates conditions sufficient for asymptotic normality under the null. Assumption A5(iii)-(v) are a strengthened version of A2(iv),(iii),(v) respectively, required under the alternative. Such a strengthening is due to the fact that under the alternative $\sigma_t^2 I_t$ is not necessarily bounded. Overall, we need to jointly control three sequences, the discrete interval $\Delta_{N,T}$, the time span $T$ and the bandwidth $\xi_{n,T}$. It is natural to define $\xi_{n,T}$ as a function of both $\Delta_{n,T}$ and $T$. Then, it is immediate to note that all rate condition, but A2($i$), are satisfied provided $n$ grows faster enough with respect to $T$. On the other hand, A2($i$) requires that $T$ grows sufficiently fast with respect to $n$. For example, if we set $\xi_{n,T} = \Delta_{n,T}^{1/5} T^{-1/5}$, then all rate conditions in A2 and A5 are satisfied provided $T^5/n \to \infty$ and $\max\left\{T^{1+\gamma}/n, T^{2(1+\delta)}/n\right\} \to 0$, where $\gamma > 1 + \delta$ and $\delta < 1/2$.

Assumption A3 impose memory and regularity conditions on $X_t$, while A5(i)(ii) imposes memory and moment conditions on $\sigma_t^2$, under the alternative of stochastic volatility. Finally, A4 states standard conditions on the kernel function.

We have the following result.

**Theorem 1.**
*(i) Let A1-A4 hold. Then, under* $\mathrm{H}_0$*:*

$$\widetilde{Z}_{n,T} \xrightarrow{d} \mathrm{N}(0,1).$$

*(ii) Let Let A1-A5 hold. Then, under* $\mathrm{H}_A$*, there exists* $\epsilon, \varepsilon > 0$*, such that:*

$$\Pr\left(T^{1/2-\epsilon}\left|\widetilde{Z}_{n,T}\right| > \varepsilon\right) \to 1.$$

It is immediate to see that under the null, the statistic converges to a standard normal, while under the alternative it diverges at rate $\sqrt{T}$. Hence, the test provides a simple way of discriminating between one factor and stochastic volatility *classes* of models.

# 3 Monte Carlo results

In this section, the small sample performance of the testing procedure proposed in the previous section is assessed through a Monte-Carlo experiment. Under the null hypothesis, we consider a data generating process process given by the Cox, Ingersoll and Ross (1985) model:

$$\mathrm{d}X_t = \kappa(\mu - X_t)\mathrm{d}t + \eta\sqrt{X_t}\mathrm{d}W_{1,t}. \tag{5}$$

We simulate a discretized version of the continuous trajectory of $X_t$ under (5). We use a Milstein scheme in order to approximate the trajectory, following Pardoux and Talay (1985). The initial value is fixed to the unconditional mean of $X_t$, and the first 1000 observations are then discarded. The process is executed for for $\kappa = 0.2$, $\eta = .01$, $\mu = 0.05$, ensuring positivity of the process, and is repeated for $S = 10,000$ replications.

We consider $T \in (200, 300, 400)$ and $n \approx T^b$, $b \in (2.5, 2.75, 3)$. We set the bandwidth parameters for the kernel estimator $\xi_{n,T} = (4/3)^{1/5} SD(X_s) n^{-1/5}$ (where $SD(\cdot)$ is standard deviation and $X_s$ is the path obtained at the $s$-th simulation), and for the variance estimator of our test statistic $l_T = (T/2)^{1/3}$. Finally, we set $\gamma = 1.25$ and choose $I_t$ such that we discard those estimators for which $X_t$ lies either in the lower 10% (resp. 5%) tail or in the upper 10% (resp. 5%) tail distribution.

The empirical sizes of the test are reported in Table 1, columns 2 to 5. Inspection of the Table reveals an overall good small sample behaviour of the considered test statistics. The reported empirical sizes are everywhere very close to the nominal ones, with the exception of the case where $n \approx T^{2.5}$, which leads to substantial over-rejection. Therefore, because of the implied size distortion, we do not report those results.

Trimming plays an important role for the size of the test. In fact, in the absence of trimming the statistic becomes highly oversized, mainly because we fail to control the denominator in $\widehat{\sigma}^2_{n,T}(X_t)$. Nevertheless, a too "heavy" trimming further reduces the effective sample size. From Table 1, it appears that by simply trimming the smallest and largest 5% of observations we have a size very close to the nominal one, even for very sample samples. Not surprisingly, for a given $T$, a smaller discrete interval, as implied by a larger $n$, has a slight positive effect on size. In fact, the smaller the interval, the faster the estimation error vanishes.

Under the alternative hypothesis, the following model has been considered,

$$\begin{aligned} \mathrm{d}X_t &= \kappa(\mu - X_t)\mathrm{d}t + \exp\left(\beta_0 + \beta_1 \sigma_t^2\right)\left(\sqrt{1-\rho^2}\mathrm{d}W_{1,t} + \rho \mathrm{d}W_{2,t}\right) \\ \mathrm{d}\sigma_t^2 &= \kappa_1 \sigma_t^2 \mathrm{d}t + \mathrm{d}W_{2,t}. \end{aligned} \tag{6}$$

A discretized version of (6) has been simulated using a Milstein scheme as above, with $\kappa_1 = -0.1$, $\beta_0 = 0$, $\beta_1 = 0.125$. Then, using the obtained values of $\sigma_t^2$, the series for $X_t$ has been generated for different values of $\rho$. Notice that we keep the values of the drift used to generate the observable $X_t$ equal under both hypotheses, hence the difference is only about how volatility is generated.

The results are reported in Table 2. The experiment reveals that the proposed test has overall good power properties. Note that the power is driven by the bias term in (A.9), which makes the statistic diverge at rate $\sqrt{T/2}$. On the other hand, the rate conditions in A5(iii)-(v) control the order of magnitude of the other terms, and ensure that $\widetilde{\sigma}^2_{n,T,t}$ converges fast enough to the "true" spot volatility, under both hypotheses, but has no direct effect on the divergence of the statistic. Hence, it is not surprising that choosing a larger $n$ for large $T$, and therefore a smaller discrete interval, has an unambiguously positive effect on the power.

Also, the presence of leverage does not reduce the power. This is because, when $X_t$ and $V_t$ are independent, only the second term in (A.9) contributes to the power, as the first one has mean zero. On the other hand, when $X_t$ and $V_t$ are correlated, both terms in (A.9) can contribute to the power. Heuristically, we might expect a nonlinear effect of the degree of leverage on the power, first increasing for nonzero leverage and then decreasing for $\rho$ close to one.

Finally, it appears that at least for smaller values of $n$, the power in some cases is not monotonic. This can be possibly attributed to the trimming. In fact, it is possible that observations which potentially most contribute to the divergence of the statistic are trimmed away, and also that the scaling factor is "inflated" by the presence of zeroes. In this sense, the effect of a "small" increase in $T$ may be offset by the trimming device. If we choose a larger value for $n$, power is everywhere monotonically increasing in $T$.

# 4 Concluding remarks

This paper provides a testing procedure which allows to discriminate between one factor and stochastic volatility models, under minimal assumptions. Apart from standard regularity conditions, no assumptions are made on the functional forms of either the drift or the diffusion term.

The suggested test statistic compares two different estimators of the spot volatility. Both of them are consistent under the null hypothesis, while under the alternative only the localized version of realized volatility is. This ensures consistency of the associated test.

As the Monte carlo experiment shows, the tests perform well in finite samples.

Table 1: Size of the test based on $\widehat{Z}_{n,T}$ for different values of $n$ and $T$

| | 10% trimming | | 5% trimming | |
|---|---|---|---|---|
| | 5% level | 10% level | 5% level | 10% level |
| | $T = 200$ | | | |
| $n \approx T^{2.75}$ | 0.04 | 0.07 | 0.07 | 0.12 |
| $n \approx T^{3}$ | 0.02 | 0.04 | 0.06 | 0.09 |
| | $T = 300$ | | | |
| $n \approx T^{2.75}$ | 0.04 | 0.08 | 0.07 | 0.11 |
| $n \approx T^{3}$ | 0.03 | 0.06 | 0.06 | 0.10 |
| | $T = 400$ | | | |
| $n \approx T^{2.75}$ | 0.04 | 0.09 | 0.06 | 0.11 |
| $n \approx T^{3}$ | 0.04 | 0.08 | 0.05 | 0.11 |

Table 2: Power of the test based on $\widehat{Z}_{n,T}$ for different values of $n$, $T$ and $\rho$

| | 10% trimming | | 5% trimming | |
|---|---|---|---|---|
| | 5% level | 10% level | 5% level | 10% level |
| | $\rho = -0.3$, $T = 200$ | | | |
| $n \approx T^{2.75}$ | 0.30 | 0.39 | 0.33 | 0.47 |
| $n \approx T^{3}$ | 0.29 | 0.42 | 0.32 | 0.42 |
| | $\rho = -0.3$, $T = 300$ | | | |
| $n \approx T^{2.75}$ | 0.30 | 0.49 | 0.39 | 0.52 |
| $n \approx T^{3}$ | 0.32 | 0.44 | 0.36 | 0.42 |
| | $\rho = -0.3$, $T = 400$ | | | |
| $n \approx T^{2.75}$ | 0.18 | 0.36 | 0.22 | 0.36 |
| $n \approx T^{3}$ | 0.36 | 0.45 | 0.37 | 0.47 |
| | $\rho = 0$, $T = 200$ | | | |
| $n \approx T^{2.75}$ | 0.19 | 0.31 | 0.26 | 0.37 |
| $n \approx T^{3}$ | 0.13 | 0.24 | 0.16 | 0.32 |
| | $\rho = 0$, $T = 300$ | | | |
| $n \approx T^{2.75}$ | 0.21 | 0.32 | 0.22 | 0.38 |
| $n \approx T^{3}$ | 0.27 | 0.38 | 0.30 | 0.36 |
| | $\rho = 0$, $T = 400$ | | | |
| $n \approx T^{2.75}$ | 0.19 | 0.28 | 0.20 | 0.32 |
| $n \approx T^{3}$ | 0.31 | 0.47 | 0.37 | 0.49 |
| | $\rho = 0.3$, $T = 200$ | | | |
| $n \approx T^{2.75}$ | 0.23 | 0.31 | 0.26 | 0.41 |
| $n \approx T^{3}$ | 0.31 | 0.39 | 0.35 | 0.39 |
| | $\rho = 0.3$, $T = 300$ | | | |
| $n \approx T^{2.75}$ | 0.21 | 0.25 | 0.20 | 0.30 |
| $n \approx T^{3}$ | 0.33 | 0.42 | 0.33 | 0.44 |
| | $\rho = 0.3$, $T = 400$ | | | |
| $n \approx T^{2.75}$ | 0.15 | 0.24 | 0.17 | 0.28 |
| $n \approx T^{3}$ | 0.35 | 0.43 | 0.34 | 0.46 |

# Appendix

The proof of Theorem 1 requires the following Lemma.

**Lemma 1.** *Let Assumption A1(i) hold.*
*(i) Under* $\mathrm{H}_0$*:*

$$\sup_{0\le t,s\le T;|t-s|\le\Delta_{n,T}} |X_t - X_s|\, I_t = O_{a.s.}\left(\sqrt{\Delta_{n,T}\log(\Delta_{n,T}^{-1})}\right).$$

*(ii) Under* $\mathrm{H}_A$*:*

$$\sup_{0\le t,s\le T;|t-s|\le\Delta_{n,T}} |X_t - X_s|\, I_t = O_{a.s.}\left(T^{\delta}\sqrt{\Delta_{n,T}\log(\Delta_{n,T}^{-1})}\right).$$

# Proof of Theorem 1:

$(i)$ First, we want to show that:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T/2}\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2 I_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T/2}\left(\sigma^2(X_t)\right)^2 I_t + o_p(1). \tag{A.1}$$

It suffices to show that:

$$\sup_t \sqrt{T}\left|\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2 I_t - \left(\sigma^2(X_t)\right)^2 I_t\right| = o_p(1).$$

We know that:

$$\begin{aligned}
&\sup_t \sqrt{T}\left|\left(\widehat{\sigma}^2_{n,T}(X_t)\right)^2 I_t - \left(\sigma^2(X_t)\right)^2 I_t\right| \\
\le\ &\sup_t \left|\widehat{\sigma}^2_{n,T}(X_t)I_t + \sigma^2(X_t)I_t\right| \sup_t \sqrt{T}\left|\widehat{\sigma}^2_{n,T}(X_t)I_t - \sigma^2(X_t)I_t\right|.
\end{aligned}$$

By Itô's formula, for each $t$:

$$\begin{aligned}
&\sqrt{T}I_t\left(\widehat{\sigma}^2_{n,T}(X_t) - \sigma^2(X_t)\right) \\
=\ &\underbrace{\sqrt{T}\left(I_t\frac{\sum_{j=1}^{n-1}K\left(\frac{X_t - X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s - X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s}{\sum_{i=1}^{n-1}K\left(\frac{X_t - X_{i\Delta_{n,T}}}{\xi_{n,T}}\right)}\right)}_{A^{(1)}_{n,T}} \\
&\underbrace{+\sqrt{T}\left(I_t\frac{\sum_{j=1}^{n-1}K\left(\frac{X_t - X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s - X_{j\Delta_{n,T}}\right)\mu(X_s)\mathrm{d}s}{\sum_{i=1}^{n-1}K\left(\frac{X_t - X_{i\Delta_{n,T}}}{\xi_{n,T}}\right)}\right)}_{A^{(2)}_{n,T}}
\end{aligned}$$

$$+\underbrace{\sqrt{T}\left(I_t\frac{\sum_{j=1}^{n-1}K\left(\frac{X_t-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)\Delta_{n,T}^{-1}\left(\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(\sigma^2(X_s)-\sigma^2(X_{j\Delta_{n,T}})\right)\mathrm{d}s\right)}{\sum_{i=1}^{n-1}K\left(\frac{X_t-X_{i\Delta_{n,T}}}{\xi_{n,T}}\right)}\right)}_{A_{n,T}^{(3)}}$$

$$+\underbrace{\sqrt{T}\left(I_t\frac{\sum_{j=1}^{n-1}K\left(\frac{X_t-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)\left(\sigma^2(X_{j\Delta_{n,T}})-\sigma^2(X_t)\right)}{\sum_{i=1}^{n-1}K\left(\frac{X_t-X_{i\Delta_{n,T}}}{\xi_{n,T}}\right)}\right)}_{A_{n,T}^{(4)}}.$$

As for $A_{n,T}^{(4)}$, given A1 and A4:

$$\begin{aligned}
\frac{1}{\sqrt{T}}A_{n,T}^{(4)} &\leq \sup_{x\in B}\left|\frac{\sum_{j=1}^{n-1}K\left(\frac{x-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)\left(\sigma^2(X_{j\Delta_{n,T}})-\sigma^2(x)\right)}{\sum_{i=1}^{n-1}K\left(\frac{x-X_{i\Delta_{n,T}}}{\xi_{n,T}}\right)}\right| \\
&= \sup_{x\in B}\left|\frac{\frac{1}{T}\int_0^T\frac{1}{\xi_{n,T}}K\left(\frac{X_t-x}{\xi_{n,T}}\right)\sigma^2(X_t)\mathrm{d}t-\sigma^2(x)+o_{a.s.}(1)}{\frac{1}{T}\int_0^T\frac{1}{\xi_{n,T}}K\left(\frac{X_t-x}{\xi_{n,T}}\right)\mathrm{d}t+o_{a.s.}(1)}\right| \\
&= \sup_{x\in B}\left|\frac{\int_{-\infty}^{\infty}\frac{1}{\xi_{n,T}}K\left(\frac{x-a}{\xi_{n,T}}\right)\left(\sigma^2(a)-\sigma^2(x)\right)f(a)\mathrm{d}a+o_{a.s.}(1)}{\int_{-\infty}^{\infty}\frac{1}{\xi_{n,T}}K\left(\frac{x-a}{\xi_{n,T}}\right)f(a)\mathrm{d}a+o_{a.s.}(1)}\right| \\
&= \sup_{x\in B}\left|\frac{\int_{-\infty}^{\infty}K(u)\left(\sigma^2(x+u\xi_{n,T})-\sigma^2(x)\right)f(x+u\xi_{n,T})\mathrm{d}u+o_{a.s.}(1)}{\int_{-\infty}^{\infty}K(u)f(x+u\xi_{n,T})\mathrm{d}u+o_{a.s.}(1)}\right| \\
&= \sup_{x\in B}\left|\xi_{n,T}^2\int u^2K(u)\mathrm{d}u\left(\frac{1}{2}\sigma^{2''}(x)+\sigma^{2'}(x)\frac{f'(x)}{f(x)}\right)\right|+o_{a.s.}\left(\xi_{n,T}^2\right)+o_{a.s.}(1),
\end{aligned}$$

where the order of the $o_{a.s.}(1)$ terms above is $o_{a.s.}\left(\xi_{n,T}^{-1}\sqrt{\log\left(\Delta_{n,T}^{-1}\right)\Delta_{n,T}}\right)$ because of (A.3) below, and the second equality follows from the proof of Theorem 3 in Bandi and Phillips (2003), by noting that in the stationary ergodic case the speed density coincides with the stationary density. Hence, $A_{n,T}^{(4)} = O\left(\sqrt{T}\xi_n^2\right)+o_{a.s.}\left(\sqrt{T}\xi_{n,T}^{-1}\sqrt{\log\left(\Delta_{n,T}^{-1}\right)\Delta_{n,T}}\right)=o_{a.s.}(1)$, given A2(ii) and A2(iv). Because of Lemma 1(i), $A_{n,T}^{(3)} = O_{a.s.}\left(\sqrt{T\Delta_{n,T}\log\left(\frac{1}{\Delta_{n,T}}\right)}\right) = o_{a.s.}(1)$, given A2(iii). Finally, $A_{n,T}^{(2)}$ is of a smaller order of probability than $A_{n,T}^{(1)}$.

We now want to show that $A_{n,T}^{(1)}=o_p(1)$. For this, it suffices to show that:

$$\sup_{x\in B}\left|\frac{\sqrt{T}}{n\xi_{n,T}}\sum_{j=1}^{n-1}I_t\frac{K\left(\frac{x-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s-X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s}{\frac{1}{n\xi_{n,T}}\sum_{j=1}^{n-1}K\left(\frac{x-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)}\right|=o_p(1).$$

We first need to bound the denominator. Hereafter, let $\widehat{f}_{n,T}(x)=1/(n\xi_{n,T})\sum_{j=1}^{n-1}K\left(\frac{x-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)$

and $\widetilde{f}_T(x) = 1/(T\xi_{n,T}) \int_0^T K\left(\frac{X_t - x}{\xi_{n,T}}\right) \mathrm{d}t$. We can write:

$$\sup_{x\in B}\left|\widehat{f}_{n,T}(x) - f(x)\right| \leq \sup_{x\in B}\left|\widehat{f}_{n,T}(x) - \widetilde{f}_T(x)\right| + \sup_{x\in B}\left|\widetilde{f}_{n,T}(x) - f(x)\right|. \qquad \text{(A.2)}$$

We begin with the first term on the right hand side (rhs) of (A.2):

$$\begin{aligned}
&\left|\frac{1}{n\xi_{n,T}}\sum_{j=1}^{n-1} K\left(\frac{X_{j\Delta_{n,T}} - x}{\xi_{n,T}}\right) - \frac{1}{T}\int_0^T \frac{1}{\xi_{n,T}} K\left(\frac{X_t - x}{\xi_{n,T}}\right)\mathrm{d}t\right| \\
= &\left|\frac{1}{T\xi_{n,T}}\sum_{j=1}^{n-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(K\left(\frac{X_{j\Delta_{n,T}} - x}{\xi_{n,T}}\right) - K\left(\frac{X_t - x}{\xi_{n,T}}\right)\right)\mathrm{d}t\right| \\
\leq &\ \frac{1}{T\xi_{n,T}}\sum_{j=1}^{n-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left|K'\left(\frac{\widetilde{X}_{j,t} - x}{\xi_{n,T}}\right)\right|\mathrm{d}t\left(\max_{j=1,\ldots,n}\ \sup_{j\Delta_{n,T}\leq t\leq (j+1)\Delta_{n,T}}\frac{\left|X_{j\Delta_{n,T}} - X_t\right|}{\xi_{n,T}}\right) \\
\leq &\ \frac{1}{T\xi_{n,T}}\sum_{j=1}^{n-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left|K'\left(\frac{X_t - x}{\xi_{n,T}} + o_{a.s}\left(\xi_{n,T}^{-1}\sqrt{\log\left(\Delta_{n,T}^{-1}\right)\Delta_{n,T}}\right)\right)\right|\mathrm{d}t \\
&\times o_{a.s.}\left(\xi_{n,T}^{-1}\sqrt{\log\left(\Delta_{n,T}^{-1}\right)\Delta_{n,T}}\right) \qquad \text{(A.3)} \\
= &\ \frac{1}{T}\int_{-\infty}^{\infty}\left|K'\left(q + o_{a.s}(1)\right)\right| L_X\left(T, q\xi_{n,T} + x\right)\mathrm{d}q \times o_{a.s.}\left(\xi_{n,T}^{-1}\sqrt{\log\left(\Delta_{n,T}^{-1}\right)\Delta_{n,T}}\right) \quad \text{(A.4)} \\
= &\ O_{a.s.}(1)o_{a.s.}(1) = o_{a.s.}(1),
\end{aligned}$$

given A2(iv). Note that (A.3) follows from Lemma 1(i), and (A.4) follows from the occupation density formula (p.267 in Bandi and Phillips, 2003). $L_X\left(T, q\xi_{n,T} + x\right)$ denotes the local time of the diffusion, i.e. how much time, between 0 and $T$ has been spent around $q\xi_{n,T} + x$, and in the stationary ergodic case, $L_X\left(T, x\right)/T = O_{a.s.}(1)$. As for the second term on the rhs of (A.2), given A3(i)-(iv), from Corollary 4.6 in Bosq (1998):

$$\sup_{x\in B}\left|\frac{1}{T}\int_0^T \frac{1}{\xi_{n,T}} K\left(\frac{X_t - x}{\xi_{n,T}}\right)\mathrm{d}t - f(x)\right| = O_{a.s.}\left(\sqrt{\frac{\log T}{T\xi_{n,T}}}\right) + O\left(\xi_{n,T}^2\right) = o_{s.s.}(1),$$

given A2(i). Hence, given A3(v), $\mathrm{plim}_{n,T\to\infty}\left|\widehat{f}_{n,T}(x)\right| < \varepsilon_{n,T} = 0$, where:

$$\varepsilon_{n,T} = \varepsilon - \left(\xi_{n,T}^{-1}\sqrt{\log\left(\Delta_{n,T}^{-1}\right)\Delta_{n,T}} + \xi_{n,T}^2 + \sqrt{\frac{\log T}{T\xi_{n,T}}}\right).$$

Then, it suffices to show that, for $n, T \to \infty$:

$$\Pr\left(\sup_{x\in B}\left|\frac{\sqrt{T}}{n\xi_{n,T}}\sum_{j=1}^{n-1} K\left(\frac{x - X_{j\Delta_{n,T}}}{\xi_{n,T}}\right) 2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s - X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s\right| > \varepsilon\right) \to 0.$$

Without loss of generality, let $B = [-D, D]$. We cover the interval $[-D, D]$ with $Q_{n,T} = 2DT^{2/3}\xi_{n,T}^{-2}$ balls $S_i$, centered at $s_i$, of radius $\xi_{n,T}^2/T^{2/3}$. We can bound the term as follows:

$$\begin{aligned}
&\sup_{x\in B}\left|\frac{\sqrt{T}}{n\xi_{n,T}}\sum_{j=1}^{n-1}K\left(\frac{x-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s-X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s\right| \\
\leq\ & \max_{i=1,\ldots,Q_{n,T}}\left|\frac{\sqrt{T}}{n\xi_{n,T}}\sum_{j=1}^{n-1}K\left(\frac{s_i-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)\right. \\
&\qquad\left.\times 2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s-X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s\right| \qquad (A.5)\\
&+\max_{i=1,\ldots,Q_{n,T}}\sup_{x\in S_i}\left|\frac{\sqrt{T}}{n\xi_{n,T}}\sum_{j=1}^{n-1}2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s-X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s\right. \\
&\qquad\left.\times\left(K\left(\frac{x-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)-K\left(\frac{s_i-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)\right)\right|. \qquad (A.6)
\end{aligned}$$

First, notice that, given A4 and because of Lemma 1(i):

$$\begin{aligned}
&\max_j\sup_{j\Delta_{n,T}\leq s\leq(j+1)\Delta_{n,T}}\left|\left(K\left(\frac{x-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)-K\left(\frac{s_j-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)\right)\right. \\
&\qquad\left.\times 2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s-X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s\right| \\
\leq\ & C\xi_{n,T}^{-1}\max_j|x-s_j|\,\Delta_{n,T}^{-1}\sup_{j\Delta_{n,T}\leq s\leq(j+1)\Delta_{n,T}}\left|X_s-X_{j\Delta_{n,T}}\right| \\
&\times\max_j\sup_{j\Delta_{n,T}\leq s\leq(j+1)\Delta_{n,T}}\left|W_s-W_{j\Delta_{n,T}}\right|\sup_{j\Delta_{n,T}\leq s\leq(j+1)\Delta_{n,T}}|\sigma(X_s)| \\
=\ & O_{a.s.}\left(\log\Delta_{n,T}^{-1}\right)T^{-2/3}\xi_{n,T}.
\end{aligned}$$

Hence, (A.6) is $O_p\left(\frac{T^{1/2}}{\xi_{n,T}}\frac{\xi_{n,T}}{T^{2/3}}\right) = o_p(1)$. Note that $\sup_{j\Delta_{n,T}\leq s\leq(j+1)\Delta_{n,T}}|\sigma(X_s)|$ is bounded, because $X_s$ is close to $X_{j\Delta_{n,T}}$, which in turn is close to the evaluation point $x$, which belongs to a bounded set.

For $j = 1,\ldots,n$, $K\left(\frac{s_i-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s-X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s$ is a zero mean martingale difference sequence, with respect to $\mathcal{F}_{j\Delta_{n,T}}^X = \sigma\left(X_{i\Delta_{n,T}},\ i = 1,\ldots,j\right)$, and:

$$\begin{aligned}
&\mathrm{E}\left(\left(K\left(\frac{s_i-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s-X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s\right)^2\middle|\mathcal{F}_{j\Delta_{n,T}}^X\right) \\
=\ & K\left(\frac{s_i-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)^2\sigma^2(X_{j\Delta_{n,T}})\left(1+o_{a.s.}(1)\right).
\end{aligned}$$

Because - for all $i$ - $s_i$ belongs to a bounded set, and - for all $j$ - $X_{j\Delta_{n,T}}$ is in a neighborhood of $s_i$, it follows that for some finite generic constant $C$:

$$\Pr\left(\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n-1}K\left(\frac{s_i-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)^2\sigma^2(X_{j\Delta_{n,T}})>C\right)=0. \tag{A.7}$$

Also, given A1, A4 and Lemma 1(i):

$$\begin{aligned}
&\max_j \sup_{j\Delta_{n,T}\le s\le (j+1)\Delta_{n,T}}\left|K\left(\frac{s_i-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s-X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_{1,s}\right|\\
\le\ & \max_j K\left(\frac{s_i-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)2\Delta_{n,T}^{-1}\sup_{j\Delta_{n,T}\le s\le (j+1)\Delta_{n,T}}\left|X_s-X_{j\Delta_{n,T}}\right|\\
&\times\max_j \sup_{j\Delta_{n,T}\le s\le (j+1)\Delta_{n,T}}\left|W_s-W_{j\Delta_{n,T}}\right|\sup_{j\Delta_{n,T}\le s\le (j+1)\Delta_{n,T}}|\sigma(X_s)|\\
\le\ & C\left(\log\Delta_{n,T}^{-1}\right)\ a.s..
\end{aligned}\tag{A.8}$$

Let $\eta=C\sqrt{\frac{T\log(n)}{n\xi_{n,T}^2}}$, and note that given A2(iv), $\eta\to 0$ as $n,T\to\infty$. Then, given the bounds in (A.7) and (A.8), using an exponential inequality for martingale difference sequences (e.g. Theorem 1.2A in De La Pena, 1999) on (A.5), for $0<\Delta_1,\Delta_2<\infty$:

$$\begin{aligned}
&\Pr\left(\max_{i=1,\dots,Q_{n,T}}\left|\frac{\sqrt{T}}{n\xi_{n,T}}\sum_{j=1}^{n-1}K\left(\frac{s_i-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)\right.\right.\\
&\qquad\left.\left.\times 2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s-X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s\right|>\eta\right)\\
\le\ & Q_{n,T}\Pr\left(\left|\sum_{j=1}^{n-1}K\left(\frac{s_i-X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)2\Delta_{n,T}^{-1}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(X_s-X_{j\Delta_{n,T}}\right)\sigma(X_s)\mathrm{d}W_s\right|>\eta\frac{n\xi_{n,T}}{\sqrt{T}}\right)\\
\le\ & Q_{n,T}\exp\left(-\frac{\eta^2n^2\xi_{n,T}^2T^{-1}}{\Delta_1 n+\eta\Delta_2 n\xi_{n,T}T^{-1/2}\log\left(\Delta_{n,T}^{-1}\right)}\right)=Q_{n,T}\exp\left(-\frac{\eta^2n^2\xi_{n,T}^2T^{-1}}{\Delta_1 n(1+o(1))}\right)\\
=\ & Q_{n,T}\exp\left(-C\log(n)\right)\le 2DT^{2/3}\xi_{n,T}^{-2}n^{-C}\to 0,
\end{aligned}$$

for $C$ sufficiently large. This concludes the proof of (A.1).

Moving to the second estimator of spot volatility, we need to show that:

$$\frac{1}{\sqrt{T}}\sum_{t=T/2+1}^{T}\widetilde{\sigma}_{n,T,t}^2 I_t=\frac{1}{\sqrt{T}}\sum_{t=T/2+1}^{T}\sigma_t^2 I_t+o_p(1),$$

with $\sigma_t^2=\sigma^2(X_t)$. Because:

$$\begin{aligned}
&\sup_t\sqrt{T}\left|\left(\widetilde{\sigma}_{n,T,t}^2\right)^2 I_t-\left(\sigma_t^2\right)^2 I_t\right|\\
\le\ & \sup_t\left|\widetilde{\sigma}_{n,T,t}^2 I_t+\sigma_t^2 I_t\right|\sup_t\sqrt{T}\left|\widetilde{\sigma}_{n,T,t}^2 I_t-\sigma_t^2 I_t\right|,
\end{aligned}$$

it suffices to show that $\sup_t \sqrt{T}\left|\widetilde{\sigma}^2_{n,T,t} I_t - \sigma^2_t I_t\right| = o_p(1)$. By Itô's Lemma, for each $t$:

$$\begin{aligned}
& \sqrt{T}\left(\widetilde{\sigma}^2_{n,T,t} - \sigma^2_t\right) I_t \\
= & \underbrace{\sqrt{T} I_t \sum_{j=0}^{\lfloor T^{-\gamma}\Delta^{-1}_{n,T}\rfloor - 1} 2T^{\gamma} \int_{t+j\Delta_{n,T}}^{t+(j+1)\Delta_{n,T}} \left(X_s - X_{t+j\Delta_{n,T}}\right) \sigma_s \mathrm{d}W_s}_{B^{(1)}_{n,T}} \\
& \underbrace{+\sqrt{T} I_t \sum_{j=0}^{\lfloor T^{-\gamma}\Delta^{-1}_{n,T}\rfloor - 1} 2T^{\gamma} \int_{t+j\Delta_{n,T}}^{t+(j+1)\Delta_{n,T}} \left(X_s - X_{t+j\Delta_{n,T}}\right) \mu_s \mathrm{d}s}_{B^{(2)}_{n,T}} \\
& \underbrace{+\sqrt{T} I_t \left(T^{\gamma} \int_t^{t+T^{-\gamma}} \left(\sigma^2_s - \sigma^2_t\right) \mathrm{d}s\right)}_{B^{(3)}_{n,T}}.
\end{aligned}$$

By Lemma 1(i):

$$B^{(3)}_{n,T} \le C\sqrt{T} \sup_{t,s\le T, |t-s|\le T^{-\gamma}} \left|\sigma^2_s - \sigma^2_t\right| = O_{a.s.}\left(\sqrt{T}\sqrt{T^{-\gamma}\log\left(T^{\gamma}\right)}\right) = o_{a.s.}(1),$$

because of A2(vi). Because $B^{(2)}_{n,T}$ is of a smaller order of probability than $B^{(1)}_{n,T}$, it suffices to show that $B^{(1)}_{n,T} = o_p(1)$. First, note that for all $t < \tau$:

$$\begin{aligned}
& \mathrm{E}\Bigg( I_t \sum_{j=0}^{\lfloor T^{-\gamma}\Delta^{-1}_{n,T}\rfloor - 1} T^{\gamma} \int_{t+j\Delta_{n,T}}^{t+(j+1)\Delta_{n,T}} \left(X_s - X_{t+j\Delta_{n,T}}\right) \sigma_s \mathrm{d}W_s \\
& \quad \times I_{\tau} \sum_{j=0}^{\lfloor T^{-\gamma}\Delta^{-1}_{n,T}\rfloor - 1} T^{\gamma} \int_{\tau+j\Delta_{n,T}}^{\tau+(j+1)\Delta_{n,T}} \left(X_s - X_{\tau+j\Delta_{n,T}}\right) \sigma_s \mathrm{d}W_s \Bigg) \\
= & \mathrm{E}\Bigg( I_t \sum_{j=0}^{\lfloor T^{-\gamma}\Delta^{-1}_{n,T}\rfloor - 1} T^{\gamma} \int_{t+j\Delta_{n,T}}^{t+(j+1)\Delta_{n,T}} \left(X_s - X_{t+j\Delta_{n,T}}\right) \sigma_s \mathrm{d}W_s \\
& \quad \times I_{\tau} \mathrm{E}\left(\left. \sum_{j=0}^{\lfloor T^{-\gamma}\Delta^{-1}_{n,T}\rfloor - 1} T^{\gamma} \int_{\tau+j\Delta_{n,T}}^{\tau+(j+1)\Delta_{n,T}} \left(X_s - X_{\tau+j\Delta_{n,T}}\right) \sigma_s \mathrm{d}W_s \right) \right| \mathcal{F}^X_{\tau+j\Delta_{n,T}} \Bigg) \\
= & 0,
\end{aligned}$$

with $\mathcal{F}^X_{\tau} = \sigma(X_{\tau+j\Delta_{n,T}}, j = 0, \ldots, \lfloor T^{-\gamma}\Delta^{-1}_{n,T}\rfloor - 1)$. Thus:

$$\mathrm{E}\left( \frac{1}{\sqrt{T}} \sum_{t=T/2+1}^{T} I_t \sum_{j=0}^{\lfloor T^{-\gamma}\Delta^{-1}_{n,T}\rfloor - 1} 2T^{\gamma} \int_{t+j\Delta_{n,T}}^{t+(j+1)\Delta_{n,T}} \left(X_s - X_{t+j\Delta_{n,T}}\right) \sigma_s \mathrm{d}W_s \right)^2$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=T/2+1}^{T} \mathrm{E}\left(I_t \sum_{j=0}^{\lfloor T^{-\gamma}\Delta_{n,T}^{-1}\rfloor-1}\left(2T^{\gamma}\int_{t+j\Delta_{n,T}}^{t+(j+1)\Delta_{n,T}}\left(X_s - X_{t+j\Delta_{n,T}}\right)\sigma_s \mathrm{d}W_s\right)^2\right)\\
&= \frac{1}{T}\sum_{t=T/2+1}^{T} \mathrm{E}\left(I_t \sum_{j=0}^{\lfloor T^{-\gamma}\Delta_{n,T}^{-1}\rfloor-1}\mathrm{E}\left(\left(2T^{\gamma}\int_{t+j\Delta_{n,T}}^{t+(j+1)\Delta_{n,T}}\left(X_s - X_{t+j\Delta_{n,T}}\right)\sigma_s \mathrm{d}W_s\right)^2\middle|\mathcal{F}_{t+j\Delta_{n,T}}^{X}\right)\right)\\
&= \frac{1}{T}\sum_{t=T/2+1}^{T} \mathrm{E}\left(I_t 2T^{2\gamma}\Delta_{n,T}^2 \sum_{j=0}^{\lfloor T^{-\gamma}\Delta_{n,T}^{-1}\rfloor-1}\sigma_{t+j\Delta_{n,T}}^4\right)(1+o(1))\\
&= O\left(T^{\gamma}\Delta_{n,T}\right) + o(1) = o(1),
\end{aligned}
$$

given A2(v). Hence:

$$
\widetilde{Z}_{n,T} = \widehat{V}_{n,T}^{-1/2}\sqrt{T}\left(\frac{1}{T}\left(\sum_{t=1}^{T/2}\sigma^2(X_t)\right)^2 I_t - \frac{1}{T}\sum_{t=T/2+1}^{T}\left(\sigma^2(X_t)\right)^2 I_t\right) + o_p(1).
$$

Let $\widetilde{V}_T$ be the infeasible version of (4), constructed replacing both $\widehat{\sigma}_{n,T}^2(X_t)$ and $\widetilde{\sigma}_{n,T,t}^2$ by $\sigma^2(X_t)$. Given A1-A4, as $n, T \to \infty$:

$$
\widehat{V}_{n,T} - \widetilde{V}_T = o_p(1).
$$

Finally, given A3(i) and A2(vii), $\widetilde{V}_T - V = o_p(1)$, where:

$$
V = \mathrm{avar}\left(\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T/2}\left(\sigma^2(X_t)\right)^2 I_t - \frac{1}{T}\sum_{t=T/2+1}^{T}\left(\sigma^2(X_t)\right)^2 I_t\right)\right).
$$

The statement in the theorem then follows straightforwardly from the central limit theorem for geometric ergodic processes.

$(ii)$ By a similar argument as in Part $(i)$:

$$
\frac{2}{\sqrt{T}}\sum_{t=T/2+1}^{T}\left(\widetilde{\sigma}_{n,T,t}^2\right)^2 I_t = \frac{2}{\sqrt{T}}\sum_{t=T/2+1}^{T}\left(\sigma_t^2\right)^2 I_t + o_p(1).
$$

The difference here is that $\sigma_t^2 I_t$ is not necessarily bounded. This affects the order of probability of $B_{n,T}^{(3)}$. However, given Lemma 1(ii), $B_{n,T}^{(3)} = O_{a.s.}\left(\sqrt{T}\sqrt{T^{-\gamma+\delta}\log\left(T^{\gamma+\delta}\right)}\right) = o_{a.s.}(1)$, because of A5(v).

Hence, we need to show that:

$$
\operatorname*{plim}_{n,T\to\infty}\frac{1}{T}\sum_{t=1}^{T/2}\left(\left(\widehat{\sigma}_{n,T}^2(X_t)\right)^2 - \left(\sigma_t^2\right)^2\right) I_t \neq 0.
$$

By Lemma 1(ii), $A_{n,T}^{(3)} = O_{a.s.}\left(T^{\delta}\sqrt{T\Delta_{n,T}\log\left(\Delta_{n,T}^{-1}\right)}\right) = o_{a.s.}(1)$, given A5(iv). $A_{n,T}^{(2)}$ is of smaller probability order than $A_{n,T}^{(1)}$. $A_{n,T}^{(1)} = o_p(1)$ by a similar argument as in part $(i)$, by simply setting $\eta = C\sqrt{\frac{\log(n)T^{1+2\delta}}{n\xi_{n,T}^2}}$, which is $o(1)$, by A5(iii). Thus, we need to show that:

$$\operatorname*{plim}_{n,T\to\infty} \frac{1}{T}\sum_{t=1}^{T/2}\left(I_t \frac{\sum_{j=1}^{n-1} K\left(\frac{X_t - X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)\left(\sigma^2_{j\Delta_{n,T}} - \sigma^2_t\right)}{\sum_{i=1}^{n-1} K\left(\frac{X_t - X_{\Delta_{n,T}}}{\xi_{n,T}}\right)}\right)^2 \neq 0.$$

We can rearrange terms as follows:

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T/2}\left(I_t \frac{\sum_{j=1}^{n-1} K\left(\frac{X_t - X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)\left(\sigma^2_{j\Delta_{n,T}} - \sigma^2_t\right)}{\sum_{i=1}^{n-1} K\left(\frac{X_t - X_{\Delta_{n,T}}}{\xi_{n,T}}\right)}\right)^2 \\
= \;&\frac{1}{T}\sum_{t=1}^{T/2}\left(I_t \frac{\sum_{j=1}^{n-1} K\left(\frac{X_t - X_{j\Delta_{n,T}}}{\xi_{n,T}}\right)\left(\sigma^2_{j\Delta_{n,T}} - \sigma^2\right)}{\sum_{i=1}^{n-1} K\left(\frac{X_t - X_{\Delta_{n,T}}}{\xi_{n,T}}\right)} - I_t\left(\sigma^2_t - \sigma^2\right)\right)^2, \quad \text{(A.9)}
\end{aligned}$$

where $\sigma^2 = \mathrm{E}(\sigma^2(V_t))$. The statement of the theorem follows straightforwardly by the law of large numbers. ■

# Proof of Lemma 1:

$(i)$ Given A1(i),

$$\begin{aligned}
&\sup_{0\le t,s\le T;|t-s|\le\Delta_{n,T}} |X_t - X_s|\, I_t \\
\le\;&\sup_{u\in(t,s),|t-s|\le\Delta_{n,T}} |\mu'(X_u)|\, I_t\Delta_{n,T} + \sup_{u\in(t,s),|t-s|\le\Delta_{n,T}} |\sigma'(X_u)|\, I_t \sup_{0\le t,s\le T;|t-s|\le\Delta_{n,T}} |W_t - W_s|\,.
\end{aligned}$$

Because of the modulus of continuity of the Brownian motion, for a finite $T < \overline{T}$:

$$\sup_{0\le t,s\le T;|t-s|\le\Delta_{n,T}} |W_t - W_s| = O_{a.s.}\left(\Delta_{n,T}^{1/2}\log\left(\Delta_{n,T}^{-1}\right)\right),$$

see e.g. Karatzas and Shreve (1991, p.114). We now have to take into account the case of $\overline{T} \le t, s \le T$, for $T\to\infty$. Define $\widetilde{W}_t = tW_{1/t}$ and $\widetilde{W}_s = sW_{1/s}$ and note that $\widetilde{W}_t$ and $\widetilde{W}_s$ are also Brownian motions (e.g. Corollary 9.4 in Karatzas and Shreve 1991). Hence:

$$\begin{aligned}
&\sup_{\overline{T}\le t,s\le T;|t-s|\le\Delta_{n,T}} \left|\widetilde{W}_t - \widetilde{W}_s\right| \\
=\;&\sup_{1/T\le 1/t,1/s\le 1/\overline{T};ts\left|\frac{1}{t}-\frac{1}{s}\right|\le\Delta_{n,T}} \left|tW_{\frac{1}{t}} - sW_{\frac{1}{s}}\right| \\
\le\;&\sup_{1/T\le 1/t,1/s\le 1/\overline{T};ts\left|\frac{1}{t}-\frac{1}{s}\right|\le\Delta_{n,T}} t\left|\left(W_{\frac{1}{t}} - W_{\frac{1}{s}}\right)\right| + \Delta_{n,T}\sup_{1/T\le 1/s\le 1/\overline{T}}\left|W_{\frac{1}{s}}\right|
\end{aligned}$$

$$= \quad O_{a.s}\left(\sqrt{\Delta_{n,T}\log\left(\Delta_{n,T}^{-1}\right)}\right) + O_{a.s.}\left(\Delta_{n,T}^{1-\varepsilon}\right) = O_{a.s}\left(\sqrt{\Delta_{n,T}\log\left(\Delta_{n,T}^{-1}\right)}\right).$$

It remains to consider $\sup_{u\in(t,s),|t-s|\leq\Delta_{n,T}} |\sigma'(X_u)| I_t$. For a finite constant $C$:

$$\sup_{t\leq T} |\sigma'(X_t)| I_t = \sup_{x\in B} |\sigma'(x)| \leq C,$$

because a continuous function over a bounded set has a maximum. Because $u \in (t,s)$, and $|t-s| \leq \Delta_{n,T}$, $X_u$ also belongs to a compact set, so that $\sup_{u\in(t,s),|t-s|\leq\Delta_{n,T}} |\sigma'(X_u)| I_t$ is bounded, almost surely.

$(ii)$ Given A1(i):

$$\begin{aligned} &\sup_{0\leq t,s\leq T;|t-s|\leq\Delta_{n,T}} |X_t - X_s| I_t \\ \leq\; &\sup_{u\in(t,s),|t-s|\leq\Delta_{n,T}} |\mu'(X_u)| I_t\Delta_{n,T} + \sup_{u\in(t,s),|t-s|\leq\Delta_{n,T}} |\sigma'(V_u)| I_t \sup_{0\leq t,s\leq T;|t-s|\leq\Delta_{n,T}} |W_t - W_s|. \end{aligned}$$

While $X_t$ is constrained to lie in a bounded set $B$, $V_t$ is not. Hence, we can't assume that $\sup_{u\in(t,s),|t-s|\leq\Delta_{n,T}} |\sigma'(V_u)|$ is bounded. However, given A5(ii), it follows immediately, using Markov inequalities, that $\sup_t |\sigma'(V_u)| = O_{a.s.}(T^{\delta})$. ■

# References


Aït-Sahalia, Y. (1996). Nonparametric Pricing of Interest Rate Derivative Securities. *Econometrica*, 64, 527-561.

Ang, A. and M. Piazzesi (2003). A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics*, 50, 745-787.

Bakshi, G., C. Cao and Z.W. Chen (2000). Do Call Prices and the Underlying Stock Always Move in the Same Direction? *Review of Financial Studies*, 13, 549-584.

Bandi, F.M. and P.C.B. Phillips (2003). Fully Nonparametric Estimation of Scalar Diffusion Models. *Econometrica*, 71, 241-283.

Bandi, F.M. and P.C.B. Phillips (2007). A Simple Approach to the Parametric Estimation of Potentially Nonstationary Diffusions. *Journal of Econometrics*, 137, 354-395.

Bergman, Y.Z., B.D. Grundy and Z. Wiener (1996). General Properties of Option Prices. *Journal of Finance*, LI, 1573-1610.

Bibkov, R. and M. Chernov (2006). No-Arbitrage Macroeconomic Determinants of the Yield Curve. Working Paper, London Business School.

Black, F., E. Derman and W. Toy (1990). A One-Factor Model of Interest Rates and its Application to treasury Bond Options. *Financial Analyst Journal*, 46, 33-39.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* 2nd Edition, Springer, New York.

Buraschi, A. and J. Jackwerth (2001). The Price of a Smile: Hedging and Spanning in Option Markets. *Review of Financial Studies*, 14, 495-527.

Buraschi, A. and A. Jiltsov (2006). Model Uncertainty and Option Markets with Heterogeneous Beliefs. *Journal of Finance*, LXI, 2841-2897.

Corradi, V., W. Distaso and A. Mele (2009). Macroeconomic Determinants of Stock Market Volatility and Volatility Risk-Premiums. Working Paper, Imperial College Business School, London.

Cox, J.C., J.E. Ingersoll and S.A. Ross (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53, 385-407.

Florens-Zmirou, D. (1993). On Estimating the Diffusion Coefficient from Discrete Observations. *Journal of Applied Probability*, 30, 790-804.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6, 327-343.

Hull, J.C. and A.D. White (1987). The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance*, XLII, 281-300.

Karatzsas, I. and S.E. Shreve (1991). *Brownian Motion and Stochastic Calculus.* Springer Verlag, New York.

Pardoux, E. and D. Talay (1985). Discretization and Simulation of Stochastic Differential Equations. *Acta Applicandae Mathematicae*, 3, 23-47.

de la Pena, V.H. (1999). A General Class of Exponential Inequalities for Martingales and Ratios. *Annals of Probability*, 27, 537-564.

Piazzesi, M. (2005). Bond Yields and the Federal Reserve. *Journal of Political Economy*, 113, 311-344.

Stein, E. and J.C. Stein (1991). Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *Review of Financial Studies*, 4, 727-752.

Vasicek, O.A. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5, 177-188.