# Testing Whether Volatility Can Be Written

# As a Function of the Asset Price

Yacine Aït-Sahalia
Princeton University and NBER

Jean Jacod
Université de Paris 6

# 1. Stochastic vs. Local Volatility

- Discrete-time ARCH, then GARCH: time-varying volatility is an important empirical feature in economic and financial data (Engle (1982))

- In continuous-time, models with stochastic volatility became prevalent
  - Require the use of an additional state variable $Y$, often latent, to drive the volatility of the variable of interest, $X$
  - $X$ on its own is typically no longer Markovian
  - Costly implications: higher dimension, plus $Y$ requires filtering or learning

- Local volatility models
  - Make the volatility of $X$ a function of $X$ itself
  - Retains the time variation of the volatility of $X$, and its Markov character, but without the additional stochasticity of $Y$
  - Such models, although more restrictive, remain very popular in financial applications

- This paper
  - Tests using high frequency observations whether the volatility of $X$ can be written as a function of $X$ only
  - Or whether some additional variable $Y$ is needed
  - Can

$$
\begin{aligned}
dX_t &= b_t dt + Y_t dW_t \\
dY_t &= m_t dt + v_t dW'_t
\end{aligned}
$$

be reduced to

$$dX_t = b_t dt + a(X_t) dW_t \text{ ?}$$

- Related question: whether $X$ can be written on its own as a Markov process
  - If $X$ is continuous, then $X$ will be Markovian if its drift and volatility functions are functions of $X$ (and time) only
  - Time-series tests examine the necessary (but not sufficient) condition that the conditional density of the process satisfies the Chapman-Kolmogorov condition (see Aït-Sahalia (1996) and Aït-Sahalia et al. (2010)); whether conditional independence between consecutive durations holds (see de Matos and Fernandes (2007)); and whether the conditional characteristic function satisfies the requirement imposed by the Markov property (see Chen and Hong (2012)).

- This paper tests directly whether the volatility of $X$ is a function of $X$ only
  - At high frequency over a finite time span, the drift is not identified
  - So only properties regarding the volatility of $X$ can be tested anyways
  - If $X$ can jump, additional separate restrictions on the structure of the jump component of $X$ are imposed by the Markov property.

# 2. The Model

$$X_t = X_0 + \int_0^t b_s \, ds + \int_0^t c_s^{1/2} \, dW_s + J_t,$$
$$J_t = \sum_{n \geq 1} \Upsilon_m 1_{\{T_m \leq t\}}.$$

- $W$ is a Brownian motion, $T_m$ jump times, $\Upsilon_m$ jump sizes
- Can be extended to Itô semimartingale with degree of activity strictly less than 1
- The local volatility case ($\mathrm{H}_0$):

  $$c_t = a(X_t)$$

  for some positive $\mathcal{C}^1$ function $a$ on $\mathbb{R}$
- Markov adds: $b_t = \mu(X_t)$ and a special structure on the jumps

- Observations on a finite $[0, T]$, sampling interval $\Delta_n \to 0$ as $n \to \infty$

- With $a(x) > 0$ is assumed, and when the jump process $J_t$ vanishes, the law of $X$ under ($\mathsf{H}_0$) in restriction to any finite time interval $[0, t]$ is equivalent to the law of the Markov process

$$\bar{X}_t = X_0 + \int_0^t a(\bar{X}_s)^{1/2} \, dW_s.$$

  Then, since $X$ is observed on a finite $[0, T]$ only, it is impossible to discriminate between $X$ and $\bar{X}$, hence ($\mathsf{H}_0$) is equivalent for the econometrician to assuming the Markov property, plus the smoothness of $a$

- Microstructure noise: white noise plus rounding

# 3. Testing the Null Hypothesis

## 3.1. Without Noise

- We wish to determine whether the spot variance $c_t$ can be written in the form $c_t = a(X_t)$ for some smooth function $a(\cdot)$
- Estimators localized in $x$
  - Choose a nonnegative kernel function $f$ on $\mathbb{R}$ with support in $[-1, 1]$ and with Lebesgue integral equal to 1, bandwidths $h_n$ and thresholds $v_n > 0$ satisfying, for some $\varepsilon > 0$:

$$\frac{h_n^3}{\Delta_n} \to 0, \qquad \frac{\Delta_n}{h_n^2} \to 0, \qquad v_n \to 0, \qquad \frac{\Delta_n^{1/2-\varepsilon}}{v_n} \to 0,$$

– Approximation of the Dirac mass at 0

$$f_n(x) = \frac{1}{h_n} f\left(\frac{x}{h_n}\right).$$

- Discrete increments (e.g., log-returns)

$$\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$$

- Define the statistics for integers $p$, typically in the set $\mathcal{P} = \{0, 2, 4\}$

$$U(x,p)_t^n = \Delta_n^{1-p/2} \frac{1}{m_p} \sum_{i=1}^{[t/\Delta_n]} f_n(X_{(i-1)\Delta_n} - x)\, |\Delta_i^n X|^p \, 1_{\{|\Delta_i^n X| \leq v_n\}}.$$

where $m_p =$ the $p$th absolute moment of $\mathcal{N}(0,1)$

- Truncation at level $v_n$ to get rid of the jumps, if any
  - If we know that $X$ is continuous, omit the indicator function above
  - The truncation can be kept even if we know that $X$ is continuous, in which case it is asymptotically irrelevant

- Local time

  - The limiting behavior of $U$ involves the local times $L^x$ of $X$ at $x$

$$L_t^x = |X_t - x| - |X_0 - x| - \int_0^t \mathrm{sign}(X_s - x)\, dX'_s \\ - \sum_{n \geq 1} \Big( |X_{T_n} - x| - |X_{T_n -} - x| \Big)\, 1_{\{T_n \leq t\}}.$$

  - $X' = X - J$ is the continuous part of $X$
  - $L_t^x > 0$ a.s. if the process $X$ has visited the point $x$ within $[0, T]$
  - This is the semimartingale version of the local time, which differs from the Markov local time by the factor $a(x)$
  - Has the advantage of being defined even when $(\mathrm{H}_0)$ fails

- Theorem:

$$U(x,p)_t^n \overset{\text{u.c.p.}}{\Longrightarrow} \int_0^t c_s^{p/2-1}\, dL_s^x.$$

- With $p = 2$ : $U(x,2)_t^n \overset{\text{u.c.p.}}{\Longrightarrow} L_t^x$

- Generic limitations
  - As in any other high frequency testing problem:
    - A single path of $X$ over $[0, T]$ is observed
    - Nothing about what happens after $t = T$ can be tested
  - Furthermore:
    - Nothing can be said about what happens to the volatility when the process is outside the (random) set of points visited by $X$ on $[0, T]$

## 3.2. The Idea Behind the Test

- Testing first at a single visited point $\mathcal{X} = \{x\}$
- Under the null, we have $c_s = c_{s'}$ for all $s, s' \in [0, t]$ with $X_s = X_{s'} = x$ and therefore

$$\int_0^t \int_0^t (c_s - c_{s'})^2 \, (c_s)^{-1} \, (c_{s'})^{-1} \, dL_s^x \, dL_{s'}^x = 0$$

- The alternative hypothesis ($\mathrm{H}_1$)
  - There is at least one pair $s, s' \in [0, T]$ such that $X_s = X_{s'}$ and $c_{s'} \neq c_s$.
  - This set is random, as often in high frequency
- The same quantity is strictly positive under the alternative
- This does not lend itself directly to a test since we have no good way of estimating $\int_0^t \int_0^t (c_s - c_{s'})^2 \, (c_s)^{-1} \, (c_{s'})^{-1} \, dL_s^x \, dL_{s'}^x$

- However, under the null we also have

$$\left(\int_0^t c_s \, dL_s^x\right) \left(\int_0^t c_s^{-1} \, dL_s^x\right) - (L_t^x)^2 = 0$$

- Using the theorem

$$\begin{aligned} & \bar{U}(x)_t^n := U(x, 4)_t^n \, U(x, 0)_t^n - (U(x, 2)_t^n)^2 \\ \xrightarrow{\mathbb{P}} \; & \bar{U}(x)_t := \left(\int_0^t c_s \, dL_s^x\right) \left(\int_0^t c_s^{-1} \, dL_s^x\right) - (L_t^x)^2 \end{aligned}$$

- So

$$\bar{U}(x)_t^n \xrightarrow{\mathbb{P}} \begin{cases} \bar{U}(x)_t = 0 & \text{under } (\mathsf{H}_0) \\ \bar{U}(x)_t > 0 & \text{under } (\mathsf{H}_1) \end{cases}$$

- Testing at an arbitrary finite set of visited points $\mathcal{X}$
  - Sum of $\bar{U}(x)_t^n$ over $x \in \mathcal{X}$, possibly weighted
  - The weight assigned to any given $x$ should reflect the time spent by $X$ at or around $x$, which is roughly proportional to the value of $L_t^x$, which itself is estimated by $U(x, 2)_t^n$
  - So we will consider a statistic of the form

$$\mathcal{U}(\mathcal{X}, r)_t^n = \sum_{x \in \mathcal{X}} \bar{U}(x)_t^n \left(U(x, 2)_t^n\right)^r$$

    for some $r$
  - We have

$$\mathcal{U}(\mathcal{X}, r)_t^n \xrightarrow{\mathbb{P}} \begin{cases} \mathcal{U} = 0 & \text{under } (\mathrm{H}_0) \\ \mathcal{U} > 0 & \text{under } (\mathrm{H}_1) \end{cases}$$

# 3.3. With Noise

- Plug-in $\tilde{X}_{i\Delta_n}$ instead of $X_{i\Delta_n}$ into the statistic, to form $U^{\text{noisy}}(x,p)_t^n$
  - However, even in the simplest white noise case we have

$$\frac{\Delta_n^{p/2}}{v_n^{p+1}} U^{\text{noisy}}(x,p)_t^n \stackrel{\text{u.c.p.}}{\Longrightarrow} \begin{cases} \frac{2}{p m_p} \int_0^t \phi(X_s - x)\, ds & \text{if } p > 0 \\ 2 \int_0^t \phi(X_s - x)\, ds & \text{if } p = 0, \end{cases}$$

  - The limit here has no connection with the volatility of $X$
  - Hence these statistics cannot be used for testing ($\text{H}_0$)

- Instead, we need to resort to a preliminary de-noising procedure
  - We use pre-averaging and plug-in the pre-averaged values of $\tilde{X}$
  - Then the same convergence holds

# 3.4. The Central Limit Theorem

- Theorem

$$\Big(\sqrt{h_n/k_n\Delta_n}\,\bar{U}(x)_t^n\Big)_{x\in\mathcal{X}} \stackrel{\mathcal{L}\text{-s}}{\to} (\bar{Z}(x)_t)_{x\in\mathcal{X}},$$

  where $\bar{Z}$ is $\mathcal{F}$–conditionally centered Gaussian with convariance

$$\tilde{\mathbb{E}}(\bar{Z}(x)_t\,\bar{Z}(y)_t \mid \mathcal{F}) = \frac{8\,\beta}{3}\,1_{\{x=y\}}a(x)\,(L_t^x)^3.$$

  and $\beta = \int_{-1}^{+1} f(x)^2 dx$.

- The variables $\bar{Z}(x)_t$ for distinct values of $x$ are $\mathcal{F}$–conditionally independent.

- Optimal Rate
  - The (not fully achievable, but almost) optimal rates are $1/\Delta_n^{1/3}$ without noise and $1/\Delta_n^{1/6}$ with noise, respectively.
  - In the no-noise case, this is the typical non-parametric rate
  - And, as in other situations, the presence of noise result in replacing the rate without noise by its square-root

# 3.5. Construction of the Test

- From the theorem, under (H$_0$):

$$\left(\frac{h_n}{k_n \Delta_n}\right)^{1/2} \mathcal{U}(\mathcal{X}, r)^n \overset{\mathcal{L}\text{-s}}{\Longrightarrow} \mathcal{Z}(\mathcal{X}, r) := (L^x)^r \sum_{x \in \mathcal{X}} \bar{Z}(x)$$

- $\mathcal{Z}(\mathcal{X}, r)_t$ is $\mathcal{F}$–conditionally centered Gaussian with conditional variance:

$$\Sigma(\mathcal{X}, r)_t = \frac{8\,\beta}{3} \sum_{x \in \mathcal{X}} a(x)(L^x_t)^{3+2r}$$

- Under (H$_0$) an estimator of $\Sigma(\mathcal{X}, r)_t$ is

$$\Sigma(\mathcal{X}, r)^n_t = \frac{8\,\beta}{3} \sum_{x \in \mathcal{X}} U(x, 0)^n_t \; (U(x, 4)^n_t)^2 \; (U(x, 2)^n_t)^{2r}$$

- Standardized test statistic

$$\mathcal{T}_t^n = \left(\frac{h_n}{k_n \Delta_n}\right)^{1/2} \frac{\mathcal{U}(\mathcal{X}, r)_t^n}{\left(\Sigma(\mathcal{X}, r)_t^n\right)^{1/2}}$$

- Which is $\mathcal{N}(0,1)$ under the null
- We use the set $\mathcal{P} = \{0, 2, 4\}$ of values of $p$, other choices are possible.

## 3.6. Tuning parameters

- The localization kernel $f$
  - Rather immaterial in practice, we suggest

$$f(x) = \frac{15}{16}\left(x^2 - 1\right)^2$$

  - In which case $\beta = 5/7$.

- The bandwidth $h_n$, plus the pre-averaging window $k_n$ in the noisy case
  - The rate of convergence for $\mathcal{U}(\mathcal{X}, r)_t^n$ gets faster when we increase $h_n$ and/or when we decrease $k_n$
  - A good choice: $h_n \asymp \Delta_n^{1/3} / \log(1/\Delta_n)$ in the no-noise case, leading to the rate $(\log(1/\Delta_n))^{1/2} / \Delta_n^{1/3}$
  - In the noisy case we cannot use $k_n \asymp \Delta_n^{1/2}$, which is the optimal choice in pre-averaging for estimating the volatility for example
  - A good choice: take $k_n \asymp \Delta_n^{1/2} \log(1/\Delta_n)$ and $h_n \asymp \Delta_n^{1/6}$, which leads to the rate $(\log(1/\Delta_n))^{1/2} / \Delta_n^{1/6}$

- The jump truncation threshold $v_n$
  - Choose $v_n \approx \gamma\Delta_n^{1/2}$ if there is no noise
  - And $v_n \approx \gamma\,(k_n\Delta_n)^{1/2}$ in the presence of noise, where $\gamma$ is 3 to 5 times a rough average of the volatility.

- The weighting power $r$
  - The choice of $r$ is asymptotically irrelevant
  - Increasing $r$ puts more weight on values visited often
  - Small sample discrepancies affect sites not visited often, so to avoid this problem $r$ should be taken relatively large
  - Increasing $r$ typically decreases the power of the test
  - Under the alternative, $\bar{U}(x)_t^n$ roughly varies as $(L_t^x)^2$, so "equal weights" means taking $r = -2$, which is good for the power of the test, but bad for the null
  - A good compromise is perhaps simply to take $r = 0$: then $\mathcal{U}_t^n$ would be a proxy for

$$\frac{1}{h_n^2} \int_0^t \int_0^t (c_s - c_{s'})^2 \, 1_{\{|X_s - X_{s'}| \leq h_n\}} \, ds \, ds'$$

    which is exactly the type of quantity which we want to test whether it vanishes or not

- The set where we test the property
  - When the cardinal of $\mathcal{X}$ is small, the procedure tests the property only at a limited number of values of $x$
  - This lowers the power of the test
  - So take $\mathcal{X}$ relatively large, but the minimal distance between two points of $\mathcal{X}$ should be bigger than $2h_n$
  - In practice, use a grid

# 4. Monte Carlo Simulations

- Null model

$$dX_t = \kappa(\alpha - X_t)\,dt + \sigma X_t^{1/2} dW_t$$

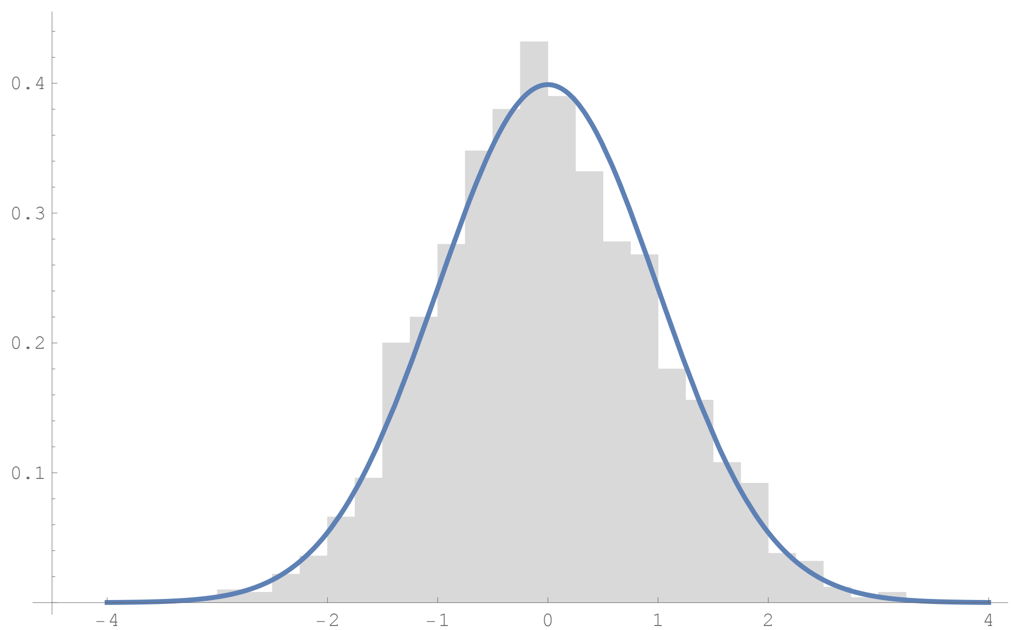
- Alternative

$$\begin{aligned} dX_t &= \kappa(\alpha - X_t)\,dt + \sigma(X_t + Y_t)^{1/2}\,dW_t \\ dY_t &= \beta(\omega - Y_t)\,dt + \eta Y_t^{1/2} dW_t' \\ \mathbb{E}\left[dW_t dW_t'\right] &= \rho dt \end{aligned}$$

- Paths simulated at 5 sec increments, data subsampled 1 to 5 mn, $T = 1$ year

- $\alpha = 10$, parameters give a stationary distribution with 99% of the mass on $[0, 30]$

- $X =$ uniform grid on $[0, 30]$ in increments of 1



# 5. Empirical Results: Interest Rates vs. Stock Prices

- Data
  - ES (S&P500 eMini Futures) 1998-2017 1mn frequency
  - ED (Eurodollar CME Futures) 1982-2017 5mn frequency, converted to interest rate equivalent
- Results
  - Clear rejection of the null for stocks
  - Not so for interest rates
  - Variation over time

# 6. Conclusions

- Testing whether

$$dX_t = b_t dt + Y_t dW_t$$
$$dY_t = m_t dt + v_t dW_t'$$

can be reduced to

$$dX_t = b_t dt + a(X_t) dW_t$$

- Based on a simple standardized statistic, involving sums of powers of increments with truncation (for jumps) and localization (at a given $x$), summed over $x$

- Asymptotically $\mathcal{N}(0, 1)$ under the null

- Interesting contrast of empirical results: stochastic volatility necessary to model stock prices, much less so for interest rates

# References

Aït-Sahalia, Y., 1996. Do interest rate really follow continuous-time Markov diffusions? Tech. rep., University of Chicago Working Paper.

Aït-Sahalia, Y., Fan, J., Jiang, J., 2010. Nonparametric tests of the Markov hypothesis in continuous-time models. Annals of Statistics 38, 3129–3163.

Chen, B., Hong, Y., 2012. Testing the Markov property in time series. Econometric Theory 28, 130–178.

de Matos, J. A., Fernandes, M., 2007. Testing the Markov property with high frequency data. Journal of Econometrics 141, 44–64.

Engle, R. F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica 50, 987–1008.