Testing the rank of the volatility process: A random perturbation approach

Mark Podolskij

joint work with J. Jacod

Heidelberg University, Germany

Okinawa Meeting 2013

Outline

- Formulation of the statistical problem
- Random perturbation of the original data
- Asymptotic theory
- Testing procedure

Continuous diffusion processes

- We consider a d-dimensional continuous diffusion process of the form

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t \sigma_s dW_s, \qquad t \in [0,1].$$

  In this model
  - σ is a $\mathbb{R}^{d\times d}$-valued volatility process
  - a is a d-dimensional drift process
  - W is a d-dimensional Brownian motion

- We observe high frequency data

$$X_0, X_{\Delta_n}, X_{2\Delta_n}, \ldots, X_{[1/\Delta_n]\Delta_n}$$

  with $\Delta_n \to 0$.

The statistical problem

- Let r denote the rank of the matrix $c = \sigma\sigma^\star$, i.e.

$$r_t = \mathrm{rank}(c_t), \qquad t \in [0,1].$$

Our aim is to estimate/test the maximal rank of c during a given trading day $[0,1]$. Hence, our object of interest is given via

$$R = \sup_{t\in[0,1)} r_t.$$

- We remark that the set $\{t \in [0,1) |\ r_t = R\}$ has positive Lebesgue measure if σ is a continuous process.

- Warning: The random variable R has no connection to the integrated volatility matrix, i.e.

$$R \neq \mathrm{rank}\Big(\int_0^1 c_s ds\Big).$$

A perturbation method

We explain the basic idea on a non-random problem. Let $A \in \mathbb{R}^{d \times d}$ be a given positive semidefinite matrix with $r = \text{rank}(A)$. We consider a positive definite matrix $B \in \mathbb{R}^{d \times d}$ and a number $h \searrow 0$. By the multilinearity of the determinant we obtain the identity

$$\det(A + hB) = h^{d-r}\gamma_{A,B} + O(h^{d-r+1})$$

with $\gamma_{A,B} := \sum_{C \in \mathcal{M}_{A,B}} \det(C)$ and

$$\mathcal{M}_{A,B} := \{C \in \mathbb{R}^{d \times d} : \ C_i = A_i \text{ or } C_i = B_i, \text{ A and C share } r \text{ joint columns}\},$$

where $A = (A_1, \ldots, A_d)$ and $B = (B_1, \ldots, B_d)$. When $\gamma_{A,B} \neq 0$, we deduce that

$$\frac{\det(A + 2hB)}{\det(A + hB)} \to 2^{d-r} \qquad \text{as} \quad h \searrow 0.$$

A simple example

Let $A = (A_1, A_2, A_3), B = (B_1, B_2, B_3) \in \mathbb{R}^{3\times 3}$ and $r = \text{rank}(A) = 1$. Then it holds that

$$\begin{aligned} \det(A + hB) = {} & \underbrace{\det(A_1, A_2, A_3)}_{=0} \\ & + h \underbrace{[\det(A_1, A_2, B_3) + \det(A_1, B_2, A_3) + \det(B_1, A_2, A_3)]}_{=0} \\ & + h^2 \underbrace{[\det(A_1, B_2, B_3) + \det(B_1, A_2, B_3) + \det(B_1, B_2, A_3)]}_{=\gamma_{A,B}} \\ & + \underbrace{h^3 \det(B_1, B_2, B_3)}_{=O(h^3)} \end{aligned}$$

Some remarks

- Although our application of random perturbation method to diffusion is new, there exist similar ideas in other fields. Let us mention the following
  - Functional analysis $\longrightarrow$ Tikhonov regularization
  - Statistical inverse problems
  - Linear regression $\longrightarrow$ ridge regression
  - Signal-noise models

- When the matrix A is not directly observed, we cannot choose a matrix B such that

$$\gamma_{A,B} \neq 0.$$

  However, in the stochastic setting the choice $B = (B_1, \ldots, B_d)$, (B_i) iid $\mathcal{N}_d(0, I_d)$, independent of A, guarantees that $\gamma_{A,B} \neq 0$ almost surely.

The test statistic

- First, we introduce a random perturbation of the original diffusion process

$$Z_t^n = X_t + \sqrt{\Delta_n}\ \widehat{W}_t,$$

where $\widehat{W}$ is a new Brownian motion independent of everything.

- The test statistic $S(Z^n, \Delta_n)$ is defined via

$$S(Z^n, \Delta_n) = \sum_{i=1}^{[1/\Delta_n]-d+1} \det{}^2 \left(\Delta_i^n Z^n / \sqrt{\Delta_n}, \ldots, \Delta_{i+d-1}^n Z^n / \sqrt{\Delta_n} \right)$$

with $\Delta_i^n Z^n = Z_{i\Delta_n}^n - Z_{(i-1)\Delta_n}^n$.

- The test statistic $\Delta_n S(X, \Delta_n)$ has been used in Jacod, Lejay and Talay (2008) to test for the full rank.

Assumptions

- In contrast to the usual asymptotic theory for high frequency data, we require some stronger assumptions:

$$a_t = a_0 + \int_0^t a_s^{(1)} ds + \int_0^t a_s^{(2)} dW_s,$$

$$\sigma_t = \sigma_0 + \int_0^t \sigma_s^{(1)} ds + \int_0^t \sigma_s^{(2)} dW_s.$$

- Furthermore, the process $\sigma^{(2)} \in \mathbb{R}^{d \times d \times d}$ must be diffusions of the same type as X, a and σ.

An asymptotic decomposition

- We define for any $1 \leq l \leq d$

$$\alpha_{i,l}^n = \Delta_n^{-1/2} \sigma_{(i-1)\Delta_n} \Delta_{i+l-1}^n W$$

$$\beta_{i,l}^n = \Delta_n^{-1/2} \Delta_{i+l-1}^n \widehat{W} + a_{(i-1)\Delta_n}$$

$$+ \Delta_n \sigma_{(i-1)\Delta_n}^{(2)} \int_{(i-l)\Delta_n}^{(i-l+1)\Delta_n} (W_s - W_{(i-l)\Delta_n}) dW_s$$

- With the notation $\alpha_i^n = (\alpha_{i,1}^n, \ldots, \alpha_{i,d}^n)$, $\beta_i^n = (\beta_{i,1}^n, \ldots, \beta_{i,d}^n)$, we deduce the asymptotic relation

$$(\Delta_i^n Z^n / \sqrt{\Delta_n}, \ldots, \Delta_{i+d-1}^n Z^n / \sqrt{\Delta_n}) = \underbrace{\alpha_i^n}_{=A} + \underbrace{\Delta_n^{1/2} \beta_i^n}_{=hB} + O_{\mathbb{P}}(\Delta_n).$$

Consistency

Theorem: Under the aforementioned assumptions, we obtain the following results.

(i) As $n \to \infty$

$$\Delta_n^{1+R-d} \sum_{i=1}^{[1/\Delta_n]-d+1} \det{}^2 \left(\Delta_i^n Z^n / \sqrt{\Delta_n}, \ldots, \Delta_{i+d-1}^n Z^n / \sqrt{\Delta_n} \right)$$

$$\xrightarrow{\mathbb{P}} S = \int_0^1 \Gamma(a_s, \sigma_s, \sigma_s^{(2)}) ds > 0.$$

(ii) In particular, we deduce that

$$T_n := \frac{S(Z^n, 2\Delta_n)}{S(Z^n, \Delta_n)} \xrightarrow{\mathbb{P}} 2^{d-R}.$$

Central limit theorem

Theorem: Under the aforementioned assumptions, we obtain the following results.

(i) As $n \to \infty$

$$\Delta_n^{-1/2}\Big(\Delta_n^{1+R-d} S(Z^n, \Delta_n) - S, (2\Delta_n)^{1+R-d} S(Z^n, 2\Delta_n) - S\Big)$$

$$\xrightarrow{d_{st}} MN\Big(0, \int_0^1 V(a_s, \sigma_s, \sigma_s^{(2)}) ds\Big).$$

(ii) In particular, we deduce that

$$\Delta_n^{-1/2}\Big(d - \frac{\log T_n}{\log 2} - R\Big) \xrightarrow{d_{st}} MN\Big(0, \int_0^1 \Sigma(a_s, \sigma_s, \sigma_s^{(2)}) ds\Big),$$

where the asymptotic variance $\int_0^1 \Sigma(a_s, \sigma_s, \sigma_s^{(2)}) ds$ can be consistently estimated by Σ_n.

Testing procedure

- Given a number $R_0 \in \{0, \dots, d\}$, let us consider the following null/alternative hypothesis

$$H_0: \; R = R_0 \qquad \text{vs.} \qquad H_1: \; R \neq R_0.$$

- Define the test statistic R_n by

$$R_n = d - \frac{\log T_n}{\log 2}.$$

- We obtain that ($c_\alpha = \alpha$-quantile of $\mathcal{N}(0,1)$)

$$\mathbb{P}_{H_0}\left(\left| \frac{\Delta_n^{-1/2}(R_n - R_0)}{\sqrt{\Sigma_n}} \right| > c_{1-\frac{\alpha}{2}} \right) \to \alpha,$$

$$\mathbb{P}_{H_1}\left(\left| \frac{\Delta_n^{-1/2}(R_n - R_0)}{\sqrt{\Sigma_n}} \right| > c_{1-\frac{\alpha}{2}} \right) \to 1.$$

Other applications

- The hypothesis testing

$$H_0 : \; R = 0 \qquad \text{vs.} \qquad H_1 : \; R > 0$$

corresponds to the test

$$H_0 : \; \text{X= integrated diffusion} \quad \text{vs.} \quad H_1 : \; \text{X= diffusion}$$

Integrated diffusions appear naturally in the engineering science. We refer to the work of M. Sorensen and A. Gloter for statistical methods.

- With some more work our method can be applied to the model $Y = (X, \sigma)$, where X is a one-dimensional diffusion with volatility σ. Testing

$$H_0 : \; R \leq 1 \qquad \text{vs.} \qquad H_1 : \; R = 2$$

is related to testing the local volatility assumption (see Podolskij and Rosenbaum (2011)). These type of questions are also important in the theory of financial bubbles developed by P. Protter.

Simulation design

- We consider the model $dX_t = a_t dt + \sigma_t dW_t$ with

$$\sigma_t = (1 + (2t-1)^2) \begin{pmatrix} \cos(tA\pi/2) & \cos(tA\pi/2) \\ \sin(tA\pi/2) & \sin(tA\pi/2) \end{pmatrix},$$

$$a_t = B \begin{pmatrix} 1 + \sin(tA\pi/2) \\ 1 + \cos(tA\pi/2) \end{pmatrix}$$

- The frequency is given as

$$\Delta_n = \frac{1}{25000}$$

- We perform 5000 replications to uncover the finite sample properties.

Simulation results at level $\alpha = 0.05$

B	A	2d moment	4th moment	$R = 0$	$R = 1$	$R = 2$
0	0	1.01	2.99	1.00	0.050	1.00
0	5	1.00	2.94	1.00	0.049	1.00
0	10	1.01	3.36	1.00	0.062	1.00
3	0	1.02	3.26	1.00	0.050	1.00
3	5	1.02	3.20	1.00	0.052	1.00
3	10	1.01	3.03	1.00	0.049	1.00
12	0	0.98	2.88	1.00	0.049	1.00
12	3	1.02	3.13	1.00	0.052	1.00
12	10	1.01	2.99	1.00	0.050	1.00

Thank you!