A test for the rank of the volatility process: the random perturbation approach

Jean Jacod * Mark Podolskij †

July 30, 2013

Abstract

In this paper we present a test for the maximal rank of the matrix-valued volatility process in the continuous Itô semimartingale framework. Our idea is based upon a random perturbation of the original high frequency observations of an Itô semimartingale, which opens the way for rank testing. We develop the complete limit theory for the test statistic and apply it to various null and alternative hypotheses. Finally, we demonstrate a homoscedasticity test for the rank process.

Keywords: central limit theorem, high frequency data, homoscedasticity testing, Itô semimartingales, rank estimation, stable convergence.

AMS 2010 Subject Classification: 62M07, 60F05, 62E20, 60F17.

1 Introduction

In the last years asymptotic statistics for high frequency observations has received a lot of attention in the literature. This interest was mainly motivated by financial applications, where observations of stocks or currencies are available at very high frequencies. As under the no-arbitrage condition prices processes must be semimartingales, a lot of research has been devoted to statistics of high frequency data of semimartingales.

This paper is devoted to testing for the maximal rank of the matrix-valued volatility process in the continuous Itô semimartingale framework, and more specifically for a d-dimensional continuous Itô semimartingale X which is observed at equidistant times over a fixed time interval $[0, T]$: we observe $(X_{i\Delta_n})_{0 \le i \le [T/\Delta_n]}$, and the high-frequency approach consists in assuming $\Delta_n \to 0$.

A continuous Itô semimartingale can be written as

$$dX_t = b_t dt + \sigma_t dW_t, \tag{1.1}$$

*Institut de Mathématiques de Jussieu, 4 Place Jussieu, 75 005 Paris, France (CNRS – UMR 7586, and Université Pierre et Marie Curie), Email: jean.jacod@upmc.fr

†Department of Mathematics, Heidelberg University, INF 294, 69120 Heidelberg, Germany, Email: m.podolskij@uni-heidelberg.de.

where W is a Brownian motion, and there are many representations of this form, with different Brownian motions W and, accordingly, different volatility processes σ. What is "intrinsic" is the drift coefficient b_t and the diffusion coefficient ("squared volatility") $c_t = \sigma_t \sigma_t^*$, in the sense that they are uniquely determined by X, up to a Lebesgue-null set of times (throughout the paper σ_t^* denotes the transpose of the matrix σ).

For modeling purposes and economical interpretation we would like to find, and often choose, the smallest possible dimension of the Brownian motion W in the representation (1.1). Assuming further that $t \mapsto c_t$ is continuous, this smallest possible dimension is the supremum in time of the rank of the $\mathbb{R}^{d\times d}$-valued process c over the time interval $[0, T]$. We are further interested in homoscedasticity testing for the rank process.

Our method is based upon a random perturbation of the original data and determinant expansions. The main idea can be described as follows: if we compute $\det(c_t + he_t)$ for a positive definite $d \times d$ matrix e_t independent of c_t and $h \downarrow 0$, then, under appropriate conditions, its rate of decay to 0 depends on the unknown rank of c_t. Hence, the ratio $\det(c_t + 2he_t)/\det(c_t + he_t)$ asymptotically identifies the rank of c_t. Indeed, our main statistic is a partial sum of squared determinants of matrices build from d consecutive increments of the process X and the random perturbation is performed by a properly scaled Brownian motion W', which is independent of all ingredients of X. We remark that perturbation methods (and matrix expansions as well) find applications in various fields of mathematics.

2 The model and main results

2.1 The setting and testing hypotheses

Our process of interest is a d-dimensional continuous Itô semimartingale X, given on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. In vector form, and with W denoting a q-dimensional Brownian motion, it can be written as

$$X_t = X_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s, \tag{2.1}$$

where b_t is a d-dimensional drift process and σ_t is a $\mathbb{R}^{d\times q}$-valued volatility process, assumed to be continuous in time (and indeed much more, see Assumption (H) below). We set

$$c_t = \sigma_t \sigma_t^*, \qquad r_t = \operatorname{rank}(c_t), \qquad R_t = \sup_{s\in[0,t)} r_s. \tag{2.2}$$

We remark that the maximal rank R_T is not bigger than the rank of the integrated volatility $\int_0^T c_t dt$, but may be strictly smaller. As already mentioned, it is suitable to use the smallest possible dimension for W, on the time interval $[0, T]$. This is the $\mathbb{P}$-essential supremum of $\omega \mapsto R_T(\omega)$, but, since a single path $t \mapsto X_t(\omega)$ is (partially) observed, the only available information is R_T itself. So the problem really boils down to finding the behavior of the process r_t, and for this the choice of the dimension of W in (2.1) is irrelevant.

The rank r_t is the biggest integer $r \leq d$ such that the sum of the determinants of the matrices $(c_t^{ij})_{i,j \in J}$, where J runs through all subsets of $\{1, \cdots, d\}$ with r points, is positive (with the convention that a 0×0 matrix has determinant 1). Since c_t is continuous, this implies that for any r the random set $\{t : r_t(\omega) > r\}$ is open in $[0, T)$, so the mapping $t \mapsto r_t$ is lower semi-continuous. In particular, the set $\{t \in [0, T) : r_t(\omega) = R_T(\omega)\}$ is a non-empty open subset. These properties also yield that the process r_t is predictable and that the following subsets of Ω, which later will be the "testing hypotheses", are $\mathcal{F}_T$-measurable:

$$\begin{aligned} \Omega_T^r &= \{\omega : R_T(\omega) = r\} \\ \Omega_T^= &= \{\omega : r_t(\omega) = R_T(\omega) \text{ for all } t \in [0, T]\} \\ \Omega_T^{\neq} &= \{\omega : t \mapsto r_t(\omega) \text{ has finitely many discontinuities and is} \\ &\qquad\qquad \text{not Lebesgue-a.s. constant on } [0, T]\}. \end{aligned} \tag{2.3}$$

Notice that we impose that $r_T = R_T$ in $\Omega_T^=$, whereas the lower semi-continuity only implies in general that $r_T \leq R_T$. Observe also that *a priori* $t \mapsto r_t$ may be Lebesgue-a.s. constant and still have discontinuities (even infinitely many) on $[0, T]$. So, $\Omega_T^=$ and $\Omega_T^{\neq}$ are disjoint but $\Omega_T^= \cup \Omega_T^{\neq} \neq \Omega$ in general. The main aim of this paper is testing the null hypothesis Ω_T^r against $\Omega_T^{\neq r} = \cup_{r' \neq r, 0 \leq r' \leq d} \Omega_T^{r'}$ (and related hypotheses) and testing the null hypothesis of $\Omega_T^=$ against $\Omega_T^{\neq}$.

2.2 Matrix perturbation

In order to explain the main idea of our method, we need to introduce some notation. Recall that d and q are the dimensions of X and W, respectively. Then $\mathcal{M}$ is the set of all $d \times d$ matrices, $\mathcal{M}_r$ for $r \in \{0, \cdots, d\}$ is the set of all matrices in $\mathcal{M}$ with rank r, and $\mathcal{M}'$ is the set of all $d \times q$ matrices. For any matrix A we denote by A_i the ith column of A; for any vectors $x_1, \ldots, x_d$ in $\mathbb{R}^d$, we write $\text{mat}(x_1, \cdots, x_d)$ for the matrix in $\mathcal{M}$ whose ith column is the column vector x_i. For $r \in \{0, \cdots, d\}$ and $A, B \in \mathcal{M}$ we define

$$\mathcal{M}_{A,B}^r = \{G \in \mathcal{M} : G_i = A_i \text{ or } G_i = B_i \text{ with } \#\{i : G_i = A_i\} = r\}. \tag{2.4}$$

In other words, $\mathcal{M}_{A,B}^r$ is the set of all matrices $G \in \mathcal{M}$ with r columns equal to those of A (at the same places), and the remaining $d - r$ ones equal to those of B. Let us define

$$\gamma_r(A, B) = \sum_{G \in \mathcal{M}_{A,B}^r} \det(G). \tag{2.5}$$

We demonstrate the main ideas for a deterministic problem first. Let $A \in \mathcal{M}$ be an unknown matrix with rank r. Assume that, although A is unknown, we have a way of computing $\det(A + hB)$ for all $h > 0$ and some given matrix $B \in \mathcal{M}_d$. The multi-linearity property of the determinant implies the following asymptotic expansion

$$\det(A + hB) = h^{d-r} \gamma_r(A, B) + O(h^{d-r+1}), \tag{2.6}$$

which is the core of our method. Thus, if $\gamma_r(A,B) \neq 0$, we have

$$\frac{\det(A+2hB)}{\det(A+hB)} \to 2^{d-r} \qquad \text{as} \quad h \downarrow 0. \tag{2.7}$$

and this convergence identifies the parameter r. However, it is impossible to choose a matrix $B \in \mathcal{M}$ which guarantees $\gamma_r(A,B) \neq 0$ for all $A \in \mathcal{M}_r$. To solve this problem we can use a random perturbation. As we will show later, for any $A \in \mathcal{M}_r$ we have $\gamma_r(A,B) \neq 0$ a.s. when B is the random matrix whose entries are independent standard normal. This idea will be the core of our testing procedure.

2.3 Assumptions and the test statistic

Before we proceed with the definition of the test statistic, we introduce the main assumptions. We need more structure than the mere Equation (2.1), namely that the processes b_t and σ_t, and also the volatility of σ_t, are continuous Itô semimartingales. In view of the previous discussion, it is no restriction to assume that all these are driven by the same q-dimensional Brownian motion, provided we take q large enough. This leads us to put

Assumption (H): The d-dimensional semimartingale X, defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, has the form

$$\begin{aligned} X_t &= X_0 + \textstyle\int_0^t b_s\,ds + \int_0^t \sigma_s\,dW_s \\ \sigma_t &= \sigma_0 + \textstyle\int_0^t a_s\,ds + \int_0^t v_s\,dW_s \\ b_t &= b_0 + \textstyle\int_0^t a'_s\,ds + \int_0^t v'_s\,dW_s \\ v_t &= v_0 + \textstyle\int_0^t a''_s\,ds + \int_0^t v''_s\,dW_s, \end{aligned} \tag{2.8}$$

where W is a q-dimensional Brownian motion, and b_t and a'_t are $\mathbb{R}^d$-valued, σ_t, a_t and v'_t are $\mathbb{R}^{d\times q}$-valued, v_t and a''_t are $\mathbb{R}^{d\times q\times q}$-valued, and v''_t is $\mathbb{R}^{d\times q\times q\times q}$-valued, all those processes being adapted. Finally, the processes a_t, v'_t, v''_t are càdlàg and the processes a'_t, a''_t are locally bounded. □

At this stage it is not quite clear why the full force of assumption (H) is required. In the standard limit theory for high frequency data of continuous Itô semimartingales only the first two representations of (2.8) are assumed. When $b_t = g_1(X_t)$, $\sigma_t = g_2(X_t)$ with $g_1 \in C^2(\mathbb{R}^d)$ and $g_2 \in C^4(\mathbb{R}^d)$, then (H) is automatically satisfied, due to Itô's formula.

Motivated by the matrix perturbation at (2.6), our tests will be based on statistics involving sums of (squared) determinants. The test function will be the nonnegative map f on $(\mathbb{R}^d)^d$ defined as

$$f(x_1,\ldots,x_d) = \det(\text{mat}(x_1,\cdots,x_d))^2. \tag{2.9}$$

Now, we introduce a random perturbation of the original process X as motivated at the end of Subsection 2.2. More specifically, we choose a non-random invertible $d\times d$ matrix $\widetilde{\sigma}$ and generate a new process

$$X'_t = \widetilde{\sigma} W'_t,$$

where W' is a d-dimensional Brownian motion independent of all processes in (2.8)(without loss of generality, for the mathematical treatment below we may assume that it is also defined on the space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$). Following the ideas of section 2.2, we add to X this new process X', with a multiplicative factor going to 0. As a matter of fact we introduce two such additions, and for $\kappa = 1$ or 2 we set

$$Z_t^{n,\kappa} = X_t + \sqrt{\kappa\Delta_n}\, X'_t. \tag{2.10}$$

Hence, with the notation of section 2.2, we use $h = \sqrt{\Delta_n}$, which leads later to the optimal rate of convergence. We define the following two basic statistics:

$$S_t^{n,1} = 2d\Delta_n \sum_{i=0}^{[t/2d\Delta_n]-1} f\left(\frac{Z^{n,1}_{(2id+1)\Delta_n} - Z^{n,1}_{2id\Delta_n}}{\sqrt{\Delta_n}}, \cdots, \frac{Z^{n,1}_{(2id+d)\Delta_n} - Z^{n,1}_{(2id+d-1)\Delta_n}}{\sqrt{\Delta_n}}\right) \tag{2.11}$$

$$S_t^{n,2} = 2d\Delta_n \sum_{i=0}^{[t/2d\Delta_n]-1} f\left(\frac{Z^{n,1}_{(2id+2)\Delta_n} - Z^{n,1}_{(2id)\Delta_n}}{\sqrt{2\Delta_n}}, \cdots, \frac{Z^{n,1}_{(2id+2d)\Delta_n} - Z^{n,1}_{(2id+2d-2)\Delta_n}}{\sqrt{2\Delta_n}}\right).$$

Notice that the statistics $S_t^{n,1}$ and $S_t^{n,2}$ are essentially the same, except $S_t^{n,2}$ is computed using the frequency $2\Delta_n$. The main result of our paper is the convergence in probability

$$\widehat{R}(n,T) := d - \frac{\log(S_T^{n,2}/S_T^{n,1})}{\log(2)} \xrightarrow{\mathbb{P}} R_T,$$

which enables us to estimate the unknown maximal rank. Furthermore, we prove the associated central limit theorem to provide a consistent testing procedure.