

On the estimation of spectral distribution of integrated covariance matrices of high dimensional stochastic processes

Aleksandr Samarin

August 2017

Abstract

In the 1-subject Master's program in Financial Mathematics of the Faculty of Mathematics and Natural Sciences of the CAU in Kiel

presented by Aleksandr Samarin

First referee: Second referee: Kiel, August 2017

(COPIED FROM THE PAPER)

We consider the estimation of integrated covariance (ICV) matrices of high dimensional diffusion processes based on high frequency observations. We start by studying the most commonly used estimator, the realized covariance (RCV) matrix. We show that in the high dimensional case when the dimension p and the observation frequency n grow in the same rate, the limiting spectral distribution (LSD) of RCV depends on the covolatility process not only through the targeting ICV, but also on how the covolatility process varies in time. We establish a Marcenko–Pastur type theorem for weighted sample covariance matrices, based on which we obtain a Marcenko–Pastur type theorem for RCV for a class C of diffusion processes. The results explicitly demonstrate how the time variability of the covolatility process affects the LSD of RCV. We further propose an alternative estimator, the time-variation adjusted realized covariance (TVARCV) matrix. We show that for processes in class C , the TVARCV possesses the desirable property that its LSD depends solely on that of the targeting ICV through the Marcenko–Pastur equation, and hence, in particular, the TVARCV can be used to recover the empirical spectral distribution of the ICV by using existing algorithms

Contents

1	Introduction	3
2	Marchenko-Pastur law	8
3	Limit theorems for non-constant covolatility processes	13
4	Simulation studies	22
5	Conclusion	28

Part 1

Introduction

Motivation

Say general words... v

Suppose that we have multiple stocks, say, p stocks, whose log price processes are denoted by $X_j(t)$ for $j = 1, \dots, p$. Let

$$\mathbf{X}(t) = (X_1(t), \dots, X_p(t))^T.$$

Then a widely used model for \mathbf{X}_t is

$$d\mathbf{X}(t) = \boldsymbol{\mu}(t)dt + \Theta(t)d\mathbf{W}(t), \quad (1)$$

where

- $\boldsymbol{\mu}(t) = (\mu_1(t), \dots, \mu_p(t))^T$ is a p -dimensional drift process,
- $\Theta(t)$ – $p \times p$ **covolatility process**,
- $\mathbf{W}(t)$ – p -dimensional standard Brownian motion.

Definition 1.1.

- (i) The **integrated covariance matrix (ICV)** is defined as follows

$$\Sigma_p := \int_0^1 \Theta(t) \cdot \Theta^T(t) dt.$$

For $p = 1$ ICV is called **integrated volatility**.

- (ii) Set time points $\{\tau_\ell\}_{0 \leq \ell \leq n}$. Then **realized covariance (RCV)** is

$$\Sigma_p^{RCV} := \sum_{\ell=1}^n \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T, \quad (2)$$

where

$$\Delta \mathbf{X}_\ell := \begin{pmatrix} \Delta X_{1\ell} \\ \vdots \\ \Delta X_{p\ell} \end{pmatrix} = \begin{pmatrix} X_1(\tau_\ell) - X_1(\tau_{\ell-1}) \\ \vdots \\ X_p(\tau_\ell) - X_p(\tau_{\ell-1}) \end{pmatrix}.$$

For $p = 1$ RCV is called **realized volatility**.

Example 1.2. (CHECK NON-SYMMETRIC THETA AND SHOW X!) Let's take $p = 3$, zero drift $\mu(t) = (0, 0, 0)^T$ and

$$\Theta(t) = \begin{pmatrix} 1 & \cos(\pi t) & \sin(\pi t) \\ \cos(\pi t) & 1 & 0 \\ \sin(\pi t) & 0 & 1 \end{pmatrix}.$$

Then

$$\Theta(t) \cdot \Theta^T(t) = \begin{pmatrix} 2 & 2\cos(\pi t) & 2\sin(\pi t) \\ 2\cos(\pi t) & \cos^2(\pi t) + 1 & \cos(\pi t)\sin(\pi t) \\ 2\sin(\pi t) & \cos(\pi t)\sin(\pi t) & \sin^2(\pi t) + 1 \end{pmatrix}$$

and corresponding ICV matrix is

$$\Sigma_3 = \begin{pmatrix} 2 & 0 & \frac{4}{\pi} \\ 0 & \frac{3}{2} & 0 \\ \frac{4}{\pi} & 0 & \frac{3}{2} \end{pmatrix}.$$

Remark 1.3.

- (i) Intuitively, one can wish zero elements of the matrix Σ_3 from Example 1.2 to show the independence of one process from another, or at least lack of correlation between them. But even if in the case of $(X_2(t), X_3(t))$ it is true, it is totally wrong for $(X_1(t), X_2(t))$. In this paper we focus on the processes, which correlation structure doesn't change in time. (SOMETHING ABOUT INTRADAY DATA)
- (ii) One of the major problems in of all the models, based on the theory of stochastic calculus and applied to financial markets, is the estimation of so-called true covariance. In our case the final goal is to find an estimator of an ICV matrix. In the next section we show how RCV matrix does the job.

The relation between RCV and ICV

Remark 1.4. Recall the basic notions from the theory of stochastic calculus (w.l.o.g. we assume that all the processes have a starting point at 0 a.s.):

- (i) An **Itô process** $X(t)$ is defined to be an adapted stochastic process that can be expressed as the sum of an integral with respect to time and an integral with respect to Brownian motion:

$$X(t) = \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW(s).$$

- (ii) Let X be a continuous local martingale. Then $[X]_t$ with

$$[X](t) = X^2(t) - 2 \int_0^t X(s)dX(s)$$

is called **quadratic variation** of X . Note that for deterministic processes quadratic variation is equal to 0.

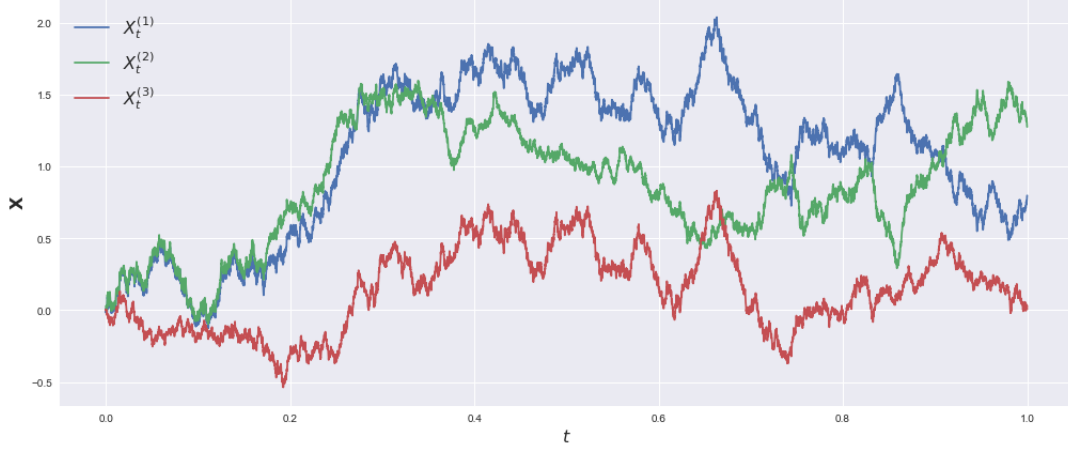


Figure 1: $\mathbf{X}(t)$ from Example 1.2

One can observe that $X_1(t)$ and $X_2(t)$ have strong positive correlation in the beginning, which decreases over time and later, when t approaches 1, turns into negative one. Meanwhile $X_3(t)$ has independent fluctuations on the borders, but in the middle it repeats the behavior of $X_1(t)$.

- (iii) The **quadratic covariation** of two continuous semimartingales X and Y is defined as follows:

$$[X, Y](t) = \frac{1}{4}([X + Y](t) - [X - Y](t)).$$

- (iv) Let X and Y be Itô processes with volatilities $\sigma_X(t)$ and $\sigma_Y(t)$ with respect to the same Brownian motion, then

$$[X, Y](t) = \int_0^t \sigma_X(s) \sigma_Y(s) ds.$$

- (v) Quadratic covariation is symmetric and bilinear map:

$$[X + Y, Z](t) = [X, Z](t) + [Y, Z](t),$$

$$[X, Y + Z](t) = [X, Y](t) + [X, Z](t),$$

$$[X, Y](t) = [Y, X](t), \quad [X](t) = [X, X](t).$$

Example 1.5.

- (i) Let us first consider two-dimensional case with $\boldsymbol{\mu}(t) = (\mu_1(t), \mu_2(t))^T$ and

$$\Theta(t) = \begin{pmatrix} \sigma_{11}(t) & \sigma_{12}(t) \\ \sigma_{21}(t) & \sigma_{22}(t) \end{pmatrix}$$

Then $\mathbf{X}(t) = (X(t), Y(t))^T$, where

$$X(t) = \int_0^t \mu_1(s)ds + \int_0^t \sigma_{11}(s)dW_1(s) + \int_0^t \sigma_{12}(s)dW_2(s)$$

and

$$Y(t) = \int_0^t \mu_2(s)ds + \int_0^t \sigma_{21}(s)dW_1(s) + \int_0^t \sigma_{22}(s)dW_2(s).$$

Using Remark 1.4 we get

$$[X](t) = \int_0^t (\sigma_{11}^2(s) + \sigma_{12}^2(s))ds, \quad [Y](t) = \int_0^t (\sigma_{21}^2(s) + \sigma_{22}^2(s))ds \quad (3)$$

and

$$[X, Y](t) = \int_0^t (\sigma_{11}(s)\sigma_{21}(s) + \sigma_{12}(s)\sigma_{22}(s))ds. \quad (4)$$

Let us look at the quadratic covariation $[X, Y]$ at time 1. It is defined as follows: let $\Pi = \{\tau_0, \dots, \tau_n\}$ be a partition of $[0, 1]$ (i.e. $0 = \tau_0 < \tau_1 < \dots < \tau_n = 1$) and set up the sampled cross variation

$$\sum_{\ell=1}^n \Delta X_\ell \Delta Y_\ell = \sum_{\ell=1}^n (X(\tau_\ell) - X(\tau_{\ell-1})) \cdot (Y(\tau_\ell) - Y(\tau_{\ell-1})). \quad (5)$$

Now let the number of partition points n go to infinity as the length of the longest subinterval $\|\Pi\| = \max_{1 \leq \ell \leq n} (\tau_\ell - \tau_{\ell-1})$ goes to zero. The sum in (5) converges in probability to $[X, Y](1)$. This limit is given by the right-hand side of (4). This assertion follows from the standard theorems in the theory of stochastic calculus. Writing formally,

$$\lim_{n \rightarrow \infty} \sum_{\ell=1}^n \Delta X_\ell \Delta Y_\ell \xrightarrow{\mathbb{P}} [X, Y](1).$$

Hence,

$$\lim_{n \rightarrow \infty} \sum_{\ell=1}^n \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T = \lim_{n \rightarrow \infty} \sum_{\ell=1}^n \begin{pmatrix} (\Delta X_\ell)^2 & \Delta X_\ell \Delta Y_\ell \\ \Delta Y_\ell \Delta X_\ell & (\Delta Y_\ell)^2 \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} [X](1) & [X, Y](1) \\ [Y, X](1) & [Y](1) \end{pmatrix}.$$

On the other hand,

$$\Theta(t) \cdot \Theta^T(t) = \begin{pmatrix} \sigma_{11}^2(t) + \sigma_{12}^2(t) & \sigma_{11}(t)\sigma_{21}(t) + \sigma_{12}(t)\sigma_{22}(t) \\ \sigma_{11}(t)\sigma_{21}(t) + \sigma_{12}(t)\sigma_{22}(t) & \sigma_{21}^2(t) + \sigma_{22}^2(t) \end{pmatrix}.$$

Taking the integral of $\Theta(t) \cdot \Theta^T(t)$ we conclude that

$$\lim_{n \rightarrow \infty} \Sigma_2^{RCV} \xrightarrow{\mathbb{P}} \Sigma_2.$$

- (ii) The convergence of realized covariance to ICV can be generalized for arbitrary p . Let's set $\boldsymbol{\mu}(t) = (\mu_1(t), \dots, \mu_p(t))^T$ and

$$\Theta(t) = \begin{pmatrix} \Theta_{11}(t) & \dots & \Theta_{1p}(t) \\ \vdots & \ddots & \vdots \\ \Theta_{p1}(t) & \dots & \Theta_{pp}(t) \end{pmatrix}.$$

Then we have $\mathbf{X}(t) = (X_1(t), \dots, X_p(t))^T$ with

$$X_j(t) = \int_0^t \mu_j(s) ds + \sum_{i=1}^p \int_0^t \Theta_{ji}(s) dW_i(s)$$

for all $1 \leq j \leq p$. By Remark 1.4 again

$$[X_j](t) = \int_0^t \left(\sum_{i=1}^p \Theta_{ji}^2(s) \right) ds$$

and

$$[X_j, X_k](t) = \int_0^t \left(\sum_{i=1}^p \Theta_{ji}(s) \cdot \Theta_{ki}(s) \right) ds$$

for all $1 \leq j, k \leq p$. We also have

$$\left(\Theta(t) \cdot \Theta^T(t) \right)_{jk} = \sum_{i=1}^p \Theta_{ji}(t) \cdot \Theta_{ki}(t).$$

Therefore, we get

$$\lim_{n \rightarrow \infty} \sum_{\ell=1}^n \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T \xrightarrow{\mathbb{P}} [\mathbf{X}, \mathbf{X}](1),$$

where

$$([\mathbf{X}, \mathbf{X}](t))_{ij} = [X_i, X_j](t),$$

and finally

$$\lim_{n \rightarrow \infty} \Sigma_p^{RCV} \xrightarrow{\mathbb{P}} \Sigma_p.$$

Remark 1.6. Let ICV matrix Σ_p be deterministic. RCV matrix Σ_p^{RCV} , on the other side, is random by definition. In Example 1.5 we used quadratic covariation as an intermediate step to prove the convergence of one matrix to another. Alternatively to Remark 1.4 (iii) quadratic covariation can be defined as a limit of the sum of squared differences, so it is also random. Having almost sure degenerate distribution of quadratic covariation is one of the major properties of Itô processes and it doesn't need to hold for all martingales in general.

Part 2

Marchenko-Pastur law

Empirical spectral distribution

Remark 2.1. In Remark 1.5 we claimed the convergence of RCV matrix to ICV for any finite p . In practice, however, it might be the case that the number of processes is approximately equal to the number of observations for each process. Theoretically, the problem arises when the dimension of p has at least the same rate of growth as n . In this case the convergence is not well defined, because the RCV matrix has the unlimited size. Then, it is more convenient to study another criteria.

Definition 2.2. Let $\{\lambda_j : j = 1, \dots, p\}$ be set of eigenvalues of matrix A , then

$$F^A(x) := \frac{\#\{j : \lambda_j \leq x\}}{p}, \quad x \in \mathbb{R},$$

is called *empirical spectral distribution (ESD)* of A .

Remark 2.3. (COPIED FROM THE PAPER) A naive estimator of the spectrum of the ICV matrix Σ_p is the spectrum of the RCV matrix Σ_p^{RCV} . In particular, one wished that the ESD $F^{\Sigma_p^{RCV}}$ of Σ_p^{RCV} would approximate F^{Σ_p} well when the value of n sufficiently high. From the large dimensional random matrix theory (LDRMT), we now understand quite well that in the high dimensional setting this good wish won't come true. For example, in the simplest case when the drift process is 0, covolatility process is constant, and observation times τ_ℓ are equally spaced, namely $\tau_\ell = \ell/n$, we are in the setting of estimating the usual covariance matrix using the sample covariance matrix, given n i.i.d. p -dimensional observations $(\Delta \mathbf{X}_\ell)_{\ell=1, \dots, n}$. From LDRMT, we know that if p/n converges to a non-zero number and the ESD $F^{\Sigma_p^{RCV}}$ of the sample covariance matrix also converges (SEE BOOKS). The relationship between the limiting spectral distribution (LSD) of Σ_p^{RCV} in this case and the LSD of Σ_p can be described by a Marchenko-Pastur equation through Stieltjes transforms, as follows.

Definition 2.4. In mathematics, the *Stieltjes transformation* $m_\mu(z)$ of a measure μ on \mathbb{R} is the function of the complex variable $z \in \mathbb{C}_+ := \{z \in \mathbb{C} : \Im(z) > 0\}$ defined by the formula

$$m_\mu(z) = \int_{t \in \mathbb{R}} \frac{1}{t - z} d\mu(t).$$

Proposition 2.5.

- (i) for $p = 1, 2, \dots$ and for $1 \leq \ell \leq n$, $\mathbf{Z}_\ell^{(p)} = (Z_{j\ell}^{(p)})_{1 \leq j \leq p}$ is a p -dimensional vector with $Z_{j\ell}^{(p)}$ i.i.d. with mean 0 and variance 1;
- (ii) $n = n(p)$ with $y_n := p/n \rightarrow y > 0$ as $p \rightarrow \infty$;

- (iii) Σ_p is a (possibly random) nonnegative definite $p \times p$ matrix such that its ESD F^{Σ_p} converges a.s. in distribution to a probability distribution H on $[0, \infty)$ as $p \rightarrow \infty$;
- (iv) Σ_p and $\mathbf{Z}_\ell^{(p)}$ are independent.

Let $\Sigma_p^{1/2}$ be the (nonnegative) square root matrix of Σ_p and

$$S_p := \frac{1}{n} \sum_{\ell=1}^n \Sigma_p^{1/2} \mathbf{Z}_\ell^{(p)} (\mathbf{Z}_\ell^{(p)})^T \Sigma_p^{1/2}.$$

Then a.s. the ESD of S_p converges in distribution to a probability distribution F , which is determined by H in that its Stieltjes transform

$$m_F(z) := \int_{\lambda \in \mathbb{R}} \frac{1}{\lambda - z} dF(\lambda), \quad z \in \mathbb{C}_+$$

solves the equation

$$m_F(z) = \int_{\tau \in \mathbb{R}} \frac{1}{\tau(1 - y(1 + zm_F(z))) - z} dH(\tau). \quad (6)$$

Proof. See Theorem 1.1 of Silverstein (1995). \square

Remark 2.6.

- (i) Note that if $y \rightarrow 0$ limiting distribution function F of S_p matches limiting distribution function H of Σ_p . This is the case, when n grows faster than p and as we stated before Σ_p^{RCV} converges to Σ_p .
- (ii) In the special case when $\Sigma_p = \sigma^2 \mathbb{I}_{p \times p}$, where $\mathbb{I}_{p \times p}$ is the $p \times p$ identity matrix, the LSD F has an analytical expression, which can be derived from the following proposition.

Proposition 2.7. Suppose that $\mathbf{Z}_\ell^{(p)}$'s are as in the previous proposition, and $\Sigma_p = \sigma^2 \mathbb{I}_{p \times p}$ for some $\sigma^2 > 0$. Then the LSD F has density

$$f(x) = \left(1 - \frac{1}{y}\right)_+ \delta_0(x) + \frac{1}{2\pi\sigma^2 xy} \sqrt{(b-x)(x-a)} \mathbf{1}_{[a,b]}(x),$$

where

$$a = \sigma^2(1 - \sqrt{y})^2 \quad \text{and} \quad b = \sigma^2(1 + \sqrt{y})^2,$$

and $\delta_0(x)$ is a Dirac delta function. The LSD F in this proposition is called the Marchenko-Pastur law with ratio index y and scale index σ^2 , and will be denoted by $\mathcal{MP}(y, \sigma^2)$.

Proof. See, e.g., Theorem 2.5 in Bai (1999). \square

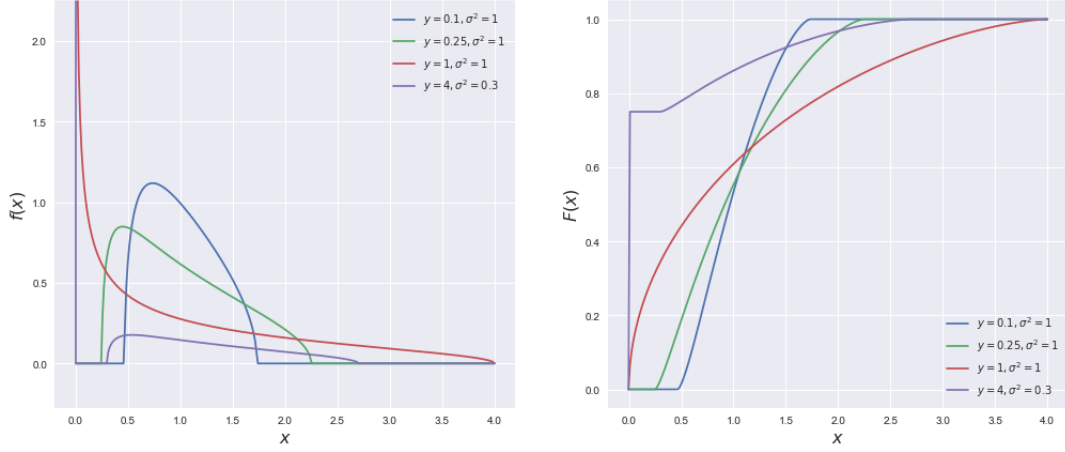


Figure 2: Marchenko-Pastur law

Non-constant volatility case

Remark 2.8. Let us set

$$\Theta^0(t) = \sqrt{\int_0^1 \Theta(s) \cdot \Theta^T(s) ds} \quad \forall t \in [0, 1]$$

and corresponding matrix $\mathbf{X}^0(t)$, such that

$$d\mathbf{X}^0(t) = \Theta^0(t) d\mathbf{W}(t).$$

Note that $\mathbf{X}(t)$ and $\mathbf{X}^0(t)$ share the same ICV matrix:

$$\Sigma_p^0 := \int_0^1 \Theta^0(t) \cdot \Theta^{0T}(t) dt = \int_0^1 dt \int_0^1 (\Theta(s) \cdot \Theta^T(s)) ds = \Sigma_p.$$

Based on $\mathbf{X}^0(t)$, we have an associated RCV matrix

$$\Sigma_p^{RCV^0} = \sum_{\ell=1}^n \Delta \mathbf{X}_\ell^0 (\Delta \mathbf{X}_\ell^0)^T.$$

Since Σ_p^{RCV} and $\Sigma_p^{RCV^0}$ are based on the same estimation method and share the same targeting ICV matrix, it is desirable that their ESDs have similar properties.

... bla bla bla discuss convergence and conclude that we can show that RCV may not converge in non-constant volatility case.

Definition 2.9. Let A, B be two symmetric $p \times p$ matrices. Then we write

$$A \succeq B$$

to denote that $A - B$ is a positive semidefinite matrix. Similarly, we say that $A \preceq B$ if $A - B$ is negative semidefinite. The partial ordering defined by \succeq is called Loewner ordering. (REPLACE WITH DIACRITIC SIGN HERE)

Lemma 2.10 (Weyl's Monotonicity Theorem). *Suppose A and B are symmetric, $p \times p$ matrices. Let $\lambda_i(A)$ be the i -th largest eigenvalue of A . If $A \preceq B$, then $\lambda_i(A) \leq \lambda_i(B)$ for all i , or, equivalently*

$$F^B(x) \leq F^A(x) \quad \forall x \geq 0.$$

Proof. Corollary 4.3.3 in Horn and Johnson (1990). \square

Proposition 2.11. Suppose that for any p , $\mathbf{X}(t)$ is a p -dimensional process satisfying

$$d\mathbf{X}(t) = \gamma(t)d\mathbf{W}(t), \quad t \in [0, 1], \quad (7)$$

where $\gamma(t) > 0$ is a nonrandom (scalar) càdlàg process. Let $\sigma^2 = \int_0^1 \gamma^2(t)dt$ and so that the ICV matrix $\Sigma_p = \sigma^2 \mathbb{I}_{p \times p}$. Assume further that the observation times τ_ℓ are equally spaced, that is, $\tau_\ell = \ell/n$, and that the RCV matrix Σ_p^{RCV} is defined by (2). Then so long as $\gamma(t)$ is not constant on $[0, 1]$, for any $\varepsilon > 0$, there exists $y_c = y_c(\gamma, \varepsilon) > 0$ such that if $\lim p/n = y \geq y_c$,

$$\limsup F^{\Sigma_p^{RCV}}(b(y) + \sigma^2 \varepsilon) < 1 \quad \text{a.s.} \quad (8)$$

In particular, $F^{\Sigma_p^{RCV}}$ doesn't converge to the Marchenko-Pastur law $\mathcal{MP}(y, \sigma^2)$.

Proof. By assumption if $\gamma(t)$ is non-constant, there exists $\delta > 0$ and an interval $[c, d] \subseteq [0, 1]$ such that

$$\gamma(t) \geq \sigma(1 + \delta) \quad \forall t \in [c, d].$$

Therefore, if $[\frac{\ell-1}{n}, \frac{\ell}{n}] \subseteq [c, d]$ for some $1 \leq \ell \leq n$, then

$$\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T \stackrel{d}{=} \int_{(\ell-1)/n}^{\ell/n} \gamma^2(t) dt \cdot \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T \succeq \frac{(1 + \delta)^2}{n} \sigma^2 \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T,$$

where $\mathbf{Z}_\ell = (Z_{1\ell}, \dots, Z_{p\ell})^T$ consists of independent standard normals. Hence, if we let $J_n = \{\ell : [\frac{\ell-1}{n}, \frac{\ell}{n}] \subseteq [c, d]\}$ and

$$\Gamma_p = \sum_{\ell \in J_n} \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T, \quad \Lambda_p = \frac{\sigma^2}{n(d - c)} \sum_{\ell \in J_n} \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T,$$

then for any $x \geq 0$, by Weyl's Monotonicity Theorem,

$$F^{\Sigma_p^{RCV}}(x) \leq F^{\Gamma_p}(x) \leq F^{\Lambda_p}\left(\frac{x}{(1 + \delta)^2(d - c)}\right).$$

Now note that $\#J_n \sim (d-c)n$, hence if $p/n \rightarrow y$, by Proposition 2.5, F^{Λ_p} will converge a.s. to the Marchenko-Pastur law with ratio index $y' = \frac{y}{d-c}$ and scale index σ^2 . By the formula of $b(\cdot)$ in Marchenko-Pastur density

$$\begin{aligned} (1+\delta)^2(d-c)b(y') &= (1+\delta)\sigma^2 \cdot (1+\delta)(d-c)(1+2\sqrt{y'}+y') \\ &= (1+\delta)\sigma^2 \cdot (1+\delta)(d-c+2\sqrt{(d-c)y}+y) \\ &:= (1+\delta)\sigma^2 \cdot g(y). \end{aligned}$$

Note that the $g(y)$ has a linear growth in y with coefficient $1+\delta$. Hence, for any $\varepsilon > 0$, there exists $y_c > 0$, such that for all $y \geq y_c$

$$g(y) \geq (1+\sqrt{y})^2 + \varepsilon,$$

that is,

$$(1+\delta)^2(d-c)b(y') \geq (1+\delta)\sigma^2 \cdot ((1+\sqrt{y})^2 + \varepsilon) = (1+\delta)(b(y) + \sigma^2\varepsilon)$$

or, equivalently,

$$\frac{b(y) + \sigma^2\varepsilon}{(1+\delta)^2(d-c)} \leq \frac{b(y')}{1+\delta}.$$

Therefore, when the above inequality holds,

$$\limsup F^{\Sigma_p^{RCV}}(b(y) + \sigma^2\varepsilon) \leq \limsup F^{\Lambda_p}\left(\frac{b(y')}{1+\delta}\right) < 1.$$

□

Part 3

Limit theorems for non-constant covolatility processes

Limiting distribution for realized covariance

Remark 3.1. In this part we will state two major theorems. The first one is a generalization of Proposition 2.5 for limiting spectral distribution of RCV matrix in the case of non-constant covolatility. The second theorem claims the convergence of ESD of alternative estimator to Marchenko-Pastur law. Both theorems require some additional assumptions.

Theorem 3.2. *Assume that all the conditions in Proposition 2.5 are satisfied. Furthermore,*

- (i) $Z_{j\ell}^{(p)}$ have finite moments of all orders;
- (ii) H has a finite second moment;
- (iii) the weights w_ℓ^n , $1 \leq \ell \leq n$, $n = 1, 2, \dots$, are all positive, and there exists $\kappa < \infty$ such that the rescaled weights (nw_ℓ^n) satisfy

$$\max_n \max_{\ell=1, \dots, n} (nw_\ell^n) \leq \kappa;$$

moreover, almost surely, there exists a càdlàg function $w(s) : [0, 1] \rightarrow \mathbb{R}_+$, such that

$$\lim_n \sum_{1 \leq \ell \leq n} \int_{(\ell-1)/n}^{\ell/n} |nw_\ell^n - w(s)| ds = 0;$$

- (iv) there exists a sequence $\eta_p = o(p)$ and a sequence of index sets \mathcal{I}_p satisfying $\mathcal{I}_p \subset \{1, \dots, p\}$ and $\#\mathcal{I}_p \leq \eta_p$ such that for all n and all ℓ , w_ℓ^n may depend on $\mathbf{Z}_\ell^{(p)}$ but only on $\{Z_{j\ell}^{(p)} : j \in \mathcal{I}_p\}$;
- (v) there exist $C < \infty$ and $\delta < 1/6$ such that for all p , $\|\Sigma_p\| \leq Cp^\delta$ a.s.

Define $S_p = \sum_{\ell=1}^n w_\ell^n \cdot \Sigma_p^{1/2} \mathbf{Z}_\ell^{(p)} (\mathbf{Z}_\ell^{(p)})^T \Sigma_p^{1/2}$. Then, almost surely, the ESD of S_p converges in distribution to a probability distribution F^w , which is determined by H and $(w(s))$ in that its Stiltjes transform $m_{F^w}(z)$ is given by

$$m_{F^w}(z) = -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{1}{\tau M(z) + 1} dH(\tau),$$

where $M(z)$, together with another function $\tilde{m}(z)$, uniquely solve the following equation in $\mathbb{C}_+ \times \mathbb{C}_+$:

$$\begin{cases} M(z) = -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{w_s}{1 + y\tilde{m}(z)w_s} ds, \\ \tilde{m}(z) = -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{\tau}{\tau M(z) + 1} dH(\tau). \end{cases}$$

Remark 3.3. If $w_\ell^n = 1/n$, then $w(s) = 1$, and Theorem 3.2 reduces to Proposition 2.5. Moreover, if $w(s)$ is not constant, that is, $w(s) \neq \int_0^1 w(t)dt$ on $[0, 1]$, then except in the trivial case when H is a delta measure at 0, the LSD $F^w \neq F$, where F is the LSD in Proposition 2.5 determined by $H(\cdot/\int_0^1 w(t)dt)$. (ADD MORE EXPLANATION FROM SUPPLEMENTARY ARTICLE)

Time-variance adjusted realized covariance

Definition 3.4. Suppose that for all p $\mathbf{X}(t) = \mathbf{X}^{(p)}(t)$ is a p -dimensional process satisfying (1), and $\Theta(t)$ is càdlàg. We say that $\mathbf{X}(t)$ belongs to *class* \mathcal{C} if, almost surely, there exist $\gamma(t) : [0, 1] \mapsto \mathbb{R}$ and Λ a $p \times p$ matrix satisfying $\text{tr}(\Lambda\Lambda^T) = p$ such that

$$\Theta(t) = \gamma(t)\Lambda. \quad (9)$$

Observe that if (9) holds, then the ICV matrix $\Sigma_p = \int_0^1 \gamma^2(t)dt \cdot \Lambda\Lambda^T$. The special case when $\Lambda = \mathbb{I}_{p \times p}$ is studied in the next part.

Example 3.5. Suppose that $X_j(t)$ satisfy

$$dX_j(t) = \mu_j(t)dt + \sigma_j(t)dW_j(t), \quad j = 1, \dots, p,$$

where $\mu_j(t), \sigma_j(t) : [0, 1] \rightarrow \mathbb{R}$ are the drift and volatility processes for stock j , and $W_j(t)$'s are (one-dimensional) standard Brownian motions. If the following conditions hold:

- the correlation matrix process of $(W_j(t))$

$$R(t) := \left(\frac{[W_j, W_k](t)}{t} \right)_{1 \leq j, k \leq p} =: (r_{jk})_{1 \leq j, k \leq p}$$

is constant in $t \in [0, 1]$.

- $r^{(jk)} \neq 0$ for all $1 \leq j, k \leq p$; and
- the correlation matrix process of $X_j(t)$

$$\left(\frac{\int_0^t \sigma_j(s)\sigma_k(s)d[W_j, W_k](s)}{\sqrt{\int_0^t \sigma_j^2(s)ds \cdot \int_0^t \sigma_k^2(s)ds}} \right)_{1 \leq j, k \leq p} =: (\rho_{jk})_{1 \leq j, k \leq p}$$

is constant in $t \in [0, 1]$;

then $\mathbf{X}(t)$ belongs to class \mathcal{C} .

Proof. For any $t \in [0, 1]$:

$$\rho_{jk} = \frac{r_{jk} \int_0^t \sigma_j(s)\sigma_k(s)ds}{\sqrt{\int_0^t \sigma_j^2(s)ds \cdot \int_0^t \sigma_k^2(s)ds}},$$

therefore

$$\frac{\rho_{jk}}{r_{jk}} = \frac{\int_0^t \sigma_j(s) \sigma_k(s) ds / t}{\sqrt{\int_0^t \sigma_j^2(s) ds / t \cdot \int_0^t \sigma_k^2(s) ds / t}}.$$

Letting $t \downarrow 0$, using l'Hôpital's rule and noting that $\sigma_j(t)$ are càdlàg, we observe that

$$\frac{\rho_{jk}}{r_{jk}} = \frac{\sigma_j(0) \sigma_k(0)}{\sqrt{\sigma_j^2(0) \cdot \sigma_k^2(0)}} = \pm 1.$$

Hence, for all $t \in [0, 1]$:

$$\left| \int_0^t \sigma_j(s) \sigma_k(s) ds \right| = \sqrt{\int_0^t \sigma_j^2(s) ds \cdot \int_0^t \sigma_k^2(s) ds}.$$

By Cauchy-Schwartz inequality, this holds only if $\sigma_j(s)$ and $\sigma_k(s)$ are proportional to each other. Therefore, almost surely, there exists a scalar process $\gamma(t) : [0, 1] \rightarrow \mathbb{R}$ and a p -dimensional vector $(\sigma_1, \dots, \sigma_p)^T$, such that

$$(\sigma_1(t), \dots, \sigma_p(t))^T = \gamma(t) \cdot (\sigma_1, \dots, \sigma_p)^T.$$

Now we show that $\mathbf{X}(t)$ belongs to class \mathcal{C} . In fact one can always find a p -dimensional standard Brownian motion $\widetilde{\mathbf{W}}(t)$, such that

$$\mathbf{W}(t) = R^{1/2} \widetilde{\mathbf{W}}(t),$$

where $\mathbf{W} = (W_1(t), \dots, W_p(t))^T$ and R is a correlation matrix of $\mathbf{W}(t)$, which is constant for all $t \in [0, 1]$ by assumption. Hence, writing $\mu(t) = (\mu_1(t), \dots, \mu_p(t))^T$, we have

$$\begin{aligned} d\mathbf{X}(t) &= \mu(t)dt + \text{diag}(\sigma_1(t), \dots, \sigma_p(t))d\mathbf{W}(t) \\ &= \mu(t)dt + \gamma(t) \cdot \text{diag}(\sigma_1, \dots, \sigma_p) R^{1/2} d\widetilde{\mathbf{W}}(t). \end{aligned}$$

□

Remark 3.6. The process in Example 1.2 doesn't belong to the class \mathcal{C} , because its correlation structure varies in time.

Definition 3.7. Suppose that a diffusion process $\mathbf{X}(t)$ belongs to class \mathcal{C} . We define *time-variation adjusted realized covariance (TVARCV) matrix* as follows:

$$\widehat{\Sigma}_p := \frac{\text{tr}(\Sigma_p^{RCV})}{n} \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T}{\|\Delta \mathbf{X}_\ell\|_2^2} = \frac{\text{tr}(\Sigma_p^{RCV})}{p} \widetilde{\Sigma}_p, \quad (10)$$

where

$$\widetilde{\Sigma}_p := \frac{p}{n} \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T}{\|\Delta \mathbf{X}_\ell\|_2^2}. \quad (11)$$

Remark 3.8. Let us explain $\tilde{\Sigma}_p$. Consider the simplest case when $\mu(t) = 0$, $\gamma(t)$ is deterministic, $\Lambda_t = \mathbb{I}_{p \times p}$, and $\tau_\ell = \ell/n$, $\ell = 0, 1, \dots, n$. In this case,

$$\Delta \mathbf{X}_\ell = \sqrt{\int_{(\ell-1)/n}^{\ell/n} \gamma^2(t) dt} \cdot \frac{\mathbf{Z}_\ell}{\sqrt{n}},$$

where $\mathbf{Z}_\ell = (Z_{1\ell}, \dots, Z_{p\ell})^T$ is a vector of i.i.d. standard normal random variables. Hence,

$$\frac{\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T}{\|\Delta \mathbf{X}_\ell\|_2^2} = \frac{\mathbf{Z}_\ell \mathbf{Z}_\ell^T}{\|\mathbf{Z}_\ell\|_2^2}.$$

However, as $p \rightarrow \infty$, $\|\mathbf{Z}_\ell\|_2^2 \sim p$, hence

$$\tilde{\Sigma}_p \sim \frac{\sum_{\ell=1}^n \mathbf{Z}_\ell \mathbf{Z}_\ell^T}{n},$$

the latter being the usual sample covariance matrix. We will show that, first, $\text{tr}(\Sigma_p^{RCV}) \sim \text{tr}(\Sigma_p)$; and second, if $\mathbf{X}(t)$ belongs to class \mathcal{C} and satisfies certain additional assumptions, then the LSD of $\tilde{\Sigma}_p$ is related to that of $\check{\Sigma}_p$ via the Marchenko-Pastur equation, where

$$\check{\Sigma}_p = \frac{p}{\text{tr}(\Sigma_p)} \Sigma_p = \Lambda \Lambda^T.$$

Hence, the LSD of $\hat{\Sigma}_p$ is also related to that of Σ_p via the same Marchenko-Pastur equation. We will now state our assumptions for the major theorem in this paper.

Assumptions 3.9.

- (i) there exists $C_0 < \infty$ such that for all p and all $j = 1, \dots, p$, $|\mu_j^{(p)}(t)| \leq C_0$ for all $t \in [0, 1]$ a.s.;
- (ii) there exist constants $C_1 < \infty$, $0 \leq \delta_1 < 1/2$, a sequence $\eta_p < C_1 p^{\delta_1}$ and a sequence of index sets \mathcal{I}_p satisfying $\mathcal{I}_p \subset \{1, \dots, p\}$ and $\#\mathcal{I}_p \leq \eta_p$ such that $\gamma^{(p)}(t)$ may depend on $\mathbf{W}^{(p)}(t)$ but only on $\{W_j^{(p)}(t) : j \in \mathcal{I}_p\}$; moreover, there exists $C_2 < \infty$ such that for all p , $|\gamma^{(p)}(t)| \in (1/C_2, C_2)$ for all $t \in [0, 1]$ a.s.;
- (iii) there exists $C_3 < \infty$ such that for all p and for all j , the individual volatilities

$$\sigma_j(t) = \gamma^{(p)}(t) \sqrt{\sum_{k=1}^p (\Lambda_{jk}^{(p)})^2} \in (1/C_3, C_3)$$

for all $t \in [0, 1]$ a.s.;

- (iv)

$$\lim_{p \rightarrow \infty} \frac{\text{tr}(\Sigma_p)}{p} = \lim_{p \rightarrow \infty} \int_0^1 (\gamma^{(p)}(t))^2 dt := \theta > 0 \quad \text{a.s.};$$

- (v) almost surely, as $p \rightarrow \infty$, the ESD F^{Σ_p} converges to a probability distribution H on $[0, \infty)$;
- (vi) there exist $C_5 < \infty$ and $0 \leq \delta_2 < 1/2$ such that for all p , $\|\Sigma_p\| \leq C_5 p^{\delta_2}$ a.s.;
- (vii) the δ_1 in (ii) and δ_2 in (vi) satisfy that $\delta_1 + \delta_2 < 1/2$;
- (viii) $p/n \rightarrow y \in (0, \infty)$ as $p \rightarrow \infty$; and
- (ix) there exists $C_4 < \infty$ such that for all n ,

$$\max_{1 \leq \ell \leq n} n \cdot (\tau_\ell - \tau_{\ell-1}) \leq C_4 \quad \text{a.s.}$$

moreover, τ_ℓ 's are independent of $\mathbf{X}(t)$.

Remark 3.10. Before stating and proving theorem for TVARCV we provide some useful lemmas.

Lemma 3.11 (Burkholder-Davis-Gundy inequality). *For any real number $p \geq 1$ there exist positive constants c_p and C_p such that, for all local martingales $X(t)$ with $X(0) = 0$, its maximum processes $M(t) = \sup_{t \leq s} |X(s)|$ and stopping times τ , the following inequalities hold*

$$c_p \mathbb{E}[|X|^{p/2}(\tau)] \leq \mathbb{E}[M^p(\tau)] \leq C_p \mathbb{E}[|X|^{p/2}(\tau)].$$

Proof. LOOK IN THE BOOKS □

Lemma 3.12. *Suppose that for each p , $\mathbf{v}_\ell^{(p)} = (v_{1\ell}^{(p)}, \dots, v_{p\ell}^{(p)})^T$ and $\mathbf{w}_\ell^{(p)} = (w_{1\ell}^{(p)}, \dots, w_{p\ell}^{(p)})^T$, $\ell = 1, \dots, n$, are all p -dimensional vectors. Define*

$$\tilde{S}_n = \sum_{\ell=1}^n (\mathbf{v}_\ell^{(p)} + \mathbf{w}_\ell^{(p)}) \cdot (\mathbf{v}_\ell^{(p)} + \mathbf{w}_\ell^{(p)})^T$$

and

$$S_n = \sum_{\ell=1}^n \mathbf{w}_\ell^{(p)} (\mathbf{w}_\ell^{(p)})^T.$$

If the following conditions are satisfied:

- (i) $\lim_{p \rightarrow \infty} p/n = y > 0$;
- (ii) there exists a sequence $\varepsilon_p = o(1/\sqrt{p})$ such that for all p and all ℓ , all the entries of $\mathbf{v}_\ell^{(p)}$ are bounded by ε_p in absolute value;
- (iii) $\limsup_{p \rightarrow \infty} \text{tr}(S_n)/p < \infty$ almost surely.

Then $L(F^{\tilde{S}_n}, F^{S_n}) \rightarrow 0$ almost surely, where for any two probability distribution functions F and G

$$L(F, G) := \inf\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \quad \forall x \in \mathbb{R}\}$$

denotes the Lévy distance between them.

Proof. □

Proposition 3.13. Under Assumptions 3.9(iv), namely, suppose that

$$\lim_{p \rightarrow \infty} \frac{\text{tr}(\Sigma_p)}{p} = \theta \quad \text{a.s.},$$

then

$$\lim_{p \rightarrow \infty} \frac{\text{tr}(\Sigma_p^{RCV})}{p} = \theta \quad \text{a.s.}$$

Proof. The full proof can be found in supplementary article. We will provide here the main idea: □

Theorem 3.14. Suppose that for all p , $\mathbf{X}(t) = \mathbf{X}^{(p)}(t)$ is a p -dimensional process in class \mathcal{C} for some drift process $\boldsymbol{\mu}^{(p)}(t) = (\mu_1^{(p)}(t), \dots, \mu_p^{(p)}(t))^T$, covolatility process $\Theta^{(p)}(t) = \gamma^{(p)}(t)\Lambda^{(p)}$ and a p -dimensional Brownian motion $\mathbf{W}_t^{(p)}$, which satisfy Assumptions 3.9 (i)-(vii) above. Suppose also that p and n satisfy Assumptions 3.9 (viii), and the observation times satisfy Assumptions 3.9(ix). Let $\tilde{\Sigma}_p$ be as in (10). Then, as $p \rightarrow \infty$, $F^{\tilde{\Sigma}_p}$ converges a.s. to a probability distribution F , which is determined by H through Stieltjes transforms via the same Marchenko-Pastur equation as in Proposition 2.5.

Proof. First we state that both $F^{\tilde{\Sigma}_p}$ and $F^{\check{\Sigma}_p}$ converge almost surely. $F^{\check{\Sigma}_p}$ converges to \check{H} , which is defined by

$$\check{H}(x) = H(\theta x) \quad \forall x \geq 0.$$

The LSD \tilde{F} of $\tilde{\Sigma}_p$ is determined by \check{H} in that its Stieltjes transform $m_{\tilde{F}}(z)$ satisfies the equation

$$m_{\tilde{F}}(z) = \int_{\tau \in \mathbb{R}} \frac{1}{\tau(1 - y(1 + zm_{\tilde{F}}(z))) - z} d\check{H}(\tau).$$

The convergence of $F^{\tilde{\Sigma}_p}$ is obvious since

$$F^{\tilde{\Sigma}_p} = F^{\Sigma_p}(\text{tr}(\Sigma_p)/p \cdot x) \quad \forall x \geq 0.$$

We now show the convergence of F^{Σ_p} . For notational ease, we shall omit the superscript p in the arguments below: thus we write $\boldsymbol{\mu}(t)$ instead of $\boldsymbol{\mu}^{(p)}(t)$, $\gamma(t)$ instead of $\gamma^{(p)}(t)$, Λ instead of $\Lambda^{(p)}$, etc. First, note that

$$\Delta \mathbf{X}_\ell = \int_{\tau_{\ell-1}}^{\tau_\ell} \boldsymbol{\mu}(t) dt + \Lambda \cdot \int_{\tau_{\ell-1}}^{\tau_\ell} \gamma(t) d\mathbf{W}(t) := \sqrt{\int_{\tau_{\ell-1}}^{\tau_\ell} \gamma^2(t) dt} \cdot (\mathbf{v}_\ell + \Lambda \cdot \mathbf{Z}_\ell),$$

where

$$\mathbf{v}_\ell = (v_{1\ell}, \dots, v_{p\ell})^T = \frac{\int_{\tau_{\ell-1}}^{\tau_\ell} \boldsymbol{\mu}(t) dt}{\sqrt{\int_{\tau_{\ell-1}}^{\tau_\ell} \gamma^2(t) dt}}$$

and

$$\mathbf{Z}_\ell = (Z_{1\ell}, \dots, Z_{p\ell})^T = \frac{\int_{\tau_{\ell-1}}^{\tau_\ell} \gamma(t) d\mathbf{W}(t)}{\sqrt{\int_{\tau_{\ell-1}}^{\tau_\ell} \gamma^2(t) dt}}$$

By performing an orthogonal transformation if necessary, w.l.o.g., we may assume that the index set $\mathcal{I}_p \subset \{1, \dots, \eta_p\}$. Then by Assumptions 3.9 (ii) and (ix), for $j > \eta_p$, $Z_{j\ell}$ are i.i.d. $\sim \mathcal{N}(0, 1)$. Write

$$\mathbf{U}_\ell = (Z_{1\ell}, \dots, Z_{\eta_p\ell})^T$$

and

$$\mathbf{D}_\ell = (Z_{\eta_p+1\ell}, \dots, Z_{p\ell})^T.$$

With the above notation, $\tilde{\Sigma}_p$ can be rewritten as

$$\tilde{\Sigma}_p = y_n \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T}{\|\Delta \mathbf{X}_\ell\|_2^2} = y_n \sum_{\ell=1}^n \frac{(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)^T}{\|\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell\|_2^2} \quad (12)$$

By Assumptions 3.9 (i), (ii) and (ix), there exists $C > 0$, such that $|v_{j\ell}| \leq C/\sqrt{n}$ for all j and ℓ , hence $\|\mathbf{v}_\ell\|_2$'s are uniformly bounded. We will show that

$$\max_{\ell=1, \dots, n} \left| \frac{\|\Lambda \mathbf{Z}_\ell\|_2^2}{p} - 1 \right| = \max_{\ell=1, \dots, n} \left| \frac{\mathbf{Z}_\ell^T \check{\Sigma}_p \mathbf{Z}_\ell}{p} - 1 \right| \rightarrow 0 \quad \text{a.s.}, \quad (13)$$

which clearly implies that

$$\max_{\ell=1, \dots, n} \left| \frac{\|\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell\|_2^2}{p} - 1 \right| \rightarrow 0 \quad \text{a.s.} \quad (14)$$

To prove (13) let us write

$$\check{\Sigma}_p = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix},$$

where A , B and C are $\eta_p \times \eta_p$, $\eta_p \times (n - \eta_p)$ and $(n - \eta_p) \times (n - \eta_p)$ matrices, respectively. Then

$$\mathbf{Z}_\ell^T \check{\Sigma}_p \mathbf{Z}_\ell = \mathbf{U}_\ell^T A \mathbf{U}_\ell + 2 \mathbf{D}_\ell^T B^T \mathbf{U}_\ell + \mathbf{D}_\ell^T C \mathbf{D}_\ell.$$

By a well-known fact about the spectral norm,

$$\|A\| \leq \|\check{\Sigma}_p\|, \quad \|B\| \leq \|\check{\Sigma}_p\| \quad \text{and} \quad \|C\| \leq \|\check{\Sigma}_p\|.$$

In particular, by Assumptions 3.9 (ii), (vi) and (vii),

$$0 \leq \text{tr}(A) \leq \eta_p \cdot \|\check{\Sigma}_p\| \leq Cp^{\delta_1 + \delta_2} = o(p),$$

hence

$$\frac{\text{tr}(C)}{p} = \frac{\text{tr}(\check{\Sigma}_p) - \text{tr}(A)}{p} \rightarrow 1.$$

Now using the fact that \mathbf{D}_ℓ consists of i.i.d. standard normals and by the same proof as that for (READ THE PROOF AND WRITE IT HERE!!!) we get

$$\max_{\ell=1,\dots,n} \left| \frac{\mathbf{D}_\ell^T C \mathbf{D}_\ell}{p} - 1 \right| \rightarrow 0 \quad \text{a.s.}$$

To complete the proof of (13), it then suffices to show that

$$\max_{\ell=1,\dots,n} \frac{|\mathbf{U}_\ell^T A \mathbf{U}_\ell|}{p} \rightarrow 0 \quad \text{and} \quad \max_{\ell=1,\dots,n} \frac{|\mathbf{D}_\ell^T B \mathbf{U}_\ell|}{p} \rightarrow 0 \quad \text{a.s.}$$

We shall only prove the first convergence, the second one can be proved similarly. We have

$$|\mathbf{U}_\ell A \mathbf{U}_\ell| \leq \|A\| \cdot \|\mathbf{U}_\ell\|_2^2 \leq C_5 p^{\delta_2} \cdot \|\mathbf{U}_\ell\|_2^2 \quad (15)$$

Observe that for all $1 \leq i \leq \eta_p$, by Assumptions 3.9 (ii),

$$|Z_\ell^{(i)}|^2 = \frac{|\int_{\tau_{\ell-1}}^{\tau_\ell} \gamma_t d\mathbf{W}_t|^2}{\int_{\tau_{\ell-1}}^{\tau_\ell} \gamma_t^2 dt} \leq \frac{C_2^2}{\tau_\ell - \tau_{\ell-1}} \cdot \left| \int_{\tau_{\ell-1}}^{\tau_\ell} \gamma_t d\mathbf{W}_t \right|^2.$$

By Lemma 3.11, we then get that for any $k \in \mathbb{N}$, there exists $\lambda_k > 0$ such that

$$\mathbb{E}[|Z_\ell^{(i)}|^{2k}] \leq \lambda_k C_2^{4k}. \quad (16)$$

Now we are ready to show that $\max_{\ell=1,\dots,n} \frac{|\mathbf{U}_\ell^T A \mathbf{U}_\ell|}{p} \rightarrow 0$. In fact, for any $\varepsilon > 0$, for any $k \in \mathbb{N}$, by Markov's inequality¹, (15), Hölder's inequality and (16),

$$\begin{aligned} \mathbb{P}\left(\max_{\ell=1,\dots,n} |\mathbf{U}_\ell^T A \mathbf{U}_\ell| \geq p\varepsilon\right) &\leq \sum_{\ell=1}^n \mathbb{P}(|\mathbf{U}_\ell^T A \mathbf{U}_\ell| \geq p\varepsilon) \\ &\leq \sum_{\ell=1}^n \frac{\mathbb{E}[|\mathbf{U}_\ell^T A \mathbf{U}_\ell|^k]}{p^k \varepsilon^k} \\ &\leq \sum_{\ell=1}^n \frac{C_5^k p^{k\delta_2} \cdot \eta_p \cdot \lambda_k C_2^{4k} \cdot \eta_p^{k-1}}{p^k \varepsilon^k} \\ &\leq C p^{1+k\delta_2+k\delta_1-k}. \end{aligned}$$

By Assumptions 3.9 (vii), $\delta_1 + \delta_2 < 1/2$, hence by choosing k to be large enough, the right hand side will be summable in p :

$$\sum_{p=1}^{\infty} \mathbb{P}\left(\max_{\ell=1,\dots,n} |\mathbf{U}_\ell^T A \mathbf{U}_\ell| \geq p\varepsilon\right) < \infty.$$

¹Markov's inequality states that for any non-negative random variable X , $k \in \mathbb{N}$ and $n \in \mathbb{R}_+$ the following inequality holds:

$$\mathbb{P}(X \geq n) \leq \frac{\mathbb{E}[X^k]}{n^k}.$$

Hence by Borel–Cantelli lemma², almost surely, $\max_{\ell=1,\dots,n} \frac{|\mathbf{U}_\ell^T A \mathbf{U}_\ell|}{p} \rightarrow 0$. We now get back to $\tilde{\Sigma}_p$. By (14), for any $\varepsilon > 0$, almost surely, for all n sufficiently large, for all $\ell = 1, \dots, n$,

$$p(1 - \varepsilon) \leq \|\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell\|_2^2 \leq p(1 + \varepsilon).$$

Hence, almost surely, for all n sufficiently large,

$$\frac{1}{1 + \varepsilon} \tilde{S}_p \preceq \tilde{\Sigma}_p = y_n \sum_{\ell=1}^n \frac{(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)^T}{\|\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell\|_2^2} \preceq \frac{1}{1 - \varepsilon} \tilde{S}_p,$$

where

$$\tilde{S}_p = \frac{1}{n} \sum_{\ell=1}^n (\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)^T.$$

Hence, by Weyl’s Monotonicity theorem,

$$F^{\tilde{S}_p}((1 + \varepsilon)x) \geq F^{\tilde{\Sigma}_p}(x) \geq F^{\tilde{S}_p}((1 - \varepsilon)x) \quad \forall x \geq 0. \quad (17)$$

Next, by Lemma 3.12 \tilde{S}_p has the same LSD as

$$S_p := \frac{1}{n} \sum_{\ell=1}^n \Lambda \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T \Lambda^T.$$

Moreover, by using the same trick as in the beginning of the proof of THEOREM 1 (WHICH TRICK???), F^{S_p} has the same limit as $F^{S'_p}$, where

$$S'_p := \frac{1}{n} \sum_{\ell=1}^n \Lambda \tilde{\mathbf{Z}}_\ell (\tilde{\mathbf{Z}}_\ell)^T \Lambda^T$$

and $\tilde{\mathbf{Z}}_\ell$ consists of i.i.d. standard normals. For $F^{S'_p}$, it follows easily from Proposition 2.5 that it converges to \tilde{F} . Moreover, by BY SOME THEOREMS IN THIS F** BOOKS, \tilde{F} is differentiable and in particular continuous at all $x > 0$. It follows from (17) that $F^{\tilde{\Sigma}_p}$ must also converge to \tilde{F} . \square

²Let X_1, X_2, \dots be a sequence of random variables in some probability space and $\varepsilon > 0$ be arbitrary. Then

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geq \varepsilon) < \infty \implies X_n \xrightarrow{a.s.} 0.$$

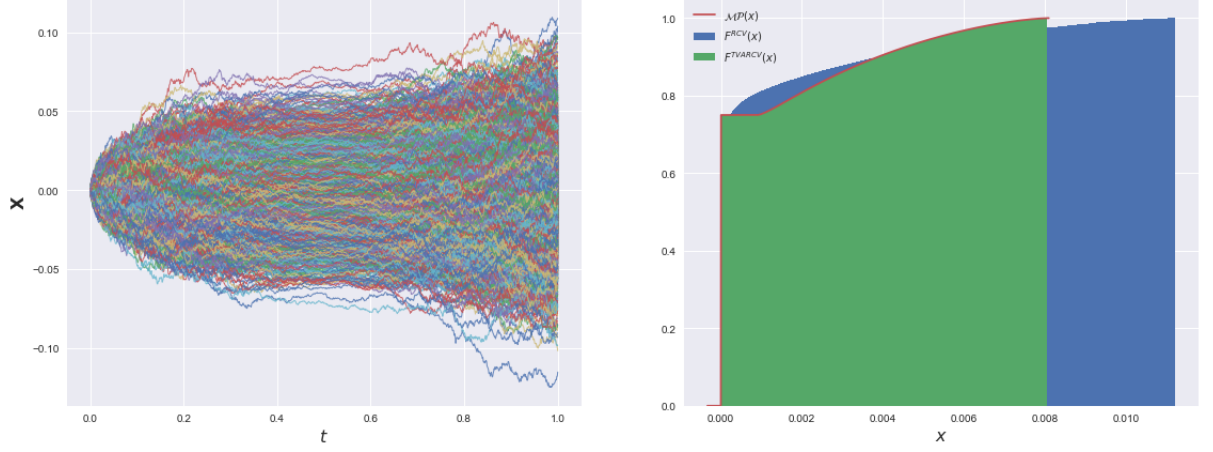


Figure 3: $\mathbf{X}(t)$ from Example 4.1, $n = 1000$, $p = 4000$, $\sigma^2 = 0.0009$

In this setting $y = p/n = 4$. Note point mass equal to 0.75 at the origin. Also note, that the domain of F^{RCV} is larger than the domain of Marchenko-Pastur law, as it was proved in Proposition 2.11.

Part 4

Simulation studies

Non-equidistant observation times

Example 4.1. Let's take an example from the paper...

$$\gamma(t) = \sqrt{0.0009 + 0.0008 \cos(2\pi t)}.$$

Then

$$\sigma^2 = 0.0009 + \int_0^1 0.0008 \cos(2\pi t) dt = 0.0009.$$

Example 4.2. Let us set time points $\{\tau_\ell\}_{0 \leq \ell \leq n}$, where $\tau_0 = 0$ a.s. and interarrival times are independent and obey the exponential distribution with rate λ , that is

$$\mathbb{P}(\tau_\ell - \tau_{\ell-1} > x) = e^{-\lambda x}, \quad 1 \leq \ell \leq n.$$

Then by Strong Law of Large Numbers

$$\frac{\tau_n}{n} = \frac{1}{n} \sum_{\ell=1}^n (\tau_\ell - \tau_{\ell-1}) \xrightarrow{a.s.} \mathbb{E}[\tau_1] = \lambda.$$

Define

$$\mathcal{T} = \{\tau_\ell : \tau_\ell \in [0, 1]\},$$

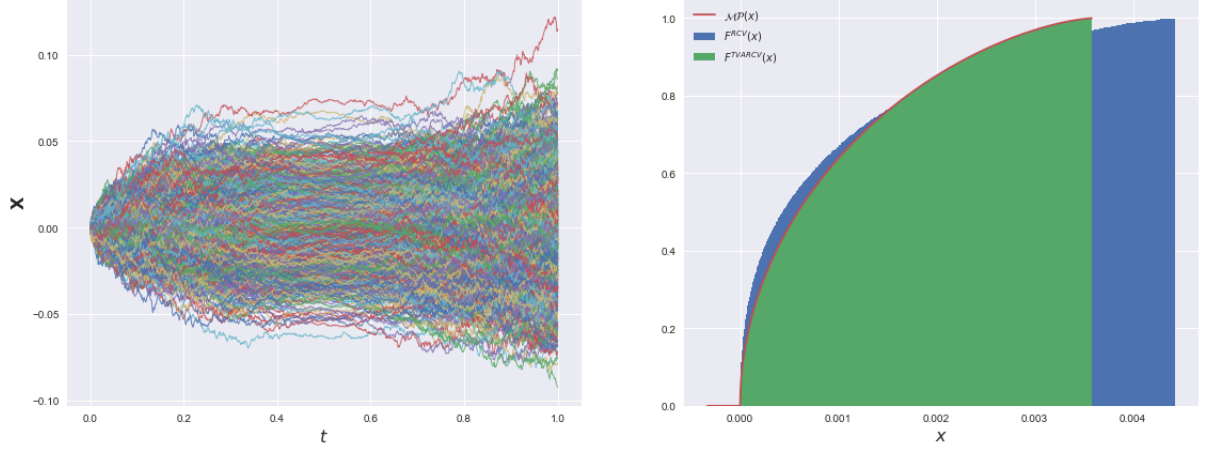


Figure 4: $\mathbf{X}(t)$ from Example 4.1, $n = 1000$, $p = 1000$
In this setting $y = p/n = 1...$

then by setting $\lambda = \frac{1}{n}$, we get $\#\mathcal{T} \sim n$. Let's take volatility γ_t as in Example 4.1. (CONSIDER WHEN TAU DEPENDS ON X, SO THE LAST ASSUMPTION WON'T BE SATISFIED)

Stochastic volatility case

Definition 4.3.

- (i) The stochastic process $Y(t)$ is called **Cox-Ingersoll-Ross (CIR) process**, if it is determined by the stochastic differential equation (SDE):

$$dY(t) = \beta(\alpha - Y(t))dt + \xi\sqrt{Y(t)}d\bar{W}(t), \quad (18)$$

where α , β and ξ are positive constants, and $\bar{W}(t)$ is a standard Brownian motion. Conditional on parameters, time t and initial value $Y(0)$, scaled CIR process has a non-central chi-squared distribution with k degrees of freedom and non-centrality parameter λ :

$$cY(t) \sim \chi_k^2(\lambda),$$

where

$$k = \frac{4\alpha\beta}{\xi^2}, \quad \lambda = Y(0)ce^{-\beta t} \text{ and } c = \frac{4\beta}{\xi^2(1 - e^{-\beta t})}.$$

CIR process never goes below 0. Moreover, it is a.s. positive if the condition

$$2\alpha\beta \geq \xi^2$$

is met. This property makes CIR process a perfect tool for volatility (or interest rate) modeling.

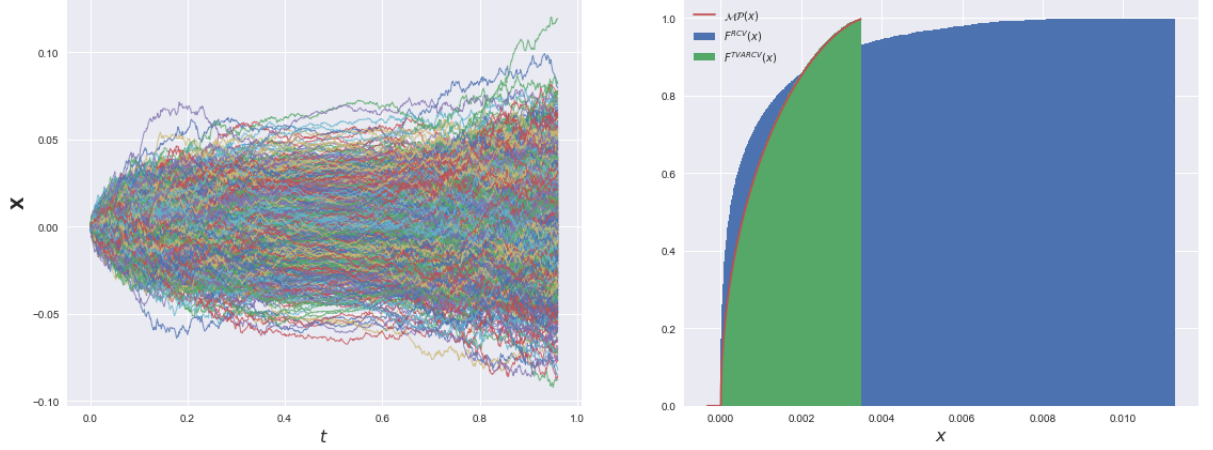


Figure 5: $\mathbf{X}(t)$ from Example 4.2, $n \simeq 1000$, $p = 1000$

We see that if observation times are not equidistant, the ESD of RCV changes drastically, even if visually there is no difference in realizations of $\mathbf{X}(t)$.

- (ii) The standard **Heston model** assumes that one-dimensional log-price process X_t is determined by the following SDE:

$$dX(t) = \mu dt + \sqrt{Y(t)} dW(t), \quad (19)$$

where $Y(t)$ satisfies SDE (18) and $W(t)$ is another standard Brownian motion, such that

$$[W, \bar{W}](t) = \int_0^t \rho(s) ds \quad (20)$$

holds with a (stochastic) process $\rho(t)$, taking values strictly between -1 and 1 . Sometimes equality (20) is informally written as

$$dW(t) \cdot d\bar{W}(t) = \rho(t) dt.$$

Example 4.4. Let's assume

$$\gamma(t) = \sqrt{Y(t)},$$

where Y_t is a CIR process, determined by (18). Then it is easy to see that

$$\sigma^2 = \beta\alpha - \beta \int_0^1 \gamma^2(t) dt + \xi \int_0^1 \gamma(t) d\bar{W}(t).$$

Therefore, as long as $\xi > 0$, the value of σ^2 is random and for different realizations of γ_t we have different LSD. Moreover, let

$$\bar{W}(t) = \frac{1}{\sqrt{\eta_p}} \sum_{j=1}^{\eta_p} W_j(t),$$

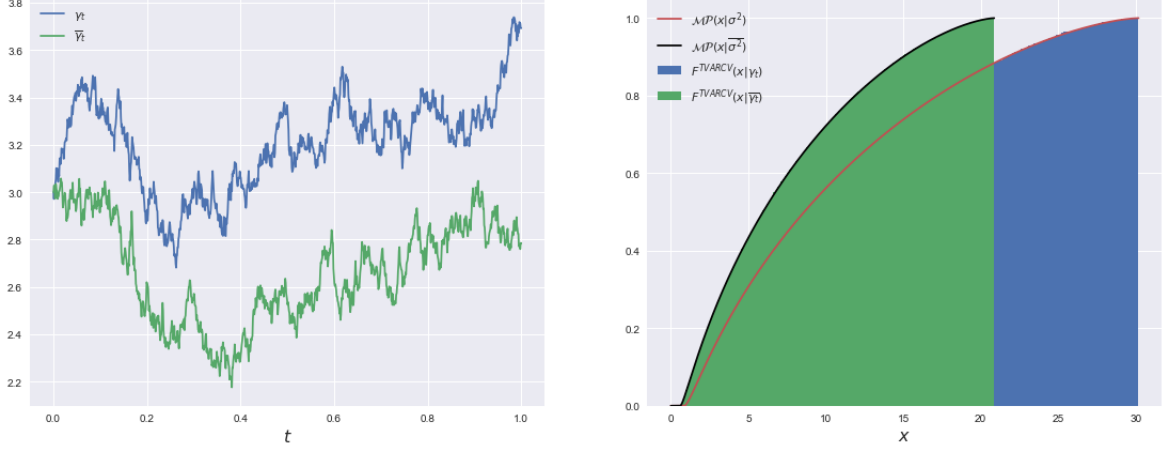


Figure 6: $\eta_p = p, y = 0.5, \alpha = 5, \beta = 1, \gamma_0 = 3, \xi = 2$

We see that even with the same parameters of volatility model, we have the different scale parameter in asymptotic spectral distribution.

where η_p is determined by Assumption 3.9 (ii). The process $\overline{W}(t)$ is still the standard Brownian motion, because it is a.s. continuous, has independent increments and

$$\overline{W}(t) - \overline{W}(s) = \frac{1}{\sqrt{\eta_p}} \sum_{j=1}^{\eta_p} (W_j(t) - W_j(s)) \sim \mathcal{N}(0, t - s) \quad \forall t > s.$$

We also take $\eta_p = p$, that means that $\overline{W}(t)$ is a sample average of $\mathbf{W}(t)$ and depends on every its component $W_j(t)$. Simulations show that even in this case, when Assumptions 3.9 (ii) are not satisfied, Theorem 3.14 still holds and F^{TVARCV} converges to Marchenko-Pastur law. Intuitively it can be explained in the following way: For any $0 \leq i \leq p$ we have

$$dW_i(t) \cdot d\overline{W}(t) = dW_i(t) \cdot \frac{1}{\sqrt{\eta_p}} \sum_{j=1}^{\eta_p} dW_j(t) = \frac{dt}{\sqrt{\eta_p}} \mathbf{1}_{\{i \leq \eta_p\}},$$

or equivalently

$$[W_i, \overline{W}](t) = \frac{t}{\sqrt{\eta_p}} \mathbf{1}_{\{i \leq \eta_p\}} \leq \frac{1}{\eta_p}.$$

Hence, with growing η_p the dependence of \overline{W}_t on ... becomes smaller and for $\eta_p = p \rightarrow \infty$ correlation between them goes to 0.

Jump-diffusion processes

Remark 4.5. (CHECK TEXT HERE!) Before we discussed pure diffusion processes, namely that these processes are continuous almost sure. It is quite useful to consider

so-called jump-diffusion processes... How should we change our estimator then? ALSO MENTION ONE- AND FINITE-DIMENSIONAL PROCESSES, THEN GO FOR $p > \infty$ ALSO WE ONLY CONSIDER EQUIDISTANT TIMES AND ONLY SIGMA EQUAL TO IDENTITY, WHICH MEANS INDEPENDENCE AND $\mu = 0$

Example 4.6.

- (i) Simplest model: let $U \sim \mathcal{U}[0, 1]$ be standard uniformly distributed random variable and

$$d\mathbf{X}(t) = \gamma(t)d\mathbf{W}(t) + \Delta_U(t), \quad (21)$$

where $\Delta_U(t) = (\delta_U(t), \dots, \delta_U(t))^T$ is a vector of Dirac delta functions shifted to U . In other words, all the components of $\mathbf{X}(t)$ make a jump of size 1 at the same random time $t = U$. Let us denote by $\mathbf{X}^c(t)$ the continuous part of $\mathbf{X}(t)$. Moreover, let ℓ' denote the index of interval, where U is located:

$$U \in \left[\frac{\ell' - 1}{n}, \frac{\ell'}{n} \right).$$

Then

$$\begin{aligned} \Sigma_p^{RCV} &= \sum_{\ell=1, \ell \neq \ell'}^n \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T + \Delta \mathbf{X}_{\ell'} (\Delta \mathbf{X}_{\ell'})^T \\ &= \sum_{\ell=1, \ell \neq \ell'}^n \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T + (\Delta \mathbf{X}^c_{\ell'} + \mathbb{U}_{p \times 1}) (\Delta \mathbf{X}^c_{\ell'} + \mathbb{U}_{p \times 1})^T \\ &= \sum_{\ell=1}^n \Delta \mathbf{X}^c_\ell (\Delta \mathbf{X}^c_\ell)^T + \Delta \mathbf{X}^c_{\ell'} \cdot \mathbb{U}_{1 \times p} + (\Delta \mathbf{X}^c_{\ell'} \cdot \mathbb{U}_{1 \times p})^T + \mathbb{U}_{p \times p}, \end{aligned}$$

where $\mathbb{U}_{k \times j}$ stands for $k \times j$ matrix filled with ones. Note that the elements of $\Delta \mathbf{X}^c_{\ell'} \cdot \mathbb{U}_{1 \times p}$ are i.i.d. $\sim \mathcal{N}(0, \Gamma_{\ell'}/n)$ with $\Gamma_{\ell'} = \int_{(\ell'-1)/n}^{\ell'/n} \gamma^2(t) dt$. Therefore, if we set $n \rightarrow \infty$, $\text{tr}(\Sigma_p^{RCV})$ will converge to the same value as before plus p . Also,

$$\|\Delta \mathbf{X}_{\ell'}\|_2^2 = \|\Delta \mathbf{X}^c_{\ell'} + \mathbb{U}_{p \times 1}\|_2^2 = \sum_{j=1}^p (\Delta X_{\ell'}^c)^2 + 2 \sum_{j=1}^p \Delta X_{\ell'}^c + p.$$

We have asymptotically:

$$\sum_{j=1}^p (\Delta X_{\ell'}^c)^2 \sim \frac{p}{n} \Gamma_{\ell'}$$

as a sum of p i.i.d. normally distributed random variables with variance $\Gamma_{\ell'}/n$ and

$$\sum_{j=1}^p \Delta X_{\ell'}^c \sim \mathcal{N}\left(0, \frac{p}{n} \Gamma_{\ell'}\right).$$

As long as p/n converges to some finite constant, $\Gamma_{\ell'}$ is bounded and p converges to infinity, we have

$$\frac{\Delta \mathbf{X}_{\ell'} (\Delta \mathbf{X}_{\ell'})^T}{\|\Delta \mathbf{X}_{\ell'}\|_2^2} \rightarrow 0.$$

Hence, if we know that all components jumped at the same time and the size of the jump is equal to 1, new estimator will be

$$\Sigma_p^{TVARCV} = \frac{\text{tr}(\Sigma_p^{RCV}) - p}{n} \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_{\ell} (\Delta \mathbf{X}_{\ell})^T}{\|\Delta \mathbf{X}_{\ell}\|_2^2}.$$

(ii) Now let U_1, \dots, U_p be i.i.d. $\sim \mathcal{U}[0, 1]$, $\mathbf{U} = (U_1, \dots, U_p)^T$ and

$$d\mathbf{X}(t) = \gamma(t)d\mathbf{W}(t) + \Delta_{\mathbf{U}}(t)$$

with $\Delta_{\mathbf{U}}(t) = (\delta_{U_1}(t), \dots, \delta_{U_p}(t))^T$.

Remark 4.7. In Example 4.6 we assumed that in the interval $[0, 1]$ each component of the process $\mathbf{X}(t)$ had exactly one jump. In practice, jump-diffusion processes are often modeled using Poisson processes, where one cannot be sure how many jumps can happen in given interval.

Definition 4.8.

(i) To construct a Poisson process, we begin with a sequence of arrival times $\tau_1^N, \tau_2^N, \dots$, where

$$\tau_{j+1}^N - \tau_j^N \sim \text{Exp}(\lambda) \quad \forall j \geq 1.$$

This setting is similar to the one in Example 4.2, except that now τ_j^N stands for the time of j -th jump. **The Poisson process** $N(t)$ counts the number of jumps that occur at or before time t . More precisely,

$$N(t) = \sum_{j \geq 1} \mathbf{1}_{\{\tau_j^N \leq t\}}.$$

Because the expected time between jumps is $1/\lambda$, the jumps are arriving at an average rate of λ per unit time. We say the Poisson process $N(t)$ has intensity λ .

(ii) Let $N(t)$ be a Poisson process with intensity λ , and let Y_1, Y_2, \dots be a sequence of i.i.d. random variables with mean

$$\beta = \mathbb{E}[Y_1].$$

We assume the random variables Y_1, Y_2, \dots are also independent of $N(t)$. We define the **compound Poisson process**

$$Q(t) = \sum_{i=1}^{N(t)} Y_i, \quad t \geq 0.$$

The jumps in $Q(t)$ occur at the same times as the jumps in $N(t)$, but whereas the jumps in $N(t)$ are always of size 1, the j -th jump in $Q(t)$ is of random size Y_j .

Part 5

Conclusion