

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

MASTERARBEIT

IM 1-FACH MASTERSTUDIENGANG FINANZMATHEMATIK DER  
MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT DER  
CHRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL

VORGELEGT VON  
ALEKSANDR SAMARIN

ERSTGUTACHTER: PROF. DR. MATHIAS VETTER

ZWEITGUTACHTER: PROF. DR. JAN KALLSEN

KIEL, AUGUST 2017



# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Motivation . . . . .	5
1.2	The relation between RCV and ICV . . . . .	7
<b>2</b>	<b>Marchenko-Pastur law</b>	<b>13</b>
2.1	Empirical spectral distribution . . . . .	13
2.2	Non-constant volatility case . . . . .	17
<b>3</b>	<b>Non-constant covolatility process</b>	<b>21</b>
3.1	Limiting distribution for realized covariance . . . . .	21
3.2	Time-variance adjusted realized covariance . . . . .	30
<b>4</b>	<b>Stochastic volatility case</b>	<b>39</b>
<b>5</b>	<b>Jump-diffusion processes</b>	<b>41</b>
<b>6</b>	<b>Simulation studies</b>	<b>45</b>
<b>A</b>	<b>Distribution function for Marchenko-Pastur law</b>	<b>53</b>
<b>B</b>	<b>Python code</b>	<b>55</b>
	<b>Bibliography</b>	<b>63</b>



# Chapter 1

## Introduction

### 1.1 Motivation

**Remark 1.1.** Continuous-time stochastic processes (e.g. Itô diffusions) are broadly accepted as a representation of (log-)price processes of stocks, currency or other financial assets. For example, suppose that we have multiple stocks, say,  $p$  stocks, whose price processes are denoted by  $S_j(t)$  and  $X_j(t) := \log S_j(t)$  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

$$d\mathbf{X}(t) = \boldsymbol{\mu}(t)dt + \Theta(t)d\mathbf{W}(t), \quad (1.1)$$

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.

If  $\Theta(t) = \sigma(t) \cdot \mathbb{I}_{p \times p}$ , we are in the case, where all the price processes are independent of each other and we can study them separately. In practice price on one good or stock can strongly depend on the price of the other. In this paper we consider the special class of possibly dependent processes, which interrelation stays stable in time. And since we are dealing with a high frequency data, we assume all the processes to live on a fixed time span, namely  $t \in [0, 1]$ .

**Definition 1.2.**

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

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

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

- (ii) Assume that we can observe the processes  $X_j(t)$  at high frequency synchronously, say at time points  $\{\tau_\ell\}_{0 \leq \ell \leq n}$ :

$$X_j(\tau_\ell) = \log S_j(\tau_\ell), \quad \ell = 0, 1, \dots, n, j = 1, \dots, p.$$

Then *realized covariance matrix (RCV)* is

$$\Sigma_p^{RCV} := \sum_{\ell=1}^n \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T, \quad (1.2)$$

where

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

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

**Example 1.3.** Let's take  $p = 3$ , zero drift  $\boldsymbol{\mu}(t) = (0, 0, 0)^T$  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 our process satisfies

$$d\mathbf{X}(t) = \begin{pmatrix} dW_1(t) + \cos(\pi t)dW_2(t) + \sin(\pi t)dW_3(t) \\ \cos(\pi t)dW_1(t) + dW_2(t) \\ \sin(\pi t)dW_1(t) + dW_3(t) \end{pmatrix}.$$

We have

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

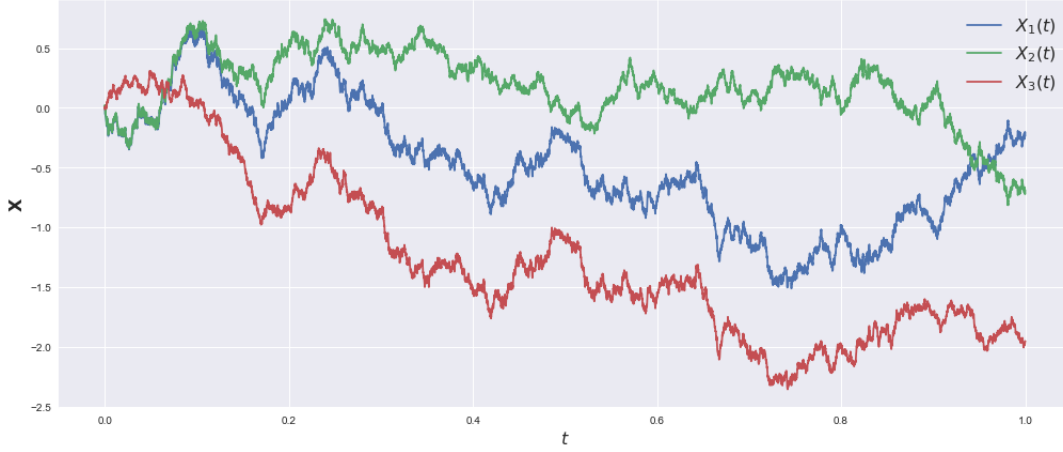
and corresponding ICV matrix is

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

The process  $\mathbf{X}(t)$  is shown on the Figure 1.1.

**Remark 1.4.**

- (i) Intuitively, one can wish zero elements of the ICV matrix to show the independence of one process from another. But in Example 1.3, where all the processes are related with each other in some sense, we still have zero values in  $\Sigma_3$ . It should be noted that if  $\mathbf{X}(t)$  is a vector of log-price processes, then over short time period, say, one day, it is reasonable to assume that the correlation structure doesn't change.

Figure 1.1:  $\mathbf{X}(t)$  from Example 1.3

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  $X_1(t)$ .

- (ii) One of the major problems in most of the models, based on the theory of stochastic calculus and applied to financial markets, is the estimation of true covariance. Without knowing how volatile is the price process, one can never estimate the fair price of the asset. In our case the final goal is to find a reasonable estimator of an ICV matrix. In the next section we show how RCV matrix does the job.

## 1.2 The relation between RCV and ICV

**Definition 1.5.** Recall the basic notions from the theory of stochastic calculus (without loss of generality we assume that all the processes have a starting point at 0 almost surely):

- (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:

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

Here,  $\mu(s)$  denotes a (locally bounded) drift process, whereas  $\sigma(s)$  is assumed to be a càdlàg volatility process.

- (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.

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

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

Note also that

$$[X, X](t) = \frac{1}{4}[2X](t) = \frac{1}{4}(4X^2(t) - 8 \int_0^t X(s)dX(s)) = [X](t).$$

**Lemma 1.6.**

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

$$[X](t) = \int_0^t \sigma_X^2(s)ds, \quad [Y](t) = \int_0^t \sigma_Y^2(s)ds$$

and

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

- (ii) Quadratic covariation is symmetric and bilinear map. Let  $X, Y, Z$  be continuous semimartingales, then

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

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

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

*Proof.*

- (i) The equality

$$[X, X](t) = \int_0^t \sigma_X^2(s)ds$$

is proven in most of the books about stochastic calculus (see, e.g., Lemma 4.4.4 in Shreve (2004) [1]). Then it follows

$$\begin{aligned} [X, Y](t) &= \frac{1}{4} \left( \left[ \int_0^\cdot (\sigma_X(s) + \sigma_Y(s))dW(s) \right](t) - \left[ \int_0^\cdot (\sigma_X(s) - \sigma_Y(s))dW(s) \right](t) \right) \\ &= \frac{1}{4} \left( \int_0^t (\sigma_X(s) + \sigma_Y(s))^2 ds - \int_0^t (\sigma_X(s) - \sigma_Y(s))^2 ds \right) \\ &= \int_0^t \sigma_X(s)\sigma_Y(s)ds. \end{aligned}$$



- (ii) It is easy to derive from the definition of quadratic covariation the so-called partial integration formula

$$[X, Y](t) = (XY)(t) - \int_0^t X(s) dY(s) - \int_0^t Y(s) dX(s).$$

Then we have

$$\begin{aligned} [X + Z, Y](t) &= ((X + Z)Y)(t) - \int_0^t (X + Z)(s) dY(s) - \int_0^t Y(s) d(X + Z)(s) \\ &= (XY)(t) - \int_0^t X(s) dY(s) - \int_0^t Y(s) dX(s) + (ZY)(t) - \int_0^t Z(s) dY(s) - \int_0^t Y(s) dZ(s) \\ &= [X, Y](t) + [Z, Y](t). \end{aligned}$$

The other equations can be proved in a similar way.

□

**Example 1.7.**

- (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 Definition 1.5 and Lemma 1.6 we get

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

and

$$[X, Y](t) = \int_0^t (\sigma_{11}(s)\sigma_{21}(s) + \sigma_{12}(s)\sigma_{22}(s)) ds. \quad (1.3)$$

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

$$\sum_{\ell=1}^n \Delta X_\ell \Delta Y_\ell = \sum_{\ell=1}^n (X(\tau_\ell) - X(\tau_{\ell-1})) \cdot (Y(\tau_\ell) - Y(\tau_{\ell-1})). \quad (1.4)$$

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 (1.4) converges in probability to  $[X, Y](1)$ . This limit is given by the right-hand side of (1.3). 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 Definition 1.5 and Lemma 1.6 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

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

and finally

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

**Remark 1.8.** Let ICV matrix  $\Sigma_p$  be deterministic. RCV matrix  $\Sigma_p^{RCV}$ , on the other side, is random by definition. In Example 1.7 we used quadratic covariation as an intermediate step to prove the convergence of one matrix to another. Alternatively to Definition 1.5 (iii) quadratic covariation can be defined as a limit of the sum of squared random differences, that means, that 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.



## Chapter 2

# Marchenko-Pastur law

### 2.1 Empirical spectral distribution

**Remark 2.1.** In Example 1.7 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  $\mathbf{A}$  be a  $p \times p$  Hermitian matrix and  $\{\lambda_j : j = 1, \dots, p\}$  be a set of its eigenvalues, then

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

where  $\#\{\cdot\}$  denotes the number of elements in the set indicated, is called *empirical spectral distribution (ESD)* of  $A$ .

**Remark 2.3.** A simple fact is that the  $h$ -th moment of  $F^{\mathbf{A}}$  can be written as

$$\int x^h dF^{\mathbf{A}}(x) = \frac{\text{tr}(\mathbf{A}^h)}{p}.$$

**Example 2.4.** Take the simplest case with  $\boldsymbol{\mu}(t) = 0$  and  $\Theta(t) = \gamma(t) \cdot \mathbb{I}_{p \times p}$  with some deterministic process  $\gamma(t)$ . Let also  $\sigma^2 = \int_0^1 \gamma^2(t) dt$ , then  $\Sigma_p = \sigma^2 \mathbb{I}_{p \times p}$  and its ESD

$$F^{\Sigma_p}(x) = 1_{[\sigma^2, \infty)}(x)$$

is a simple step function. In general case, where  $\Sigma_p$  have different eigenvalues, ESD is a combination of such step functions.

**Remark 2.5.** A naive estimator of the spectral distribution of the ICV matrix  $\Sigma_p$  is the spectral distribution of the RCV matrix  $\Sigma_p^{RCV}$ . In particular, one might wish that the

ESD  $F^{\Sigma_p^{RCV}}$  of  $\Sigma_p^{RCV}$  would approximate  $F^{\Sigma_p}$  well when the value of  $n$  sufficiently high, namely

$$\lim_{n \rightarrow \infty} F^{\Sigma_p^{RCV}} \rightarrow F^{\Sigma_p}.$$

From the large dimensional random matrix theory (LDRMT) it follows that in the high dimensional setting this good wish won't always come true. For example, in the simplest case when the drift process  $\boldsymbol{\mu}(t) \equiv 0$ , covolatility process  $\Theta(t)$  is constant, and observation times  $\tau_\ell$  are equally spaced ( $\tau_\ell = \ell/n$ ), 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, it is well known that if  $p/n$  converges to a non-zero number and the ESD  $F^{\Sigma_p}$  of the true covariance matrix converges, then the ESD  $F^{\Sigma_p^{RCV}}$  of the sample covariance matrix also converges (see e.g. Marchenko and Pastur (1967) [2], Silverstein (1995) [3]). However, the relationship between the limiting spectral distribution (LSD) of  $\Sigma_p^{RCV}$  in this case and the LSD of  $\Sigma_p$  still can be described, namely by a Marchenko-Pastur equation through Stieltjes transforms.

**Definition 2.6.** In mathematics, the *Stieltjes transformation*  $m_\mu(z)$  of a measure  $\mu$  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

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

The inversion formula is given by

$$\mu([a, b]) + \frac{\mu(\{a\}) + \mu(\{b\})}{2} = \frac{1}{\pi} \lim_{\eta \rightarrow 0^+} \int_a^b \Im(m_\mu(z + i\eta)) dz. \quad (2.1)$$

**Proposition 2.7.**

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

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

$$m_F(z) = \int_{\tau \in \mathbb{R}} \frac{1}{\tau(1 - y(1 + zm_F(z))) - z} dH(\tau). \quad (2.2)$$

*Proof.* See Theorem 1.1 of Silverstein (1995) [3].  $\square$

**Remark 2.8.**

- (i) It is easy to see that for ESD  $F$  its Stieltjes transform can be found by formula

$$m_F(z) = \frac{1}{p} \sum_{j=1}^p \frac{1}{\lambda_j - z} = \frac{1}{p} \text{tr}(S_p - z\mathbb{I}_{p \times p})^{-1}. \quad (2.3)$$

- (ii) Schematically Proposition 2.7 can be depicted as follows

$$(F^{\Sigma_p^{RCV}}, F^{\Sigma_p}) \xrightarrow{p, n \rightarrow \infty} (F, H).$$

Our goal is to find  $F^{\Sigma_p}$  for sufficiently large  $p$ . First, we approximate  $F$  with  $F^{\Sigma_p^{RCV}}$  and calculate  $m_F(z)$ , using, for example, (2.3). Then we derive  $H$  via Marchenko-Pastur equation (2.2), and use it as approximation of  $F^{\Sigma_p}$ . HOW TO DERIVE H???

- (iii) Note that if  $y \rightarrow 0$  limiting distribution function  $F$  of  $S_p$  matches limiting distribution function  $H$  of  $\Sigma_p$ . This is the case, when  $n$  grows faster than  $p$  and as we stated before  $\Sigma_p^{RCV}$  converges to  $\Sigma_p$ .
- (iv) 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 2.9.** Suppose that  $\mathbf{Z}_\ell^{(p)}$ 's are as in the Proposition 2.7, 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  $\delta_0(x)$  is a Dirac delta function and

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

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

*Proof.* See, e.g., Theorem 2.5 in Bai (1999) [4].  $\square$

**Remark 2.10.**

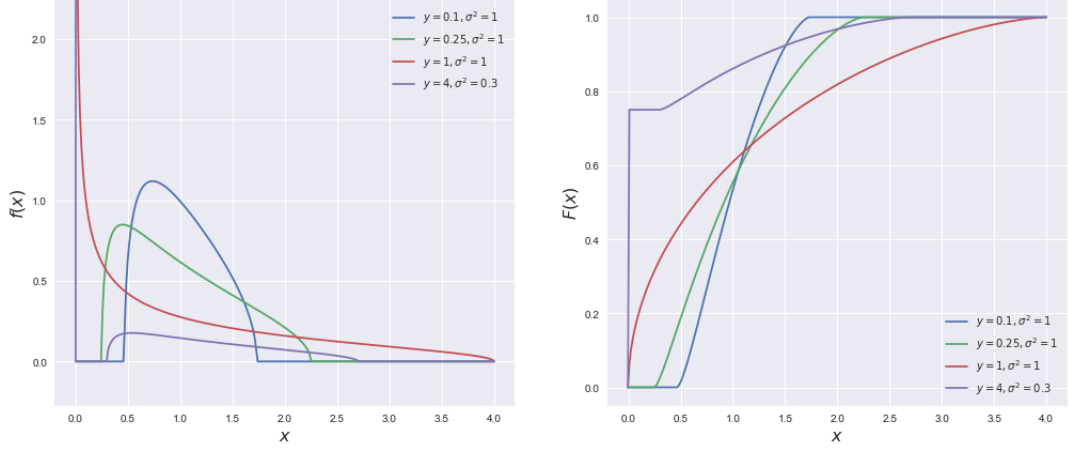


Figure 2.1: Marchenko-Pastur law

The left panel shows graphs of the density function  $f(x)$  for different parameters  $y$  and  $\sigma^2$ . On the right there are graphs of corresponding cumulative distribution functions  $F(x)$ . Notice that as  $y$  becomes larger the eigenvalues of  $\Sigma_p$  spread out instead of concentrating near  $\sigma^2$  and the length of the density's support increases.

- (i) We have for  $\Sigma_p = \sigma^2 \mathbb{I}_{p \times p}$  the density of limiting spectral distribution

$$dH(x) = \delta_{\sigma^2}(x).$$

Define  $m_{\mathcal{MP}}(z)$  as the Stieltjes transform of the Marchenko-Pastur distribution. Then it satisfies the self-consistent equation

$$\begin{aligned} m_{\mathcal{MP}}(z) &= \int_{\tau \in \mathbb{R}} \frac{1}{\tau(1-y-zm_{\mathcal{MP}}(z)) - z} \delta_{\sigma^2}(x) dx \\ &= \frac{1}{\sigma^2(1-y-zm_{\mathcal{MP}}(z)) - z}. \end{aligned}$$

- (ii) Note again that for  $y \rightarrow 0$  the density in Proposition 2.9 is

$$f(x) = \delta_{\sigma^2}(x),$$

which coincides with the density in Example 2.4 for finite value of  $p$  and with density  $dH(x)$ , because in this case  $\Sigma_p^{RCV}$  converges to  $\Sigma_p$ .

- (iii) One can show that

$$m_{\mathcal{MP}}(z) = \frac{\sigma^2(1-y) - z + i\sqrt{(b-z)(z-a)}}{2yz}.$$

Therefore, one can find  $m_{\mathcal{MP}}(z)$  through (2.3) and derive  $\sigma^2$  from it easily.

*Proof.* See, e.g., Lemma 2.18 in Ben Adlam [5]. □



## 2.2 Non-constant volatility case

**Remark 2.11.** In practice, the covolatility process is typically neither constant, nor even deterministic. Furthermore, stock volatility might also depend on the price process itself. Generally, for any time-varying covolatility process  $\Theta(t)$ , we associate it with a constant covolatility process given by the square root of the ICV matrix

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

Let  $\mathbf{X}^0(t)$  be 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 at time 1:

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

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  $\Sigma_p^{RCV}$  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. In particular, based on the results in LDRMT and the discussion about constant covolatility case, we have the following convergence property for  $\Sigma_p^{RCV^0}$ :

$$\lim_{p, n \rightarrow \infty} F^{\Sigma_p} = H \quad \implies \quad \lim_{p, n \rightarrow \infty} F^{\Sigma_p^{RCV^0}} = F.$$

Furthermore, their limits are related to each other via the Marchenko-Pastur equation (2.2). The question is: does this property hold for ESD of  $\Sigma_p^{RCV}$ ? Later we will see that even in the most ideal case, when the covolatility process has the form  $\Theta(t) = \gamma(t) \mathbb{I}_{p \times p}$  for some deterministic scalar function  $\gamma(t)$ , such convergence results may *not* hold for  $\Sigma_p^{RCV}$ . Moreover, the limit of  $F^{\Sigma_p^{RCV}}$  (when it exists) changes according to how the covolatility process evolves over time.

**Definition 2.12.** 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 **Löwner ordering**.

**Lemma 2.13** (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) [6].  $\square$

**Proposition 2.14.** Suppose that for any  $p$ ,  $\mathbf{X}(t)$  is a  $p$ -dimensional process satisfying

$$d\mathbf{X}(t) = \gamma(t)d\mathbf{W}(t), \quad t \in [0, 1], \quad (2.4)$$

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  $\tau_\ell$  are equally spaced, that is,  $\tau_\ell = \ell/n$ , and that the RCV matrix  $\Sigma_p^{RCV}$  is defined by (1.2). 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  $\lim p/n = y \geq y_c$ ,

$$\limsup F^{\Sigma_p^{RCV}}(b(y) + \sigma^2 \varepsilon) < 1 \quad \text{a.s.} \quad (2.5)$$

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 2.7,  $F^{\Lambda_p}$  will converge a.s. to the Marchenko-Pastur law with ratio index  $y' = \frac{y}{d-c}$  and scale index  $\sigma^2$ . 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,

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

□

**Remark 2.15.** In the proof of Proposition 2.14 we saw that more volatility process varies in time, the bigger value of  $\delta$  can be found, and hence more the maximum value of domain of  $F^{\Sigma_p^{RCV}}$  goes away from  $b(y)$ . Therefore ESD of  $\Sigma_p^{RCV}$  cannot be considered as consistent estimator anymore. In the next chapter we will study different estimator, which is based on the full behavior of volatility process.



## Chapter 3

# Non-constant covolatility process

### 3.1 Limiting distribution for realized covariance

**Remark 3.1.** In this chapter we will state two major theorems. The first one is a generalization of Proposition 2.7 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.

**Definition 3.2.** Suppose that for all  $p$   $\mathbf{X}(t) = \mathbf{X}^{(p)}(t)$  is a  $p$ -dimensional process satisfying (1.1), 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  $\Lambda$  a  $p \times p$  matrix satisfying  $\text{tr}(\Lambda\Lambda^T) = p$  such that

$$\Theta(t) = \gamma(t)\Lambda. \quad (3.1)$$

Observe that if (3.1) holds, then the ICV matrix  $\Sigma_p = \int_0^1 \gamma^2(t)dt \cdot \Lambda\Lambda^T$ . The special case when  $\Lambda = \mathbb{I}_{p \times p}$  will be studied in simulations.

**Example 3.3.** 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 the 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 the 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 3.4.**

- (i) In particular, if all Brownian motions in Example 3.3 are independent of each other, then  $R(t)$  is an identity matrix  $\mathbb{I}_{p \times p}$ . Then correlation matrix process of  $X_j(t)$  is a diagonal matrix and we need only to require  $\rho_{jj}$  to be constant in  $t$ .
- (ii) The process in Example 1.3 doesn't belong to the class  $\mathcal{C}$ , because its correlation structure varies in time.
- (iii) Observe that if a diffusion process  $\mathbf{X}(t)$  belongs to the class  $\mathcal{C}$ , the drift process  $\boldsymbol{\mu}(t) = 0$ , and  $\tau_\ell$ 's and  $\gamma(t)$  are independent of  $\mathbf{W}(t)$ , then

$$\Delta \mathbf{X}_\ell = \int_{\tau_{\ell-1}}^{\tau_\ell} \gamma(t) \Lambda d\mathbf{W}(t) \stackrel{d}{=} \sqrt{\int_{\tau_{\ell-1}}^{\tau_\ell} \gamma^2(t) dt} \cdot \bar{\Sigma}^{1/2} \cdot \mathbf{Z}_\ell,$$

where  $\bar{\Sigma}^{1/2}$  is the nonnegative square root matrix of

$$\bar{\Sigma} := \Lambda \Lambda^T$$

and  $\mathbf{Z}_\ell = (Z_{1\ell}, \dots, Z_{p\ell})^T$  consists of independent standard normals. Therefore, the RCV matrix

$$\Sigma_p^{RCV} = \sum_{\ell=1}^n \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T \stackrel{d}{=} \sum_{\ell=1}^n w_\ell^n \cdot \bar{\Sigma}^{1/2} \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T \bar{\Sigma}^{1/2}$$

with

$$w_\ell^n = \int_{\tau_{\ell-1}}^{\tau_\ell} \gamma^2(t) dt.$$

This is similar to the  $S_p$  in Proposition 2.7, except that here the weights  $w_\ell^n$  may vary in  $\ell$ , while in Proposition 2.7 the weights are constantly  $1/n$ .

**Remark 3.5.** In order to avoid confusion, we should specify the designations. For any vector  $A = (A_1, \dots, A_n)^T$ ,  $\|A\|_2$  denotes its  $L^2$ -norm, that is

$$\|A\|_2 = \sqrt{\sum_{\ell=1}^n A_\ell^2}.$$

Note that

$$\|A\|_2^2 = A^T A.$$

Let  $\mathbf{A}$  be  $m \times n$  matrix and  $\mathbf{A}^* = \overline{\mathbf{A}^T}$  its conjugate transpose. Let also  $\lambda_{\max}(\cdot)$  be the operator, taking the largest eigenvalue. Then

$$\|\mathbf{A}\| = \sqrt{\lambda_{\max}(\mathbf{A}^* \mathbf{A})}$$

denotes a spectral norm of  $\mathbf{A}$ . The value  $\lambda_{\max}(\mathbf{A}^* \mathbf{A})$  is also called singular value of the matrix  $\mathbf{A}$ .

**Proposition 3.6.** Suppose that  $\mathbb{P}_n$  are real probability measures with Stieltjes transforms  $m_n(z)$ . Let  $K \subseteq \mathbb{C}_+$  be an infinite set with a limit point in  $\mathbb{C}_+$ . If  $\lim_{n \rightarrow \infty} m_n(z) = m(z)$  exists for all  $z \in K$ , then there exists a probability measure  $\mathbb{P}$  with Stieltjes transform  $m(z)$  if and only if

$$\lim_{v \rightarrow \infty} iv \cdot m(iv) = -1, \quad (3.2)$$

in which case  $\mathbb{P}_n \rightarrow \mathbb{P}$  in distribution.

*Proof.* See, e.g., Theorem 2 of Geronimo and Hill (2003) [7].  $\square$

**Remark 3.7.** The idea of Proposition 3.6 is the stating of necessary and sufficient condition for limiting spectral distribution to exist. If (3.2) holds, then we can claim the convergence of RCV to some law.

**Theorem 3.8.** Assume that all the conditions in Proposition 2.7 are satisfied. Furthermore,

- (i)  $Z_{j\ell}^{(p)}$  have finite moments of all orders;
- (ii)  $H$  has a finite second moment;
- (iii) the weights  $w_\ell^n$ ,  $1 \leq \ell \leq n$ ,  $n = 1, 2, \dots$ , are all positive, and there exists  $\kappa < \infty$  such that the rescaled weights  $(nw_\ell^n)$  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  $\ell$ ,  $w_\ell^n$  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 Stieltjes transform  $m_{F^w}(z)$  is given by

$$m_{F^w}(z) = -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{1}{\tau M(z) + 1} dH(\tau) \quad (3.3)$$

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

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



*Proof.* For notational ease, we shall sometimes omit the sub/superscripts  $p$  and  $n$  in the arguments below: thus, we write  $\mathbf{Z}_\ell$  instead of  $Z_\ell^{(p)}$ ,  $w_\ell$  instead of  $w_\ell^n$ ,  $\Sigma_p$  instead of  $\Sigma_p$ ,  $S$  instead of  $S_p$ , etc. Also, recall that  $y_n = p/n$ , which converges to  $y > 0$ . By assumption (iv) we may, without loss of generality, assume that the weights  $w_\ell$  are independent of  $\mathbf{Z}_\ell$ 's. This is because, if we let  $\tilde{Z}_\ell$  be the result of replacing  $Z_{j_\ell}^{(p)}$ ,  $j \in \mathcal{I}_p$ , with independent random variables with the same distribution that are also independent of  $w_\ell$ , and

$$\tilde{S} := \sum_{\ell=1}^n w_\ell \cdot \Sigma^{1/2} \tilde{Z}_\ell (\tilde{Z}_\ell)^T \Sigma^{1/2},$$

then  $\text{rank}(\tilde{S} - S) \leq 2\eta_p$ , and so by the rank inequality

$$\|F^A - F^B\| \leq \frac{\text{rank}(A - B)}{p}$$

for any  $A, B$   $p \times p$  Hermitian matrices (see, e.g. Lemma 2.2 in Bai (1999) [4]),  $\tilde{S}$  and  $S$  must have the same LSD.

We proceed according to whether  $H$  is a delta measure at 0 or not. If  $H$  is a delta measure at 0, we claim that  $F^w$  is also a delta measure at 0, and the conclusion of the theorem holds. The reason is as follows. By assumption (iii)

$$S = \sum_{\ell=1}^n w_\ell \cdot \Sigma^{1/2} \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T \Sigma^{1/2} \leq \frac{\kappa}{n} \sum_{\ell=1}^n \Sigma^{1/2} \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T \Sigma^{1/2} =: \kappa \bar{S}.$$

Hence, by Lemma 2.13 again, for any  $x \geq 0$

$$F^S(x) \geq F^{\bar{S}}(x/\kappa).$$

However, it follows easily from Proposition 2.7 that  $F^{\bar{S}}$  converges to the delta measure at 0, hence so does  $F^S$ .

Below we assume that  $H$  is not a delta measure at 0. We will skip most of the technical details here, but provide main ideas of the proof instead. The major steps are

- Define the sequences  $(m_n(z))_n$  and  $(M_n(z))_n$  in the similar way as  $m_{F^w}(z)$  and  $M(z)$ , but with respect to the realization of  $\mathbf{X}(t)$  for some  $n$ .
- Show the convergence of such sequences to  $m(z)$  and  $M(z)$ .
- Conclude that the limit is unique.
- Using Proposition 3.6 state the convergence of probability measure.

Starting from the first step we rewrite equation (3.3) as follows:

$$m_{F^w}(z) = \frac{\text{tr}(-zM(z)\Sigma - z\mathbb{I}_{p \times p})^{-1}}{p}.$$

Let us define

$$m_n(z) = \frac{\text{tr}(S - z\mathbb{I}_{p \times p})^{-1}}{p}$$

and

$$M_n(z) = -\frac{1}{z} \sum_{\ell=1}^n \frac{w_\ell}{1 + w_\ell \mathbf{r}_\ell^T (S_{(\ell)} - z\mathbb{I}_{p \times p})^{-1} \mathbf{r}_\ell},$$

where

$$S_{(\ell)} := S - w_\ell \mathbf{r}_\ell \mathbf{r}_\ell^T$$

and

$$\mathbf{r}_\ell := \Sigma^{1/2} \mathbf{Z}_\ell.$$

The idea is to show the convergence

$$\frac{\text{tr}(-zM_n(z)\Sigma - z\mathbb{I}_{p \times p})^{-1}}{p} \rightarrow m_n \quad (3.4)$$

first. Using some lemmas from linear algebra and prior assumptions, we claim that (3.4) holds. Then we state that if  $m_n(z)$  converges to  $m(z)$ , then

$$M_n(z) \rightarrow M(z) := -\frac{1}{z} \int_0^1 \frac{w(s)}{1 + y\tilde{m}(z)w(s)} ds \neq 0, \quad (3.5)$$

where  $\tilde{m}(z)$  is the unique solution in  $\mathbb{Q}_1 = \{z \in \mathbb{C} : \Re(z) \geq 0, \Im(z) \geq 0\}$  to the following equation:

$$\int_0^1 \frac{1}{1 + y\tilde{m}(z)w(s)} ds = 1 - y(1 + zm(z)). \quad (3.6)$$

BLA BLA BLA EXPLAIN IT!!! Therefore, by convergence of  $F^{\Sigma_p}$  to  $H$

$$\frac{\text{tr}(-zM_n(z)\Sigma - z\mathbb{I}_{p \times p})^{-1}}{p} \rightarrow -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{1}{\tau M(z) + 1} dH(\tau).$$

But (3.4) implies that

$$m(z) = -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{1}{\tau M(z) + 1} dH(\tau). \quad (3.7)$$

Observing that  $M(z) \neq 0$ ,  $\Re(M(z)) \geq 0$ , and  $H$  is not a delta measure at 0, we obtain that  $|m(z)| < \frac{1}{|z|}$ . Hence  $1 + zm(z) \neq 0$ , and by (3.6),  $\tilde{m}(z) \neq 0$ . Based on this, we can get another expression for  $M(z)$ , as follows. By (3.5) we have

$$\begin{aligned} M(z) &= -\frac{1}{z} \int_0^1 \frac{w(s)}{1 + y\tilde{m}(z)w(s)} ds \\ &= \frac{1}{z} \cdot \frac{1}{y\tilde{m}(z)} \cdot \left(1 - \int_0^1 \frac{1}{1 + y\tilde{m}(z)w(s)} ds\right) \\ &= \frac{1}{z} \cdot \frac{1}{y\tilde{m}(z)} \cdot (1 - (1 - y(1 + zm(z)))) \\ &= \frac{1}{z} \cdot \frac{1 + zm(z)}{\tilde{m}(z)}. \end{aligned} \quad (3.8)$$

We can derive another formula for  $\tilde{m}(z)$ . By (3.7) and the fact that  $H$  is a probability distribution,

$$1 + zm(z) = 1 - \int_{\tau \in \mathbb{R}} \frac{1}{\tau M(z) + 1} dH(\tau) = M(z) \int_{\tau \in \mathbb{R}} \frac{\tau}{\tau M(z) + 1} dH(\tau).$$

Dividing both sides by  $-z\tilde{m}(z) \neq 0$  and using (3.8) yield

$$M(z) = -\frac{1 + zm(z)}{z\tilde{m}(z)} = -\frac{M(z)}{z\tilde{m}(z)} \cdot \int_{\tau \in \mathbb{R}} \frac{\tau}{\tau M(z) + 1} dH(\tau),$$

and since  $M(z) \neq 0$ ,

$$\tilde{m}(z) = -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{\tau}{\tau M(z) + 1} dH(\tau). \quad (3.9)$$

For the third step we show that for those  $z = iv$  with  $v$  sufficiently large, there is at most one triple

$$(m(z), M(z), \tilde{m}(z)) \in \mathbb{Q}_1 \times \mathbb{Q}_1 \times \mathbb{Q}_1,$$

that solves the equations (3.5), (3.7) and (3.9). THAT CAN BE SHOWN HOW??? For the fourth step it remains to verify that  $\lim_{v \rightarrow \infty} iv \cdot m(iv) = -1$ . In fact, using (3.7) we get that

$$iv \cdot m(iv) = - \int_{\tau \in \mathbb{R}} \frac{1}{1 + \tau M(iv)} dH(\tau). \quad (3.10)$$

Since  $\Re(M(iv)) \geq 0$ ,

$$\left| \frac{1}{1 + \tau M(iv)} \right| \leq 1 \quad \forall \tau \geq 0.$$

Moreover, by (3.5) and that  $\Re(\tilde{m}(z)) \geq 0$ ,

$$M(iv) \leq \frac{1}{v} \int_0^1 w(s) ds,$$

hence by the dominated convergence theorem, the right-hand side of (3.10) converges to  $-1$  as  $v \rightarrow \infty$ .  $\square$

**Remark 3.9.** If  $w_\ell^n = 1/n$ , then  $w(s) = 1$ , and Theorem 3.8 reduces to Proposition 2.7. 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 2.7 determined by  $H(\cdot / \int_0^1 w(t) dt)$ . The reason is as follows: firstly, for any  $w = w(s)$ , the family of distributions  $\{F^{cw} : c > 0\}$  is a scale family, i.e. for any  $x \geq 0$ ,

$$F^{cw}(x) = F^w(x/c).$$

Hence, it suffices to show that for any positive  $w$  such that  $w(s) \not\equiv 1$  on  $[0, 1]$  and  $\int_0^1 w(s) ds = 1$ ,  $F^w \neq F$ . In fact,  $F^w = F$  if and only if

$$m_{F^w}(z) = m_F(z) \quad \forall z \in \mathbb{C}_+,$$

and by the uniqueness of the solution to the equation (2.2) this can be true only if

$$M(z) = -\frac{1 - y(1 + zm_F(z))}{z}.$$

Suppose this is true, then by (3.8)

$$\frac{1 + zm_F(z)}{\tilde{m}(z)} = 1 - y(1 + zm_F(z))$$

or

$$\frac{1}{1 + y\tilde{m}(z)} = \frac{1 + zm_F(z)}{\tilde{m}(z)}.$$

By (3.6), the last equation implies that

$$\int_0^1 \frac{1}{1 + y\tilde{m}(z)w(s)} ds = \frac{1}{1 + y\tilde{m}(z)}.$$

Since both sides are analytical in  $\tilde{m}(z)$ , we get

$$\int_0^1 \frac{1}{1 + uw(s)} ds = \frac{1}{1 + u}, \quad \forall u \in \mathbb{C} \text{ such that } \Re(u) > 0.$$

In particular, focusing on  $u > 0$ , multiplying both sides by  $u$ , letting  $u \rightarrow \infty$  and applying the monotone convergence theorem to the left hand side yield

$$\int_0^1 \frac{1}{w(s)} ds = 1. \tag{3.11}$$

However, by the Cauchy-Schwarz inequality, since  $\int_0^1 w(s) ds = 1$  (3.11) holds only if  $w(s) \equiv 1$  on  $[0, 1]$ .

**Proposition 3.10.** Suppose that for all  $p$ ,  $X^{(p)}(t)$  is a  $p$ -dimensional process in class  $\mathcal{C}$  with some drift process  $\boldsymbol{\mu}(t) = (\mu_1^{(p)}(t), \dots, \mu_p^{(p)}(t))^T$ , covolatility process  $\Theta^{(p)}(t) = \gamma^{(p)}(t)\Lambda^{(p)}$  and  $p$ -dimensional Brownian motion  $\mathbf{W}^{(p)}(t) = (W_1^{(p)}(t), \dots, W_t^{(p)}(t))^T$ . Suppose further that:

- (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  $t \in [0, 1]$  almost surely;
- (ii)  $\check{\Sigma}_p = \Lambda^{(p)}(\Lambda^{(p)})^T$  is (a possibly random) positive semi-definite  $p \times p$  matrix such that its ESD  $F^{\check{\Sigma}_p}$  converges almost surely in distribution to a probability distribution  $\check{H}$  on  $[0, \infty)$  as  $p \rightarrow \infty$ ; moreover,  $\check{H}$  has a finite second moment;
- (iii) there exist  $C < \infty$  and  $\delta < 1/6$  such that for all  $p$ ,  $\|\check{\Sigma}_p\| \leq Cp^\delta$  almost surely;

- (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  $\gamma^{(p)}(t)$  may depend on  $\mathbf{W}^{(p)}(t)$  but only on  $\{W_j^{(p)}(t) : j \in \mathcal{I}_p\}$ ; moreover, there exists  $C_1 < \infty$  such that for all  $p$ ,  $|\gamma^{(p)}(t)| \leq C_1$  for all  $t \in [0, 1]$  almost surely; additionally, almost surely, there exists  $\gamma(t) : [0, 1] \rightarrow \mathbb{R}$  such that

$$\lim_{p \rightarrow \infty} \int_0^1 |\gamma^{(p)}(t) - \gamma(t)| dt = 0;$$

- (v) the observation times  $\tau_\ell$  are independent of  $\mathbf{X}(t)$ ; moreover, there exists  $\kappa < \infty$  such that the observation durations  $\Delta\tau_\ell := \tau_\ell - \tau_{\ell-1}$  satisfy

$$\max_n \max_{\ell=1, \dots, n} (n \cdot \Delta\tau_\ell) \leq \kappa;$$

additionally, almost surely, there exists a process  $v(s) : [0, 1] \rightarrow \mathbb{R}_+$  such that

$$\lim_{n \rightarrow \infty} \tau_{[ns]} = \Upsilon(s) := \int_0^s v(r) dr \quad \forall 0 \leq s \leq 1,$$

where for any  $x$ ,  $[x]$  stands for its integer part.

Then, as  $p \rightarrow \infty$ ,  $F^{\Sigma_p^{RCV}}$  converges almost surely to a probability distribution  $F^w$  as specified in Theorem 3.8 for

$$w(s) = \gamma^2(\Upsilon(s)) \cdot v(s).$$

**Remark 3.11.**

- (i) In particular, if we have equally spaced observation times, that is  $\tau_\ell = \frac{\ell}{n}$  for all  $0 \leq \ell \leq n$ , then

$$v(s) = 1 \quad \text{and} \quad \Upsilon(s) = s.$$

Therefore,  $w(s) = \gamma^2(s)$ .

- (ii) In the case of constant volatility  $\gamma(s) = 1$  we have  $w(s) = v(s)$ . Hence, RCV will converge to Marchenko-Pastur law only if observation times are equally spaced.
- (iii) Proposition 3.10 demonstrates explicitly how  $\gamma(s)$  and  $\tau_\ell$  affect LSD of RCV matrix. If  $\gamma(s)$  and hence  $w(s)$  are known, then in principle Theorem 3.8 can be used to recover  $F^{\Sigma_p}$ . However, in general  $\gamma(s)$  is unknown and estimating the process  $\gamma(s)$  can be challenging and will bring in more complication in the inference. Moreover, equations in Theorem 3.8 are different from and more complicated than the classical Marchenko-Pastur equation in Proposition 2.7 and in order to recover  $F^{\Sigma_p}$ , based on these equations, one has to extend existing algorithms, which are defined for equation (2.2). We shall instead propose an alternative estimator, which overcomes these difficulties.

### 3.2 Time-variance adjusted realized covariance

**Definition 3.12.** Suppose that a diffusion process  $\mathbf{X}(t)$  belongs to class  $\mathcal{C}$ . We define *time-variation adjusted realized covariance (TVARCV) matrix* as follows:

$$\widehat{\Sigma}_p := \frac{\text{tr}(\Sigma_p^{RCV})}{n} \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T}{\|\Delta \mathbf{X}_\ell\|_2^2} = \frac{\text{tr}(\Sigma_p^{RCV})}{p} \widetilde{\Sigma}_p, \quad (3.12)$$

where

$$\widetilde{\Sigma}_p := \frac{p}{n} \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T}{\|\Delta \mathbf{X}_\ell\|_2^2}. \quad (3.13)$$

**Remark 3.13.** Let us explain  $\widetilde{\Sigma}_p$ . 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

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

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

**Assumptions 3.14.**

- (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  $F^{\Sigma_p}$  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  $\delta_1$  in (ii) and  $\delta_2$  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,  $\tau_\ell$ 's are independent of  $\mathbf{X}(t)$ .

**Lemma 3.15** (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  $\tau$ , 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.* See, e.g., Revuz and Yor (1998) [8]. □

**Definition 3.16.** Let  $F$  and  $G$  be two cumulative distribution functions. Then the **Lévy distance** between them is

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

**Lemma 3.17.** *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  $\ell$ , all the entries of  $\mathbf{v}_\ell^{(p)}$  are bounded by  $\varepsilon_p$  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.

*Proof.* For notational ease we shall omit the superscript  $p$  and write  $\mathbf{v}_\ell = \mathbf{v}_\ell^{(p)}$ ,  $\mathbf{w}_\ell = \mathbf{w}_\ell^{(p)}$ ,  $v_{j\ell} = v_{j\ell}^{(p)}$ , etc. For each  $p$  set

$$\mathbf{A} = \mathbf{A}_p = (\mathbf{v}_1 + \mathbf{w}_1, \dots, \mathbf{v}_n + \mathbf{w}_1)$$

and

$$\mathbf{B} = \mathbf{B}_p = (\mathbf{w}_1, \dots, \mathbf{w}_n).$$

Then

$$\tilde{S}_n = \mathbf{A}\mathbf{A}^T \quad \text{and} \quad S_n = \mathbf{B}\mathbf{B}^T.$$

By Lemma 2.7 of Bai (1999)

$$L^4(F^{\tilde{S}_n}, F^{S_n}) \leq \frac{2}{p^2} \text{tr}((\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^T) \cdot \text{tr}(\tilde{S}_n + S_n).$$

By condition (ii),

$$\text{tr}((\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^T) = \sum_{\ell=1}^n \sum_{j=1}^p v_{j\ell}^2 \leq np \cdot \varepsilon_p^2.$$

Moreover,

$$\begin{aligned} \text{tr}(\tilde{S}_n) &= \sum_{\ell=1}^n \sum_{j=1}^p (v_{j\ell} + w_{j\ell})^2 \\ &\leq 2 \sum_{\ell=1}^n \sum_{j=1}^p v_{j\ell}^2 + 2 \sum_{\ell=1}^n \sum_{j=1}^p w_{j\ell}^2 \\ &\leq 2np \cdot \varepsilon_p^2 + 2 \text{tr}(S_n). \end{aligned}$$

Therefore,

$$L^4(F^{\tilde{S}_n}, F^{S_n}) \leq 4n^2 \varepsilon_p^4 + 6n \varepsilon_p^2 \frac{\text{tr}(S_n)}{p}.$$

The conclusion follows since  $\varepsilon_p = o(1/\sqrt{p})$  and hence  $o(1/\sqrt{n})$ .  $\square$

**Lemma 3.18.** For  $\mathbf{X} = (X_1, \dots, X_p)^T$ , where  $X_j$ 's are i.i.d. random variables such that  $\mathbb{E}[X_1] = 0$ ,  $\text{Var}(X_1) = 1$  and  $\mathbb{E}[|X_1|^{2k}] < \infty$  for some  $2 \leq k \in \mathbb{N}$ , there exists  $C_k \geq 0$ , depending only on  $k$ ,  $\mathbb{E}[|X_1|^4]$  and  $\mathbb{E}[|X_1|^{2k}]$ , such that for any  $p \times p$  nonrandom matrix  $\mathbf{A}$ ,

$$\mathbb{E}[|\mathbf{X}^* \mathbf{A} \mathbf{X} - \text{tr}(\mathbf{A})|^{2k}] \leq C_k (\text{tr}(\mathbf{A} \mathbf{A}^*))^k \leq C_k p^k \|\mathbf{A}\|^{2k}$$



*Proof.* See, e.g., Lemma 2.7 of Bai and Silverstein (1998) [9].  $\square$

**Lemma 3.19.** *Under Assumptions 3.14 (iv), namely, suppose that*

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

*then*

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

*Proof.* The full proof can be found in supplementary article of Zheng and Li (2011) [10]. We will provide here the main idea: it suffices to show that

$$\frac{\text{tr}(\Sigma_p^{RCV}) - \text{tr}(\Sigma_p)}{p} \xrightarrow[p \rightarrow \infty]{} 0 \quad a.s.$$

Let us first consider one-dimensional process

$$dX(t) = \mu(t)dt + \sigma(t)dW(t).$$

Under slight assumptions it can be shown that there exist constants  $c$ ,  $C_1$  and  $C_2$ , such that for all  $0 \leq x \leq c\sqrt{n}$

$$\mathbb{P}\left(\sqrt{n}\left|[X, X](1) - \int_0^1 \sigma^2(t)dt\right| \geq x\right) \leq C_1 \exp\{-C_2 x^2\}.$$

Therefore, generalizing it for  $p$ -dimensional processes, we have for every  $1 \leq j \leq p$

$$\mathbb{P}(\sqrt{n}|(\Sigma_p^{RCV})_{jj} - (\Sigma_p)_{jj}| > x) \leq C_1 \exp\{-C_2 x^2\}.$$

Hence, for any  $0 < \varepsilon < c$  and for all  $p$ ,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{\text{tr}(\Sigma_p^{RCV}) - \text{tr}(\Sigma_p)}{p}\right| \geq \varepsilon\right) &\leq \sum_{j=1}^p \mathbb{P}(\sqrt{n}|(\Sigma_p^{RCV})_{jj} - (\Sigma_p)_{jj}| \geq \varepsilon) \\ &\leq p \cdot C_1 \exp\{-C_2 \varepsilon^2 \cdot n\}. \end{aligned}$$

The almost sure convergence then follows from the Borel-Cantelli lemma<sup>1</sup>.  $\square$

---

<sup>1</sup>Let  $X_1, X_2, \dots$  be a sequence of random variables in some probability space and  $\varepsilon > 0$  be arbitrary. Then Borel-Cantelli lemma states that

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

**Corollary 3.20.** *The main results of the Lemma 3.19 are the convergence of  $\check{\Sigma}_p$*

$$\lim_{p \rightarrow \infty} \check{\Sigma}_p \rightarrow \frac{1}{\theta} \Sigma_p$$

*and of TVARCV estimator*

$$\lim_{p \rightarrow \infty} \hat{\Sigma}_p \rightarrow \theta \check{\Sigma}_p.$$

**Remark 3.21.** Using Lemma 3.19 we state that ESD of  $\check{\Sigma}_p$  converges almost surely:

$$\lim_{p, n \rightarrow \infty} F^{\check{\Sigma}_p} \rightarrow \check{H} \quad \text{a.s.},$$

where  $\check{H}$  is defined by

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

It is obvious, since

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

Now it is left to prove the following:

$$\lim_{p, n \rightarrow \infty} F^{\check{\Sigma}_p} \rightarrow \check{F} \quad \text{a.s.},$$

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

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

Then the convergence of  $\hat{\Sigma}_p$  will follow from the Corollary 3.20.

**Theorem 3.22.** *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 3.14 (i)-(vii) above. Suppose also that  $p$  and  $n$  satisfy Assumptions 3.14 (viii), and the observation times satisfy Assumptions 3.14 (ix). Let  $\hat{\Sigma}_p$  be as in (3.12). Then, as  $p \rightarrow \infty$ ,  $F^{\hat{\Sigma}_p}$  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 2.7.*

*Proof.* We only need to show the convergence of  $F^{\check{\Sigma}_p}$ . 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  $\gamma^{(p)}(t)$ ,  $\Lambda$  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, without loss of generality, we may assume that the index set  $\mathcal{I}_p \subset \{1, \dots, \eta_p\}$ . Then by Assumptions 3.14 (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

$$\tilde{\Sigma}_p = y_n \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T}{\|\Delta \mathbf{X}_\ell\|_2^2} = y_n \sum_{\ell=1}^n \frac{(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)^T}{\|\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell\|_2^2} \quad (3.14)$$

By Assumptions 3.14 (i), (ii) and (ix), there exists  $C > 0$ , such that  $|v_{j\ell}| \leq C/\sqrt{n}$  for all  $j$  and  $\ell$ , hence  $\|\mathbf{v}_\ell\|_2$ 's are uniformly bounded. We will show that

$$\max_{\ell=1, \dots, n} \left| \frac{\|\Lambda \mathbf{Z}_\ell\|_2^2}{p} - 1 \right| = \max_{\ell=1, \dots, n} \left| \frac{\mathbf{Z}_\ell^T \check{\Sigma}_p \mathbf{Z}_\ell}{p} - 1 \right| \rightarrow 0 \quad \text{a.s.}, \quad (3.15)$$

which clearly implies that

$$\max_{\ell=1, \dots, n} \left| \frac{\|\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell\|_2^2}{p} - 1 \right| \rightarrow 0 \quad \text{a.s.} \quad (3.16)$$

To prove (3.15) let us write

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

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

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

By a well-known fact about the spectral norm,

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

In particular, by Assumptions 3.14 (ii), (vi) and (vii),

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

hence

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

Now using the fact that  $\mathbf{D}_\ell$  consists of i.i.d. standard normals, Assumptions 3.14 and Lemma 3.18 we get for any  $k \in \mathbb{N}$

$$\mathbb{E}[|\mathbf{D}_\ell^T \mathbf{C} \mathbf{D}_\ell - \text{tr}(\mathbf{C})|^{2k}] \leq C_k p^k \cdot C_5 p^{2\delta k} = C'_k p^{(2\delta+1)k}.$$

Using Markov's inequality<sup>2</sup> we get that for any  $\varepsilon > 0$ ,

$$\mathbb{P}(|\mathbf{D}_\ell^T \mathbf{C} \mathbf{D}_\ell - \text{tr}(\mathbf{C})| \geq p\varepsilon) \leq \frac{C'_k p^{(2\delta+1)k}}{p^{2k} \varepsilon^{2k}} = C'_k p^{(2\delta-1)k} \varepsilon^{-2k} \quad \forall 1 \leq \ell \leq n.$$

Hence, choosing  $k > \frac{2}{1-2\delta}$ , using Borel-Cantelli and that  $n = O(p)$  yield

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

To complete the proof of (3.15), it then suffices to show that

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

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

$$|\mathbf{U}_\ell^T \mathbf{A} \mathbf{U}_\ell| \leq \|\mathbf{A}\| \cdot \|\mathbf{U}_\ell\|_2^2 \leq C_5 p^{\delta_2} \cdot \|\mathbf{U}_\ell\|_2^2 \quad (3.17)$$

Observe that for all  $1 \leq i \leq \eta_p$ , by Assumptions 3.14 (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 3.15, we then get that for any  $k \in \mathbb{N}$ , there exists  $\lambda_k > 0$  such that

$$\mathbb{E}[|Z_\ell^{(i)}|^{2k}] \leq \lambda_k C_2^{4k}. \quad (3.18)$$

Now we are ready to show that  $\max_{\ell=1, \dots, n} \frac{|\mathbf{U}_\ell^T \mathbf{A} \mathbf{U}_\ell|}{p} \rightarrow 0$ . In fact, for any  $\varepsilon > 0$ , for any

---

<sup>2</sup>Markov's inequality states that for any non-negative random variable  $X$ ,  $k \in \mathbb{N}$  and  $n \in \mathbb{R}_+$  the following inequality holds:

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

$k \in \mathbb{N}$ , by Markov's inequality, (3.17), Hölder's inequality and (3.18),

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

By Assumptions 3.14 (vii),  $\delta_1 + \delta_2 < 1/2$ , hence by choosing  $k$  to be large enough, the right hand side will be summable in  $p$ :

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

Hence by Borel–Cantelli lemma, almost surely,  $\max_{\ell=1,\dots,n} \frac{|\mathbf{U}_\ell^T \mathbf{A} \mathbf{U}_\ell|}{p} \rightarrow 0$ . We now get back to  $\tilde{\Sigma}_p$ . By (3.16), 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,

$$F^{\tilde{S}_p}((1 + \varepsilon)x) \geq F^{\tilde{\Sigma}_p}(x) \geq F^{\tilde{S}_p}((1 - \varepsilon)x) \quad \forall x \geq 0. \quad (3.19)$$

Next, by Lemma 3.17  $\tilde{S}_p$  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 3.8,  $F^{S_p}$  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 2.7 that it converges to  $\tilde{F}$ , because  $\check{\Sigma}_p = \Lambda\Lambda^T$ . Moreover, by Theorems 1.1 and 2.1 in Silverstein and Choi (1995),  $\tilde{F}$  is differentiable and in particular continuous at all  $x > 0$ . It follows from (3.19) that  $F^{\tilde{\Sigma}_p}$  must also converge to  $\tilde{F}$ .  $\square$

## Chapter 4

# Stochastic volatility case

**Remark 4.1.** Stochastic volatility models describe the behavior of volatility of a log-price  $\mathbf{X}(t)$  by a function of some auxiliary (possibly multi-dimensional) stochastic process. In this chapter we give definitions of the models which are the most widely used in a quantitative sphere.

**Definition 4.2.**

- (i) Let  $\overline{W}(t)$ ,  $t \geq 0$  be a standard Brownian motion. The *Vasicek model* for stochastic process is defined by the stochastic differential equation (SDE)

$$dZ(t) = \beta(\alpha - Z(t))dt + \xi \overline{W}(t), \quad (4.1)$$

where  $\alpha$ ,  $\beta$  and  $\xi$  are positive constants. The solution can be determined in closed form and is

$$Z(t) = Z(0)e^{-\beta t} + \alpha(1 - e^{-\beta t}) + \xi e^{-\beta t} \int_0^t e^{\beta s} d\overline{W}(s).$$

The properties of  $Z(t)$  are the following:

- $Z(t)$  is normally distributed with mean

$$\mathbb{E}[Z(t)] = e^{-\beta t} Z(0) + \alpha(1 - e^{-\beta t})$$

and variance

$$\text{Var}(Z(t)) = \frac{\xi^2}{2\beta}(1 - e^{-2\beta t}).$$

- The model is mean-reverting with long-run mean  $\alpha$ . That means,  $Z(t)$  tends to reach  $\alpha$ , and once it does it, it starts to fluctuate near its vicinity. Parameter  $\beta$  represents the speed of mean reversion and  $\xi$  is a parameter, which determines the volatility of the variance process.
- There is a probability that  $Z(t)$  will go below zero:

$$\mathbb{P}(Z(t) < 0) > 0.$$

Due to the last property it is not always convenient to consider this process for volatility modeling.

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

$$dY(t) = \beta(\alpha - Y(t))dt + \xi\sqrt{Y(t)}d\bar{W}(t), \quad (4.2)$$

where  $\alpha$ ,  $\beta$  and  $\xi$  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  $\lambda$ :

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

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. Also,  $Y(t)$  has the same expectation as in Vasicek model, and its variance is

$$\text{Var}(Y(t)) = \frac{\xi^2}{\beta}Y(0)(e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha\xi^2}{2\beta}(1 - 2e^{-\beta t} + e^{-2\beta t}).$$

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

$$dX(t) = \mu dt + \sqrt{Y(t)}dW(t), \quad (4.3)$$

where  $Y(t)$  satisfies SDE (4.2) and  $W(t)$  is another standard Brownian motion, such that

$$[W, \bar{W}](t) = \int_0^t \rho(s)ds \quad (4.4)$$

holds with a (stochastic) process  $\rho(t)$ , taking values strictly between  $-1$  and  $1$ . Sometimes equality (4.4) is informally written as

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

**Remark 4.3.** If we are in a Heston model with  $\rho(t) \equiv 0$ , then we have a closed-form solution for  $X(t)$ :

$$X(t) \sim \mathcal{N}(\mu t, Y(t) \cdot t),$$

where  $Y(t)$  is an independent CIR process and therefore conditionally on  $t$  and other parameters has a scaled non-central chi-squared distribution. In the case of nonzero correlation between  $Y(t)$  and  $X(t)$  one can simulate  $X(t)$ , using, for example, either Euler discretization, or the exact simulation algorithm (see, e.g., Broadie and Kaya (2004) [11]).



## Chapter 5

# Jump-diffusion processes

**Remark 5.1.** Before we discussed pure diffusion processes, namely that these processes are continuous almost sure. In this chapter we study diffusion processes with jumps. The reason is, that it doesn't matter how volatile or random is  $\gamma^2(t)$ , until it is finite or at least integrable, the process  $\mathbf{X}(t)$  will be always continuous. In practice, things can be different: during previous century we observed numerous situations, when the prices of some assets dropped on a huge value instantly, causing worldwide market crises. The question now is: if it is useful to consider the so-called jump-diffusion processes, how should we change our estimator?

**Example 5.2.** Simplest model: let  $U \sim \mathcal{U}[0, 1]$  be standard uniformly distributed random variable and

$$d\mathbf{X}(t) = \Theta(t)d\mathbf{W}(t) + a\Delta_U(t)dt, \quad (5.1)$$

where  $a$  is a known positive constant and  $\Delta_U(t) = (\delta_U(t), \dots, \delta_U(t))^T$  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  $a$  at the same random time  $t = U$ . Let us denote by  $\mathbf{X}^c(t)$  the continuous part of  $\mathbf{X}(t)$ :

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

Moreover, assume that we have equidistant observation times  $\tau_\ell = \ell/n$  and let  $\ell'$  be the index of interval, where  $U$  is located:

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

Then

$$\begin{aligned} \Sigma_p^{RCV} &= \sum_{\ell=1, \ell \neq \ell'}^n \Delta\mathbf{X}_\ell (\Delta\mathbf{X}_\ell)^T + \Delta\mathbf{X}_{\ell'} (\Delta\mathbf{X}_{\ell'})^T \\ &= \sum_{\ell=1, \ell \neq \ell'}^n \Delta\mathbf{X}_\ell (\Delta\mathbf{X}_\ell)^T + (\Delta\mathbf{X}_{\ell'}^c + a\mathbb{U}_{p \times 1}) (\Delta\mathbf{X}_{\ell'}^c + a\mathbb{U}_{p \times 1})^T \\ &= \sum_{\ell=1}^n \Delta\mathbf{X}_\ell^c (\Delta\mathbf{X}_\ell^c)^T + a\Delta\mathbf{X}_{\ell'}^c \cdot \mathbb{U}_{1 \times p} + a(\Delta\mathbf{X}_{\ell'}^c \cdot \mathbb{U}_{1 \times p})^T + a^2\mathbb{U}_{p \times p}, \end{aligned}$$

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  $\text{tr}(\Sigma_p)/p$  converges to  $\theta$  as in Lemma 3.19, then

$$\lim_{n \rightarrow \infty} \frac{\text{tr}(\Sigma_p^{RCV})}{p} \rightarrow \theta + a^2 \quad \text{a.s.}$$

Also,

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

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  $a$ , new estimator will be

$$\widehat{\Sigma}'_p = \frac{\text{tr}(\Sigma_p^{RCV}) - a^2 p}{n} \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_{\ell} (\Delta \mathbf{X}_{\ell})^T}{\|\Delta \mathbf{X}_{\ell}\|_2^2}.$$

**Remark 5.3.** In Example 5.2 we assumed that in the interval  $[0, 1]$  each component of the process  $\mathbf{X}(t)$  had exactly one jump and exactly at the same time. In practice, jump-diffusion processes are often modeled using Poisson processes, where one cannot be sure how many jumps and when can happen in given interval.

**Definition 5.4.**

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

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

Value  $\tau_j^N$  defines 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,

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

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

- (ii) Let  $N(t)$  be a Poisson process with intensity  $\lambda$ , 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  $Y_j$ .

- (iii) Let  $J(t)$  be an adapted, right-continuous **pure jump process** with  $J(0) = 0$ . By right-continuous we mean that  $J(t) = \lim_{s \downarrow t} J(s)$  for all  $t \geq 0$ . In other words, if  $J$  has a jump at time  $t$ , then  $J(t)$  is the value of  $J$  immediately. A Poisson process and compound Poisson processes are the examples of pure jump processes. A process  $X(t)$  of the form

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

where  $\mu(s), \sigma(s)$  are adapted processes and  $W(s)$  is a standard Brownian motion is called a **jump-diffusion process**.

**Proposition 5.5.** Let  $X_1(t) = \int_0^t \mu_1(s) ds + \int_0^t \sigma_1(s) dW(s) + J_1(t)$  be a jump-diffusion process. Then

$$[X_1, X_1](t) = \int_0^t \sigma_1^2(s) ds + \sum_{0 < s \leq t} (\Delta J_1(s))^2.$$

Let also  $X_2(t) = \int_0^t \mu_2(s) ds + \int_0^t \sigma_2(s) dW(s) + J_2(t)$ , then

$$[X_1, X_2](t) = \int_0^t \sigma_1(s) \sigma_2(s) ds + \sum_{0 < s \leq t} \Delta J_1(s) \Delta J_2(s).$$

*Proof.* See Theorem 11.4.7 in Shreve (2004) [1]. □

**Remark 5.6.** Let us again look at the case with fixed  $p$ . Let also  $\Theta(t)$  for simplicity be constant in  $t$  and observation times equally spaced. In Example 5.2 we then have

$$J_j(t) = a \mathbf{1}_{[U, 1]}(t) \quad \forall 1 \leq j \leq p$$

and

$$\Sigma_p^{RCV} \xrightarrow{\mathbb{P}} \Sigma_p + a^2 \mathbb{U}_{p \times p}.$$

Let us consider the process, determined by the following SDE:

$$dX_j(t) = \sigma dW_j(t) + \Delta N_j(t) \quad \forall 1 \leq j \leq p,$$

where  $N_j(t)$  are independent Poisson processes. The probability that two jumps happen simultaneously is equal to 0, hence

$$[X_i, X_j](t) = \sigma^2 \quad \forall 1 \leq i \neq j \leq p.$$

Recall that the size of the jumps of Poisson process is equal to 1 and therefore,

$$\Sigma_p^{RCV} \xrightarrow{\mathbb{P}} \Sigma_p + \Sigma_p^N$$

with

$$\Sigma_p^N := \begin{pmatrix} \sum_{i \geq 1} \mathbf{1}_{[0,1]}(\tau_i^{N_1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{i \geq 1} \mathbf{1}_{[0,1]}(\tau_i^{N_p}) \end{pmatrix}.$$

The limit of  $\Sigma_p^{RCV}$  is no longer deterministic, because diagonal elements of  $\Sigma_p^N$  are random.

Now let us set  $p \rightarrow \infty$  with the same rate as  $n$ . Matrix  $\Sigma_p^N$  won't be diagonal anymore, because the probability that we observe multiple jumps at the same time is bigger than 0. Moreover, all its elements are non-negative, which means that the limit of  $\Sigma_p^{RCV}$  will be bigger, than  $\Sigma_p$  and their ESD will be different. Let us replace Poisson process  $N(t)$  with compound Poisson process  $Q(t)$  with  $\beta = \mathbb{E}[Y_1] = 0$ . Then  $\mathbf{X}(t)$  stays a martingale and the natural question then is, will ESD of RCV or TVARCV converge to Marchenko-Pastur law? We will study this case in the next chapter.

## Chapter 6

# Simulation studies

**Remark 6.1.** BLA BLA BLA Here we do simulations

**Example 6.2.** Let's take an example from the paper... Assume we have equidistant observation times  $\tau_\ell = \ell/n$ ,  $\boldsymbol{\mu}(t) \equiv 0$  and

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

Then

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

Note that

$$\Delta X_{jl} = X_j(\ell/n) - X_j((\ell-1)/n) \sim \mathcal{N}(0, \Gamma_\ell/n),$$

for all  $1 \leq j \leq p$ , where

$$\Gamma_\ell = \int_{(\ell-1)/n}^{\ell/n} \gamma^2(t) dt = 0.0009 + 0.0008 \cdot \frac{\sin(2\pi\ell/n) - \sin(2\pi(\ell-1)/n)}{2\pi}.$$

The simulations show that TVARCV estimator converges asymptotically to Marchenko-Pastur law, while RCV doesn't.

Firstly, we are in a setting  $y = 4$ . Then  $a = \sigma^2 = 0.0009$  and  $b = 9\sigma^2 = 0.0081$ . The process  $\mathbf{X}(t)$ , LSD histograms and plot of the density of Marchenko-Pastur law are shown on Figure 6.1.

In the second case, we run simulations for  $y = 1$ . Then  $a = 0$  and  $b = 4\sigma^2 = 0.0036$ . We will compare it with the next setting.

**Example 6.3.** 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  $\lambda$ , that is

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

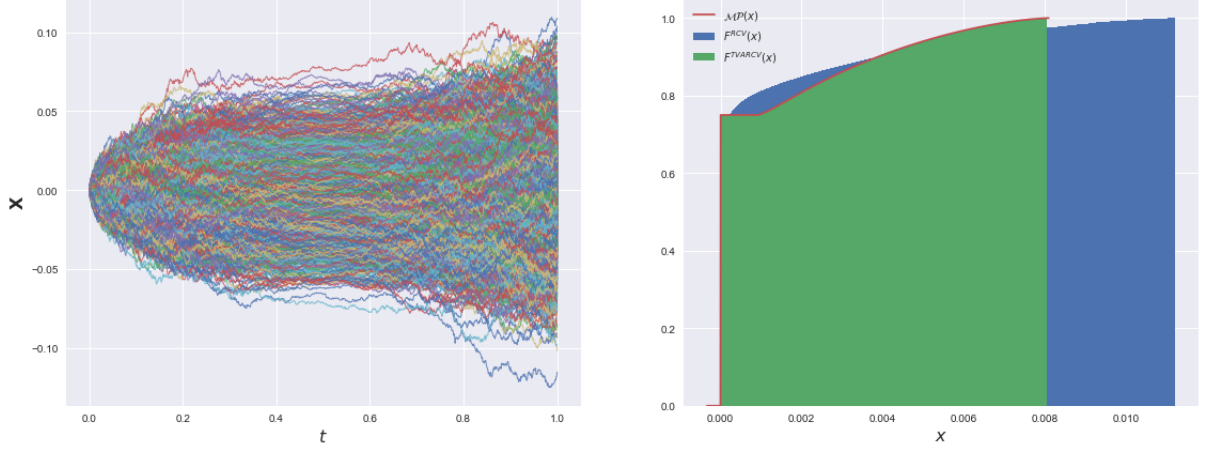


Figure 6.1:  $\mathbf{X}(t)$  from Example 6.2,  $n = 1000$ ,  $p = 4000$ ,  $\sigma^2 = 0.0009$

In this setting  $y = p/n = 4$ . On the left panel one can see paths of four thousands independent stochastic processes. On the right blue histogram shows ESD of  $\Sigma_p^{RCV}$ , green histogram – ESD of  $\hat{\Sigma}_p$  and red curve – Marchenko-Pastur law. Note that distribution function makes jump at 0 to  $(1 - 1/y) = 0.75$  and then stays flat until it reaches the value of  $a = 0.0009$ .

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

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

then by setting  $\lambda = \frac{1}{n}$ , we get  $\#\mathcal{T} \sim n$ . Let's take volatility  $\gamma_t$  as in Example 6.2. Then we have

$$\Delta X_{jl} = X_j(\tau_\ell) - X_j(\tau_{\ell-1}) \sim \mathcal{N}(0, \Gamma_\ell \cdot (\tau_\ell - \tau_{\ell-1})),$$

for all  $1 \leq j \leq p$ , with

$$\Gamma_\ell = \int_{\tau_{\ell-1}}^{\tau_\ell} \gamma^2(t) dt = 0.0009 + 0.0008 \cdot \frac{\sin(2\pi\tau_\ell) - \sin(2\pi\tau_{\ell-1})}{2\pi}.$$

According to Proposition 3.10 different intervals between observation times will affect RCV as strong as fluctuating volatility, and we observe it on the Figure 6.3.

**Example 6.4.** Let's assume

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

where  $Y_t$  is a CIR process, determined by (4.2). Then it is easy to see that

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

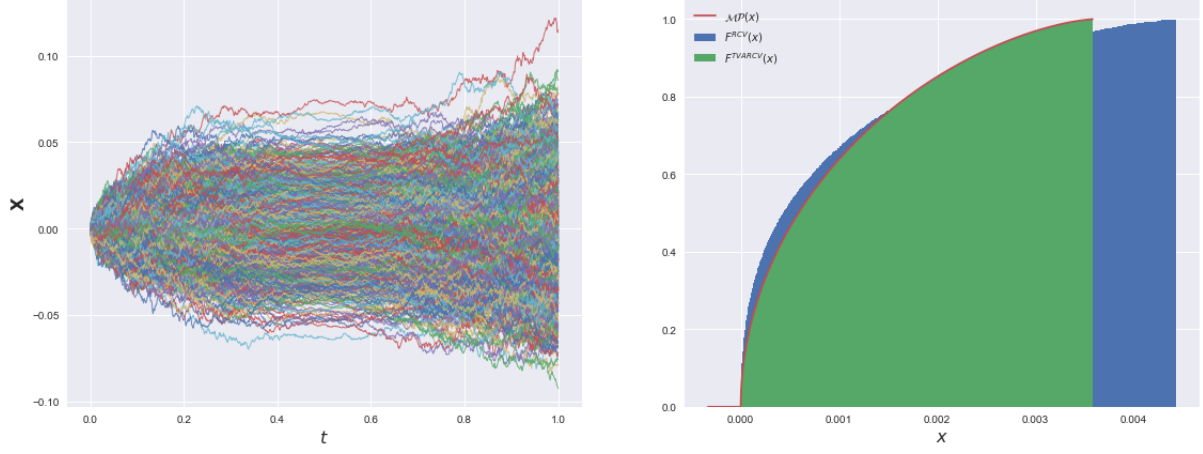


Figure 6.2:  $\mathbf{X}(t)$  from Example 6.2,  $n = 1000$ ,  $p = 1000$

In this setting  $y = p/n = 1$ . Note, that the domain of  $F^{RCV}$  is larger than the domain of Marchenko-Pastur law, as it was proved in Proposition 2.14.

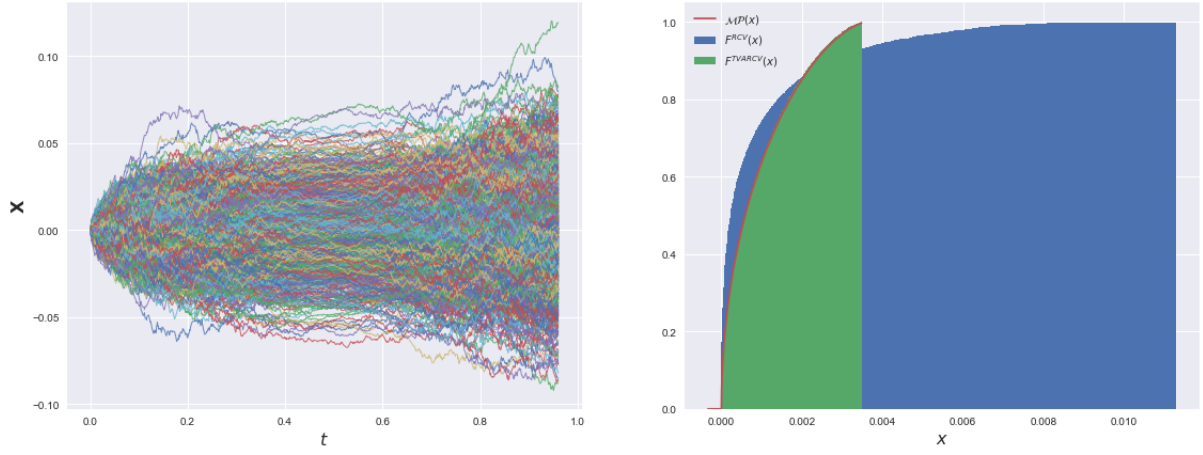


Figure 6.3:  $\mathbf{X}(t)$  from Example 6.3,  $n \simeq 1000$ ,  $p = 1000$

We see that if observation times are not equidistant, the ESD of RCV changes drastically, even if visually there is no difference for paths of  $\mathbf{X}(t)$  and volatility process  $\gamma(t)$  stays the same.

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

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

where  $\eta_p$  is determined by Assumption 3.14 (ii). The process  $\overline{W}(t)$  is still the standard Brownian motion, because it is a.s. continuous, has independent increments and

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

**Remark 6.5.** Theorem 3.22 states that as long  $\eta_p = o(p)$ , TVARCV will converge to Marchenko-Pastur law. But let us take  $\eta_p = p$ , that means that  $\overline{W}(t)$  is a sample average of  $\mathbf{W}(t)$  and depends on every of its component  $W_j(t)$ . Simulations show that even in this case, when Assumptions 3.14 (ii) are not satisfied, Theorem 3.22 still holds and ESD of  $\widehat{\Sigma}_p$  converges to Marchenko-Pastur law. Intuitively it can be explained in the following way:

For any  $1 \leq i \leq p$  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

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

Hence, with growing  $\eta_p$   $\overline{W}_t$  is based on bigger amount of components of  $\mathbf{W}(t)$ , but the dependence becomes smaller and for  $\eta_p = p \rightarrow \infty$  correlation between  $\overline{W}_t$  and any component of  $\mathbf{W}(t)$  goes to 0.

**Example 6.6.** Let  $Y_1, Y_2, \dots$  be i.i.d. random variables  $\sim \mathcal{N}(0, 1)$ . Let also

$$\mathbf{N}(t) = (N_1(t), \dots, N_p(t))^T$$

be a vector of independent Poisson processes all with intensity  $\lambda$  and

$$d\mathbf{X}(t) = d\mathbf{W}(t) + \mathbf{Q}(t),$$

where  $\mathbf{Q}(t) = (Q_1(t), \dots, Q_p(t))^T$  is a vector of compound Poisson processes, whose components are

$$Q_j(t) = \sum_{i=1}^{N_j(t)} Y_i \quad \forall 1 \leq j \leq p.$$



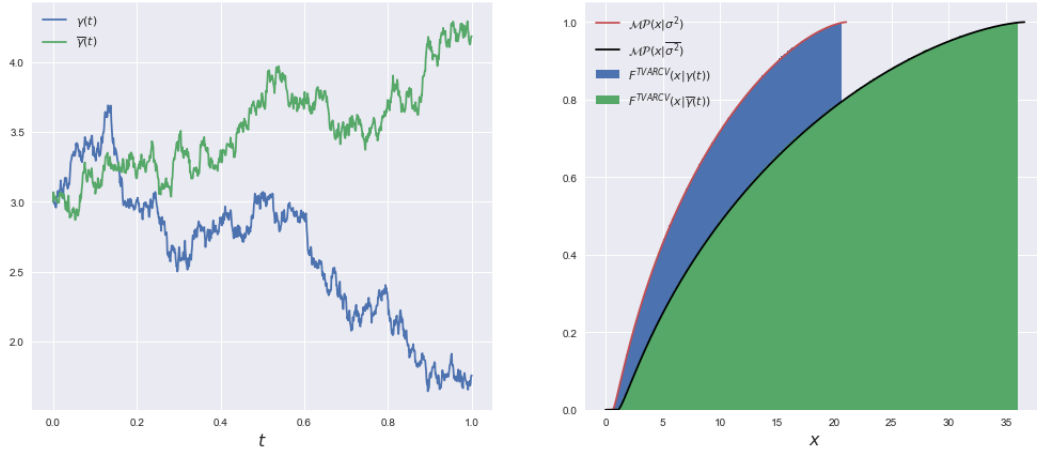


Figure 6.4:  $\eta_p = p, y = 1, \alpha = 5, \beta = 1, \gamma(0) = 3, \xi = 2$

Left panel shows two different realizations of  $\gamma(t)$ . On the right panel blue and green histograms are ESD of  $\hat{\Sigma}_p$  for corresponding volatility processes, red and black curve – Marchenko-Pastur law for different realizations of  $\sigma^2$ . We see that even with the same parameters of volatility model, we have the different scale parameter in asymptotic spectral distribution for different realizations.

Note, that  $Q_j(t)$  are martingales.

Let us first be in a setting  $\lambda = 1$ . Then we can see on a figure that both RCV and TVARCV fail.

However, if we set  $\lambda \rightarrow 0$ , jumps will happen too frequently and  $\mathbf{Q}(t)$  will converge to the simple diffusion process. Due to normality of  $Y_i$  in this example, we get

$$d\mathbf{Q}(t) \simeq \frac{1}{\sqrt{dt}} d\overline{\mathbf{W}}(t),$$

where  $\overline{\mathbf{W}}(t)$  is a  $p$ -dimensional standard Brownian motion. Then, roughly speaking,

$$\Sigma_p \simeq n\mathbb{I}_{p \times p}.$$

**Remark 6.7.** For TVARCV to converge to Marchenko-Pastur law, it is sufficient, but not necessary, that  $\mathbf{X}(t)$  belongs to the class  $\mathcal{C}$ . It is noteworthy to show that there can be the case when we have convergence even for RCV.

**Example 6.8.** Let drift process be as usual  $\boldsymbol{\mu}(t) \equiv 0$  and  $\Theta(t)$  be diagonal matrix, which elements are:

$$\left(\Theta(t)\right)_{jj} = \sin(\pi t)\mathbf{1}_{\{j \leq p/2\}} + \cos(\pi t)\mathbf{1}_{\{j > p/2\}}, \quad 1 \leq j \leq p.$$

In other words, first  $p/2$  elements of the diagonal of  $\Theta(t)$  are  $\cos(\pi t)$  and the others are  $\sin(\pi t)$ . It is obvious, that  $\mathbf{X}(t)$  doesn't belong to the class  $\mathcal{C}$ . The ICV matrix is in this

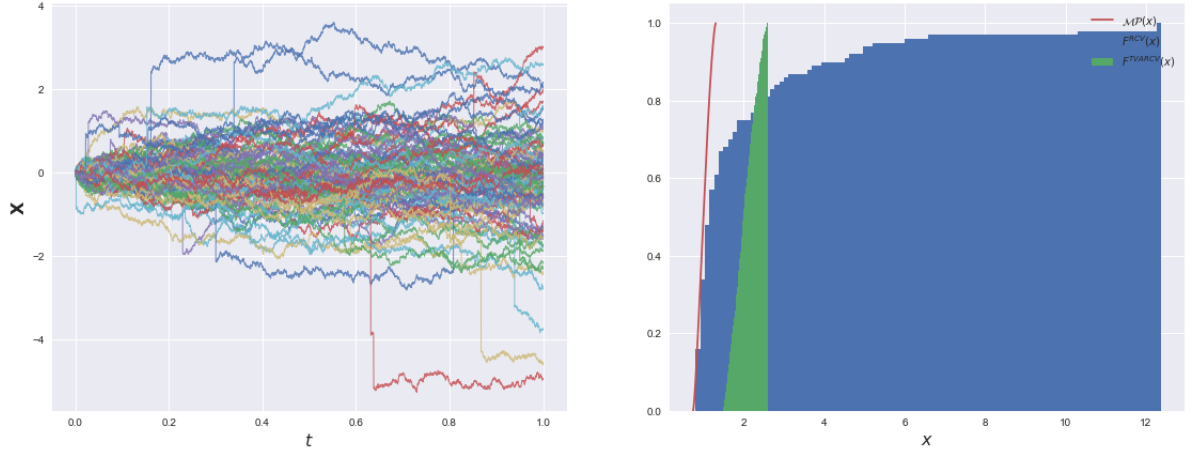


Figure 6.5:  $n = 5000$ ,  $p = 100$ ,  $\lambda = 1$ ,  $Y_1 \sim \mathcal{N}(0, 1)$   
 Jump-diffusion process with jumps, based on compound Poisson process.

case

$$\Sigma_p = \begin{pmatrix} \int_0^1 \cos(\pi t)^2 dt & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \int_0^1 \sin(\pi t)^2 dt \end{pmatrix} = \frac{1}{2} \mathbb{I}_{p \times p}.$$

WHAT DO WE SEE?

**Example 6.9.** Let us change slightly the structure of  $\Theta(t)$ :

$$\left( \Theta(t) \right)_{jj} = \sin(\pi t) \mathbf{1}_{\{j \leq p/4\}} + \cos(\pi t) \mathbf{1}_{\{j > p/4\}}, \quad 1 \leq j \leq p.$$

The ICV matrix  $\Sigma_p$  remains the same. But we can observe from Figure 6.9 that both estimators RCV and TVARCV do not converge to Marchenko-Pastur law anymore. Note, however, that if we fill diagonals  $\Theta(t)$  with only  $\cos(\pi t)$  or  $\sin(\pi t)$ , then  $\mathbf{X}(t)$  belongs to class  $\mathcal{C}$  and TVARCV (but not RCV) converges to  $\mathcal{MP}(y, \sigma^2)$ .

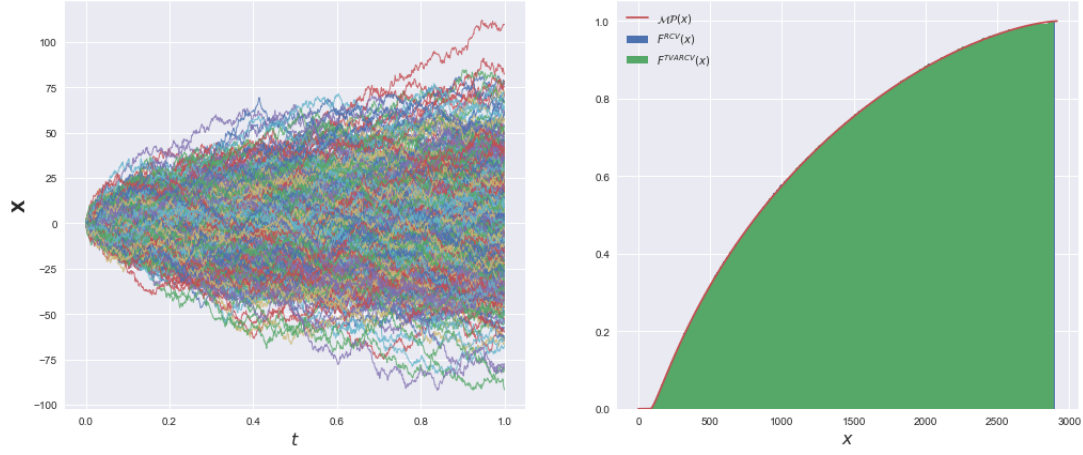


Figure 6.6:  $n = 1000$ ,  $p = 500$ ,  $\lambda = 10000$ ,  $Y_1 \sim \mathcal{N}(0, 1)$

We observe that jump-diffusion process converges to pure-diffusion process.

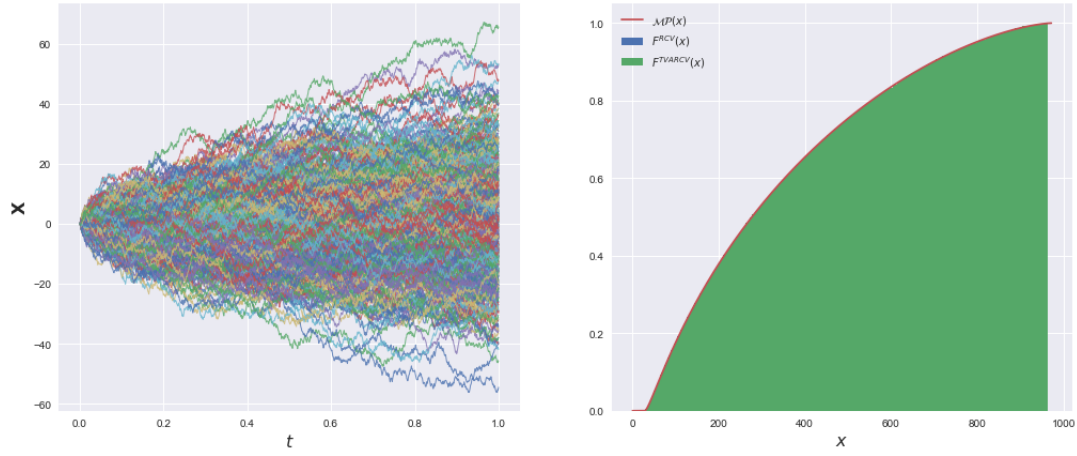


Figure 6.7:  $n = 1000$ ,  $p = 500$ ,  $\lambda = 10000$ ,  $Y_1 \sim \mathcal{U}(0, 1)$

Jump-diffusion process still converges to pure diffusion process due to Central Limit Theorem, as long as the variance of  $Y_i$  is finite.

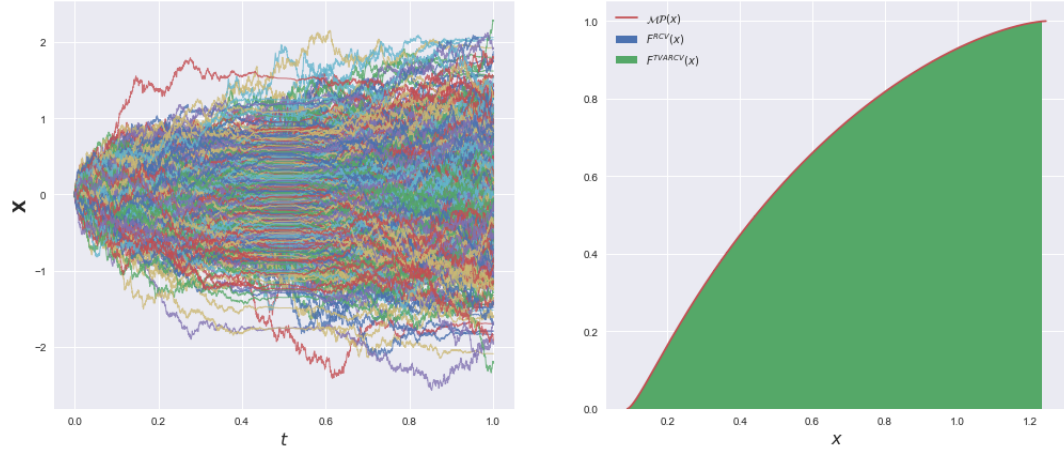


Figure 6.8:  $\mathbf{X}(t)$  from Example 6.8,  $n = 3000$ ,  $p = 1000$   
 Blue and green histograms coincide with each other and Marchenko-Pastur law.

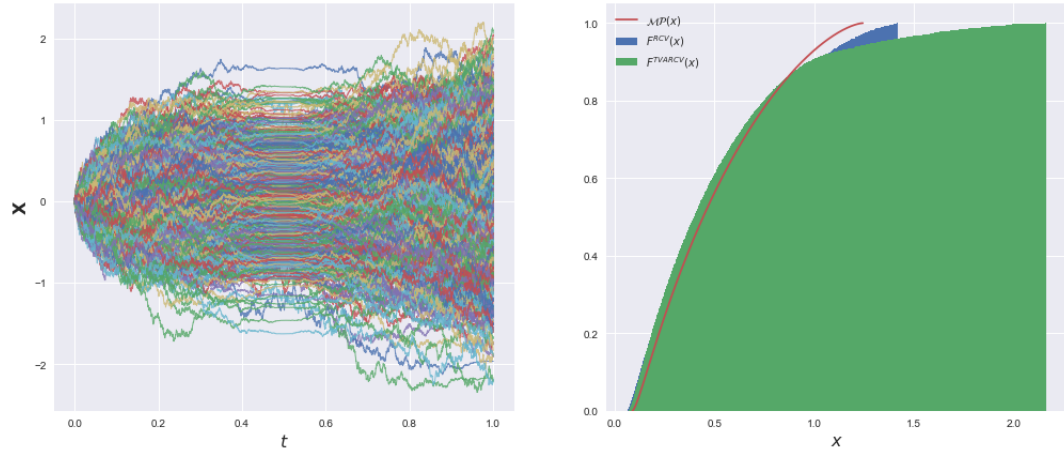


Figure 6.9:  $\mathbf{X}(t)$  from Example 6.9,  $n = 3000$ ,  $p = 1000$   
 TVARCV and RCV give different results, which are both distinct from Marchenko-Pastur law.

## Appendix A

# Distribution function for Marchenko-Pastur law

First let us consider the case  $0 < y < 1$ . Then the domain of the density  $f(x)$  is the interval  $[a, b]$  and hence  $F(x) = 0$  for  $x < a$  and  $F(x) = 1$  for  $x \geq b$ . Inside the interval the formula is

$$F(x) = \frac{1}{2\pi y} \left( \pi y + \sqrt{-x^2 - (-1+y)^2 + 2x(1+y)} - \right. \\ (1+y) \arctan \left( \frac{1-x+y}{\sqrt{4y - (1-x+y)^2}} \right) + \\ \left. (-1+y) \arctan \left( \frac{(-1+y)^2 - x(1+y)}{(-1+y)\sqrt{4y - (1-x+y)^2}} \right) \right).$$

For  $y \geq 1$  we have  $f(0) = (1 - 1/y) \cdot \delta_0(0)$ . Therefore,  $F(x) = 0$  for  $x < 0$ ,  $F(x) = 1 - 1/y$  for  $0 \leq x < a$  and  $F(x) = 1$  for  $x \geq b$ . For  $x \in [a, b)$  we have:

$$F(x) = 1 - \frac{1}{2\pi y} \left( \pi + \sqrt{4y - (1-x+y)^2} - \right. \\ (1+y) \arctan \left( \frac{1-x+y}{\sqrt{4y - (1-x+y)^2}} \right) + \\ \left. (-1+y) \arctan \left( \frac{-(-1+y)^2 + x(1+y)}{(-1+y)\sqrt{4y - (1-x+y)^2}} \right) \right).$$



## Appendix B

### Python code

---

```
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib
import numpy as np

# Marchenko–Pastur probability density function
def MPPdf(x, y, sigmaSq):
    if x == 0:
        if y > 1:
            return np.inf
        return 0
    sqrtY = np.sqrt(y)
    a = sigmaSq * (1 - sqrtY) * (1 - sqrtY)
    b = sigmaSq * (1 + sqrtY) * (1 + sqrtY)
    if x < a or x > b:
        return 0
    pdf = np.sqrt((b - x) * (x - a))
    pdf /= (y * x)
    pdf /= (2 * np.pi * sigmaSq)
    return pdf

# Marchenko–Pastur cumulative distribution function
def MPCdf(x, y, sigmaSq):
    # Marchenko–Pastur cdf for y > 1
    def MPCdfLargeRatio(x, y):
        y1 = 1 - x + y
        temp = np.sqrt(4 * y - y1 * y1)
        y1 /= temp
```

```

    y1 = (1 + y) * np.arctan(y1)
    y2 = x * (1 + y)
    y2 -= (y - 1) * (y - 1)
    y2 /= (temp * (y - 1))
    y2 = (y - 1) * np.arctan(y2)
    cdf = np.pi - temp + y1 + y2
    cdf /= (2 * np.pi * y)
    return 1 - cdf

# Marchenko-Pastur cdf for y <= 1
def MPCdfSmallRatio(x, y):
    y1 = 1 - x + y
    temp = np.sqrt(4 * y - y1 * y1)
    y1 /= temp
    y1 = (1 + y) * np.arctan(y1)
    if y == 1:
        y2 = 0
    else:
        y2 = x * (1 + y)
        y2 -= (y - 1) * (y - 1)
        y2 /= (temp * (1 - y))
        y2 = (y - 1) * np.arctan(y2)
    cdf = np.pi * y + temp - y1 + y2
    cdf /= (2 * np.pi * y)
    return cdf

x0 = x / sigmaSq
sqrtY = np.sqrt(y)
a = (1 - sqrtY) * (1 - sqrtY)
b = (1 + sqrtY) * (1 + sqrtY)
if x0 >= b:
    return 1
if x0 < 0:
    return 0
if y > 1:
    if x0 <= a:
        return 1 - 1 / y
    return MPCdfLargeRatio(x0, y)
if x0 <= a:
    return 0
return MPCdfSmallRatio(x0, y)

```



```

# get sorted eigenvalues
def get_sorted_eigvals(Sigma):
    return np.sort(np.linalg.eigvals(Sigma))

# get empirical spectral distribution
def ESD(x, sorted_eig_values):
    numerator = len(sorted_eig_values[sorted_eig_values <= x])
    denominator = len(sorted_eig_values)
    return numerator / denominator

# TVARCV
def TVARCV(dX, Sigma_RCV):
    (p, n) = dX.shape
    Sigma = np.zeros((p, p))
    for l in range(n):
        dX_l = np.matrix(dX[:, l]).transpose()
        norm = np.linalg.norm(dX_l, ord=2)
        Sigma += dX_l.dot(np.transpose(dX_l)) / (norm * norm)
    Sigma *= np.trace(Sigma_RCV) / n
    return Sigma

# RCV
def RCV(dX):
    (p, n) = dX.shape
    Sigma = np.zeros((p, p))
    for l in range(n):
        dX_l = np.matrix(dX[:, l]).transpose()
        Sigma += dX_l.dot(np.transpose(dX_l))
    return Sigma

# function to plot stochastic processes in Figure 1.1
def plot_exmp_1():
    n = 10000
    time = np.linspace(0, 1, n)
    mu = np.array([0, 0, 0])
    X = np.zeros((n, 3))
    dt = time[1] - time[0]
    sqrt_dt = np.sqrt(dt)
    dX = np.random.normal(0, sqrt_dt, 3)

```

```

X[1] = X[0] + dX
for i in range(1, n - 1):
    dW = np.random.normal(0, sqrt_dt, 3)
    Theta = np.identity(3)
    Theta[0][1] = Theta[1][0] = np.cos(np.pi * time[i])
    Theta[0][2] = Theta[2][0] = np.sin(np.pi * time[i])
    dX_l = np.dot(Theta, dW)
    dX_l += mu * dt
    X[i + 1] = X[i] + dX_l
    dX = np.c_[dX, dX_l]

plt.plot(time, X)
plt.legend(["$X_1(t)$", "$X_2(t)$", "$X_3(t)$"], fontsize = 16)
plt.xlabel('$t$', fontsize = 16)
plt.ylabel('$\mathbf{X}$', fontsize = 16)
plt.show()

# function to plot graphs of Marchenko–Pastur law (Figure 2.1)
def plot_MP_law
    y = [0.1, 0.25, 1, 4]
    sigma_sq = [1, 1, 1, 0.3]
    m = 300
    xpdf = np.linspace(1/m, 4, m)
    xcdf = np.linspace(-1/m, 4, m)
    cdf = np.zeros(m)
    pdf = np.zeros(m)

    plt.subplot(1, 2, 1)
    for ratio, vol in zip(y, sigma_sq):
        for i in range(m):
            pdf[i] = MPPdf(xpdf[i], ratio, vol)
        plt.plot(xpdf, pdf)
    plt.ylim(ymax=2.25)
    # plot infinity line for y > 1
    plt.plot((0, 0), (0, 3), color='C3')
    plt.legend(["$y=0.1, \sigma^2=1$", "$y=0.25, \sigma^2=1$", \
        "$y=1, \sigma^2=1$", "$y=4, \sigma^2=0.3$"])
    plt.xlabel('$x$', fontsize = 16)
    plt.ylabel('$f(x)$', fontsize = 16)

    plt.subplot(1, 2, 2)
    for ratio, vol in zip(y, sigma_sq):

```

```

        for i in range(m):
            cdf[i] = MPCdf(xcdf[i], ratio, vol)
        plt.plot(xcdf, cdf)
    plt.legend(["$y=0.1$, \sigma^2=1$", "$y=0.25$, \sigma^2=1$", \
        "$y=1$, \sigma^2=1$", "$y=4$, \sigma^2=0.3$"])
    plt.xlabel('$x$', fontsize = 16)
    plt.ylabel('$F(x)$', fontsize = 16)

plt.show()

# function that returns volatility as CIR process
# and corresponding integrated covariance
def get_CIR(t, alpha, beta, xi, x0):
    dt = t[1] - t[0] # we assume times are equally spaced
    exp_mbeta_dt = np.exp(-beta * dt)
    c = 4 * beta / (xi * xi * (1.0 - exp_mbeta_dt))
    degree = 4 * alpha * beta / (xi * xi)
    Y = x0 * np.ones(len(t))
    sigma_sq = alpha * alpha * dt
    for i in range(1, len(gamma)):
        k = c * Y[i - 1] * exp_mbeta_dt
        Y[i] = np.random.noncentral_chisquare(degree, k) / c
        sigma_sq += Y[i] * Y[i] * dt
    return Y, sigma_sq

# function that plots figures for different data
# parameter get_data is a function that accepts parameters n and p
# and returns observed data
def plot_figures(get_data, n, p):
    X, dX, time, sigma_sq = get_data(n, p)

    plt.subplot(1, 2, 1)
    plt.plot(time, X, alpha=0.8, linewidth=1)
    plt.xlabel('$t$', fontsize = 16)
    plt.ylabel('$\mathbf{X}$', fontsize = 16)

    plt.subplot(1, 2, 2)
    Sigma_RCV = RCV(dX)
    sorted_eigvals_RCV = get_sorted_eigvals(Sigma_RCV)
    plt.hist(sorted_eigvals_RCV, p, cumulative=True, normed=True)

```

```

sorted_eigvals_TVARCHV = get_sorted_eigvals(TVARCHV(dX, Sigma_RCV)
plt.hist(sorted_eigvals_TVARCHV, p, cumulative=True, normed=True)

cdf_len = 3000
cdf = np.zeros(cdf_len)
sqrt_ratio = np.sqrt(p / n)
a = 1 - sqrt_ratio
b = 1 + sqrt_ratio
a *= sigma_sq * a
b *= sigma_sq * b
x = np.linspace(a, b, cdf_len)
for i in range(cdf_len):
    cdf[i] = MPCdf(x[i], p / n, sigma_sq)
plt.plot(x, cdf)

plt.xlabel('$x$', fontsize = 16)
plt.legend(['$\mathcal{MP}(x)$', \
            '$F^{RCV}(x)$', '$F^{TVARCHV}(x)$'])
plt.show()

# function that returns data with non-constant volatility
def get_data_for_volatile_gamma(n, p):
    time = np.zeros(n + 1)
    X = np.zeros((n + 1, p))
    delta_time = 1/n
    dX = 0.03 * np.random.normal(0, np.sqrt(delta_time), p)
    X[1] = X[0] + dX
    time[1] += delta_time
    newSin = 0
    for i in range(1, n):
        delta_time = np.exp(np.mean(X[i, :])) / n
        time[i + 1] = time[i] + delta_time
        dW = np.random.normal(0, np.sqrt(delta_time), p)
        oldSin = newSin
        newSin = np.sin(2 * np.pi * (i + 1) / n)
        dSin = newSin - oldSin
        Gamma = 0.0009 + n * 0.0008 * dSin / (2 * np.pi)
        Gamma = np.sqrt(Gamma)
        dX_1 = Gamma * dW
        X[i + 1] = X[i] + dX_1
        dX = np.c_[dX, dX_1]
    return X, dX, time, 0.0009

```

*# function that returns data for non-equidistant observation times ,  
 # note that time[n-1] may be not equal to 1*

```
def get_data_poisson_times(n, p):
    time = np.zeros(n + 1)
    X = np.zeros((n + 1, p))
    rate = 1 / n
    delta_time = np.random.exponential(rate)
    dX = 0.03 * np.random.normal(0, np.sqrt(delta_time), p)
    X[1] = X[0] + dX
    time[1] += delta_time
    newSin = 0
    for i in range(1, n):
        delta_time = np.random.exponential(rate)
        time[i + 1] = min(time[i] + delta_time, 1)
        dW = np.random.normal(0, np.sqrt(delta_time), p)
        oldSin = newSin
        newSin = np.sin(2 * np.pi * time[i+1])
        dSin = newSin - oldSin
        Gamma = 0.0009 + 0.0008 * dSin / (2 * np.pi * delta_time)
        Gamma = np.sqrt(Gamma)
        dX_l = Gamma * dW
        X[i + 1] = X[i] + dX_l
        dX = np.c_[dX, dX_l]
    return X, dX, time, 0.0009
```

*# function that returns data for cos/sin case*

```
def get_data_for_cossin(n, p):
    time = np.zeros(n + 1)
    X = np.zeros((n + 1, p))
    delta_time = 1/n
    dW = np.random.normal(0, np.sqrt(delta_time), p)
    Theta = np.identity(p)
    for j in range(0, int(p/2)):
        Theta[j][j] = 0
    dX = np.dot(Theta, dW)
    X[1] = X[0] + dX
    time[1] += delta_time
    for i in range(1, n):
        time[i + 1] = time[i] + delta_time
        dW = np.random.normal(0, np.sqrt(delta_time), p)
```

```

        nsin = np.sin(np.pi * (i + 1) / n)
        ncos = np.cos(np.pi * (i + 1) / n)
        Theta = np.identity(p) * ncos
        for j in range(0, int(p/2)):
            Theta[j][j] = nsin
        dX_l = np.dot(Theta, dW)
        X[i + 1] = X[i] + dX_l
        dX = np.c_[dX, dX_l]
    return X, dX, time, 0.5

# function that simulates jump-diffusion process
# with volatility 1, compound Poisson process with rate 1
# and jumps with standard normal distribution
def get_data_jumps(n, p):
    X = np.zeros((n + 1, p))
    time = np.linspace(0, 1, n + 1)
    dt = time[1] - time[0]
    gamma = 1
    sqrt_dt = np.sqrt(dt)
    dX = gamma * np.random.normal(0, sqrt_dt, p)
    X[1] = X[0] + dX
    rate = 1
    jump_time = np.random.exponential(1 / rate, p)
    for i in range(1, n):
        dX_l = np.random.normal(0, sqrt_dt, p)
        dX_l *= gamma
        for j in range(p):
            if jump_time[j] <= i / n:
                dX_l[j] += np.random.normal(0, 1)
                jump_time[j] += np.random.exponential(1 / rate)
        X[i + 1] = X[i] + dX_l
        dX = np.c_[dX, dX_l]
    return X, dX, time, gamma

# function that simulates data with gamma as CIR process
# with parameters alpha = 5, beta = 1, xi = 2 and gamma(0) = 3
def get_data_stochastic_gamma(n, p):
    sqrt_p = np.sqrt(p)
    X = np.zeros((n + 1, p))
    time = np.linspace(0, 1, n + 1)
    dt = time[1] - time[0]

```

```

alpha = 5
beta = 1
xi = 2
gamma0 = 3
gamma_sq = gamma0 * gamma0
gamma = gamma0
sigma_sq = alpha * alpha * dt
volatility = gamma0 * np.ones(len(time))
sqrt_dt = np.sqrt(dt)
dX = gamma * np.random.normal(0, sqrt_dt, p)
X[1] = X[0] + dX
for i in range(1, n):
    dX_l = np.random.normal(0, sqrt_dt, p)
    dW = np.sum(dX_l) / sqrt_p
    dg = beta * (alpha - gamma_sq) * dt + xi * gamma * dW
    gamma_sq += dg
    gamma = np.sqrt(gamma_sq)
    sigma_sq += gamma_sq * dt
    volatility[i] = gamma
    dX_l *= gamma
    X[i + 1] = X[i] + dX_l
    dX = np.c_[dX, dX_l]
return X, dX, time, sigma_sq

```

---





# Bibliography

- [1] Shreve Steven E. Stochastic calculus for finance II. Continuous-time models. Pittsburg: Springer, 2004.
- [2] V.A. Marchenko, L.A. Pastur. Distribution of eigenvalues in certain sets of random matrices. 1967.
- [3] J.W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. J. Multivariate analysis, 1995.
- [4] Z.D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. Statistica Sinica, 1999.
- [5] Adlam Ben. The Local Marchenko-Pastur Law for Sparse Covariance Matrices. 2013.
- [6] R.A. Horn, C.R. Johnson. Matrix analysis. 1990.
- [7] J.S. Geronimo, T.P. Hill. Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform.
- [8] Revuz D., Yor M. Continuous martingales and Brownian motion. Berlin: Springer, 1998.
- [9] Z.D. Bai, W. Silverstein J. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. 1998.
- [10] Zheng Xinghua, Li Yingying. On the estimation of integrated covariance matrices of high dimensional diffusion processes. Hong Kong: Institute of Mathematical Statistics, 2011.
- [11] Broadie Mark, Kaya Özgür. Exact simulation of stochastic volatility and other affine jump diffusion processes. 2004.

## **Erklärung**

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Weiterhin versichere ich, dass diese Arbeit noch nicht als Abschlussarbeit an anderer Stelle vorgelegen hat.