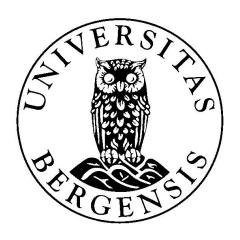# Stochastic unit-root processes

Master thesis in Statistics

Sigve Brix Dyrhovden

Supervisor

Hans Arnfinn Karlsen

Department of Mathematics
University of Bergen
June 2016

Contents

1 Introduction

This thesis primary focus is stochastic unit-root processes. We also study standard unit-root processes and cointegration. When the root of the characteristic equation of the AR(1) model is 1 or -1, the process is said to have a unit root. The best known example of unit-root processes is the random-walk model. The term random walk was introduced by Karl Pearson in 1905. He used the model to describe a mosquito infestation in a forest. The theory of random walk was also developed by Louis Bachelier around the same time. He proposed the random-walk model as the fundamental model for financial time series and in this way he was years ahead of his time, see [Rycroft, page 1]. The modern literature on unit roots dates back to White [1958]. However, by the early 1980s, only a handful papers had been written about unit roots, mostly by Professor Wayne Fuller and his coauthors. But the last two decades have seen significant developmets in the literature on unit roots, see [Choi, 2015, page 1] It was Nelson and Plosser [1982] who brought the issue of nonstationarity to the forefront of economic research, see [Choi, 2015, page 5]. The concept of cointegration was introduced by Engle and Granger [1987]. In cointegration analysis, the first step is to test whether the variables of interest have a unit root, and thus, unit root tests are used in application of cointegration, see [Choi, 2015, page 6].

In practice there are many types of processes that have near unit roots and are very difficult to distinguish from standard unit-root processes, given a finite sample. In Granger and Swanson [1997] a particular class of such processes are introduced. They have a root that is stochastic and varying around unity. In our thesis we study the properties of such processes, also called **STUR** processes. They are divided into two classes, **STURA** and **STURB**. Some important results are listed below:

1. **STURA** is stationary.

2. Testing done in Granger and Swanson [1997] support the proposition that **STUR** processes are generic unit root processes.

3. Testing done in Granger and Swanson [1997] shows that the **STUR** model performs very well in multi-step ahead forecasting.

2 Preliminaries

In this introductory chapter we will state definitions and well-known results which may be found in Brockwell and Davis [2002] and Tsay [2010]. We begin with defining time series and time series models as in [Brockwell and Davis, 2002, page 1] and [Brockwell and Davis, 2002, page 7].

Definition 2.1. *A **time series** is a set of observations $\{x_t\}$, each one being recorded at a spesific time t.*

Definition 2.2. *A **time series model** for the observed data $\{x_t\}$ is a specification of the joint distribution (or possibly only the means and covariances) of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.*

2.1 Some simple time series models

In this section we will introduce some simple time series models, see [Brockwell and Davis, 2002, page 8-9] and [Brockwell and Davis, 2002, page 16-17].

Example 2.3. We are given a sequence of random variables $X_1, X_2, \ldots$ If the variables are independent and identically distributed (iid) random variables with zero mean and variance σ^2 we refer to such a sequence as **iid noise**. We use the notation

$$\{X_t\} \sim \text{IID}(0, \sigma^2).$$

Example 2.4. If we have a sequence of uncorrelated random variables $X_1, X_2, \ldots$, each with zero mean and variance σ^2, we refer to such a sequence as **white noise**. We use the notation

$$\{X_t\} \sim \text{WN}(0, \sigma^2).$$

Example 2.5. We are given a sequence of iid random variables $X_1, X_2, \ldots$, each with zero mean and variance σ^2. We define $S_0 = 0$ and

$$S_t = X_1 + X_2 + \cdots + X_t$$

for $t = 1, 2, \ldots$. Then $\{S_t\}$ is called a **random walk**.

2.2   Stationary processes

In this section we will introduce stationary processes. We begin by defining the **mean function** of $\{X_t\}$ and the **covariance function** of $\{X_t\}$, see [Brockwell and Davis, 2002, page 15].

Definition 2.6. Let $\{X_t\}$ be a time series with $\mathrm{E}(X_t^2) < \infty$. The **mean function** of $\{X_t\}$ is

$$\mu_X(t) = \mathrm{E}(X_t).$$

Definition 2.7. Let $\{X_t\}$ be a time series with $\mathrm{E}(X_t^2) < \infty$. The **covariance function** of $\{X_t\}$ is

$$\gamma_X(r,s) = \mathrm{Cov}(X_r, X_s) = \mathrm{E}[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

for all integers r and s.

We now define the terms **weakly stationary** and **stationary**, see [Brockwell and Davis, 2002, page 15].

Definition 2.8. Let $\{X_t\}$ be a time series with $\mathrm{E}(X_t^2) < \infty$. Then $\{X_t\}$ is **weakly stationary** if (i) $\mu_X(t)$ is independent of t, and (ii) $\gamma_X(t+h, t)$ is independent of t for each h.

Definition 2.9. Let $\{X_t\}$ be a time series. Then $\{X_t\}$ is **stationary** if $(X_1, \ldots, X_n)$ and $(X_{1+h}, \ldots, X_{n+h})$ have the same joint distributions for all integers h and $n > 0$.

If $\{X_t\}$ is stationary and $\mathrm{E}(X_t^2) < \infty$ for all t, then $\{X_t\}$ is also weakly stationary. Since $\gamma_X(t+h, t)$ is independent of t for each h whenever $\{X_t\}$ is weakly stationary we introduce the following function:

$$\gamma_X(h) \overset{def}{=} \gamma_X(h, 0) = \gamma_X(t+h, t) = \mathrm{Cov}(X_{t+h}, X_t).$$

We will now formalize this and define the **autocovariance function** (ACVF) and the **autocorrelation function** (ACF) of a weakly stationary series as in [Brockwell and Davis, 2002, page 16].

Definition 2.10. Let $\{X_t\}$ be a weakly stationary time series. The **autocovariance function** (ACVF) of $\{X_t\}$ at lag h is

$$\gamma_X(h) \overset{def}{=} \gamma_X(h, 0) = \gamma_X(t+h, t) = \mathrm{Cov}(X_{t+h}, X_t).$$

Definition 2.11. *Let $\{X_t\}$ be a weakly stationary time series. The **autocorrelation function** (ACF) of $\{X_t\}$ at lag h is*

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)} = \text{Cor}(X_{t+h}, X_t).$$

We now look at Example 2.3 and Example 2.4 again. We see immediately that iid noise is stationary and iid noise with finite second moment is weakly stationary. White noise may not be stationary, but white noise with finite second moment is weakly stationary. A multivariate Gaussian distribution is fully characterized by its first two moments. A Gaussian white noise series is therefore stationary. However random walk as defined in Example 2.5 is not stationary. With the notation as in the example we see that $\text{E}(S_t) = 0$ for all t, but

$$\gamma_S(t + h, t) = \text{Cov}(S_{t+h}, S_t) = \text{Cov}(S_t + X_{t+1} + \cdots + X_{t+h}, S_t) = \text{Cov}(S_t, S_t) = t\sigma^2,$$

which depends on t.

2.3   Linear time series

In this section we look closer at **linear time series**, see [Brockwell and Davis, 2002, page 51-53] and [Tsay, 2010, page 36-37]. We begin by defining linear time series as in [Tsay, 2010, page 36].

Definition 2.12. *The time series $\{X_t\}$ is said to be **linear** if it can be written as*

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i Z_{t-i}, \tag{1}$$

*where μ is the mean of X_t, $\psi_0 = 1$, and $\{Z_t\}$ is a white noise series with zero mean and variance σ^2. The time series is also called a **moving-average** or $\boldsymbol{MA(\infty)}$.*

If $\{X_t\}$ is weakly stationary, we see that $\text{Var}(X_t) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2$, where σ^2 is the variance of Z_t. The variance of X_t is finite when X_t is weakly stationary. This implies that $\{\psi_i^2\}$ must be a convergent sequence, i.e. $\psi_i^2 \to 0$ as $i \to \infty$.

The ACVF of $\{X_t\}$ at lag h is

$$\gamma_X(h) = \text{Cov}(X_t, X_{t-h}) = \text{E}\left[\left(\sum_{i=0}^{\infty} \psi_i Z_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j Z_{t-h-j}\right)\right] = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}.$$

Hence the ACF of $\{X_t\}$ at lag h is

$$\rho_X(h) = \frac{\sum\limits_{i=0}^{\infty} \psi_i \psi_{i+h}}{1 + \sum\limits_{i=1}^{\infty} \psi_i^2},$$

see [Tsay, 2010, page 37, equation (2.6) and (2.7)].

2.4   AR models

In this section we study autoregressive (AR) models. We begin by looking closer at an AR model of order 1 or simply an AR(1) model, see [Brockwell and Davis, 2002, page 53-55] and [Tsay, 2010, page 37-40].

Definition 2.13. *The general representation of an autoregressive (AR) model of order* 1 *or simply an **AR(**1**)** model is*

$$X_t = \phi_0 + \phi_1 X_{t-1} + Z_t, \tag{2}$$

where $\{Z_t\}$ is assumed to be a white noise series with zero mean and variance σ^2.

We proceed as in [Tsay, 2010, page 37-39]. By assuming that the series defined by (2) is weakly stationary we have $\mathrm{E}(X_t) = \mu$, $\mathrm{Var}(X_t) = \gamma_X(0)$, and $\mathrm{Cov}(X_t, X_{t-h}) = \gamma_X(h)$, where μ and $\gamma_X(0)$ are constants and $\gamma_X(h)$ is a function of h, not t. From (2) we have

$$\mu = \phi_0 + \phi_1 \mathrm{E}(X_{t-1}).$$

Since the weakly stationary assumption gives $\mathrm{E}(X_{t-1}) = \mu$, we obtain

$$\mu = \frac{\phi_0}{1 - \phi_1}.$$

We see that the mean of $\{X_t\}$ exists if $\phi_1 \neq 1$. Using $\phi_0 = (1 - \phi_1)\mu$, (2) can be rewritten as

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + Z_t. \tag{3}$$

By repeated substitutions (3) implies that

$$X_t - \mu = \sum_{i=0}^{\infty} \phi_1^i Z_{t-i}. \tag{4}$$

We observe that (4) expresses an AR(1) model as a linear time series, where $\psi_i = \phi_1^i$ in (1). Since $X_{t-1} - \mu$ is a linear function of Z_{t-i} for $i \geq 1$, we have that $\text{Cov}(X_{t-1}, Z_t) = \text{E}[(X_{t-1} - \mu)Z_t] = 0$. If we are taking the square of (3) and then the expectation of the result we obtain

$$\text{Var}(X_t) = \phi_1^2 \text{Var}(X_{t-1}) + \sigma^2.$$

Under the weakly stationary condition, $\text{Var}(X_t) = \text{Var}(X_{t-1})$, hence

$$\text{Var}(X_t) = \frac{\sigma^2}{1 - \phi_1^2},$$

provided that $\phi_1^2 < 1$. Hence, the weakly stationarity of an AR(1) model implies that $|\phi_1| < 1$. It is easy to verify that if $|\phi_1| < 1$ the AR(1) model is weakly stationary, see [Tsay, 2010, page 39]. We summarize this in the following result, see [Tsay, 2010, page 39].

Theorem 2.14. *The necessary and sufficient condition for the AR(1) model in (2) to be weakly stationary is $|\phi_1| < 1$.*

We will now obtain the ACF for a weakly stationary AR(1) model in (2), see [Tsay, 2010, page 39-40]. If we multiply (3) by Z_t and then take the expectation of the result we obtain

$$\text{E}[Z_t(X_t - \mu)] = \phi_1 \text{E}[Z_t(X_{t-1} - \mu)] + \text{E}(Z_t^2) = \text{E}(Z_t^2) = \sigma^2.$$

If we multiply (3) by $X_{t-h} - \mu$ and then take the expectation of the result we obtain

$$\gamma_X(h) = \phi_1 \gamma_X(h - 1),$$

for $h > 0$. The ACF of $\{X_t\}$ must clearly satisfy

$$\rho_X(h) = \phi_1 \rho_X(h - 1),$$

for $h > 0$. This lead us to the next result, see [Tsay, 2010, page 40].

Theorem 2.15. *The autocorrelation function or the ACF for a weakly stationary AR(1) model in (2) is*

$$\rho_X(h) = \phi_1^h.$$

We observe that the ACF of the series starts at the value $\rho_X(0) = 1$ and decays exponentially with rate ϕ_1.

A generalization of the AR(1) model above is the AR(p) model given by

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \tag{5}$$

where p is a nonnegative integer and $\{Z_t\}$ is assumed to be a white noise series with zero mean and variance σ^2. When the series is weakly stationary it is easy to verify that the mean is given by

$$\mathrm{E}(X_t) = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p},$$

provided that the denominator is not zero, see [Tsay, 2010, page 46].

The **backward shift operator** is defined by

$$BX_t = X_{t-1}.$$

Powers of the operator B is defined by

$$B^j(X_t) = X_{t-j}.$$

Another way to express (5) is therefore

$$\phi(B)X_t = \phi_0 + Z_t, \tag{6}$$

where

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p. \tag{7}$$

If we set the last expression equal to zero we have the **characteristic equation** of the AR(p) model above. The next theorem is an important result given in [Tsay, 2010, page 46].

Theorem 2.16. *If all the solutions of the characteristic equation of the AR(p) model in (5) are greater than 1 in modulus, then the series $\{X_t\}$ is stationary.*

2.5   MA models and ARMA models

In this section we look at moving-average (MA) models and autoregressive moving-average (ARMA) models. We begin by defining an MA(q) model:

Definition 2.17. *The general representation of an moving-average (MA) model of order q or simply an **MA(q)** model is*

$$X_t = \phi_0 + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{8}$$

where $\{Z_t\}$ is assumed to be a white noise series with zero mean and variance σ^2.

Moving-average models are weakly stationary because they are finite linear combinations of a white noise sequence. We easily see that for the MA(q) model in (8) we have

$$\mathrm{E}(X_t) = \phi_0$$

and

$$\mathrm{Var}(X_t) = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma^2.$$

It can be shown that for the MA(q) model in (8) the ACF of $\{X_t\}$ at lag q is not zero, but $\rho_X(h) = 0$ for $h > q$, see [Tsay, 2010, page 59].

We will now define the general ARMA(p,q) model:

Definition 2.18. *The general representation of an autoregressive moving-average (ARMA) model of order (p,q) or simply an **(ARMA)(p,q)** model is*

$$X_t = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t + \sum_{i=1}^{q} \theta_i Z_{t-i}, \tag{9}$$

where $\{Z_t\}$ is assumed to be a white noise series with zero mean and variance σ^2, and p and q are nonnegative integers.

We see that the AR(p) and MA(q) models are special cases of the ARMA(p,q) model. If we use the backward shift operator we can write the model in (9) as

$$(1 - \phi_1 B - \cdots - \phi_p B^p)X_t = \phi_0 + (1 + \theta_1 B + \cdots + \theta_q B^q)Z_t.$$

The polynomial $1 - \phi_1 B - \cdots - \phi_p B^p$ is the AR polynomial of the model and the polynomial $1 + \theta_1 B + \cdots + \theta_q B^q$ is the MA polynomial of the model. We require that there are no common factors between the AR and MA polynomials - otherwise the order of the model can be reduced. The AR polynomial introduces the characteristic equation of an ARMA model. If all the solutions of this equation are greater than 1 in modulus, then the series is weakly stationary, see [Tsay, 2010, page 66]. In this case, the mean of the series is given by

$$\mathrm{E}(X_t) = \frac{\phi_0}{(1 - \phi_1 - \cdots - \phi_p)}.$$

2.6   Forecasting

In this section we look at forecasting, and we consider the $\mathrm{AR}(p)$ model in (5), as our example, see [Tsay, 2010, page 54-57]. We suppose that we are at time index h and are interested in forecasting X_{h+l}, where $l \geq 1$. Let $\mathcal{F}_h$ be the collection of information available at the forecast origin h. Let $\widehat{X}_h(l)$ be the forecast of X_{h+l} using the minimum squared error loss function. That is, the forecast $\widehat{X}_h(l)$ is chosen such that

$$\mathrm{E}\{[X_{h+l} - \widehat{X}_h(l)]^2 | \mathcal{F}_h\} \leq \min_g \mathrm{E}[(X_{h+l} - g)^2 | \mathcal{F}_h],$$

where g is a function of the information available at time h (inclusive), that is, a function of $\mathcal{F}_h$. We will refer to $\widehat{X}_h(l)$ as the l-step ahead forecast of X_t at the forecast origin h, see [Tsay, 2010, page 54].

We begin by looking at **1-Step-Ahead Forecast**.

For the $\mathrm{AR}(p)$ model in (5), we have

$$X_{h+1} = \phi_0 + \phi_1 X_h + \cdots + \phi_p X_{h+1-p} + Z_{h+1}.$$

Under the minimum squared error loss function we have

$$\widehat{X}_h(1) = \mathrm{E}(X_{h+1} | \mathcal{F}_h) = \phi_0 + \sum_{i=1}^{p} \phi_i X_{h+1-i},$$

and the associated forecast error is

$$e_h(1) = X_{h+1} - \widehat{X}_h(1) = Z_{h+1}.$$

The variance of the 1-step-ahead forecast error is therefore σ^2.

We now look at **Multistep-Ahead Forecast**.

For the AR(p) model in (5), we have

$$X_{h+l} = \phi_0 + \phi_1 X_{h+l-1} + \cdots + \phi_p X_{h+l-p} + Z_{h+l}.$$

Under the minimum squared error loss function we have

$$\widehat{X}_h(l) = \mathrm{E}(X_{h+l}|\mathcal{F}_h) = \phi_0 + \sum_{i=1}^{p} \phi_i \widehat{X}_h(l-i),$$

where it is understood that $\widehat{X}_h(i) = X_{h+i}$ if $i \leq 0$. The l-step-ahead forecast error is $e_h(l) = X_{h+l} - \widehat{X}_h(l)$. It can be shown that for a stationary AR(p) model $\widehat{X}_h(l)$ converges to $\mathrm{E}(X_t)$ as $l \to \infty$, see [Tsay, 2010, page 56]. For such a series we therefore have that its long-term point forecast approaches its unconditional mean. We refer to this property as *mean reversion*.

2.7 Spectral Densities

In this section we define the **spectral density** of $\{X_t\}$. We look at some of its properties and give two simple examples, see [Brockwell and Davis, 2002, page 112-119].

Definition 2.19. *The spectral density of $\{X_t\}$ is the function $f(\cdot)$ defined by*

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h), \quad -\infty < \lambda < \infty.$$

Since cos and sin have period 2π, so also does f, and it suffices to look at f on the interval $(-\pi, \pi]$. Moreover it can be shown that f is **even** and **nonnegative**, see [Brockwell and Davis, 2002, page 112-113]. We look at to examples:

Example 2.20. If $\{X_t\} \sim \mathrm{WN}(0, \sigma^2)$, then $\gamma(0) = \sigma^2$ and $\gamma(h) = 0$ for all $|h| > 0$. Hence the process has a flat spectral density given by

$$f(\lambda) = \frac{\sigma^2}{2\pi}, \quad -\pi \leq \lambda \leq \pi.$$

We see that each frequency in the spectrum contributes equally to the variance of the process and for a good reason the process is called **white noise**.

Example 2.21. Consider the AR(1) process given in (2) with $\phi_0 = 0$. From the definition above we have:

$$
\begin{aligned}
f(\lambda) &= \frac{\sigma^2}{2\pi(1-\phi_1^2)} \left(1 + \sum_{h=1}^{\infty} \phi_1^h (e^{-ih\lambda} + e^{ih\lambda}) \right) \\
&= \frac{\sigma^2}{2\pi(1-\phi_1^2)} \left(1 + \frac{\phi_1 e^{i\lambda}}{1-\phi_1 e^{i\lambda}} + \frac{\phi_1 e^{-i\lambda}}{1-\phi_1 e^{-i\lambda}} \right) \\
&= \frac{\sigma^2}{2\pi(1-\phi_1^2)} \left(\frac{1-\phi_1^2}{(1-\phi_1 e^{i\lambda})(1-\phi_1 e^{-i\lambda})} \right) \\
&= \frac{\sigma^2}{2\pi} \left(1 - \phi_1 e^{-i\lambda} - \phi_1 e^{i\lambda} + \phi_1^2 \right)^{-1} \\
&= \frac{\sigma^2}{2\pi} \left(1 - \phi_1 [\cos(-\lambda) + i\sin(-\lambda) + \cos(\lambda) + i\sin(\lambda)] + \phi_1^2 \right)^{-1} \\
&= \frac{\sigma^2}{2\pi} (1 - 2\phi_1 \cos\lambda + \phi_1^2)^{-1}.
\end{aligned}
$$

3 Technical Tools

In this chapter we will introduce some technical tools that have been used in the literature on unit roots, see [Choi, 2015, page 12-14]. We begin with Brownian motion and stochastic integrals.

3.1 Brownian motion and Stochastic Integrals

In this section we introduce Brownian motion and stochastic integrals. We get our information from [Bjork, 2009, chapter 4]. The following definition is found in [Bjork, 2009, page 40].

Definition 3.1. *A stochastic process W is called Brownian motion if the following conditions hold.*

1. *$W(0) = 0$.*

2. *The process W has independent increments.*

3. *For $s < t$ the stochastic variable $W(t) - W(s)$ has the Gaussian distribution with zero mean and variance $t - s$.*

4. *W has continuous trajectories.*

We now take a closer look at Brownian motion, see [Bjork, 2009, page 50-51]. One can show that, with probability 1, the trajectory of the process will be a continuous function of time which is nondifferentiable at every point. This lack of smoothness gives rise to an odd property of the quadratic variation of the process which we now define. Fix a point in time t and subdivide the interval $[0, t]$ into n equally large subintervalls of the form $[k\frac{t}{n}, (k+1)\frac{t}{n}]$, where $k = 0, 1, \ldots, n - 1$. Given this subdivision, we now define the quadratic variation of the process by S_n, i.e.

$$S_n = \sum_{i=1}^{n} \left[W\left(i\frac{t}{n} \right) - W\left((i-1)\frac{t}{n} \right) \right]^2,$$

and we want to see what happens to S_n as the subdivision becomes finer, i.e. as $n \to \infty$. It can be shown (as in [Bjork, 2009, page 51]) that $\mathrm{E}[S_n] = t$ and $\mathrm{Var}[S_n] = \frac{2t^2}{n}$. Hence $\mathrm{Var}[S_n] \to 0$ as $n \to \infty$ and we see that S_n tends to the **deterministic** limit t. This motivates us to write

$$\int_0^t [dW]^2 = t,$$

or, equivalently,

$$[dW]^2 = dt.$$

We now want to construct the stochastic integral. Let X be any given stochastic process. We need to define "the information generated by X" as time goes by, see [Bjork, 2009, page 43].

Definition 3.2. *The symbol $\mathcal{F}_t^X$ denotes "the information generated by X on the interval $[0,t]$". If the value of a given stochastic variable Z can be completely determined given observations of the trajectory $X(s)$, where $0 \le s \le t$, we write this as*

$$Z \in \mathcal{F}_t^X.$$

If Y is a stochastic process such that we have

$$Y(t) \in \mathcal{F}_t^X$$

*for all $t \ge 0$ then we say that Y is **adapted** to the **filtration** $\{\mathcal{F}_t^X\}_{t\ge 0}$.*

Let W be Brownian motion and we also consider as given another stochastic process g. The following construction of the stochastic integral is found in [Bjork, 2009, page 44-45]. First we need some integrability conditions on g in order to guarantee the existence of the stochastic integral. The class $\mathcal{L}^2$ turns out to be natural.

Definition 3.3. *We say that the process g belongs to the class $\mathcal{L}^2[a,b]$ if the following conditions are satisfied.*

1. $\int_a^b \mathrm{E}[g^2(s)]ds < \infty$.

2. *The process g is adapted to the $\mathcal{F}_t^W$-filtration.*

We say that the process g belongs to the class $\mathcal{L}^2$ if $g \in \mathcal{L}^2[0,t]$ for all $t > 0$.

Our goal is to define the stochastic integral $\int_a^b g(s)dW(s)$ for a process $g \in \mathcal{L}^2[a,b]$. This is now carried out in two steps. In the first step we assume that $g \in \mathcal{L}^2[a,b]$ is **simple**. That is, we assume that there exist deterministic points in time $a = t_0 < t_1 < \cdots < t_n = b$, such that $g(s) = g(t_k)$ for $s \in [t_k, t_{k+1})$. We define the stochastic integral in this case by the following formula.

$$\int_a^b g(s)dW(s) = \sum_{k=0}^{n-1} g(t_k)[W(t_{k+1}) - W(t_k)].$$

In the second step we assume that $g \in \mathcal{L}^2[a,b]$ is a general process (not necessarily simple). We proceed in three steps.

1. Approximate g with a sequence $\{g_n\}$ of simple processes such that

$$\int_a^b \mathrm{E}\left[(g_n(s) - g(s))^2\right] ds \to 0.$$

16

2. For each n the integral $\int_a^b g_n(s)dW(s)$ is a well defined stochastic variable Z_n, and it is possible to prove that there exists a stochastic variable Z such that $Z_n \to Z$ (in $\mathcal{L}^2$) as $n \to \infty$.

3. We now define the stochastic integral by

$$\int_a^b g(s)dW(s) = \lim_{n\to\infty} \int_a^b g_n(s)dW(s).$$

We end this section with the main result in the theory of stochastic calculus - the Itô formula, see [Bjork, 2009, page 51].

Theorem 3.4. *Assume that the process X has a stochastic differential given by*

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

where μ and σ are adapted processes, and let f be a $C^{1,2}$-function. Define the process Z by $Z(t) = f(t, X(t))$. Then Z has a stochastic differential given by

$$df(t, X(t)) = \left(\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}\right)dt + \sigma\frac{\partial f}{\partial x}dW(t).$$

3.2 Continuous-Mapping Theorem

The following result is for instance found in [Choi, 2015, page 13]. Suppose that $X_T \Rightarrow X$ as $T \to \infty$, where X_T is a sequence of random vectors and X a random vector (here $\Rightarrow$ denotes weak convergence, i.e., convergence in distribution). The continuous mapping theorem states that $g(X_T) \Rightarrow g(X)$ as $T \to \infty$ where the function $g(\cdot)$ is continuous with probability one.

3.3 Functional Central Limit Theorem

For this theorem we proceed as in [Tsay, page 1-2]. We are given a weakly stationary time series $\{Y_t\}$, with certain properties to be defined shortly. Define $S_0 = 0$ and

$$S_t = Y_1 + Y_2 + \cdots + Y_t$$

for $t = 1, 2, \dots$. The average variance of S_T is given by

$$\tau^2 = \lim_{T\to\infty} \mathrm{E}(T^{-1}S_T^2).$$

Define the function

$$X_T(r) = \frac{1}{\sqrt{T}\tau}S_{[Tr]}, \ 0 \le r \le 1,$$

where $[Tr]$ denotes the integer part of Tr. The properties we need for our result is listed below.

1. $E(Y_t) = 0$ for all t.

2. $\sup_t E|Y_t|^\beta < \infty$ for some $\beta > 2$.

3. The average variance τ^2 defined above exists and is positive.

4. $\{Y_t\}$ is strongly mixing, i.e. the serial dependence between Y_t and Y_{t-h} approaches zero as h increases.

Theorem 3.5. *If $\{Y_t\}$ satisfies the four properties listed above, then $X_T(r) \Rightarrow W(r)$, where $W(r)$ is a standard Brownian motion for $r \in [0, 1]$ and $\Rightarrow$ denotes weak convergence, i.e. convergence in distribution.*

4 Unit-root

In this chapter we will introduce unit-root nonstationary time series.

4.1 Random Walk

The best known example of unit-root nonstationary time series is the random-walk model, see [Tsay, 2010, page 72]. We have already seen the random-walk model as our third example of simple time series models in the first chapter, see Example 2.5.

Definition 4.1. *A time series $\{X_t\}$ is a random walk if it satisfies*

$$X_t = X_{t-1} + Z_t, \ t = 1, 2, \dots, \tag{10}$$

where X_0 is a fixed real number, and $\{Z_t\}$ is assumed to be a white noise series with zero mean and variance σ^2.

This is an AR(1) model with $\phi_0 = 0$ and $\phi_1 = 1$ in the notation of Definition 2.13. The characteristic equation of the model is

$$1 - z = 0.$$

The root of this equation is 1 and the process is said to have a unit root, see [Choi, 2015, page 4]. From Theorem 2.14 the necessary and sufficient condition for the AR(1) model in (2) to be weakly stationary is $|\phi_1| < 1$. In our random-walk model we have $\phi_1 = 1$ and the process is nonstationary. We will now look closer at some of the properties of the random-walk model. We will also look closer at another important model - random walk with drift. We proceed as in [Tsay, 2010, page 72-74].

We begin by looking at the 1-step-ahead forecast of model (10) at the forecast origin h. It is given by

$$\widehat{X}_h(1) = \mathrm{E}(X_{h+1}|\mathcal{F}_h) = X_h.$$

It is easy to verify that for any forecast horizon $l > 0$, we have

$$\widehat{X}_h(l) = X_h,$$

see [Tsay, 2010, page 72]. Thus, for all forecast horizons, point forecasts of a random-walk model are simply the value of the series at the forecast origin. Hence, the process is **not mean reverting**. We can write (10) in the following way:

$$X_t = X_0 + Z_t + Z_{t-1} + Z_{t-2} + \cdots + Z_1.$$

We see that the l-step-ahead forecast error is given by

$$e_h(l) = Z_{h+l} + \cdots + Z_{h+1}.$$

19

Hence $\mathrm{Var}[e_h(l)] = l\sigma^2$ (where σ^2 is the variance of Z_t), which diverges to infinity as $l \to \infty$. As in [Tsay, 2010, page 73] we conclude that the model is **not predictable**. In fact, theoretically, X_t can assume any real value for a sufficiently large t. With our random-walk model we have $\psi_i = 1$ in Definition 2.12. Hence, the impact of any past shock Z_{t-i} on X_t does not decay over time. The series has a **strong memory** and it **remembers all of the past shocks**.

We will now look at a related process - random walk with drift, see [Tsay, 2010, page 73-74].

Definition 4.2. *A time series $\{X_t\}$ is a random walk with drift if it satisfies*

$$X_t = \phi_0 + X_{t-1} + Z_t, \ t = 1, 2, \ldots, \tag{11}$$

where X_0 is a fixed real number, and $\{Z_t\}$ is assumed to be a white noise series with zero mean and variance σ^2.

We see that $\phi_0 = \mathrm{E}[X_t - X_{t-1}]$. The constant term ϕ_0 of model (11) is very important and is referred to as the **drift** of the model. We can easily verify that

$$X_t = t\phi_0 + X_0 + Z_t + Z_{t-1} + \cdots + Z_1,$$

see [Tsay, 2010, page 73]. The last equation shows that the process consists of a time trend $t\phi_0$ and a random-walk process. The conditional standard deviation of X_t is $\sqrt{t}\sigma$ (where σ^2 is the variance of Z_t), which grows at a slower rate than the conditional expectation of X_t. Therefore, if we graph X_t against the time index t, we have a time trend with slope ϕ_0. In other words, for a random walk with drift, the constant term becomes the time slope of the series.

4.2 Some properties of unit-root processes

Consider now the follwing AR(1) model:

$$X_t = \phi X_{t-1} + Z_t,$$

where $\{Z_t\}$ is assumed to be a white noise series with zero mean and variance σ^2. Discussions in this section revolve around the case $\phi = 1$, see [Choi, 2015, page 4]. The properties of $\{X_t\}$ with $\phi = 1$ are very different from those of $\{X_t\}$ with $|\phi| < 1$. We have the following summary in [Choi, 2015, page 4]:

1. When $\phi = 1$, $\mathrm{Var}(X_t) \to \infty$ as $t \to \infty$. When $|\phi| < 1$, however, $\mathrm{Var}(X_t) = (1 - \phi^2)^{-1}\sigma^2$ for all t.

2. When $\phi = 1$, the impact of any past shock Z_{t-i} on X_t does not decay over time. When $|\phi| < 1$, an innovation will lose its effect on the value of X_t eventually as we move forward into the future.

3. When $\phi = 1$, $f(0) = \infty$ where $f(\cdot)$ denotes the spectral density of $\{X_t\}$. This means that $\{X_t\}$ has a strong long-run component. When $|\phi| < 1$, the spectral density is finite at all frequencies.

4. When $\phi = 1$, the expected time between crossings of $y = 0$ is infinite. When $\phi < 1$, the expected time between crossings of $y = 0$ is finite.

5. When $\phi = 1$, the theoretical autocorrelation at lag h converges to 1 for all h as $t \to \infty$. When $|\phi| < 1$, the autocorrelation decreases steadily in magnitude as h increases.

6. When $\phi = 1$, the process is a null recurrent Markov chain.

4.3   Some basic results of a unit-root process

In this section we follow closely [Tsay, page 1-5]. Let $\{Y_t\}$ be a weakly stationary time series with weak serial dependence. The properties of $\{Y_t\}$ is as in Section 3.3:

1. $\mathrm{E}(Y_t) = 0$ for all t.

2. $\sup_t \mathrm{E}|Y_t|^\beta < \infty$ for some $\beta > 2$.

3. The average variance τ^2 exists and is positive.

4. Y_t is strongly mixing, i.e. the serial dependence between Y_t and Y_{t-h} approaches zero as h increases.

We see that if $\{Y_t\}$ is a white noise series with finite second moment it satisfies the above properties. We will however consider the general case. Let $\{X_t\}$ be a time series given by

$$X_t = \pi X_{t-1} + Y_t, \ t = 1, 2, \ldots, \tag{12}$$

where $\pi = 1$, X_0 is a fixed real number, and Y_t is as given above. As in Section 3.3 we define $S_0 = 0$ and

$$S_t = Y_1 + Y_2 + \cdots + Y_t$$

for $t = 1, 2, \ldots$. The average variance of S_T is (as we know) given by

$$\tau^2 = \lim_{T \to \infty} \mathrm{E}(T^{-1} S_T^2),$$

21

and as before we define the function $X_T(r)$ by

$$X_T(r) = \frac{1}{\sqrt{T}\tau} S_{[Tr]}, \ 0 \leq r \leq 1,$$

where $[Tr]$ denotes the integer part of Tr.

The ordinary least squares estimate of π in (12) is

$$\hat{\pi} = \frac{\sum\limits_{t=1}^{T} X_{t-1}X_t}{\sum\limits_{t=1}^{T} X_{t-1}^2},$$

and its variance is estimated by

$$\mathrm{Var}(\hat{\pi}) = \frac{S^2}{\sum\limits_{t=1}^{T} X_{t-1}^2},$$

where S^2 is the residual variance given by

$$S^2 = \frac{1}{T-1} \sum_{t=1}^{T} (X_t - \hat{\pi}X_{t-1})^2,$$

see [Tsay, page 2-3].

The usual t-ratio for testing the null hypothesis $H_0 : \pi = 1$ versus $H_a : \pi < 1$ is given by

$$t_\pi = \left(\sum_{t=1}^{T} X_{t-1}^2 \right)^{1/2} \frac{\hat{\pi} - 1}{S} = \frac{\sum\limits_{t=1}^{T} X_{t-1}Y_t}{S\sqrt{\sum\limits_{t=1}^{T} X_{t-1}^2}}, \tag{13}$$

see [Tsay, page 3].

Define σ_Y^2 as

$$\sigma_Y^2 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathrm{E}(Y_t^2).$$

We now have the following result given in [Tsay, page 3].

Theorem 4.3.

1. $T^{-2} \sum\limits_{t=1}^{T} X_{t-1}^2 \Rightarrow \tau^2 \int_0^1 W(r)^2 dr.$

2. $T^{-1} \sum\limits_{t=1}^{T} X_{t-1}Y_t \Rightarrow \frac{\tau^2}{2}(W(1)^2 - \frac{\sigma_Y^2}{\tau^2}).$

3. $T(\hat{\pi} - 1) \Rightarrow (\frac{1}{2})(W(1)^2 - (\frac{\sigma_Y^2}{\tau^2}))(\int_0^1 W(r)^2 dr)^{-1}.$

4. $\hat{\pi}$ converges to 1 in probability.

5. $t_\pi \Rightarrow (\frac{\tau}{2\sigma_Y})(W(1)^2 - (\frac{\sigma_Y^2}{\tau^2}))(\int_0^1 W(r)^2 dr)^{-1/2}.$

Here $\Rightarrow$ denotes convergence in distribution. The proof of the theorem is given in [Tsay, page 3-4]. We give the first part of the proof with our notation.

Proof of part 1 of Theorem 4.3.

$$
\begin{aligned}
T^{-2} \sum_{t=1}^{T} X_{t-1}^2 &= T^{-2} \sum_{t=1}^{T} (S_{t-1} + X_0)^2 \\
&= T^{-2} \sum_{t=1}^{T} (S_{t-1}^2 + 2X_0 S_{t-1} + X_0^2) \\
&= \tau^2 \sum_{t=1}^{T} \left(\frac{1}{\tau\sqrt{T}} S_{t-1}\right)^2 \frac{1}{T} + 2X_0 \tau T^{-1/2} \sum_{t=1}^{T} \left(\frac{1}{\tau\sqrt{T}} S_{t-1}\right) \frac{1}{T} + T^{-1} X_0^2 \\
&= \tau^2 \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \left(\frac{1}{\tau\sqrt{T}} S_{[Tr]}\right)^2 dr + 2X_0 \tau T^{-1/2} \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \frac{1}{\tau\sqrt{T}} S_{[Tr]} dr + T^{-1} X_0^2 \\
&= \tau^2 \int_0^1 X_T^2(r) dr + 2X_0 \tau T^{-1/2} \int_0^1 X_T(r) dr + T^{-1} X_0^2 \\
&\Rightarrow \tau^2 \int_0^1 W(r)^2 dr, \; T \to \infty.
\end{aligned}
$$

In the last step we have used Theorem 3.5. We notice that from part 3 of Theorem 4.3 $\hat{\pi}$ converges to 1 at the rate of T^{-1}, not the usual rate $T^{-1/2}$. This is referred to as the **super consistency** in the theory of unit-root.

Example 4.4. Consider the random walk $\{X_t\}$ where

$$X_t = X_{t-1} + Z_t, \; t = 1, 2, \ldots, \tag{14}$$

where X_0 is a fixed real number, and $\{Z_t\}$ is assumed to be a white noise series with mean zero and variance σ^2. In this model we have $\tau^2 = \sigma^2 = \sigma_Y^2$. Hence our result above gives:

1. $T^{-2} \sum_{t=1}^{T} X_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 W(r)^2 dr.$

2. $T^{-1} \sum_{t=1}^{T} X_{t-1} Y_t \Rightarrow \frac{\sigma^2}{2}(W(1)^2 - 1).$

3. $T(\hat{\pi} - 1) \Rightarrow (\frac{1}{2})(W(1)^2 - 1)(\int_0^1 W(r)^2 dr)^{-1}.$

4. $t_{\pi} \Rightarrow (\frac{1}{2})(W(1)^2 - 1)(\int_0^1 W(r)^2 dr)^{-1/2}.$

The critical values of t_{π} has been tabulated by several authors. The 0.01, 0.05, and 0.10 quantiles of the limit distribution of t_{π} in the above example are for instance given by -2.58, -1.95, and -1.62, see [Tsay, page 7].

We now look at the general AR(p) case and follow closely [Tsay, page 5-6]. We start with the AR(2) case where $(1 - B)(1 - \phi B)X_t = Z_t$, where $|\phi| < 1$ and $\{Z_t\}$ is assumed to be a white noise series with mean zero and variance σ^2. The model can be written as

$$X_t = X_{t-1} + Y_t, \ Y_t = \phi Y_{t-1} + Z_t.$$

For the weakly stationary AR(1) process Y_t, it can be shown that $\sigma_Y^2 = (1 - \phi^2)^{-1}\sigma^2$ and $\tau^2 = (1 - \phi)^{-2}\sigma^2$. Hence the limiting distributions discussed depend on the AR(1) coefficient ϕ. The t-ratio of $\hat{\pi}$ can be obtained by Theorem 4.3 and it becomes

$$t_{\pi} \Rightarrow \frac{1}{2}\sqrt{\frac{1 + \phi}{1 - \phi}}\left(W(1)^2 - \frac{1 - \phi}{1 + \phi}\right)\left(\int_0^1 W(r)^2 dr\right)^{-1/2}.$$

This dependence on ϕ makes it difficult to use t_{π} in unit-root testing and the dependence continues to hold for the general AR(p) process. A test that can overcome this difficulty is the **augmented Dickey-Fuller** test which we will discuss in the next section.

4.4 Augmented Dickey-Fuller test

In this section we look at the **augmented Dickey-Fuller** test and we follow closely [Tsay, page 7-8]. We consider an AR(p) process defined by $\phi(B)X_t = Z_t$, where $\{Z_t\}$ is assumed to be a white noise series. We look closer at the case where $\phi(B) = \phi^*(B)(1 - B)$. Here $\phi^*(B)X_t = Z_t$ defines a stationary model, see Theorem 2.16. Let $\phi^*(B) = 1 - \sum_{i=1}^{p-1} \phi_i^* B^i$. The model then becomes

$$\phi(B)X_t = \phi^*(B)(1 - B)X_t = (1 - B)X_t - \sum_{i=1}^{p-1} \phi_i^*(1 - B)X_{t-i} = Z_t.$$

We see that testing for a unit root in $\phi(B)$ is equivalent to testing $\pi = 1$ in the following model:

$$X_t = \pi X_{t-1} + \sum_{j=1}^{p-1} \phi_j^*(X_{t-j} - X_{t-j-1}) + Z_t.$$

Or equivalently, the same as testing for $\pi - 1 = 0$ in the following model:

$$\Delta X_t = (\pi - 1)X_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta X_{t-j} + Z_t,$$

where $\Delta X_t = X_t - X_{t-1}$. In practise, the linear model

$$\Delta X_t = \beta X_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta X_{t-j} + Z_t, \tag{15}$$

where $\beta = \pi - 1$, is used. The least squares estimate of β can then be used in unit-root testing. We have that testing $H_0 : \pi = 1$ versus $H_a : \pi < 1$ is equivalent to testing $H_0 : \beta = 0$ versus $H_a : \beta < 0$. The t-ratio of $\hat{\beta}$ has the same limiting distribution as t_π in the random-walk case. Hence, for an $\text{AR}(p)$ model with $p > 1$, by including the lagged variables of ΔX_t in the linear regression of (15), one can remove the nuisance parameters in unit-root testing. This is the well-known **augmented Dickey-Fuller** unit-root test.

4.5 Differencing and Random Walk

The idea of transforming a nonstationary time series $\{X_t\}$ into a stationary one by considering its change series $\{C_t\}$ defined by $C_t = X_t - X_{t-1}$ is called **differencing**. For the random walk in (10) we have

$$C_t = X_t - X_{t-1} = (X_{t-1} + Z_t) - X_{t-1} = Z_t,$$

which is a weakly stationary process. We say that the random walk is integrated of order one and we call it an $I(1)$ process. In Chapter 6 we will look further into this concept.

5 Stochastic unit-root processes

In this chapter we are going to introduce what Granger and Swanson call stochastic unit-root processes in Granger and Swanson [1997]. From the *Abstract* in this article we read: "*A class of nonlinear processes which have a root that is not constant, but is stochastic, and varying around unity is introduced. The process can be stationary for some periods, and mildly explosive for others.*", see [Granger and Swanson, 1997, page 35]. Before studying the process in Granger and Swanson [1997], we look at the following stochastic difference equation.

5.1 A stochastic equation

In this section we look closer at the stochastic equation

$$X_t = A_t X_{t-1} + B_t, \ t = 1, 2, \ldots, \tag{16}$$

where X_0 is a finite random variable.

We can solve (16) in the following way:

$$\begin{aligned}
X_t &= A_t X_{t-1} + B_t \\
&= A_t (A_{t-1} X_{t-2} + B_{t-1}) + B_t \\
&= B_t + A_t B_{t-1} + A_t A_{t-1} X_{t-2} \\
&= \ldots \\
&= \left[\prod_{j=0}^{t-1} A_{t-j} \right] X_0 + \sum_{k=0}^{t-1} \left[\prod_{j=0}^{k-1} A_{t-j} \right] B_{t-k},
\end{aligned}$$

where it is understood that $\prod_{j=0}^{-1} A_{t-j} = 1$.

We consider two different settings and later apply some of the results to the process described in Granger and Swanson [1997].

CASE 1. We assume that $\{(A_t, B_t)\}$ are stationary and ergodic as is the case in Brandt [1986]. If $\{X_t\}$ is stationary and ergodic with $\mathrm{E}(X_t) = \mu$, then the time series average, $T^{-1} \sum_{t=1}^{T} X_t$, converges to μ. Note that stationarity itself does not guarantee ergodicity. The main result for this situation is found in [Brandt, 1986, page 212].

Theorem 5.1. *The stochastic equation*

$$X_t = A_t X_{t-1} + B_t, \ t = 1, 2, \ldots,$$

where X_0 is a finite random variable and $\{(A_t, B_t)\}$ are stationary and ergodic, has a unique stationary and ergodic solution if and only if

$$-\infty \leq \mathrm{E}(\log|A_1|) < 0$$

and

$$\mathrm{E}(\log^+|B_1|) < \infty.$$

Moreover the solution is given by

$$X_t = \sum_{j=0}^{\infty} \left[\prod_{i=t-j}^{t-1} A_i \right] B_{t-j-1}, \ t \in \mathbb{Z}. \tag{17}$$

Note that $\omega^+ = \max(\omega, 0)$ and $\omega^- = -\min(\omega, 0)$.

Remark. We can write

$$X_t = \left[\prod_{j=1}^{t} A_j \right] X_0 + S_t, \ S_t = \sum_{k=0}^{t-1} \left[\prod_{j=0}^{k-1} A_{t-j} \right] B_{t-k}. \tag{18}$$

In the stationary case $|X_t - S_t| \overset{a.s.}{\to} 0$. We can neither say that $X_t \overset{a.s.}{\to} S_t$ nor $X_t \overset{a.s.}{\to} S_\infty$. What we can say is that $X_t \Rightarrow S_\infty$, where $\Rightarrow$ denotes convergence in distribution.

CASE 2. We assume that $\{(A_t, B_t)\}$ are independent and identically distributed. This setup is studied in [Vervaat, 1979, page 752-758]. Let (A, B) denote a generic random pair with the same distribution as (A_t, B_t). Considering the limiting behaviour of $\{\sum_{k=1}^{t} \log|A_k|\}$, $t = 1, 2, \ldots$, we have that one and only one of the following three cases occurs.

1. $\sum_{k=1}^{t} \log|A_k| \Rightarrow -\infty$, where $\Rightarrow$ denotes convergence in distribution.

2. $\log|A| = 0$ with probability 1.

3. $\limsup \mathrm{P}\left(\sum_{k=1}^{t} \log|A_k| > 0 \right) > 0$.

Let $\nu \overset{def}{=} \mathrm{E}(\log|A|)$ in case the expectation exists, finite or infinite. In order to get a valid expression (not an expression like "$\infty - \infty$"), at least one of $\mathrm{E}(\log^+|A|)$ and $\mathrm{E}(\log^-|A|)$ has to be finite. We now have the following situation.

28

- If $-\infty \leq \nu < 0$, then the first case occurs.

- If $0 < \nu \leq \infty$, then the third case occurs.

- If $\nu = 0$ then all three cases occur.

- If $E(\log|A|)$ does not exist, then the first case and the third case occur.

From [Vervaat, 1979, page 757] we have the following important result.

Theorem 5.2. *The stochastic equation*

$$X_t = A_t X_{t-1} + B_t, \ t = 1, 2, \ldots,$$

where X_0 is a finite random variable and $\{(A_t, B_t)\}$ are independent and identically distributed, has a solution if and only if

$$-\infty \leq \mathrm{E}(\log|A|) < 0$$

and

$$\mathrm{E}(\log^+|B|) < \infty.$$

The next result is found in [Vervaat, 1979, page 753]. If $X_t \Rightarrow X$ (where X_t is given as in Theorem 5.2) for some random variable X, then X satisfies the stochastic equation

$$X \stackrel{d}{=} AX + B, \tag{19}$$

where X and (A, B) are independent and $\stackrel{d}{=}$ denotes equality in distribution. A distributional solution of (19) in the situation of Theorem 5.2 is now given by

$$X \sim \sum_{k=1}^{\infty} A_1 A_2 \cdots A_{k-1} B_k.$$

We also have the following result considering the tail of a distributional solution of (19), see [Goldie, 1991, page 135]:

Theorem 5.3. *If there exists some $\kappa > 0$ such that*

$$\mathrm{E}|A|^\kappa = 1,$$

$$\mathrm{E}(|A|^\kappa \log^+|A|) < \infty,$$

$$0 < \mathrm{E}|B|^\kappa < \infty,$$

and moreover $\frac{B}{1-A}$ is non-degenerate and the conditional distribution of $\log|A|$ given $A \neq 0$ is non-lattice, then there are constants c_+ and c_-, at least one of them positive, such that

$$s^\kappa P(X > s) \to c_+,$$

$$s^\kappa P(X < -s) \to c_-,$$

as $s \to \infty$, where X is a distributional solution of (19).

The important requirement in this theorem is that $\mathrm{E}|A|^\kappa = 1$ for some $\kappa > 0$.

5.2 The process in the article by Granger and Swanson

In this section we study the theory of the process described in [Granger and Swanson, 1997, page 37-40]. Note that we will use a different notation than the notation used in Granger and Swanson [1997].

We will consider the following process in this section:

$$X_t = A_t X_{t-1} + B_t, \ t = 1, 2, \ldots, \tag{20}$$

where X_0 is a finite random variable. We assume the following.

1. X_0 is independent of $\{A_t, B_t, t \geqslant 1\}$.

2. $\{B_t\}$ is iid noise with zero mean and finite variance σ_B^2.

3. $\{B_t\}$ is independent of $\{A_t\}$.

4. $A_t = \exp(C_t)$.

5. $\{C_t\}$ is a Gaussian stationary process with mean μ_C, variance $\sigma_C^2 > 0$, and spectral density f_C. We also require $\sum_h |h||\gamma_C(h)| < \infty$ for later results.

We call this process a **STUR** process. If we allowed $C_t \equiv 0$ in the **STUR** model, then $\{X_t\}$ would be a standard unit-root process. Now as an example, let C_t be given by the following AR(1) process:

$$C_t = \phi_0 + \phi_1 C_{t-1} + Z_t, \tag{21}$$

where $|\phi_1| < 1$, and Z_t is iid normally distributed with zero mean and variance σ_Z^2 and is independent of the series B_t. For this model we will assume that C_t is generated exogenously from X_t. This means that

$$\mathrm{E}(C_{t+1}|\mathcal{F}_t^C \vee \mathcal{F}_t^X) = \mathrm{E}(C_{t+1}|\mathcal{F}_t^C).$$

We get the following results from Section 2.4:

$$\mu_C = \frac{\phi_0}{1 - \phi_1},$$

$$\sigma_C^2 = \frac{\sigma_Z^2}{1 - \phi_1^2}.$$

We will now find expressions for the expectation of X_t and the variance of X_t and first we proceed as in Granger and Swanson [1997]. We begin by defining the following:

$$S_{C,t}(j) = \sum_{i=0}^{j-1} C_{t-i},$$

with the notation $S_{C,t}(0) = 1$. Further let

$$W_{t,j} = \exp(S_{C,t}(j)).$$

We see that $S_{C,t}(1) = C_t$ and $S_{C,t}(2) = C_t + C_{t-1}$. Furter we see that $W_{t,1} = \exp(S_{C,t}(1)) = \exp(C_t) = A_t$ and $W_{t,2} = \exp(S_{C,t}(2)) = \exp(C_t + C_{t-1}) = A_t A_{t-1}$. We are led to the following relation:

$$X_t = B_t + W_{t,1}B_{t-1} + W_{t,2}B_{t-2} + \cdots + W_{t,k-1}B_{t-k+1} + W_{t,k}X_{t-k}, \tag{22}$$

for any integer k with $0 \le k \le t$. When $j > 0$ the expectation of $S_{C,t}(j)$ denoted by $\mathrm{E_j}$ is given as follows:

$$\begin{aligned}
\mathrm{E_j} &= \mathrm{E}(S_{C,t}(j)) \\
&= \mathrm{E}(C_t + C_{t-1} + \cdots + C_{t-j+1}) \\
&= j\mu_C.
\end{aligned}$$

The variance of $S_{C,t}(j)$ denoted by $\mathrm{V_j}$ is for $j = 2$ given by:

$$\begin{aligned}
\mathrm{V_2} &= \mathrm{Var}(S_{C,t}(2)) \\
&= \mathrm{Var}(C_t + C_{t-1}) \\
&= \mathrm{Var}(C_t) + \mathrm{Var}(C_{t-1}) + 2\mathrm{Cov}(C_t, C_{t-1}) \\
&= \sigma_C^2 + \sigma_C^2 + 2\rho_C(1)\sigma_C^2 \\
&= \sigma_C^2(2 + 2\rho_C(1)).
\end{aligned}$$

This is different from what we get from [Granger and Swanson, 1997, equation (2.6)]. The equation gives us:

$$\mathrm{V_2} = \sigma_C^2(2 + \rho_C(1)),$$

and clearly

$$\sigma_C^2(2 + \rho_C(1)) \ne \sigma_C^2(2 + 2\rho_C(1)).$$

By using the identity

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j),$$

we see that the correct expression for V_j for $j > 1$ is given by:

$$V_j = \mathrm{Var}(S_{C_t}(j)) = \sigma_C^2\left[j + 2\sum_{r=1}^{j-1}(j-r)\rho_C(r)\right]. \tag{23}$$

Note that the constant 2 is not included in front of the summation symbol in [Granger and Swanson, 1997, equation (2.6)]. However this mistake is not important for the results later in the article. For j large we have the following approximation:

$$V_j \approx jg_C(0),$$

where $g_C = 2\pi f_C$. There is a mistake in [Granger and Swanson, 1997, equation (2.7)], where it wrongly says that $g_C = (2\pi)^{-1}f_C$.

Remark. $g_C(0) = \sum_h \gamma_C(h)$ by the definition of the spectral density, see Section 2.7.

If a random variable X is normally distributed with mean μ and variance σ^2 we have the following standard result using the definition of the moment-generating function M of the random variable X:

$$M_X(k) \stackrel{def}{=} \mathrm{E}(\exp(kX)) = \exp(k\mu + 2^{-1}k^2\sigma^2) = \exp(\mathrm{E}(kX) + 2^{-1}\mathrm{Var}(kX)).$$

It follows that for k large

$$\begin{aligned}
\mathrm{E}(W_{t,k}) &= \mathrm{E}(\exp(S_{C,t}(k))) \\
&= \exp(\mathrm{E}(S_{C,t}(k)) + 2^{-1}\mathrm{Var}(S_{C,t}(k))) \\
&= \exp(\mathrm{E}_k + 2^{-1}V_k) \\
&\approx \exp(k\mu_C + 2^{-1}kg_C(0)) \\
&= \exp(k(\mu_C + 2^{-1}g_C(0))) \\
&= \exp(k\theta),
\end{aligned}$$

where $\theta = \mu_C + 2^{-1}g_C(0)$, see [Granger and Swanson, 1997, equations (2.8), (2.9) and (2.10)].

If we let $k = t$ in (22) we get

$$X_t = B_t + W_{t,1}B_{t-1} + W_{t,2}B_{t-2} + \cdots + W_{t,t-1}B_1 + W_{t,t}X_0.$$

If we assume that X_0 is deterministic and if we use that B_t is iid noise with zero mean, we get

$$\mathrm{E}(X_t) = X_0\mathrm{E}(W_{t,t}) = X_0\exp(t\theta). \tag{24}$$

We also have the approximation

$$\mathrm{Var}(X_t) \approx \sigma_B^2 \sum_{k=0}^{t-1} \exp(2k\phi), \tag{25}$$

where $\phi = \mu_C + g_C(0)$, see [Granger and Swanson, 1997, equation (2.11)].

Theorem 5.4. *Let X_t be a **STUR** process. Then*

$$\mathrm{E}(X_t) = X_0\exp(t\theta),$$

$$\mathrm{Var}(X_t) \approx \sigma_B^2 \sum_{k=0}^{t-1} \exp(2k\phi),$$

*where $\theta = \mu_C + 2^{-1}g_C(0)$ and $\phi = \mu_C + g_C(0)$. Here $g_C = 2\pi f_C$, where f_C is the spectral density of $\{C_t\}$, and μ_C is the mean of $\{C_t\}$, where C_t is given in the **STUR** model (20).*

When $\phi \neq 0$ we get

$$\mathrm{Var}(X_t) \approx \sigma_B^2 \sum_{k=0}^{t-1} \exp(2k\phi) = \sigma_B^2 \sum_{k=0}^{t-1} \exp(2\phi)^k = \frac{\sigma_B^2}{\exp(2\phi) - 1} \left(\exp(2\phi)^t - 1\right).$$

When $\phi = 0$ we get

$$\mathrm{Var}(X_t) \approx \sigma_B^2 \sum_{k=0}^{t-1} 1 = t\sigma_B^2.$$

Note that the conditional variance of a random walk behaves in exactly the same manner.

Let us calculate the variance of X_t in another way than in Granger and Swanson [1997]. We know that

$$X_t = \left[\prod_{j=0}^{t-1} A_{t-j}\right] X_0 + \sum_{k=0}^{t-1} \left[\prod_{j=0}^{k-1} A_{t-j}\right] B_{t-k}.$$

We have assumed that X_0 is independent of $\{A_t, B_t, t \geqslant 1\}$, and that $\{B_t\}$ is independent of $\{A_t\}$. Since $\{B_t\}$ is iid noise with zero mean, we therefore have that:

$$\mathrm{E}(X_t^2) = \mathrm{E}\left[\prod_{j=0}^{t-1} A_{t-j}^2\right] X_0^2 + \sigma_B^2 \sum_{k=0}^{t-1} \mathrm{E}\left[\prod_{j=0}^{k-1} A_{t-j}^2\right]$$

$$= \mathrm{E}\left[\prod_{j=0}^{t-1} A_{t-j}^2\right] X_0^2 + \sigma_B^2 \sum_{k=0}^{t-1} \mathrm{E}\left[\prod_{j=t-k+1}^{t} A_j^2\right]$$

$$= \mathrm{E}\left[\prod_{j=0}^{t-1} A_{t-j}^2\right] X_0^2 + \sigma_B^2 \sum_{k=0}^{t-1} \mathrm{E}\left[\prod_{j=1}^{k} A_j^2\right].$$

Now we use that $A_t = \exp(C_t)$, and get

$$\mathrm{E}\left[\prod_{j=1}^{k} A_j^2\right] = \mathrm{E}\left[\prod_{j=1}^{k} \exp(2C_j)\right] = \mathrm{E}\left[\exp\left(\sum_{j=1}^{k} 2C_j\right)\right] = \mathrm{E}\left[\exp\left(2S_k\right)\right] = \mathrm{M}_{S_k}(2),$$

where $S_k \stackrel{def}{=} \sum_{t=1}^{k} C_t$. Since C_t is Gaussian, so is S_k. If we denote the expectation of S_k by μ_k and the variance of S_k by σ_k^2, we have

$$\mathrm{E}\left[\prod_{j=1}^{k} A_j^2\right] = \exp\left(2\mu_k + 2\sigma_k^2\right).$$

Lemma 5.5. *Suppose that $\{C_t\}$ is a stationary Gaussian time series with mean μ_C and covariance function γ_C. Assume that $\sum_h |h||\gamma_C(h)| < \infty$. Let $\alpha_k = \sum_{|h|<k} \gamma_C(h)$ and $\beta_k = -\sum_{|h|<k} |h|\gamma_C(h)$. Then*

$$\sum_{t=1}^{k} C_t \sim \mathcal{N}(k\mu_C, \ k\alpha_k + \beta_k).$$

Assuming that $X_0 = 0$, Lemma 5.5 gives

$$\mathrm{Var}(X_t) = \sigma_B^2 \sum_{k=0}^{t-1} \exp\left(2\mu_k + 2\sigma_k^2\right) = \sigma_B^2 \sum_{k=0}^{t-1} \exp\left(2k\mu_C + 2k\alpha_k + 2\beta_k\right).$$

Theorem 5.6. *Let X_t be a **STUR** process. Assuming that $X_0 = 0$ we have that*

$$\mathrm{Var}(X_t) = \sigma_B^2 \sum_{k=0}^{t-1} \exp\left(2k(\mu_C + \alpha_k) + 2\beta_k\right),$$

*where $\alpha_k = \sum_{|h|<k} \gamma_C(h)$, $\beta_k = -\sum_{|h|<k} |h|\gamma_C(h)$, γ_C is the covariance function of $\{C_t\}$, and μ_C is the mean of $\{C_t\}$, where C_t is given in the **STUR** model (20).*

Note that from Theorem 5.4 we have that

$$\mathrm{Var}(X_t) \approx \sigma_B^2 \sum_{k=0}^{t-1} \exp(2k\phi) = \sigma_B^2 \sum_{k=0}^{t-1} \exp(2k(\mu_C + \sum_h \gamma_C(h))).$$

We have that $\alpha_k \to \sum_h \gamma_C(h)$ and that $\beta_k \to 0$ under the assumption $\sum_h |h||\gamma_C(h)| < \infty$. Hence the variance of X_t is determined by the expression $\mu_C + \sum_h \gamma_C(h)$. We have the following result:

Theorem 5.7. *Let $\{X_t\}$ be a **STUR** process. We have the following situations:*

1. $\mu_C + \sum_h \gamma_C(h) > 0$. The variance of X_t increases at an exponential rate.

2. $\mu_C + \sum_h \gamma_C(h) < 0$. The process is weakly stationary.

3. $\mu_C + \sum_h \gamma_C(h) = 0$. $\mathrm{Var}(X_t) \approx t\sigma_B^2$.

In Granger and Swanson [1997] two alternative characterizations of the **STUR** process are presented. They are called **STURA** and **STURB** and have the following properties.

1. **STURA**: $\mathrm{E}(A_t) = \mathrm{E}(\exp(C_t)) = 1$ so that

$$\mu_C + 2^{-1}\sigma_C^2 = 0.$$

We see that since $\sigma_C^2 > 0$ for the **STUR** model, it follows that $\mu_C < 0$. Note that if we had allowed for $\sigma_C^2 = 0$ in the **STUR** model, then $\mu_C = 0$, and hence $C_t \equiv 0$. In this case we would have a standard unit-root nonstationary time series.

2. **STURB**: $\mathrm{E}(A_t A_{t-1} \cdots A_{t-k+1}) = \mathrm{E}(W_{t,k}) = 1$ for k large so that

$$\theta = \mu_C + 2^{-1}g_C(0) = 0,$$

see [Granger and Swanson, 1997, equation (2.15) and (2.16)].

We will look closer at some of the properties of **STURA** later in this chapter, but first we give a remark considering **STURB**. Calculating gives:

$$\mathrm{E}\left[\prod_{j=1}^{k} A_j \right] = \mathrm{E}\left[\exp\left(\sum_{j=1}^{k} C_j \right) \right] = \mathrm{M}_{S_k}(1) = \exp(\mu_k + 2^{-1}\sigma_k^2).$$

Thus we have

$$\mathrm{E}\left[\prod_{j=1}^{k} A_j \right] = \exp\left(k\mu_C + 2^{-1}k\alpha_k + 2^{-1}\beta_k \right),$$

where α_k and β_k are given in Lemma 5.5.

Let us apply some of the results in this section to **STURA**. In the case of **STURA**, we have seen that since $\sigma_C^2 > 0$, then $\mu_C < 0$. It follows that

$$\mathrm{E}(\log|A_t|) = \mathrm{E}(\log|\exp(C_t)|) = \mathrm{E}(C_t) = \mu_C < 0,$$

and the first requirement in Theorem 5.1 is satisfied. This result can also be verified by *Jensen's inequality* which in our case says that for any strictly concave function $\mathrm{E}(f(X)) < f(\mathrm{E}(X))$. Since the logarithm is a strictly concave function it follows that

$$\mathrm{E}(\log|A_t|) = \mathrm{E}(\log(A_t)) < \log(\mathrm{E}(A_t)) = \log(1) = 0.$$

The second requirement in Theorem 5.1 is assumed to hold and we get the following important result.

Theorem 5.8. *STURA is stationary.*

We now consider the tail behaviour of **STURA**. With the conditions in Theorem 5.3 we have that
$$\mathrm{P}(|X| > s) \approx cs^{-\kappa},$$
where $c > 0$ is a constant. However for **STURA** we have the restrictions that

$$\mathrm{E}|A_t|^1 = \mathrm{E}(A_t) = 1,$$

$$\mathrm{E}(|A_t|^1 \log^+ |A_t|) < \infty,$$

$$0 < \mathrm{E}|B_t|^1 < \infty,$$

and further that B_t is not proportional to $1 - A_t$. We conclude therefore that the tail of a solution X in this case obeys
$$\mathrm{P}(|X| > s) \approx ds^{-1}$$

where $d > 0$ is a constant, see [Yoon, 2006, page 256].

STURA processes have very heavy tails and therefore extreme values are very likely to happen. It can further be shown that the sample autocorrelation functions of **STURA** will have in general, random limiting distibutions, see Davis and Mikosch [1998].

5.3 An augmented Dickey-Fuller test with a STUR alternativ

In this section we look closer at part 3 in [Granger and Swanson, 1997, page 41-43]. Here we discuss whether the augmented Dickey-Fuller (ADF) test has power against **STUR** alternatives. The ADF test is described in Section 4.4. It has as the null hypothesis that the series in question is distributed as $I(1)$. From [Granger and Swanson, 1997, page 41] we read that with the amount of data available in macroeconomics the ADF test is unlikely to has much power against a variety of processes which are, in some sense, near I(1). In general, processes against which unit-root tests have little power are called "generic unit-root" processes. In [Granger and Swanson, 1997, page 42] two experiments were performed in order to study the power properties of the ADF test in the presence of **STUR**. In the experiments two different **STUR** processes are generated. In the first experiment Z_t in (21) is iid normally distributed with zero mean and variance 0.0001, $\phi_0 = -0.00003125$ and $\phi_1 = 0.60$. This implies that

$$\mu_C = \frac{\phi_0}{1 - \phi_1} = \frac{-0.00003125}{1 - 0.60} = -0.000078125,$$

and

$$\sigma_C^2 = \frac{\sigma_Z^2}{1 - \phi_1^2} = \frac{0.0001}{1 - 0.60^2} = 0.00015625.$$

Hence

$$\begin{aligned}
\mathrm{E}(A_t) &= \mathrm{E}(\exp(C_t)) \\
&= \exp(\mu_C + \frac{1}{2}\sigma_C^2) \\
&= \exp(-0.000078125 + \frac{1}{2}0.00015625) \\
&= \exp(0) \\
&= 1.
\end{aligned}$$

Thus the first experiment is a **STURA** process. In the second experiment Z_t in (21) is iid normally distributed with zero mean and variance 0.000001, $\phi_0 = -0.0000003125$ and $\phi_1 = 0.60$. This also implies that $\mathrm{E}(A_t) = 1$ and thus both experiments are **STURA** processes.

In Granger and Swanson [1997] different ADF tests were run, see [Granger and Swanson, 1997, equation (3.1)]. The test in the form of the one in Section 4.4 without trend and intercept where among those who were run:

$$\Delta X_t = \beta X_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta X_{t-j} + Y_t,$$

where Y_t is a white noise series.

The relevant test statistic is the standard t-value for β, the coefficient of X_{t-1}, see [Granger and Swanson, 1997, page 41]. Monte Carlo results with 10 000 simulations in each case were performed, see Table 1 in [Granger and Swanson, 1997, page 42] for the results. The table lists the percentage of estimated t-values that lie above the theoretical 95% and 5% values in the different cases. In the table results for $\sigma_Z^2 = 0.0001$ and $\sigma_Z^2 = 0.000001$ are reported. From the table with $\sigma_Z^2 = 0.0001$ we read that using the 95% critical value of -1.95, the null of a pure unit root fails to reject 96.8% of the time. Similarily with $\sigma_Z^2 = 0.000001$ we read that again using the 95% critical value of -1.95, the null of a pure unit root fails to reject 96.4% of the time. From both experiments it is clear that the ADF test in these cases cannot distinguish between the given **STUR** model and a standard unit-root process. This is also the situation for the other ADF tests without constraints and without trend, see [Granger and Swanson, 1997, page 43]). These results support the proposition that **STUR** processes are generic unit root processes. It would be ideal if we could construct a test which has **STUR** as the null hypothesis, but this is rather difficult and is therefore not done in Granger and Swanson [1997].

5.4 Forecasting

From the *Abstract* in Granger and Swanson [1997] we read: "...*a forecast comparison of linear random walk and AR(p) models, time-varying parameter models, and STUR models suggests that this new class of processes is potentially useful, particularly when the objective is multi-step ahead forecasting.*", see [Granger and Swanson, 1997, page 35]. In [Granger and Swanson, 1997, page 51-57] one judge the relevance of the **STUR** model by comparing it to other models when analyzing actual data. The relative forecasting performance of a simple **STUR** model is compared to

1. Standard random walk with drift models.

2. AR(p) models.

3. A time-varying parameter model, $X_t = A_t X_{t-1} + Z_t$, where A_t is assumed to evolve according to an AR(1) process.

Overall, the results are rather mixed, but the **STUR** model performs very well outside the 1-Step-Ahead case. These results are encouraging according to Granger and Swanson [1997], but more light should be shed on the issue of why **STUR** perform so poorly for 1-Step-Ahead case. Further experiments might compare **STUR** forecasts with more complicated alternative forecasts.

6 Cointegration

In this chapter we look closer at cointegration. But first we introduce bivariate time series.

6.1 Bivariate time series

We will now introduce bivariate time series and we get our information from Chapter 7 in Brockwell and Davis [2002] and Chapter 8 in Tsay [2010]. A bivariate time series is a series of two-dimensional vectors $(X_{1t}, X_{2t})'$ observed at times t. The two component series could be studied independently as univariate time series, but such an approach fails to take into account possible dependence between the two component series, see [Brockwell and Davis, 2002, page 224]).

We write $\boldsymbol{X}_t = (X_{1t}, X_{2t})'$ and we further assume that $\mathrm{E}(X_{1t}^2) < \infty$ and $\mathrm{E}(X_{2t}^2) < \infty$ for all t. If all the finite-dimensional distributions of $\{X_{1t}\}$ and $\{X_{2t}\}$ are multivariate normal, then the distributional properties of $\{X_{1t}\}$ and $\{X_{2t}\}$ will be completely determined by the means

$$\mu_{it} \stackrel{def}{=} \mathrm{E}(X_{it}), \ i = 1, 2,$$

and the covariances

$$\gamma_{ij}(t + h, t) \stackrel{def}{=} \mathrm{Cov}(X_{i,t+h}, X_{jt}), \ i, j = 1, 2.$$

We define the mean vector

$$\boldsymbol{\mu}_t \stackrel{def}{=} \mathrm{E}(\boldsymbol{X}_t) = \left[\begin{array}{c} \mu_{1t} \\ \mu_{2t} \end{array} \right]$$

and covariance matrices

$$\Gamma(t + h, t) \stackrel{def}{=} \mathrm{Cov}(\boldsymbol{X}_{t+h}, \boldsymbol{X}_t) = \left[\begin{array}{cc} \gamma_{11}(t + h, t) & \gamma_{12}(t + h, t) \\ \gamma_{21}(t + h, t) & \gamma_{22}(t + h, t) \end{array} \right].$$

The bivariate series $\{\boldsymbol{X}_t\}$ is **weakly stationary** if $\boldsymbol{\mu}_t$ and $\Gamma(t+h, t)$ are both independent of t. In this case we use the notation

$$\boldsymbol{\mu} \stackrel{def}{=} \mathrm{E}(\boldsymbol{X}_t) = \left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right]$$

and

$$\Gamma(h) \stackrel{def}{=} \mathrm{Cov}(\boldsymbol{X}_{t+h}, \boldsymbol{X}_t) = \left[\begin{array}{cc} \gamma_{11}(h) & \gamma_{12}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) \end{array} \right].$$

We see immediately that $\gamma_{12}(h) = \gamma_{21}(-h)$ and it follows that $\Gamma(h) = \Gamma'(-h)$. Moreover we have that γ_{11} and γ_{22} are autocovariance functions for $\{X_{1t}\}$ and $\{X_{2t}\}$ respectively.

The function $\gamma_{ij}(\cdot), i \neq j$, is called the cross-covariance function of the two series $\{X_{it}\}$ and $\{X_{jt}\}$. In general $\gamma_{ij}(\cdot)$ is not the same as $\gamma_{ji}(\cdot)$.

We now define bivariate iid noise and bivariate white noise as in [Brockwell and Davis, 2002, page 232]:

Definition 6.1. *The bivariate series* $\{\boldsymbol{Z}_t\}$, *where* $\boldsymbol{Z}_t = (Z_{1t}, Z_{2t})'$, *is called **iid noise with mean 0 and covariance matrix** $\mathbb{A}$, written*

$$\{\boldsymbol{Z}_t\} \sim \mathrm{iid}(\boldsymbol{0}, \mathbb{A}),$$

if the random vectors $\{\boldsymbol{Z}_t\}$ *are independent and identically distributed with mean* $\boldsymbol{0}$ *and covariance matrix* $\mathbb{A}$.

Definition 6.2. *The bivariate series* $\{\boldsymbol{Z}_t\}$, *where* $\boldsymbol{Z}_t = (Z_{1t}, Z_{2t})'$, *is called **white noise with mean 0 and covariance matrix** $\mathbb{A}$, written*

$$\{\boldsymbol{Z}_t\} \sim \mathrm{WN}(\boldsymbol{0}, \mathbb{A}),$$

if $\{\boldsymbol{Z}_t\}$ *is weakly stationary with mean vector* $\boldsymbol{0}$ *and covariance matrix function* $\Gamma(h)$ *where* $\Gamma(0) = \mathbb{A}$ *and* $\Gamma(h) = 0$ *if* $h \neq 0$.

Let us generate the linear processes as in [Brockwell and Davis, 2002, page 232]:

Definition 6.3. *The bivariate series* $\{\boldsymbol{Z}_t\}$, *where* $\boldsymbol{Z}_t = (Z_{1t}, Z_{2t})'$, *is a **linear process**, if it has the representation*

$$\boldsymbol{Z}_t = \sum_{j=-\infty}^{\infty} \mathbb{C}_j \boldsymbol{Z}_{t-j}, \ \{\boldsymbol{Z}_t\} \sim \mathrm{WN}(\boldsymbol{0}, \mathbb{A}),$$

where $\{\mathbb{C}_j\}$ *is a sequence of* 2×2 *matrices whose components are absolutely summable.*

A causal **MA(∞)** process is a linear process with $\mathbb{C}_j = 0$ for $j < 0$ in the above definition, see [Brockwell and Davis, 2002, page 233].

We will now introduce bivariate ARMA processes. The following definition is found in [Brockwell and Davis, 2002, page 241]:

Definition 6.4. *The bivariate series $\{\boldsymbol{X}_t\}$, where $\boldsymbol{X}_t = (X_{1t}, X_{2t})'$, is an $\boldsymbol{ARMA}(p,q)$ process if $\{\boldsymbol{X}_t\}$ is weakly stationary and if for every t,*

$$\boldsymbol{X}_t - \Phi_1 \boldsymbol{X}_{t-1} - \cdots - \Phi_p \boldsymbol{X}_{t-p} = \boldsymbol{Z}_t + \Theta_1 \boldsymbol{Z}_{t-1} + \cdots + \Theta_q \boldsymbol{Z}_{t-q}, \tag{26}$$

where Φ_j $(j = 1, 2, \ldots, p)$ and Θ_k $(k = 1, 2, \ldots, q)$ are 2×2 matrices, and $\{\boldsymbol{Z}_t\} \sim$ WN$(\boldsymbol{0}, \mathbb{A})$.

Equations (26) can be written in the more compact form

$$\Phi(B)\boldsymbol{X}_t = \Theta(B)\boldsymbol{Z}_t, \ \{\boldsymbol{Z}_t\} \sim \text{WN}(\boldsymbol{0}, \mathbb{A}), \tag{27}$$

where $\Phi(z) \stackrel{def}{=} \mathbb{I} - \Phi_1 z - \cdots - \Phi_p z^p$ and $\Theta(z) \stackrel{def}{=} \mathbb{I} + \Theta_1 z + \cdots + \Theta_q z^q$ are matrix-valued polynomials, $\mathbb{I}$ is the 2×2 identity matrix, and B denotes the backward shift operator. We assume that the two matrix polynomials have no left common factors; otherwise the model can be simplified, see [Tsay, 2010, page 424]. Setting $q = 0$ in (26) we obtain the bivariate AR(p) process and similarly setting $p = 0$ in (26) we obtain the bivariate MA(q) process. We will now look closer at the bivariate AR(1) process, see [Brockwell and Davis, 2002, page 241-242].

Example 6.5. Setting $p = 1$ and $q = 0$ in (26) gives the defining equations

$$\boldsymbol{X}_t = \Phi \boldsymbol{X}_{t-1} + \boldsymbol{Z}_t, \ \{\boldsymbol{Z}_t\} \sim \text{WN}(\boldsymbol{0}, \mathbb{A}), \tag{28}$$

for the bivariate AR(1) series $\{\boldsymbol{X}_t\}$. From (28) it follows directly that

$$\boldsymbol{X}_t = \sum_{j=0}^{\infty} \Phi^j \boldsymbol{Z}_{t-j}, \tag{29}$$

provided that the eigenvalues of Φ are less than 1 in modulus. The reason for this requirement is that Φ^j must converge to zero as $j \to \infty$ for the dependence in (29) to be meaningful; otherwise, if not the eigenvalues of Φ are less than 1 in modulus, then Φ^j will either explode or converge to a nonzero matrix as $j \to \infty$, see [Tsay, 2010, page 402]. The condition that the eigenvalues of Φ should be less than 1 in modulus is just the bivariate analogue of the condition $|\phi_1| < 1$ in Theorem 2.14.

6.2 Cointegration

We will now introduce cointegration. In the second part of the article by Engle and Granger, cointegration is well explained, see [Engle and Granger, 1987, page 252-255]. Often time series must be differenced before they get stationary. We have discussed differencing in Section 4.5 and will now generalize this concept with the following definition found in [Engle and Granger, 1987, page 252]:

Definition 6.6. *A time series $\{X_t\}$ which has a stationary ARMA representation after differencing d times, is said to be integrated of order d, denoted $X_t \sim I(d)$.*

We have already seen in Section 4.5 that a random walk is integrated of order 1 and we are mostly interested in the cases where $d = 0$ or $d = 1$. For $d = 0$, $\{X_t\}$ will be stationary, while for $d = 1$, its change series is stationary. From Section 4.2 we know that there are substantial differences between a time series that is $I(0)$ and one that is $I(1)$. An $I(1)$ series is rather smooth, having dominant long swings, compared to an $I(0)$ series. It is always true that the sum of an $I(0)$ series and an $I(1)$ series will be an $I(1)$ series. If a and b are constants with $b \neq 0$, and if $X_t \sim I(d)$, then $a + bX_t$ is also $I(d)$. If X_{1t} and X_{2t} are both $I(d)$, then it is generally true that the linear combination

$$Z_t = X_{1t} - aX_{2t}$$

will also be $I(d)$. However, it is possible that $Z_t \sim I(d-b)$, where $b > 0$. When this occurs, a very special constraint operates on the long-run components of the series. For the case $d = b = 1$, the constant a is such that the bulk of the long run components of X_{1t} and X_{2t} cancel out. For $a = 1$ in this case, the difference of X_{1t} and X_{2t} will be $I(0)$. The vague idea is that X_{1t} and X_{2t} cannot drift too far apart. We will now formalize these ideas as in [Engle and Granger, 1987, page 253]:

Definition 6.7. *The components of the vector $\boldsymbol{X}_t = (X_{1t}, X_{2t})'$ are said to be **cointegrated of order** (d,b), denoted $\boldsymbol{X}_t \sim CI(d,b)$, if (i) both components of $\boldsymbol{X}_t$ are $I(d)$ and (ii) there exists a vector $\boldsymbol{\beta} = (\beta_1, \beta_2)' \neq \boldsymbol{0}$ such that $\boldsymbol{\beta}'\boldsymbol{X}_t = \beta_1 X_{1t} + \beta_2 X_{2t} \sim I(d-b), b > 0$. The vector $\boldsymbol{\beta}$ is called the **cointegration vector**.*

We will now give a simple example of cointegration found in [Brockwell and Davis, 2002, page 255]:

Example 6.8. Let $\boldsymbol{X}_t = (X_{1t}, X_{2t})'$, where X_{1t} is the random walk

$$X_{1t} = \sum_{j=1}^{t} Z_j, \ t = 1, 2, \ldots, \ \{Z_t\} \sim \mathrm{IID}(0, \sigma^2),$$

and X_{2t} consists of noisy observations of the same random walk,

$$X_{2t} = X_{1t} + W_t, \ t = 1, 2, \ldots, \ \{W_t\} \sim \mathrm{IID}(0, \tau^2),$$

where $\{W_t\}$ is independent of $\{Z_t\}$. Then clearly $\boldsymbol{X}_t \sim CI(1,1)$ with cointegration vector $\boldsymbol{\beta} = (1, -1)'$.

If the components of the vector $\boldsymbol{X}_t = (X_{1t}, X_{2t})'$ are $I(1)$, they are cointegrated if there is a linear combination $\beta_1 X_{1t} + \beta_2 X_{2t}$ that is stationary. This linear combination is often motivated by economic theory and referred to as **long-run equilibrium** relationship. The intuition is that two $I(1)$ time series with a long-run equilibrium relationship cannot drift too far apart from the equilibrium because economic forces will act to restore the equilibrium relationship, see [Zivot and Wang, 2006, page 435]. The cointegration vector $\boldsymbol{\beta}$ in Definition 6.7 is not unique. For any constant c, the linear combination $c\boldsymbol{\beta}'\boldsymbol{X}_t \sim I(0)$. In order to identify $\boldsymbol{\beta}$ uniquely some normalization assumption is required. We use the normalization

$$\boldsymbol{\beta} = (1, -\beta_2)'.$$

Hence the cointegration relationship can be expressed as

$$\boldsymbol{\beta}'\boldsymbol{X}_t = X_{1t} - \beta_2 X_{2t} \sim I(0)$$

or

$$X_{1t} = \beta_2 X_{2t} + Z_t,$$

where $Z_t \sim I(0)$. The error term Z_t is often referred to as the **disequilibrium error** or the **cointegrationg residual**.

6.3 Cointegration and Error Correction Models

Let $\boldsymbol{X}_t = (X_{1t}, X_{2t})'$ and assume that $\boldsymbol{X}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$ so that $\boldsymbol{\beta}'\boldsymbol{X}_t = X_{1t} - \beta_2 X_{2t}$ is $I(0)$. We have that cointegration implies the existence of an **error correction model** (ECM) of the form

$$\Delta X_{1t} = c_1 + \alpha_1(X_{1,t-1} - \beta_2 X_{2,t-1}) + \sum_j \psi_{11}^j \Delta X_{1,t-j} + \sum_j \psi_{12}^j \Delta X_{2,t-j} + Z_{1t} \qquad (30)$$

$$\Delta X_{2t} = c_2 + \alpha_2(X_{1,t-1} - \beta_2 X_{2,t-1}) + \sum_j \psi_{21}^j \Delta X_{1,t-j} + \sum_j \psi_{22}^j \Delta X_{2,t-j} + Z_{2t} \qquad (31)$$

that describes the dynamic behavior of X_{1t} and X_{2t}. Here $\{Z_{1t}\}$ and $\{Z_{2t}\}$ are iid noise series. The ECM links the long-run equilibrium relationship with the short-run dynamic adjustment mechanism, see [Zivot and Wang, 2006, page 441]. The idea is that a proportion of the disequilibrium from one period is corrected in the next period. The coefficients α_1 and α_2 are adjustment coefficients and their magnitudes determine the speed of adjustment toward equilibrium. We will now give an simple example of an ECM found in [Zivot and Wang, 2006, page 442-443] to illustrate the concept:

Example 6.9. Let $\boldsymbol{X}_t = (X_{1t}, X_{2t})'$ and assume that $\boldsymbol{X}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -1)'$. We suppose that the ECM has the form

$$\Delta X_{1t} = c_1 + \alpha_1(X_{2,t-1} - X_{1,t-1}) + Z_{1t}$$

$$\Delta X_{2t} = c_2 + \alpha_2(X_{2,t-1} - X_{1,t-1}) + Z_{2t}$$

where $c_1 > 0$ and $c_2 > 0$. Further we consider the special case where $\alpha_1 = 0.5$ and $\alpha_2 = 0$. The ECM equations then become

$$\Delta X_{1t} = c_1 + 0.5(X_{2,t-1} - X_{1,t-1}) + Z_{1t}$$

$$\Delta X_{2t} = c_2 + Z_{2t},$$

so that only $\{X_{1t}\}$ responds to the lagged disequilibrium error. We have that

$$\mathrm{E}[\Delta X_{1t} | \boldsymbol{X}_{t-1}] = c_1 + 0.5(X_{2,t-1} - X_{1,t-1})$$

and

$$\mathrm{E}[\Delta X_{2t} | \boldsymbol{X}_{t-1}] = c_2.$$

We now consider three situations:

1. $X_{2,t-1} - X_{1,t-1} = 0$. In this case $\mathrm{E}[\Delta X_{1t} | \boldsymbol{X}_{t-1}] = c_1$ and there is no expected adjustment since the model was in long-run equilibrium in the previous period.

2. $X_{2,t-1} - X_{1,t-1} > 0$. In this case $\mathrm{E}[\Delta X_{1t} | \boldsymbol{X}_{t-1}] = c_1 + 0.5(X_{2,t-1} - X_{1,t-1}) > c_1$. The model was above long-run equilibrium in the last period so the expected adjustment is downward toward equilibrium.

3. $X_{2,t-1} - X_{1,t-1} < 0$. In this case $\mathrm{E}[\Delta X_{1t} | \boldsymbol{X}_{t-1}] = c_1 + 0.5(X_{2,t-1} - X_{1,t-1}) < c_1$. The model was below long-run equilibrium in the last period so the expected adjustment is upward to equilibrium.

The discussion above illustrates why the model is called an error correcting model.

6.4 Testing for Cointegration

In this short section we suppose that the components of the vector $\boldsymbol{X}_t = (X_{1t}, X_{2t})'$ are $I(1)$ and formulate a two-step procedure for determining if the vector $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ is a cointegrating vector for $\boldsymbol{X}_t$. This procedure was originally considered by Engle and Granger in Engle and Granger [1987]. We get our information from [Zivot and Wang, 2006, page 444]. The procedure is as follows:

1. Form the cointegrating residual $\boldsymbol{\beta}' \boldsymbol{X}_t = Z_t$

2. Perform a unit root test on Z_t to determine if it is $I(0)$.

There are two cases to consider. In the first case the proposed cointegrating vector is pre-specified while in the second case the cointegrating vector is estimated from the data available. In the second case an estimate of the cointegrating residual is formed. Tests for cointegration using a pre-specified cointegrating vector are generally much more powerful than tests employing an estimated vector, see [Zivot and Wang, 2006, page 444].

6.5   The bivariate AR(p) process and the VECM

In this section we look closer at the bivariate AR(p) process which is an important model for the analysis of bivariate time series and for cointegration analysis. We begin by the following definition.

Definition 6.10. *The bivariate series* $\{\boldsymbol{X}_t\}$*, where* $\boldsymbol{X}_t = (X_{1t}, X_{2t})'$*, is an* ***AR(p) process*** *if* $\{\boldsymbol{X}_t\}$ *is weakly stationary and if for every* t*,*

$$\boldsymbol{X}_t = \Phi_1 \boldsymbol{X}_{t-1} + \cdots + \Phi_p \boldsymbol{X}_{t-p} + \boldsymbol{Z}_t \tag{32}$$

where Φ_j *(*$j = 1, 2, \ldots, p$*) are* 2×2 *matrices, and* $\{\boldsymbol{Z}_t\} \sim \mathrm{WN}(\boldsymbol{0}, \mathbb{A})$*.*

Equations (32) can be written in the more compact form.

$$\Phi(B)\boldsymbol{X}_t = \boldsymbol{Z}_t, \ \{\boldsymbol{Z}_t\} \sim \mathrm{WN}(\boldsymbol{0}, \mathbb{A}), \tag{33}$$

where $\Phi(z) \stackrel{def}{=} \mathbb{I} - \Phi_1 z - \cdots - \Phi_p z^p$ is a matrix-valued polynomial, $\mathbb{I}$ is the 2×2 identity matrix and B denotes the backward shift operator. If the solutions of $\det(\Phi(z)) = 0$ are greater than 1 in modulus, then the series $\{X_t\}$ is stationary, see [Zivot and Wang, 2006, page 456]. If $\det(\Phi(z)) = 0$ has a root on the unit circle, then one or both of $\{X_{1t}\}$ and $\{X_{2t}\}$ are $I(1)$. We now suppose that the components of $\boldsymbol{X}_t$ are possibly cointegrated (we also say that $\boldsymbol{X}_t$ is cointegrated, when we mean that the components of $\boldsymbol{X}_t$ are cointegrated). The cointegrating relation become apparent if we transform (32) to the **vector error correction model** (VECM)

$$\Delta\boldsymbol{X}_t = \Gamma\boldsymbol{X}_{t-1} + \Gamma_1\Delta\boldsymbol{X}_{t-1} + \cdots + \Gamma_{p-1}\Delta\boldsymbol{X}_{t-p+1} + \boldsymbol{Z}_t, \tag{34}$$

where $\Gamma = \Phi_1 + \cdots + \Phi_p - \mathbb{I}$ and $\Gamma_k = -\sum_{j=k+1}^{p} \Phi_j, \ k = 1, \ldots, p - 1$. The matrix Γ is called the **long-run impact matrix** and Γ_k are the **short-run impact matrices**. In the VECM (34), $\Delta\boldsymbol{X}_t$ and its lags are $I(0)$. The term $\Gamma\boldsymbol{X}_{t-1}$ is the only one which includes

potential $I(1)$ variables. But for $\Delta \boldsymbol{X}_t$ to be $I(0)$ it must be the case that $\Gamma \boldsymbol{X}_{t-1}$ is also $I(0)$. Therefore, $\Gamma \boldsymbol{X}_{t-1}$ must contain the cointegration relation if it exists, see [Zivot and Wang, 2006, page 456-457] for this and the following discussion. If the bivariate $\mathrm{AR}(p)$ process has unit roots, then, using the relation $\det(\Phi(z)) = 0$ and the definition of Γ, it is clear that Γ is a singular matrix. It follows that the rank of Γ equals 0 or 1. If the rank of Γ equals 0, then $\Gamma = 0$, and $\boldsymbol{X}_t$ is $I(1)$ and not cointegrated. The VECM (34) reduces to a bivariate $\mathrm{AR}(p-1)$ in first differences

$$\Delta \boldsymbol{X}_t = \Gamma_1 \Delta \boldsymbol{X}_{t-1} + \cdots + \Gamma_{p-1} \Delta \boldsymbol{X}_{t-p+1} + \boldsymbol{Z}_t.$$

If the rank of Γ equals 1, $\boldsymbol{X}_t$ is $I(1)$ with one cointegrating vector. Since Γ has rank 1 it can be written as the product

$$\Gamma = \boldsymbol{\alpha} \boldsymbol{\beta}'$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are 2×1 vectors. Further we have that $\boldsymbol{\beta}$ is a cointegration vector for $\boldsymbol{X}_t$ and the elements in the vector $\boldsymbol{\alpha}$ are interpreted as speed of adjustment coefficients. The VECM (34) now becomes

$$\Delta \boldsymbol{X}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \boldsymbol{X}_{t-1} + \Gamma_1 \Delta \boldsymbol{X}_{t-1} + \cdots + \Gamma_{p-1} \Delta \boldsymbol{X}_{t-p+1} + \boldsymbol{Z}_t. \tag{35}$$

We end this section by considering the case of a bivariate $\mathrm{AR}(1)$ model with one cointegration vector, see [Zivot and Wang, 2006, page 457-458] and [Tsay, 2010, page 434].

Example 6.11. Consider the bivariate $\mathrm{AR}(1)$ model. From Definition 6.10 we have

$$\boldsymbol{X}_t = \Phi \boldsymbol{X}_{t-1} + \boldsymbol{Z}_t \tag{36}$$

where Φ is a 2×2 matrix and $\{\boldsymbol{Z}_t\} \sim \mathrm{WN}(\boldsymbol{0}, \mathbb{A})$.

The VECM is

$$\Delta \boldsymbol{X}_t = \Gamma \boldsymbol{X}_{t-1} + \boldsymbol{Z}_t,$$

where $\Gamma = \Phi - \mathbb{I}$.

Assuming $\boldsymbol{X}_t$ is cointegrated, there exists a 2×1 vector $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ such that $\boldsymbol{\beta}' \boldsymbol{X}_t = \beta_1 X_{1t} + \beta_2 X_{2t}$ is $I(0)$. We use the normalization $\beta_1 = 1$ and $\beta_2 = -\beta$, and the cointegration relation becomes $\boldsymbol{\beta}' \boldsymbol{X}_t = X_{1t} - \beta X_{2t}$. Since $\boldsymbol{X}_t$ is cointegrated with one cointegration vector, the rank of Γ is 1 and Γ can be decomposed as

$$\Gamma = \boldsymbol{\alpha} \boldsymbol{\beta}' = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & -\beta \end{bmatrix} = \begin{bmatrix} \alpha_1 & -\alpha_1 \beta \\ \alpha_2 & -\alpha_2 \beta \end{bmatrix}.$$

The VECM now may be written

$$\Delta \boldsymbol{X}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{X}_{t-1} + \boldsymbol{Z}_t. \tag{37}$$

We can also write the VECM equation by equation:

$$\Delta X_{1t} = \alpha_1(X_{1,t-1} - \beta X_{2,t-1}) + Z_{1t},$$
$$\Delta X_{2t} = \alpha_2(X_{1,t-1} - \beta X_{2,t-1}) + Z_{2t}.$$

Premultiplying (37) by $\boldsymbol{\beta}'$ and moving $\boldsymbol{\beta}'\boldsymbol{X}_{t-1}$ to the right-hand side of the equation gives

$$\boldsymbol{\beta}'\boldsymbol{X}_t = (1 + \boldsymbol{\beta}'\boldsymbol{\alpha})\boldsymbol{\beta}'\boldsymbol{X}_{t-1} + \boldsymbol{\beta}'\boldsymbol{Z}_t$$

or

$$U_t = \phi U_{t-1} + V_t,$$

where $U_t = \boldsymbol{\beta}'\boldsymbol{X}_t$, $\phi = 1 + \boldsymbol{\beta}'\boldsymbol{\alpha} = 1 + \alpha_1 - \beta\alpha_2$ and $V_t = \boldsymbol{\beta}'\boldsymbol{Z}_t$. This is an AR(1) model for U_t and as we know it is weakly stationary if $|\phi| < 1$ or if $|1 + \alpha_1 - \beta\alpha_2| < 1$.

6.6   The Johansen test

The prior discussion shows that the rank of Γ in (34) determines if there is a cointegration vector. In Section 6.4 we looked at a two-step procedure test for cointegration originally considered by Engle and Granger in Engle and Granger [1987]. We now consider another procedure outlined by Johansen in Johansen [1988]. Using the information in the previous section, the two basic steps in Johansen's methodology in the bivariate case are:

1. Specify and estimate a bivariate AR(p) model for $\boldsymbol{X}_t$.

2. Construct likelihood ratio tests for the rank of Γ to determine if there is a cointegration vector.

6.7   Heteroskedastic cointegration

According to Hansen [1992] it is not clear that the model of cointegration formulated in Engle and Granger [1987] (which we from now on will refer to as the CI model) is sufficiently general to cover all nonstationary economic models of interest (see [Hansen, 1992, page 139]). From [Hansen, 1992, Introduction] we read: "*The CI regression errors differ stochastically from the regressors in that they have a fixed mean and a bounded variance. The asymmetry in variance orders is intuitively unsatisfying in some cases. One*

might expect that as the regressors increase in magnitude, the residual variance would also increase. One might also expect that the variance of the error process might change over time, due to other factors. Essentially, we may wish to allow the variance of the error to be nonstationary." The empirical relevance of this idea is illustrated in [Hansen, 1992, page 140-141]. In [Hansen, 1992, page 141] one consider the process $\{W_t\}$ generated by $W_t = A_t B_t$, where $A_t \sim I(1)$ and $B_t \sim I(0)$ and call this process a bi-integrated (BI) process. We think of $\{B_t\}$ as the stationary part of $\{W_t\}$ and $\{A_t^2\}$ as the variance part of $\{W_t\}$. If $\{Y_t\}$ is generated by $Y_t = \beta X_t + W_t$, where $X_t \sim I(1)$ and $W_t \sim BI$, then we say that (Y_t, X_t) are **heteroskedastically cointegrated**. We note that

$$\mathrm{Var}(X_t) \approx C_1 t, \ 0 < C_1 < \infty,$$

$$\mathrm{Var}(W_t) \approx C_2 t, \ 0 < C_2 < \infty,$$

and are thus of the same stochastic order. This is a substantial difference from the CI model, where the variance of the regression errors is constant, see [Hansen, 1992, page 143].

6.8   Stochastic unit-root and cointegration

In the last section in [Granger and Swanson, 1997, page 58-59] stochastic unit-root and cointegration is discussed. Granger and Swanson [1997] are somewhat unclear concerning this matter. Stochastic unit-root and cointegration is an open field and few papers have been written in this area. A stochastic unit-root in the model defining the cointegration variables introduce nonlinearity in the model. It is not obvious how a possible version of the ECM generalize for such a nonlinear cointegration model. Another possibility is to use **STUR** as a substitute for the underlying $I(1)$ process in the cointegration model. This is done in [Granger and Swanson, 1997, page 58], but Granger and Swanson [1997] find this situation not very interesting.

References

Tomas Bjork. *Arbitrage Theory in Continuous Time, Third Edition*. Oxford, 2009. ISBN 978-0-19-957474-2.

Andreas Brandt. The Stochastic Equation $Y_{n+1} = A_n Y_n + B_n$ with Stationary Coefficients. *Advances in Applied Probability*, 18(1):211–220, 1986.

Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting, Second Edition*. Springer, 2002. ISBN 978-0-387-95351-9.

In Choi. *Almost All About Unit Roots*. Cambridge, 2015. ISBN 978-1-107-48250-0.

Richard A. Davis and Thomas Mikosch. The sample autocorrelations of heavy-tailed processes with applications to ARCH. *The Annals of Statistics*, 26(5):2049–2080, 1998.

Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

Charles M. Goldie. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.*, 1(1):126–166, 1991.

Clive WJ Granger and Norman R Swanson. An introduction to stochastic unit-root processes. *Journal of Econometrics*, 80(1):35–62, 1997.

Bruce E. Hansen. Heteroskedastic cointegration. *Journal of Econometrics*, 54:139–158, 1992.

Soren Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231–254, 1988.

Charles R Nelson and Charles R Plosser. Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of monetary economics*, 10(2):139–162, 1982.

Chris H. Rycroft. Introduction to Random Walks and Diffusion. *Expository papers/Lecture notes. Available at http://ocw.mit.edu/courses/mathematics/18-366-random-walks-and-diffusion-fall-2006/lecture-notes/lec01.pdf.*

Ruey S. Tsay. Unit Root and Unit-Root Test. *Expository papers/Lecture notes. Available at http://faculty.chicagobooth.edu/ruey.tsay/teaching/uts/lec11-08.pdf.*

Ruey S. Tsay. *Analysis of Financial Time Series, Third Edition*. Wiley, 2010. ISBN 978-0-470-41435-4.

Wim Vervaat. On a Stochastic Difference Equation and a Representation of Non-Negative Infinitely Divisible Random Variables. *Advances in Applied Probability*, 11(4):750–783, 1979.

John S. White. The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case. *The Annals of Mathematical Statistics*, (4):1188–1197, 1958.

Gawon Yoon. A Note on Some Properties of STUR Processes. *Oxford Bulletin of Economics and Statistics*, 68(2):253–260, 2006.

E. Zivot and J. Wang. *Modeling Financial Time Series with S-PLUS, Second Edition*. Springer, 2006.