Technische Universiteit Delft
Faculteit Elektrotechniek, Wiskunde en Informatica
Delft Institute of Applied Mathematics

Unit Root Tests voor AR(1) Processen

(Engelse titel: Unit Root Testing for AR(1) Processes)

Verslag ten behoeve van het
Delft Institute of Applied Mathematics
als onderdeel ter verkrijging

van de graad van

BACHELOR OF SCIENCE
in
TECHNISCHE WISKUNDE

door

EMMA INGELEIV WAGNER

Delft, Nederland
Januari 2015

BSc verslag TECHNISCHE WISKUNDE

"Unit Root Tests voor AR(1) Processen"
(Engelse titel: "Unit Root Testing for AR(1) Processes")

EMMA INGELEIV WAGNER

Technische Universiteit Delft

Begeleider

Dr. I.G. Becheri

Overige commissieleden

Dr.ir. M. Keijzer

Dr.ir. F.H. van der Meulen

Dr.ir. M.B. van Gijzen

Januari, 2015

Delft

Abstract

The purpose of this study is to investigate the asymptotics of a first order auto regressive unit root process, AR(1). The goal is to determine which tests could be used to test for the presence of a unit root in a first order auto regressive process. A unit root is present when the root of the characteristic equation of this process equals unity. In order to test for the presence of a unit root, we developed an understanding of the characteristics of the AR(1) process, such that the difference between a trend stationary process and a unit root process is clear.

The first test that will be examined is the Dickey-Fuller test. The estimator of this test is based on Ordinary Least Square Regression and a t-test statistic, which is why we have computed an ordinary least square estimator and the test statistic to test for the presence of unit root in the first order auto regressive process. Furthermore we examined the consistency of this estimator and its asymptotic properties. The limiting distribution of the test statistic is known as the Dickey-Fuller distribution. With a Monte Carlo approach, we implemented the Dickey-Fuller test statistic in Matlab and computed the (asymptotic) power of this test. Under the assumption of Gaussian innovations (or shocks) the limiting distribution of the unit root process is the same as without the normality assumption been made. When there is a reason to assume Gaussianity of the innovations, the Likelihood Ratio test can be used to test for a unit root.

The asymptotic power envelope is obtained with help of the Likelihood Ratio test, since the Neyman-Pearson lemma states that the Likelihood Ratio test is the point optimal test for simple hypotheses. By calculating the likelihood functions the test statistic was obtained, such that an explicit formula for the power envelope was found. Since each fixed alternative results in a different critical value and thus in a different unit root test, there is no uniform most powerful test available. Instead we are interested in asymptotically point optimal tests and we will analyze which of these point optimal tests is the overall best performing test. By comparing the asymptotic powercurve to the asymptotic power envelope for each fixed alternative we could draw a conclusion on which fixed alternative results in the overall best performing test.

On the basis of the results of this research, it can be concluded that there does not exist a uniform most powerful test, nonetheless we can define an overall best performing test.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The statistical analysis of a stationary time series is more straightforward and more studied than the analysis of a non-stationary time series. That is why in order to perform regression analysis on the data, the raw data is often transformed to a stationary process. Roughly speaking, a stationary process is a process whose statistical properties do not change over time, that is the mean and the variance are constant. For many real applications the assumption of stationarity is not valid and therefore stationarity of the data needs to be tested. One possible reason for a time series to be non-stationary is the presence of a unit root. The following example illustrates the importance of testing if a unit root is present.

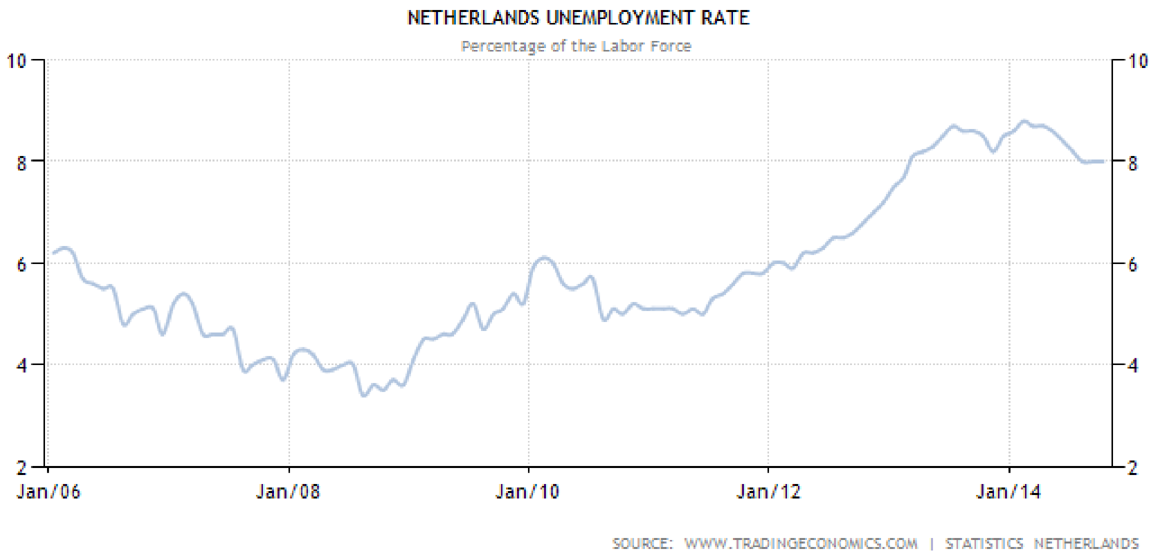


Figure 1.1: The unemployment rate of the Netherlands from 2006 to 2014.

Example 1. Consider the unemployment rate of the Netherlands. A person is considered unemployed if he or she is without a job and actively seeking for a paid job. Figure 1.1 shows the time series of the monthly unemployment rate of the Netherlands from January 2006 up till December 2014. In the field of macroeconomics there are two alternative theories on the nature of unemployment: the Natural Rate Hypothesis (NRH) and the

Hysteresis Hypothesis (HH). The NRH states that the unemployment rate fluctuates around a certain rate, the natural rate. The unemployment rate can temporarily deviate from the natural rate due to for example an exogenous shock, but in the long run it will revert to its natural rate. The NRH theory is therefore consistent with the absence of a unit root in the unemployment rate time series. Opposed to the NRH, the Hysteresis Hypothesis states that there does not exist an equilibrium level of the unemployment rate, with the consequence that a shock has a permanent effect on the unemployment rate time series. From a statistical point of view this implies that if the HH theory holds the time series contains a unit root.

Just as in the rest of the European Union, the recession starting in 2008 had a huge negative effect on the Dutch economy. One way to mitigate the economic recession in a country is to keep the inflation rate low. The positive effect of a low inflation rate is the opportunity for the labor market to adjust quickly to negative changes. Monetary authorities such as De Nederlandsche Bank (*The Dutch Bank*) have the power to keep the inflation rate stable and low. In particular, a contractionary monetary policy has the aim to reduce the growth of the money supply such that the rate of inflation will stop growing or will even shrink. However, a contractionary monetary policy does not only have the positive effect on the inflation rate. The undesirable effect of this policy is that is has the tendency to increase the unemployment rate of a country. Because of policy implications the decision whether or not the unemployment rate time series has a unit root is very important for a monetary authority. If the NRH theory holds, this means that the raise of the unemployment rate, due to a contractionary monetary policy, does not have a permanent effect and the time series will eventually revert to its natural rate. However, if the HH theory is correct, the adverse effect of the monetary policy is permanent, which will lead to a permanently higher unemployment rate in the Netherlands. Therefore we can conclude that whether or not the unemployment rate has a unit root is a key element in designing an optimal policy by The Dutch Bank. $\triangle$

Not only the unemployment rate time series could be non-stationary, in fact many time series such as exchange rates, inflation rates, real outputs etc. should be tested for stationarity. The nature of non-stationarity could differ: seasonality of the data, the presence of a deterministic trend, the presence of a stochastic trend and structural breaks are examples of non-stationarity. Seasonality, structural breaks and the presence of a deterministic trend in time series are very interesting, yet complicated on their own and the present thesis will not discuss them, since this thesis focuses on the unit root problem. Since testing for a unit root is important, this thesis will present two possible unit root tests: The Dickey-Fuller test and the Likelihood Ratio test.

If a time series is non-stationary, the standard statistical analysis is not valid. The most common methods of statistical analysis rely on the Law of Large Numbers and the Central Limit Theorem. But to be able to work with these two theorems the assumption that the time series is stationary is required, such that the standard methods to perform statistical analysis on a non-stationary time series will not be correct. This thesis is a study on the asymptotic properties of inference procedures designed for non-stationary time series. In large sample theory (asymptotic theory) the properties of an estimator and test statistic are examined when the sample size becomes indefinitely large. The idea is that the properties of an estimator when the sample size becomes arbitrarily large, are comparable to the properties of the estimator if the sample is finite.

The unit root problem is a well-studied topic in econometrics. For example Elliot, Stock and Rothenberg (1992) [4] did a study on the efficiency of tests for auto regressive unit root processes, where earlier Dickey and Fuller (1979) [1] studied the distribution of the estimators for auto regressive time series with a unit root. Of course, there are results for more complex time series than the first order auto regressive process, but they are beyond this thesis. With this thesis, we aim to examine a non standard problem in a simple case and show how non-stationarity of a time series can influence statistical analysis.

First we will give a brief introduction on the first order auto regressive process and the unit root process and will explain the difference between a trend stationary process and a unit root process. The basics of unit root testing, such as the hypotheses of interest, ordinary least square regression and the asymptotic properties of the unit root process will be dealt with in Chapter 3. The Dickey-Fuller test will be examined in Chapter 4 and Chapter 5. The Likelihood Ratio test will be dealt with in Chapter 6 as well as the computation of the asymptotic power envelope. The last chapter will summarize the results of the asymptotic properties of the Dickey-Fuller test and the Likelihood Ratio test.

Chapter 2

Introduction to Unit Root Processes

2.1 Unit Root in the AR(1) Process

In this section, we will discuss the meaning of the presence of a unit root in the first order auto regressive model, denoted as AR(1).

2.1.1 The First Order Auto Regressive Process

If the time series $\{Y_t\}$ satisfies the first order auto regressive process, denoted as AR(1) it can be described as the following iterative equation:

$$Y_t = m + \gamma t + \phi Y_{t-1} + \epsilon_t, \text{ with starting value } Y_0 \tag{2.1.1}$$

with intercept m, deterministic trend γ and innovations ϵ_t which satisfy the conditions of a White Noise process with zero mean and constant variance σ^2, i.e. $\epsilon_t \sim \text{IID } WN(0, \sigma^2)$. The ϵ_t terms are often called random shocks, innovations or errors. We assume that the ϵ_t are independent and identically distributed. We recall the definition of a White Noise Process:

Definition 2.1.1. ϵ_t is a White Noise Process if $E[\epsilon_t] = 0$, $E[\epsilon_t^2] = \sigma^2$ and $E[\epsilon_t \epsilon_s] = 0$ for all $s \neq t$.

Assumption 1. In the rest of the thesis we assume that there is no intercept present, no deterministic trend and that the initial value equals zero, i.e. $m = 0$, $\gamma = 0$ and $Y_0 = 0$.[1]

The term auto regression is being used since the equation is a linear regression model for Y_t in terms of the explanatory variable Y_{t-1}. We assume that ϵ_t is uncorrelated to the past values of the time series and thus ϵ_t represents the random contribution to Y_t. Since the equation is implicit it is useful to rewrite it to an explicit equation. First we substitute Y_{t-1} in (2.1.1) and we obtain

$$Y_t = \phi(\phi Y_{t-2} + \epsilon_{t-1}) + \epsilon_t = \phi^2 Y_{t-2} + \phi \epsilon_{t-1} + \epsilon_t. \tag{2.1.2}$$

[1] In the discussion we mention how to deal with the process in which Assumption 1 is not satisfied.

Substituting for Y_{t-2} yields

$$Y_t = \phi^2(\phi Y_{t-3} + \epsilon_{t-2}) + \phi\epsilon_{t-1} + \epsilon_t = \phi^3 Y_{t-3} + \phi^2\epsilon_{t-2} + \phi\epsilon_{t-1} + \epsilon_t. \tag{2.1.3}$$

By continuing this process we see that for $t = 0, 1, 2, ..,$

$$Y_t = \sum_{i=0}^{t} \phi^i \epsilon_{t-i}. \tag{2.1.4}$$

The factor ϕ has a strong effect on the behavior of the AR(1) process. We distinguish three cases:

- $|\phi| < 1$;
- $|\phi| > 1$;
- $\phi = 1$.

Let us consider the first case when $-1 < \phi < 1$, the observation at time t, Y_t, in (2.1.4) will be negligible when t is very large, such that the effect of shocks in the past have no significance influence on the behavior of the time series. The weight given to a shock which occured far in to the past will be extremely small. Therefore the time series has a long term mean and is stationary. On the other hand, if $|\phi| > 1$ the observation at time t, Y_t, will be large and the weight given to a shock a long time ago will be greater than the weight given to recent shocks. In the long term this process will explode. Clearly this process is non-stationary. At last let us consider the case in which $\phi = 1$. In this case the process is non-stationary and behaves as a random walk process. We will discuss this process in detail later on. Mathematically we can show (weak) stationarity of a process by computing the mean and the variance in the case $|\phi| < 1$. If $|\phi| < 1$ we can rewrite the model in the following way:

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \epsilon_t \\ Y_t &= \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots . \end{aligned} \tag{2.1.5}$$

By taking expectations we obtain

$$\begin{aligned} E[Y_t] =& E\left[\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots + \phi^t\epsilon_0\right] \\ =& E[\epsilon_t] + \phi E[\epsilon_{t-1}] + \phi^2 E[\epsilon_{t-2}] + \cdots + \phi^t E[\epsilon_0] \\ =& 0 + \phi 0 + \phi^2 0 + \cdots + \phi^t 0 \\ =& 0. \end{aligned} \tag{2.1.6}$$

Resulting the mean $E[Y_t]$ of a process in which $|\phi| < 1$ to be zero. The variance is given by

$$\begin{aligned} \mathrm{Var}(Y_t) &= \mathrm{Var}(\phi Y_{t-1} + \epsilon_t) \\ &= \mathrm{Var}(\phi Y_{t-1}) + \mathrm{Var}(\epsilon_t) \\ &= \phi^2 \mathrm{Var}(Y_{t-1}) + \sigma^2. \end{aligned} \tag{2.1.7}$$

Under the stationarity assumption $\text{Var}(Y_t) = \text{Var}(Y_{t-1})$. After substituting $\text{Var}(Y_{t-1})$ with $\text{Var}(Y_t)$ we obtain

$$\begin{aligned}\text{Var}(Y_t) &= \phi^2\text{Var}(Y_t) + \sigma^2 \Leftrightarrow (1-\phi^2)\text{Var}(Y_t) = \sigma^2 \\ \text{Var}(Y_t) &= \frac{\sigma^2}{1-\phi^2}.\end{aligned} \tag{2.1.8}$$

Since $\text{Var}(Y_t) > 0$, it follows that $1 - \phi^2 > 0$ and we see that the stationarity assumption is satisfied for $|\phi| < 1$. Therefore the process is stationary if $|\phi| < 1$. We conclude that the mean and the variance of an AR(1) process with $|\phi| < 1$ are constant and thus $\{Y_t\}$ is stationary (see Definition 2.1.2).

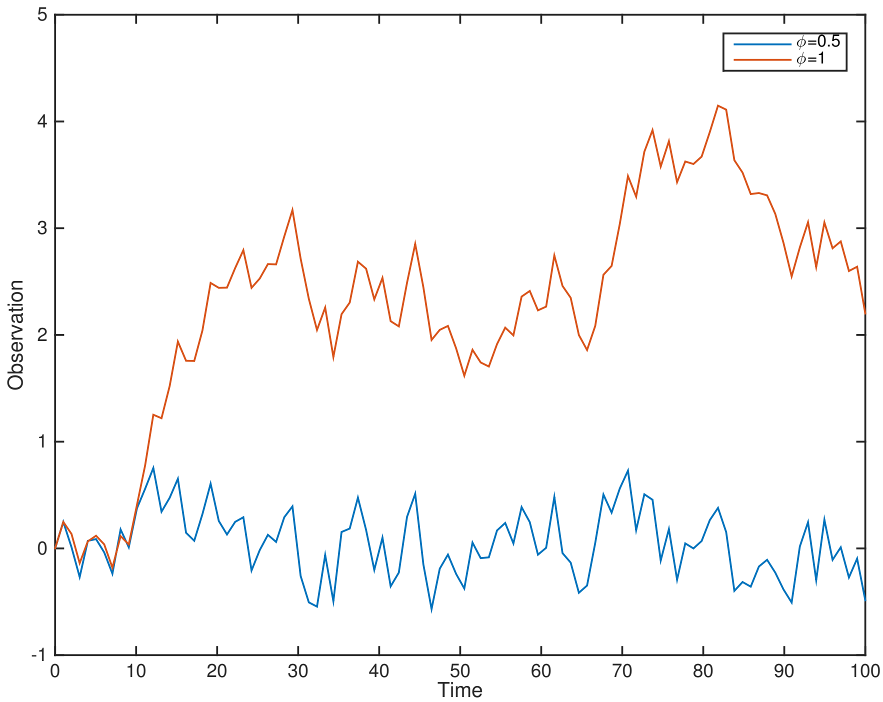


Figure 2.1: A stationary AR(1) process ($\phi = 0.5$) and a unit root process ($\phi = 1$).

Figure 2.1 indicates the difference between a stationary AR(1) process and a unit root process, since these time series are simulated with the same innovations $\{\epsilon_t\}$. In the stationary case, the shocks do not have a permanent effect, while in the unit root process, the shocks do have a permanent effect on the behavior of the process. A process can be stationary in two ways: strong/strict stationary or weak stationary. A stochastic process is weak stationary or strict stationary when it meets the following definitions:

Definition 2.1.2. A stochastic process Y_t is weak stationary when it meets the following properties:

- $E[Y_t^2] < \infty, \forall t > 0 \in \mathbb{N}$;

- $E[Y_t] = \mu, \forall t > 0 \in \mathbb{N}$;
- $Cov(Y_s, Y_t) = Cov(Y_{s+h}, Y_{t+h}), \forall s, t, h > 0 \in \mathbb{N}$.

Definition 2.1.3. A stochastic process Y_t is strict (or strong) stationary when its joint distribution $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_k})$ function equals $(Y_{t_{1+h}}, Y_{t_{2+h}}, \dots, Y_{t_{k+h}})$.

By Definition 2.1.2 of weak stationarity we conclude that the AR(1) process with $|\phi| < 1$ is a weak stationary process.

2.1.2 Presence of a Unit Root in AR(1)

In a linear model a unit root is present, if one is a root of the process' characteristic equation. In the case of a first order auto regressive process, AR(1), one way to determine if the process is stationary is to look at the roots of the characteristic equation. The characteristic equation is obtained when we express the process in lag polynomial notation. The lag operator is represented by the symbol L,

$$LY_t \equiv Y_{t-1}.$$

In the case of the AR(1) process, the lag polynomial notation results in the following equation:

$$Y_t = \phi(LY_t) + \epsilon_t \Leftrightarrow (1 - \phi L)Y_t = \epsilon_t. \tag{2.1.9}$$

The characteristic equation is obtained after substituting L by a variable (z), and setting the polynomial equation equal to zero:

$$1 - \phi z = 0. \tag{2.1.10}$$

Thus if the AR(1) process has a unit root, i.e. $z = 1$ is a root of the characteristic equation (2.1.10), ϕ must equal 1, i.e. $\phi = 1$.

$$1 - \phi z = 0 \Leftrightarrow z = \phi = 1$$

If $\phi = 1$ the process is non-stationary. This is easy to verify by computing the variance, since in the case of $\phi = 1$

$$Y_t = \sum_{i=0}^{t} \epsilon_i. \tag{2.1.11}$$

The variance of Y_t is given by

$$\text{Var}(Y_t) = \sum_{i=0}^{t} \sigma^2 = t\sigma^2. \tag{2.1.12}$$

The variance of Y_t depends on t since $\text{Var}(Y_1) = \sigma^2$, while $\text{Var}(Y_2) = 2\sigma^2$. Therefore the AR(1) process with $\phi = 1$ is a non-stationary process.

2.2 Trend Stationary Process VS Unit Root Process.

In economics, many time series are not stationary. In general we distinguish between two cases the *Trend Stationary Process* and the *Unit Root Process*:

- *Trend Stationary Process*: a non-stationary process consisting of a stationary stochastic part and a deterministic time trend.
- *Unit Root Process*: a non-stationary process with a stochastic trend or a unit root.

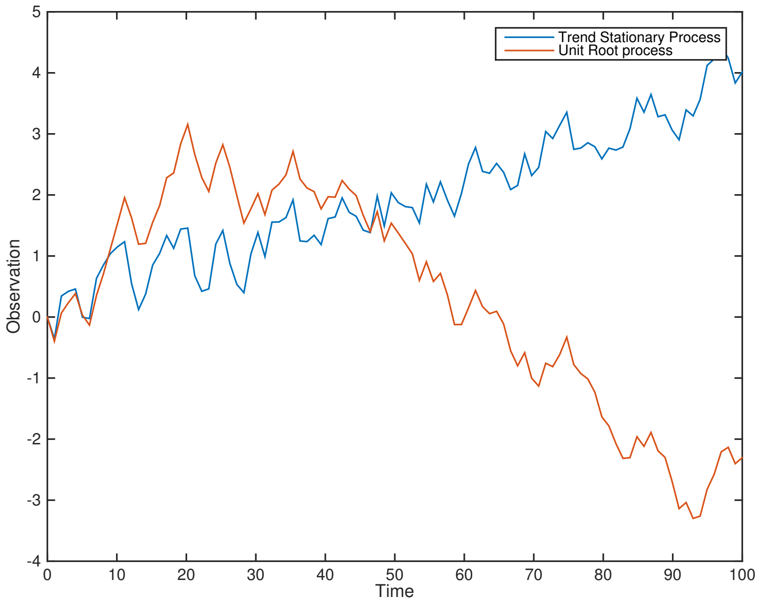


Figure 2.2: Trend stationary process compared to a unit root process.

In a time series a unit root and a deterministic trend could both be present, in that case the process satisfies equation (2.1.1) with $\phi = 1$ and $\gamma \neq 0$. In this thesis we will not analyze this special case, but in Chapter 8 we have added explanatory notes on this process. From figure 2.2 we conclude that in the trend stationary process a positive trend is present, while in the unit root process it seems that there is no deterministic trend present.

2.2.1 Trend Stationary Process

In the analysis of time series a stochastic process is called *trend stationary* if the process consists of a stochastic stationary process and a deterministic trend. Let us consider the first auto regressive process (2.1.1) with a deterministic linear trend γt and an intercept m:

$$Y_t = m + \gamma t + \phi Y_{t-1} + \epsilon_t \tag{2.2.1}$$

In the equation the additional factor γt represents the deterministic linear trend, which is independent of the stochastic term Y_t, and m represents the intercept. Let us compute the mean and variance of this trend stationary process

$$
\begin{aligned}
E[Y_t] =& E[m + \gamma t + \phi Y_{t-1} + \epsilon_t] \\
=& E[m] + E[\gamma t] + E[\phi Y_{t-1}] + E[\epsilon_t] \\
=& m + \gamma t + \phi E[Y_{t-1}] + E[\epsilon_t] \\
=& m + \gamma t + \phi E\left[\sum_{i=0}^{t-1} \phi^i \epsilon_{t-i}\right] + E[\epsilon_t] \\
=& m + \gamma t + \phi \sum_{i=0}^{t-1} \phi^i 0 + 0 \\
=& m + \gamma t
\end{aligned}
\tag{2.2.2}
$$

The mean contains a linear trend dependent on γ and it follows that the mean is not constant. We can conclude the trend stationary process is not stationary.

2.2.2 Unit Root Process

A process in which a unit root is present is called a unit root process. In particular, we say that the time series is integrated of order one, $I(1)$ (integrated of order 1) if it must be differenced 1 time in order to be stationary.

Definition 2.2.1. A time series Y_t is integrated of order 1 if $(1-L)Y_t$ is integrated of order zero, where L is the lag operator and $(1-L)$ is the first difference.

$$(1-L)Y_t = Y_t - Y_{t-1} = \Delta Y_t$$

we write $Y_t \sim I(1)$. If Y_t is stationary then Y_t is integrated of order zero, $Y_t \sim I(0)$.

For example, consider the first order auto regressive process, AR(1)

$$Y_t = \phi Y_{t-1} + \epsilon_t \text{ with starting value } Y_0.$$

If $\phi = 1$, $Y_t = Y_{t-1} + \epsilon_t$ and

$$\Delta Y_t = Y_t - Y_{t-1} = \epsilon_t.$$

Since ϵ_t has constant variance and zero mean the process is stationary, so we write $Y_t \sim I(1)$ and a unit root is present. In conclusion, we have seen that a unit root process is a non-stationary process, since the variance of this process (as obtained in equation (2.1.12)) equals $\mathrm{Var}(Y_t) = t\sigma^2$ and increases as t becomes larger.

Chapter 3

Unit Root Testing

The previous chapter discussed two types of non-stationary processes, a trend stationary process and a unit root process. In order to determine whether a time series contains a stochastic trend (a unit root) we perform a unit root test. The unit root test tests the null hypothesis of the presence of a unit root in the AR(1) process, against the alternative hypothesis that the process has no unit root and as a result is stationary. A unit root test is used to test the following hypotheses:

- H_0: $\phi = 1$ (There is a unit root present.) $\Rightarrow Y_t \sim I(1)$;
- H_1: $|\phi| < 1 \Rightarrow Y_t \sim I(0)$.

The Dickey-Fuller test statistic is based on the Ordinary Least Square estimator. In this chapter we will introduce the Ordinary Least Square (OLS) estimator for ϕ and we will specify the distribution of this OLS regressor. Furthermore we will show that this estimator is consistent.

3.1 Ordinary Least Square Regression

We consider the AR(1) process defined in (2.1.1) under Assumption 1:

$$Y_t = \phi Y_{t-1} + \epsilon_t.$$

With the method of Ordinary Least Square Regression, we are able to compute an estimator for the parameter of interest, $\hat{\phi}_T$. The method of Ordinary Least Square Regression is based on the idea that you want to minimize the sum of the squared residuals (SSR) with respect to $\hat{\phi}_T$:

$$SSR = \sum_{t=1}^{T}(Y_t - \hat{\phi}_T Y_{t-1}) = \sum_{t=1}^{T} \epsilon_t^2,$$

where T represents the sample size. The goal is to find the value $\hat{\phi}_T$ that minimizes the SSR,

$$SSR = \sum_{t=1}^{T} \epsilon_t^2 = \sum_{t=1}^{T} (Y_t - \hat{\phi}_T Y_{t-1})^2, \tag{3.1.1}$$

$$\frac{\partial(SSR)}{\partial \hat{\phi}_T} = \frac{\partial}{\partial \hat{\phi}_T} \left[\sum_{t=1}^{T} (Y_t - \hat{\phi}_T Y_{t-1})^2 \right], \tag{3.1.2}$$

$$\frac{\partial(SSR)}{\partial \hat{\phi}_T} = \sum_{t=1}^{T} \left[\frac{\partial}{\partial \hat{\phi}_T} (Y_t - \hat{\phi}_T Y_{t-1})^2 \right] = \sum_{t=1}^{T} -2\hat{\phi}_T Y_{t-1}(Y_t - \hat{\phi}_T Y_{t-1}). \tag{3.1.3}$$

In order to minimize (3.1.3) we set the partial derivative equal to zero and we can solve (3.1.3) for $\hat{\phi}_T$:

$$\frac{\partial(SSR)}{\partial \hat{\phi}_T} = 0 \Leftrightarrow \sum_{t=1}^{T} -2\hat{\phi}_T Y_{t-1}(Y_t - \hat{\phi}_T Y_{t-1}) = 0 \tag{3.1.4}$$

and we can solve (3.1.4) for $\hat{\phi}_T$

$$\sum_{t=1}^{T} -2\hat{\phi}_T Y_{t-1} Y_t - \sum_{t=1}^{T} -2\hat{\phi}_T^2 Y_{t-1}^2 = 0 \Leftrightarrow \sum_{t=1}^{T} -2\hat{\phi}_T Y_{t-1} Y_t = \sum_{t=1}^{T} -2\hat{\phi}_T^2 Y_{t-1}^2. \tag{3.1.5}$$

If we take the value $-2\hat{\phi}_T$ out of the summation, we obtain

$$-2\hat{\phi}_T \sum_{t=1}^{T} Y_{t-1} Y_t = -2\hat{\phi}_T^2 \sum_{t=1}^{T} Y_{t-1}^2. \tag{3.1.6}$$

If we divide both sides by $-2\hat{\phi}_T$ we obtain

$$\sum_{t=1}^{T} Y_{t-1} Y_t = \hat{\phi}_T \sum_{t=1}^{T} Y_{t-1}^2 \tag{3.1.7}$$

such that $\hat{\phi}_T$ equals

$$\hat{\phi}_T = \frac{\sum_{t=1}^{T} Y_{t-1} Y_t}{\sum_{t=1}^{T} Y_{t-1}^2}. \tag{3.1.8}$$

The Ordinary Least Square estimator $\hat{\phi}_T$ which minimizes the sum of the squared residuals is defined by

$$\hat{\phi}_T = \frac{\sum_{t=1}^{T} Y_{t-1} Y_t}{\sum_{t=1}^{T} Y_{t-1}^2}. \tag{3.1.9}$$

Since we calculated an estimator for ϕ by the method of Ordinary Least Square Regression, it is time to focus on the asymptotic properties of this estimator. By the Central Limit Theorem (A.1) we know that for $T \to \infty$ if the process is stationary ($\phi < 1$),

$$\sqrt{T}(\hat{\phi}_T - \phi) \xrightarrow{d} \mathcal{N}(0, (1 - \phi^2)). \tag{3.1.10}$$

Under the null hypothesis Y_t is a non-stationary process, thus we are interested in the asymptotic distribution of $\phi = 1$. If we work under the null hypothesis, we simply can not use the Central Limit Theorem in the way we used it earlier. Note that for $\phi = 1$ (3.1.10) implies

$$\sqrt{T}(\hat{\phi}_T - \phi) = \sqrt{T}(\hat{\phi}_T - 1) \to \mathcal{N}(0, 0) = 0 \tag{3.1.11}$$

and we obtain a degenerate limiting distribution. A degenerate distribution is a probability function of a discrete random variable which consists of one single value. In our case it resulted in a distribution centered around zero with zero variance. Our aim is to find a non-degenerate asymptotic distribution for $\hat{\phi}_T$ under the null hypothesis and therefore we have to make a change to the OLS estimator (3.1.9). It turns out that we need to multiply $\hat{\phi}_T$ by T rather than by $\sqrt{T}$. To show why scaling with T is needed under H_0 (i.e. $\phi = 1$), we define the difference between $\hat{\phi}_T$ and ϕ as

$$\hat{\phi}_T - \phi = \hat{\phi}_T - 1 = \frac{\sum_{t=1}^T Y_{t-1}\epsilon_t}{\sum_{t=1}^T Y_{t-1}^2}. \tag{3.1.12}$$

If we multiply (3.1.12) with T and after substituting $T = \frac{\frac{1}{T}}{\frac{1}{T^2}}$ we obtain

$$T(\hat{\phi}_T - 1) = T\left(\frac{\sum_{t=1}^T Y_{t-1}Y_t}{\sum_{t=1}^T Y_{t-1}^2} - 1\right) = \left(\frac{\frac{1}{T}\sum_{t=1}^T Y_{t-1}Y_t}{\frac{1}{T^2}\sum_{t=1}^T Y_{t-1}^2} - T\right). \tag{3.1.13}$$

If we replace 1 with

$$1 = \frac{\sum_{t=1}^T Y_{t-1}^2}{\sum_{t=1}^T Y_{t-1}^2} \tag{3.1.14}$$

and substitute in (3.1.13), we obtain

$$\begin{aligned} T(\hat{\phi}_T - 1) &= \left(\frac{\frac{1}{T}\sum_{t=1}^T Y_{t-1}Y_t}{\frac{1}{T^2}\sum_{t=1}^T Y_{t-1}^2} - T\left(\frac{\sum_{t=1}^T Y_{t-1}^2}{\sum_{t=1}^T Y_{t-1}^2}\right)\right) \\ &= \frac{\frac{1}{T}\sum_{t=1}^T Y_{t-1}[Y_t - Y_{t-1}]}{\frac{1}{T^2}\sum_{t=1}^T Y_{t-1}^2} \\ &= \frac{\frac{1}{T}\sum_{t=1}^T \Delta Y_t Y_{t-1}}{\frac{1}{T^2}\sum_{t=1}^T Y_{t-1}^2}. \end{aligned} \tag{3.1.15}$$

Recall that we work under the null hypothesis, such that we can substitute ΔY_t by ϵ_t,

$$\Delta_t = Y_t - Y_{t-1} = \epsilon_t$$

therefore (3.1.15) yields

$$T(\hat{\phi}_T - 1) = \frac{\frac{1}{T}\sum_{t=1}^{T} \epsilon_t Y_{t-1}}{\frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1}^2}. \tag{3.1.16}$$

With this result we can now examine if the limiting distribution of this estimator is non-degenerate. In the next section we will examine the asymptotic properties of the OLS estimator and we will show that by scaling $(\hat{\phi}_T - 1)$ with T we obtain a non-degenerate distribution of the OLS estimator.

3.2 Asymptotic Properties of an AR(1) Process under the Null Hypothesis

Under the null hypothesis the process Y_t is best described by a random walk. A random walk is a stochastic process that is a sequence of random steps, for example the fluctuating price of a stock is often modeled as a random walk.

$$Y_t = Y_{t-1} + \epsilon_t = Y_0 + \epsilon_1 + \epsilon_2 + \ldots + \epsilon_t. \tag{3.2.1}$$

Since we assume $Y_0 = 0$ the process is the sum of random IID innovations. The aim is to find the asymptotic properties of the AR(1) process with $\phi = 1$ by means of the asymptotic properties of a random walk. First we will introduce the Wiener Process. A Wiener Process is a continuous time stochastic process satisfying the following definition:

Definition 3.2.1 (Standard Brownian Motion (Wiener Process)). [1]
A continuous time stochastic process $\{W(t)\}_{t\geq 0}$ is called a Standard Brownian Motion (Wiener Process) when it meets the following properties:

- $W(0) = 0$;
- For any dates $0 \leq t_1 < t_2 < \cdots < t_k \leq 1$, the increments $[W(t_2) - W(t_1)], [W(t_3) - W(t_2)], \cdots, [W(t_k) - W(t_{k-1})]$ are independent and Gaussian for any collection of points $0 \leq t_1 < t_2 < \cdots < t_{k-1} < t_k$ and integer $k > 2$.
- $W(t+s) - W(t) \sim N(0, s)$ for $s > 0$.

The Wiener Process is also known as the Standard Brownian Motion. The Wiener Process is highly related to a random walk. According to Donsker's Theorem or the Functional Central Limit Theorem (A.2) the discrete random walk approaches a Standard Brownian Motion if the number of steps in the random walk increases ($t \to \infty$) and the step size becomes smaller. As a result, the Wiener Process is the scaling limit of a random walk. The following proposition holds for a random walk:

Proposition 3.2.1 (Convergence of a random walk). *Suppose ψ_t a random walk,*

$$\psi_t = \psi_{t-1} + u_t$$

where u_t is IID with zero mean and constant variance σ^2, such that the following properties hold:

[1] *Time Series Analysis*- James D. Hamilton [8].

1. $\frac{1}{T}\sum_{t=1}^{T} u_t\psi_{t-1} \xrightarrow{d} \sigma^2\int_0^1 W(t)dW(t) = \frac{1}{2}\sigma^2\left(W(1)^2-1\right)$

2. $\frac{1}{T^2}\sum_{t=1}^{T}\psi_{t-1}^2 \xrightarrow{d} \sigma^2\int_0^1 W(t)^2dt$

Where $\{W(t)\}$ defines a Wiener Process and $\xrightarrow{d}$ defines convergence in distribution.

As we have shown in (3.1.16) the deviation of the OLS estimator from the actual value ϕ satisfies

$$T(\hat{\phi}_T-1) = \frac{\frac{1}{T}\sum_{t=1}^{T}\epsilon_t Y_{t-1}}{\frac{1}{T^2}\sum_{t=1}^{T}Y_{t-1}^2}. \tag{3.2.2}$$

Under the null hypothesis Y_t describes a random walk, with IID innovations ϵ_t with zero mean and constant variance σ^2. By Proposition 3.2.1 and the Continuous Mapping Theorem A.3 we may conclude that the asymptotic distribution of (3.2.2) is defined as:

$$T(\hat{\phi}_T-1) = \frac{\frac{1}{T}\sum_{t=1}^{T}\epsilon_t Y_{t-1}}{\frac{1}{T^2}\sum_{t=1}^{T}Y_{t-1}^2} \xrightarrow{d} \frac{\frac{1}{2}\left(W(1)^2-1\right)}{\int_0^1 W(t)^2dt}. \tag{3.2.3}$$

If the assumption of Gaussian innovations is valid, $\epsilon_t \sim \mathcal{N}(0,\sigma^2)$, the asymptotic properties of $T(\phi-1)$ under the null hypothesis are the same as when the innovations satisfy a White Noise process. We will illustrate this by computing the asymptotic properties of the random walk for Gaussian innovations.

3.3 Asymptotic Properties of the AR(1) Random Walk Process with Gaussian Innovations

Consider the AR(1) process (2.1.1) under the null hypothesis satisfying Assumption 1.

$$Y_t = \sum_{s=1}^{t}\phi^s\epsilon_{t-s} = \sum_{s=1}^{t}\epsilon_{t-s} = \epsilon_t + \epsilon_{t-1}\epsilon_{t-2} + \cdots + \epsilon_1. \tag{3.3.1}$$

When we make the extra assumption of $\epsilon_t \sim \mathcal{N}(0,\sigma^2)$, the process $\{Y_t\}_{t\geq 0}$ is the sum of Gaussian random variables. Therefore (3.3.1) implies that Y_t is Gaussian with zero mean and variance $t\sigma^2$:

$$Y_t \sim \mathcal{N}(0, t\sigma^2). \tag{3.3.2}$$

Since Y_t represents a random walk, we can write the squared random walk process Y_t^2 in the following way:

$$Y_t^2 = (Y_{t-1}+\epsilon_t)^2 = Y_{t-1}^2 + 2Y_{t-1}\epsilon_t + \epsilon_t^2 \Leftrightarrow Y_{t-1}\epsilon_t = \frac{Y_t^2 - Y_{t-1}^2 - \epsilon_t^2}{2}. \tag{3.3.3}$$

We are interested in the sum of all the squared observations of the process $\{Y_t\}_{t\geq 0}$, so if we sum (3.3.3) from 1 to T we obtain

$$\begin{aligned}\sum_{t=1}^{T} Y_{t-1}\epsilon_t =& \sum_{t=1}^{T} \frac{Y_t^2 - Y_{t-1}^2 - \epsilon_t^2}{2} \\ =& \frac{\sum_{t=1}^{T} Y_t^2}{2} - \frac{\sum_{t=1}^{T} Y_{t-1}^2}{2} - \frac{\sum_{t=1}^{T} \epsilon_t}{2} \\ =& \frac{Y_1^2 + Y_2^2 + \ldots + Y_{T-1}^2 + Y_T^2}{2} - \frac{Y_0^2 + Y_1^2 + \ldots + Y_{T-1}^2}{2} - \frac{\epsilon_1 + \epsilon_2 + \ldots + \epsilon_{T-1}^2 + \epsilon_T^2}{2} \\ =& \frac{Y_T^2 - Y_0^2}{2} - \frac{\sum_{t=1}^{T} \epsilon_t}{2}.\end{aligned} \tag{3.3.4}$$

Assumption 1 states that $Y_0 = 0$, such that

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\epsilon_t = \frac{1}{2T}Y_T^2 - \frac{1}{2T}\sum_{t=1}^{T}\epsilon_t^2, \tag{3.3.5}$$

and when we divide each side of (3.3.5) by σ^2 we obtain

$$\frac{1}{\sigma^2 T}\sum_{t=1}^{T} Y_{t-1}\epsilon_t = \frac{1}{2\sigma^2 T}Y_T^2 - \frac{1}{2\sigma^2 T}\sum_{t=1}^{T}\epsilon_t^2 = \frac{1}{2}\left(\frac{Y_T}{\sigma\sqrt{T}}\right)^2 - \frac{1}{2\sigma^2 T}\sum_{t=1}^{T}\epsilon_t^2. \tag{3.3.6}$$

Since $Y_t \sim \mathcal{N}(0, \sigma^2 t)$ it implies that $\frac{Y_T}{\sigma\sqrt{T}} \sim \mathcal{N}(0,1)$. Then by definition $\left(\frac{Y_T}{\sigma\sqrt{T}}\right)^2$ follows a Chi-Squared distribution,

$$\left(\frac{Y_T}{\sigma\sqrt{T}}\right)^2 \sim \mathcal{X}^2(1).$$

Let us have a look at the term $\sum_{t=1}^{T}\epsilon_t^2$. This is a sum of squared IID normal random variables with zero mean and constant variance σ^2, such that by the Law of Large Numbers [2]:

$$\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \xrightarrow{p} \sigma^2. \tag{3.3.7}$$

Combining the previous results, we have shown that

$$\frac{1}{\sigma^2 T}\sum_{t=1}^{T} Y_{t-1}\epsilon_t \xrightarrow{d} \frac{1}{2}\left(\overline{Y} - 1\right) \tag{3.3.8}$$

where $\overline{Y} \sim \mathcal{X}^2(1)$.

As a result we have found the limiting distribution of the numerator of equation (3.2.3) for Gaussian innovations. By the definition of a Wiener Process it follows that $W(1)^2 = \mathcal{X}^2(1)$, which implies that

[2] J. Doob, Stochastic Processes, John Wiley & Sons, 1953 [3]

$$\frac{1}{2}\left(W(1)^2-1\right)=\frac{1}{2}\left(\mathcal{X}^2(1)-1\right), \tag{3.3.9}$$

such that the limiting distribution of the numerator of equation (3.2.3) for Gaussian and non Gaussian innovations are indeed the same. By a similar argument as in section 3.2 we can conclude from the Continuous Mapping Theorem (A.3) and Proposition 3.2.1 that the limiting distribution of $T(\hat{\phi}_T - 1)$ satisfies:

$$T(\hat{\phi}_T-1)=\frac{\frac{1}{T}\sum_{t=1}^{T}\epsilon_t Y_{t-1}}{\frac{1}{T^2}\sum_{t=1}^{T}Y_{t-1}^2}\xrightarrow{d}\frac{\frac{1}{2}\left(\mathcal{X}^2(1)-1\right)}{\int_0^1 W(t)^2dt}. \tag{3.3.10}$$

3.4 Consistency of $\hat{\phi}_T$

An estimator is called consistent if it has the property that if the sample size increases, the resulting estimate will converge in probability to the value being investigated. Thus if $T \to \infty$ in the AR(1) process, $\hat{\phi}_T$ is consistent if $\text{p-}\lim_{T\to\infty}\hat{\phi}_T = \phi$, i.e. $\text{p-}\lim_{T\to\infty}\hat{\phi}_T - 1 = 0$. From (3.2.3) we can conlude that $\hat{\phi}_T$ is a super consistent estimator at the rate T for $\phi = 1$. Since we have found that the asymptotic distribution for $T \to \infty$ is

$$T(\hat{\phi}_T-1)\xrightarrow{d}\frac{\int_0^1 W(t)dW(t)}{\int_0^1 W(t)^2dt}. \tag{3.4.1}$$

If we divide (3.4.1) by T and subtract $\frac{\int_0^1 W(t)dW(t)}{\int_0^1 W(t)^2dt}$ we obtain

$$(\hat{\phi}_T-1)-\frac{1}{T}\frac{\int_0^1 W(t)dW(t)}{\int_0^1 W(t)^2dt}. \tag{3.4.2}$$

Since we know that the asymptotic distribution is non-degenerate it follows that as $T \to \infty$

$$\frac{1}{T}\frac{\int_0^1 W(t)dW(t)}{\int_0^1 W(t)^2dt}\xrightarrow{d}0$$

such that $\hat{\phi}_T - 1$ from equation (3.4.1) converges in probability to zero:

$$(\hat{\phi}_T-1)-\frac{1}{T}\frac{\int_0^1 W(t)dW(t)}{\int_0^1 W(t)^2dt}\xrightarrow{d}0. \tag{3.4.3}$$

Hence we can conclude that the OLS estimator $\hat{\phi}_T$ is a super consistent estimator for the real value ϕ. With the asymptotic properties of this estimator we are able to examine the asymptotic distribution of a test statistic in order to test for the presence of a unit root. The test was defined by David Dickey and Wayne Fuller in 1979. In the next chapter we will investigate the Dickey-Fuller test and explain how we can use it to test for a unit root. For the special case in which the innovations are Gaussian, $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$, we will construct the Likelihood Ratio test and approximate the asymptotic power of this test.

Chapter 4

The Dickey-Fuller Test

In 1979 Wayne Fuller and David Dickey developed a test to examine whether there is a unit root present in a first order auto regressive process $\{Y_t\}_{t\geq 0}$. This test is named after the two statisticians and is known as the Dickey-Fuller test. The Dickey-Fuller test studies the presence of a unit root in the first order auto regressive process (2.1.1), even if Assumption 1 is not valid. The consideration to include intercept and/or trend results in three possible auto regressive processes:

- Testing for a unit root.
$$Y_t = \phi Y_{t-1} + \epsilon_t$$
- Testing for a unit root with drift.
$$Y_t = m + \phi Y_{t-1} + \epsilon_t$$
- Testing for a unit root with drift and deterministic time trend.
$$Y_t = m + \phi Y_{t-1} + \gamma t + \epsilon_t$$

Each model results in a different test statistic and different critical values for the Dickey-Fuller test. That is why it is important for practical reasons to select the correct underlying auto regressive process before performing a unit root test. Clearly, the first two processes are simplifications of the third more general auto regressive process. The AR(1) process under Assumption 1 corresponds to the first model and we will perform a Monte Carlo simulation to obtain the (asymptotic) critical values for the Dickey-Fuller test statistic introduced in the next section.

4.1 Dickey-Fuller Test Statistic

Consider the first order auto regressive process under Assumption 1:

$$Y_t = \phi Y_{t-1} + \epsilon_t \tag{4.1.1}$$

where $\epsilon_t \sim WN(0, \sigma^2)$. In this section we will examine the OLS estimator and we will introduce the test statistic that we use for testing whether there is a unit root present. In Chapter 3 we found the OLS estimator $\hat{\phi}_T$ to be

$$\hat{\phi}_T = \frac{\sum_{t=1}^{T} Y_{t-1} Y_t}{\sum_{t=1}^{T} Y_{t-1}^2}.$$

Let us define the standard t-statistic t_T

$$t_T = \frac{\hat{\phi}_T - 1}{\hat{\sigma}_{\hat{\phi}_T}} \tag{4.1.2}$$

where $\hat{\sigma}_{\hat{\phi}_T}$ is the usual OLS standard error of the estimator,

$$\hat{\sigma}_{\hat{\phi}_T} = \sqrt{\frac{s_T^2}{\sum_{t=1}^{T} Y_{t-1}^2}}$$

and s_T^2 is residual variance of the OLS estimator

$$s_T^2 = \frac{\sum_{t=1}^{T} (Y_t - \hat{\phi}_T Y_{t-1})^2}{T-1}$$

for a fixed significance level α. The test rejects the null hypothesis (the presence of a unit root) for small values of the test statistic t_T. If the test-statistic t_T is smaller than the corresponding critical value k_T^α, the null hypothesis of a unit root is rejected. Critical values k_T^α can be obtained with Monte Carlo simulation. If we simulate an AR(1) process of sample size T with a unit root, i.e. $\phi = 1$, we can compute the test statistic t_T and by repeating these steps for a total number of N replications, we will obtain an approximation to the distribution of t_T under the null hypothesis. The critical values can be found with this distribution, since the critical value k_T^α satisfies

$$Pr(t_T \leq k_T^\alpha | H_0) = \alpha \tag{4.1.3}$$

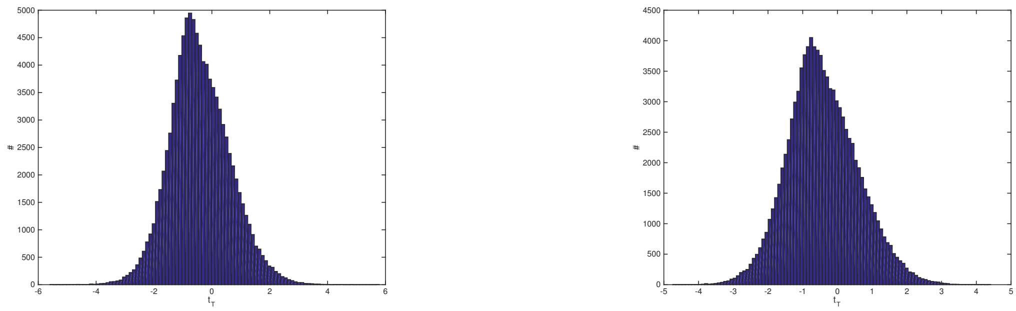
for a fixed significance level α and sample size T. Since this thesis focuses on the asymptotics of a unit root process, we are interested in the asymptotic critical values k_∞^α of the test statistic. In order to find asymptotic critical values for the test statistic t_∞ it is necessary to find the limiting distribution of the test statistic t_T. We have defined the t-statistic in the standard way, but as we know the limiting behavior is not Gaussian. However, we can use the properties obtained in paragraph 3.2 to examine the limiting distribution of the test statistic. Since $\hat{\phi}_T$ is a super consistent estimator for ϕ, it implies that $s_T^2 \xrightarrow{p} \sigma^2$ for $T \to \infty$. Such that the test statistic t_T for $T \to \infty$ has limiting distribution

$$t_T = \frac{\hat{\phi}_T - 1}{\hat{\sigma}_{\hat{\phi}_T}} \xrightarrow{d} \frac{\frac{1}{2}\sigma^2 \left(W(1)^2 - 1\right)}{\sqrt{\left(\sigma^2 \int_0^1 W(t)^2 dt\right)}\sqrt{\sigma^2}} = \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\sqrt{\int_0^1 W(t)^2 dt}}. \tag{4.1.4}$$

The distributions (3.2.3) and (4.1.4) are known as Dickey-Fuller distributions, since David Dickey and Wayne Fuller developed the asymptotics of this unit root test.

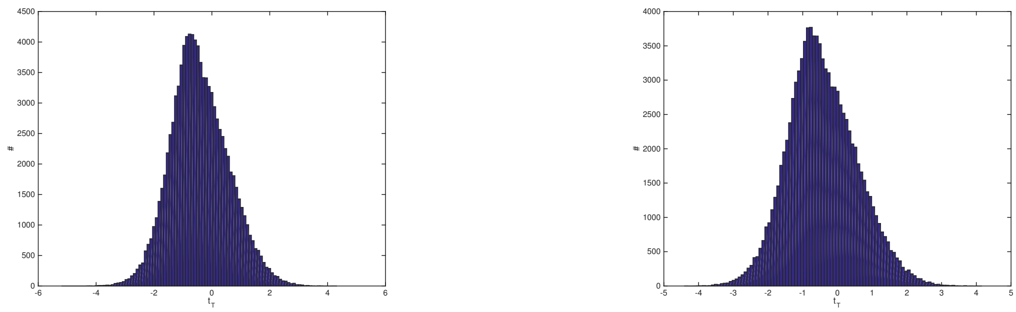
4.2 Critical Values of the Dickey-Fuller Distribution

With the Monte Carlo approach we simulated a distribution for the t-statistic (4.1.2) and are able calculate the critical values for the test statistic t_T. Monte Carlo simulation is based on the idea of repeated random sampling in order to approximate the underlying distribution. Under the null hypothesis of a unit root, we repeatedly ($N = 50,000$ times) sampled a first order auto regressive process of length T to approximate the distribution of t_T. Figure 4.1 and Figure 4.2 show the outline of the distribution of the t-statistic for the sample sizes $T = 25, 50, 100, 250$.



(a) $T = 25$ (b) $T = 50$

Figure 4.1: The approximation of the distributions of t_T for sample sizes $T = 25$ and $T = 50$.



(a) $T = 100$ (b) $T = 250$

Figure 4.2: The approximation of the distributions of t_T for sample sizes $T = 100$ and $T = 250$.

As we can see in both Figure 4.1 and Figure 4.2 the distribution of the t-statistic t_T is positively skewed. The finite sample critical values are obtained by sorting the results of the finite sample Monte Carlo simulation and determining the critical value at significance level α. Table 4.1 shows the critical values k_T^α of the test statistic t_T for several different sample sizes and significance levels α. The asymptotic distribution of the test statistic t_T is known as the Dickey-Fuller distribution (4.1.4). Instead of simulating the data from the AR(1) process, we obtain the asymptotic critical values of t_∞ by sampling a Wiener Process and approximate the asymptotic distribution of the test statistic t_∞. Figure 4.3

shows an approximation of the distribution of t_∞, which has been used to approximate the asymptotic critical values of the Dickey-Fuller distribution listed in Table 4.1.

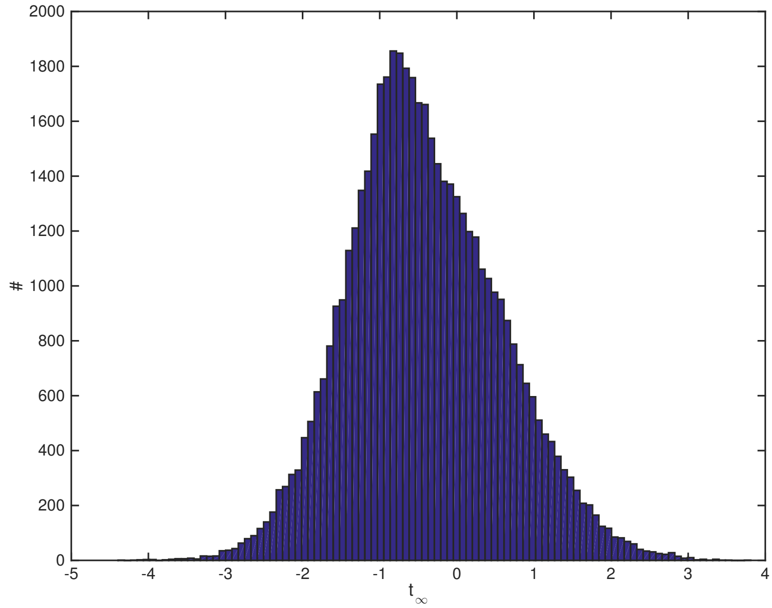


Figure 4.3: The simulated distribution of t_∞.

T	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$
25	-2.7509	-2.3280	-2.0026
50	-2.6409	-2.2697	-1.9682
100	-2.6000	-2.2592	-1.9645
250	-2.5815	-2.2403	-1.9521
∞	-2.5775	-2.2281	-1.9312

Table 4.1: Critical values k_T^α of t_T for several significance levels α and sample sizes T.

By acquiring the critical values of the Dickey-Fuller test, we are able to examine the power of the Dickey-Fuller test at a certain significance level α. The next chapter examines the power of the Dickey-Fuller test for several values of ϕ and different sample sizes T of the AR(1) process (2.1.1) under Assumption 1. These powers will be evaluated at significance level $\alpha = 0.05$, such that we will only consider the critical values listed in the last column of Table 4.1. We will not only look into the power of the finite sample process, but also examine the asymptotic power when the parameter ϕ is close to unity. Obviously, if we examine the finite sample power we will use the finite sample critical values and if we examine the asymptotic power we will be using the asymptotic critical value $k_\infty^\alpha = -1.9312$.

Chapter 5

Power of the Dickey-Fuller Test

Power analysis provides information on how good the test is to detect an effect. An effect is the difference between the value of the parameter under the null hypothesis ($\phi_0 = 1$) and the actual true value (ϕ_1). Since we will discuss the performance of the Dickey-Fuller test it is useful to check the power of the test under different circumstances. We will analyze the power of the test for several sample sizes and effect sizes. The power of a statistical test is the probability that the null hypothesis is rejected at a fixed significance level α when in fact the null hypothesis is false. Which is equivalent to correctly accepting the alternative hypothesis. Thus

	H_0 is true $\phi = 1$	H_1 is true $\phi = \phi_1 < 1$
We accept H_0	correct conclusion	type II error
We reject H_0	type I error	correct conclusion

$$\text{Power} = \Pr(H_0 \text{ is rejected } | H_0 \text{ is false}) = \Pr(H_0 \text{ is rejected } | \phi = \phi_1 < 1). \tag{5.0.1}$$

The power of the Dickey-Fuller test at significance level $\alpha = 0.05$ corresponds to

$$\begin{aligned} \text{Power} =& \Pr(t_T < k_T^{0.05} | \phi < 1) \\ =& 1 - \Pr(t_T > k_T^{0.05} | \phi < 1). \end{aligned} \tag{5.0.2}$$

The statistical power is dependent on the sample size. The power is often used to calculate the sample size needed to detect an effect of a given size α. To enlarge the power of a statistical test, one possible option is to increase the sample size. By means of the finite sample critical values listed in Table 4.1 we can obtain the power of the Dickey-Fuller test. With Monte Carlo simulation we approximated the power of the Dickey-Fuller test at significance level $\alpha = 0.05$. We calculated the power of the test for AR(1) processes under Assumption 1, for different alternative hypotheses $H_1 : \phi = \phi_1$ and several sample sizes $T = 25, T = 50, T = 100$ and $T = 250$. For the sake of clarity we define ϕ_1 as the parameter of the AR(1) process we simulated from and $\phi = 1$ as the value of the parameter under the null hypothesis.

Since the power of the test is dependent on the difference between the real parameter ϕ_1 and $\phi = 1$, we compute for a sequence of alternatives ϕ_1's the power of the Dickey-Fuller test. In Table 5.1 we see that for $\phi_1 = 0.5$ the power of the test is very close to 1, which is very high. But if we consider $\phi_1 = 0.9$ we conclude that the power of the test is low. The probability of a type II error β, satisfying Power $= 1 - \beta$, is very high, in other words the probability of failing to reject the null hypothesis is large.

ϕ_1	$T = 25$	$T = 50$	$T = 100$	$T = 250$
0.5	0.907	1.000	1.000	1.000
0.6	0.774	0.996	1.000	1.000
0.7	0.566	0.977	1.000	1.000
0.8	0.335	0.7606	0.999	1.000
0.9	0.153	0.322	0.752	1.000
1	0.050	0.057	0.045	0.051

Table 5.1: The power of the Dickey-Fuller test for finite sample sizes T at significance level $\alpha = 0.05$.

As shown in table 5.1 if the distance $|\phi_1 - \phi|$ gets smaller, the power decreases. To illustrate this decreasing power figure 5.1 shows the power of the Dickey-Fuller test at significance level $\alpha = 0.05$ for several sample sizes T.

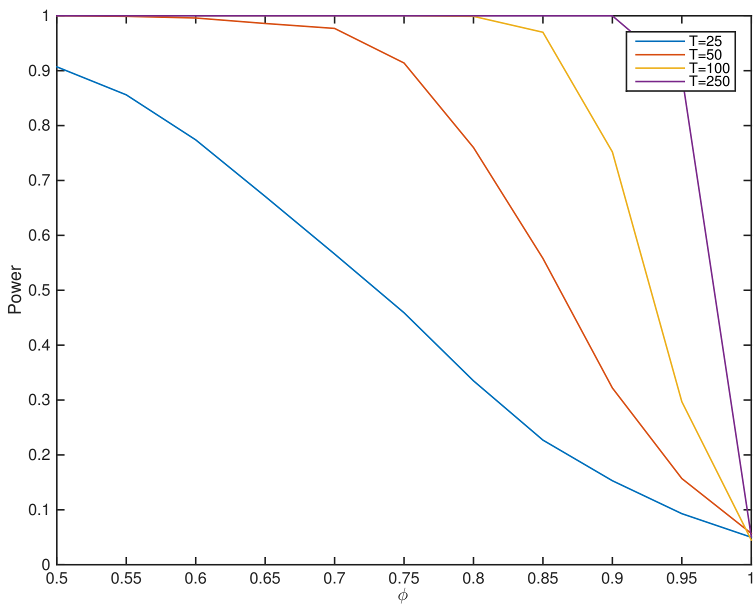


Figure 5.1: The power of the Dickey-Fuller test at significance level $\alpha = 0.05$ for $T = 25, 50, 100, 250$.

Figure 5.1 illustrates the importance of unit root testing. When the value ϕ_1 approaches unity, the Dickey-Fuller test does not perform the way we would like it to perform; the power of the test for values of ϕ_1 close to unity is low. However, if the distance

$|\phi_1 - 1|$ is large, there is no need to test for non-stationarity since the time series will show stationary properties. Therefore the area of interest is the area of ϕ_1 close to unity and we will examine the properties of the Dickey-Fuller test for a sequence of ϕ_1's close to 1. Since we have examined the asymptotic properties of the Dickey-Fuller test, we assume that we can add data such that the sample size grows indefinitely. To shrink the area of interest, a small neighborhood of unity, we introduce the local alternative framework. The local alternative framework is used to shrink the neighborhood of unity as the sample size T grows indefinitely.

5.1 Local Alternatives Framework

Since we are interested in the power of the Dickey-Fuller test for parameters ϕ_1 in a small neighborhood of unity, we will define the local alternatives framework. We define $\phi_1 = 1 + \frac{c_1}{T}$. Such that

$$\lim_{T\to\infty} \phi_1 = \lim_{T\to\infty} \left(1 + \frac{c_1}{T}\right) = 1,$$

where $c_1 < 0 \in \mathbb{Z}$ which specifies the fixed local alternative. By introducing local alternatives we obtain the power of the Dickey-Fuller test for values extremely close to unity. For example, if we work with a sample of 1000 observations[1], $T = 1000$, and consider the local alternative $c_1 = -1$, we can obtain the power of the Dickey-Fuller test for $\phi_1 = 1 - \frac{1}{1000} = 0.999$, which is very close to unity. The results of the simulation of this example for the power of the Dickey-Fuller test with ϕ close to unity is shown in Figure 5.2.

5.2 AR(1) with Gaussian Innovations

In the previous chapters and analysis we assumed the random shocks or innovations to be independent identically distributed White Noise, i.e. $\epsilon_t \sim WN(0, \sigma^2)$. Sometimes it makes sense to assume Gaussianity of the innovations. In that case we can do further power analysis. Therefore the assumption of Gaussianity gives us an advantage. Which is why it is nice to check whether the Dickey-Fuller test performs in the same way for Gaussian innovations, $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$, as for White Noise innovations $\epsilon_t \sim WN(0, \sigma^2)$. In order to compare the two types of innovations we will examine the power of the Dickey-Fuller test for standard normal innovations and innovations with a known constant variance $\sigma^2 \neq 1$. Statisticians White (1958) and Hasza (1977) provided us with the knowledge that the asymptotic properties and the limiting distribution of the estimator $\hat{\phi}$ is the same for non normal innovations and Gaussian innovations and therefore we can use the limiting distribution and the critical values we have obtained in Chapter 4. From Section 3.3 we know that the asymptotic properties of the test statistic t_T are the same for Gaussian innovations, such that we can use the same asymptotic critical values as in Table 4.1.

[1]Note that the existence of a time series in econometrics with a unit root of 1000 observations is very unlikely since the data is often listed quarterly, $T = 1000$ would result in the approximation of the asymptotic power of the Dickey-Fuller test.

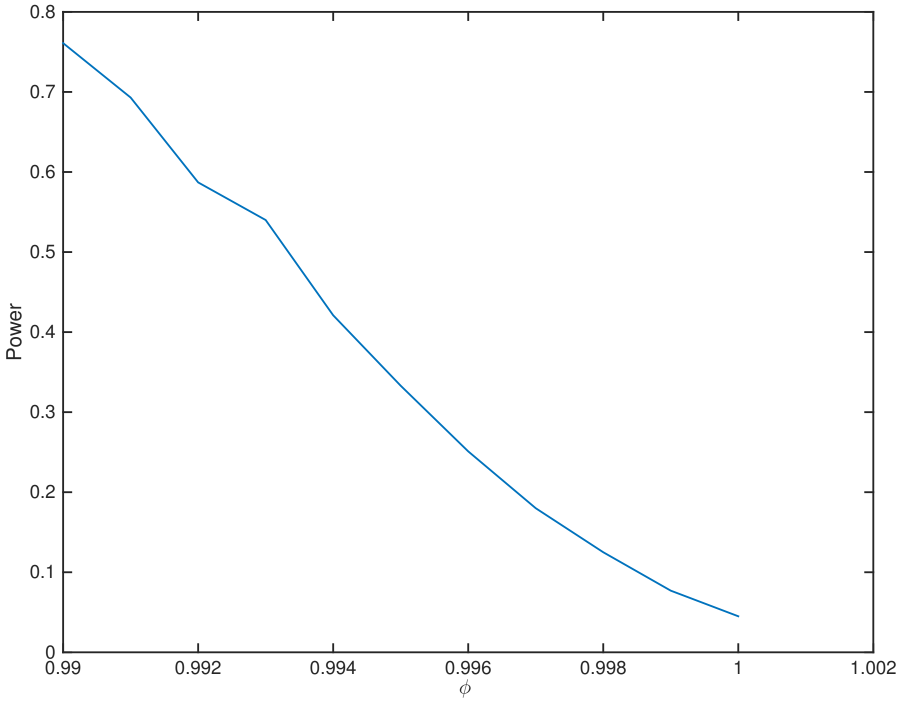


Figure 5.2: Power of the Dickey-Fuller test for local alternatives $\phi_1 = 1 + \frac{c_1}{T}$ close to unity.

5.2.1 $\epsilon_t \sim \mathcal{N}(0, 1)$

In this section, we will examine the power of the Dickey-Fuller test when the innovations are standard normally distributed, $\epsilon_t \sim \mathcal{N}(0, 1)$. Monte Carlo Simulation is used to perform the Dickey-Fuller test. To be able to compare the power of the Dickey-Fuller test simulated with Gaussian innovations with the power of the Dickey-Fuller test simulated with White Noise innovations, we will sample an AR(1) process for several local alternatives $\phi_1 = 1 + \frac{c_1}{T}$ close to unity at nominal significance level $\alpha = 0.05$ for different sample sizes T.

$-c_1$	$T = 50$	$T = 100$	$T = 250$
0	0.047	0.050	0.058
2	0.116	0.111	0.127
4	0.236	0.241	0.226
6	0.412	0.406	0.417
8	0.599	0.601	0.554
10	0.785	0.763	0.716

Table 5.2: The power of the Dickey-Fuller test for $\mathcal{N}(0, 1)$ innovations at nominal significance level $\alpha = 0.05$.

5.2.2 $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$

Since the innovations could also be normally distributed with mean zero and a known constant variance $\sigma^2 \neq 1$, it is interesting to examine the power of the Dickey-Fuller test of this special case. By Monte Carlo simulation we have obtained Table 5.3 which consists of the power of the Dickey-Fuller test for several sample sizes T and local alternatives $c_1 < 0 \in \mathbb{Z}$ such that $\phi_1 = 1 + \frac{c_1}{T}$. For simulation purposes, we performed Monte Carlo simulation with known variance $\sigma^2 = 4$.

$-c_1$	$T = 50$	$T = 100$	$T = 250$
0	0.047	0.042	0.042
2	0.117	0.111	0.126
4	0.227	0.249	0.224
6	0.418	0.420	0.425
8	0.617	0.590	0.582
10	0.748	0.740	0.760

Table 5.3: The power of the Dickey-Fuller test for $\mathcal{N}(0, \sigma^2 = 2^2)$ innovations at nominal significance level $\alpha = 0.05$.

By Monte Carlo simulation we have obtained the power for different types of innovations and we can compare these values in order to see which type of innovations produces the highest power of the Dickey-Fuller test. Table 5.4 illustrates the difference between the power of the three types of innovations for several fixed local alternatives $c_1 < 0 \in \mathbb{Z}$. From Table 5.4 we can conclude that the times series simulated with standard normal innovations leads to the least biased power estimation. We have fixed the significance level at $\alpha = 0.05$ which means that we allow the probability of falsely rejecting the null hypothesis to be 5% (probability on a type I error). As we can see in the first line in Table 5.4 the power of the Dickey-Fuller test for the three types of innovations do not correspond to the significance level $\alpha = 0.05$. Therefore we can conclude that the results are biased. Since the probability on a type I error for the AR(1) process simulated with standard normal innovations is the closest to 0.05, we conclude that the power of the Dickey-Fuller test for this type of innovations is the least biased.

If the data gives a reason to assume the innovations to be Gaussian, it gives the opportunity to implement a different unit root test. If indeed the innovations are IID and Gaussian, we are able to implement the Likelihood Ratio test for unit roots. In the next chapter we will look into the Likelihood Ratio test and we will examine the power of this unit root test. Furthermore with the Likelihood Ratio test we can compute the asymptotic power envelope for unit root tests.

$-c_1$	$WN(0, \sigma^2)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0, \sigma^2)$
0	0.042	0.0540	0.063
2	0.106	0.136	0.128
4	0.230	0.274	0.244
6	0.399	0.415	0.419
8	0.579	0.625	0.580
10	0.761	0.761	0.750

Table 5.4: The power of the Dickey-Fuller test for several innovations and large sample size $T = 500$ with nominal significance level $\alpha = 0.05$.

Chapter 6

The Likelihood Ratio Test

Section 5.2 approximates the asymptotic power of the Dickey-Fuller test under the assumption that the innovations are Gaussian. The Monte Carlo simulation of the AR(1) process is done with Gaussian innovations rather than White Noise innovations. In order to perform further power analysis on the AR(1) process with Gaussian innovations, we can implement the Likelihood Ratio test. This test can also be implemented for other types of innovations, but in that case the computation of the Likelihood Ratio would become rather difficult. Instead we stick to Gaussian innovations which are rather common in applications. By means of the Likelihood Ratio test, we can compute the asymptotic power envelope for unit root tests. The asymptotic power envelope is a great tool to compare unit root tests to the maximum asymptotically attainable power. Therefore, within this chapter the Likelihood Ratio test will be defined and it will be explained how to compute the power envelope as well as the asymptotic power envelope for unit root tests.

The Likelihood Ratio test is based on the likelihood of two models, the first model defines the model under the null hypothesis and the second model defines the model under the alternative hypothesis. Both models will be fitted to the time series' data and the likelihood functions will be calculated in order to determine which model is more likely to be true (hence the name Likelihood Ratio test). If the model under the null hypothesis is more likely to be true, it results in a large test statistic, denoted as $\Lambda_T(\cdot)$ (dependent on the sample size T) and the null hypothesis is rejected for small values of the test statistic $\Lambda_T(\cdot)$. In the unit root problem the null hypothesis and the simple alternative hypothesis are the following:

$$H_0 : \phi = 1 \text{ such that } Y_t = Y_{t-1} + \epsilon_t$$
$$H_1 : \phi = \phi_1 < 1 \text{ such that } Y_t = \phi_1 Y_{t-1} + \epsilon_t$$

The likelihood ratio test rejects the null hypothesis if

$$\Lambda_T(Y_t) = \frac{L(\phi = 1|Y_1, \dots, Y_T)}{L(\phi = \phi_1|Y_1, \dots, Y_T)} = \frac{f(Y_1, \dots, Y_T|\phi = 1)}{f(Y_1, \dots, Y_T|\phi = \phi_1)} \leq l_T^\alpha \tag{6.0.1}$$

where $L(\phi = 1|Y_1, \dots, Y_T)$ defines the likelihood function of the model under the null hypothesis, $L(\phi = \phi_1|Y_1, \dots, Y_T)$ represents the likelihood function of the model under the alternative hypothesis and l_T^α denotes the critical value at significance level α satisfying

$$\Pr(\Lambda_T(Y_t) \leq l_T^\alpha) = \alpha.$$

6.1 Computing the Likelihood Functions

First we calculate the likelihood functions: $L(\phi = \phi_0|Y_1, \ldots, Y_T)$ & $L(\phi = \phi_1|Y_1, \ldots, Y_T)$. The trick is to write the process $\{Y_t\}_{t\geq 0}$ in terms of the innovations $\epsilon_t = Y_t - \phi Y_{t-1}$. Since we assumed the innovations independent identically distributed and Gaussian with zero mean and known constant variance σ^2, we can easily compute the likelihood functions. Let us compute the likelihood functions of the hypotheses of interest:

$$\begin{aligned} f(\epsilon_1, \ldots, \epsilon_T|\phi = 1) =& \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma^2}\right) \\ =&(2\pi\sigma^2)^{-T/2} \exp\left(-\frac{\sum_{t=1}^{T} \epsilon_t^2}{2\sigma^2}\right) \\ =&(2\pi\sigma^2)^{-T/2} \exp\left(-\frac{\sum_{t=1}^{T}(Y_t - Y_{t-1})^2}{2\sigma^2}\right), \end{aligned} \tag{6.1.1}$$

$$\begin{aligned} f(\epsilon_1, \ldots, \epsilon_T|\phi = \phi_1) =& \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma^2}\right) \\ =&(2\pi\sigma^2)^{-T/2} \exp\left(-\frac{\sum_{t=1}^{T} \epsilon_t^2}{2\sigma^2}\right) \\ =&(2\pi\sigma^2)^{-T/2} \exp\left(-\frac{\sum_{t=1}^{T}(Y_t - \phi_1 Y_{t-1})^2}{2\sigma^2}\right). \end{aligned} \tag{6.1.2}$$

With these likelihood functions we will obtain the Likelihood Ratio test statistic $\Lambda_T(Y_t)$ (6.0.1):

$$\begin{aligned} \Lambda_T(\epsilon_t) =& \frac{L(\phi = 1|\epsilon_1, \ldots, \epsilon_T)}{L(\phi = \phi_1|\epsilon_1, \ldots, \epsilon_T)} = \frac{f(\epsilon_1, \ldots, \epsilon_T|\phi = 1)}{f(\epsilon_1, \ldots, \epsilon_T|\phi = \phi_1)} \\ =& \frac{(2\pi\sigma^2)^{-T/2} \exp\left(-\frac{\sum_{t=1}^{T}(Y_t - Y_{t-1})^2}{2\sigma^2}\right)}{(2\pi\sigma^2)^{-T/2} \exp\left(-\frac{\sum_{t=1}^{T}(Y_t - \phi_1 Y_{t-1})^2}{2\sigma^2}\right)} \\ =& \frac{\exp\left(-\frac{\sum_{t=1}^{T}(Y_t - Y_{t-1})^2}{2\sigma^2}\right)}{\exp\left(-\frac{\sum_{t=1}^{T}(Y_t - \phi_1 Y_{t-1})^2}{2\sigma^2}\right)} \\ =& \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{t=1}^{T}(Y_t - Y_{t-1})^2 - \sum_{t=1}^{T}(Y_t - \phi_1 Y_{t-1})^2\right]\right) \\ =& \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left[Y_t^2 - 2Y_tY_{t-1} + Y_{t-1}^2 - Y_t^2 + 2\phi_1 Y_t Y_{t-1} - \phi_1^2 Y_{t-1}^2\right]\right) \\ =& \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left[(2\phi_1 - 2)Y_tY_{t-1} + (1 - \phi_1^2)Y_{t-1}^2\right]\right). \end{aligned} \tag{6.1.3}$$

By substituting $\phi_1 = 1 + \frac{c_1}{T}$ in (6.1.3) we obtain

$$\begin{aligned}
\Lambda_T(\epsilon_t) =& \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left[(2\left(1+\frac{c_1}{T}\right)-2)Y_tY_{t-1} + (1-\left(1+\frac{c_1}{T}\right)^2)Y_{t-1}^2\right]\right) \\
=& \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left[\frac{2c_1}{T}Y_tY_{t-1} - \left(\frac{2c_1}{T}-\frac{c_1^2}{T^2}\right)Y_{t-1}^2\right]\right) \\
=& \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left[\frac{2c_1}{T}Y_{t-1}(Y_t - Y_{t-1}) - \frac{c_1^2}{T^2}Y_{t-1}^2\right]\right) \\
=& \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left[\frac{2c_1}{T}Y_{t-1}\Delta Y_t - \frac{c_1^2}{T^2}Y_{t-1}^2\right]\right).
\end{aligned} \tag{6.1.4}$$

The resulting Likelihood Ratio function statistic is defined by

$$\Lambda_T(\epsilon_t) = \exp\left(-\frac{1}{\sigma^2}\sum_{t=1}^{T}\left[\frac{c_1}{T}Y_{t-1}\Delta Y_t - \frac{1}{2}\left(\frac{c_1}{T}\right)^2 Y_{t-1}^2\right]\right). \tag{6.1.5}$$

The log Likelihood Ratio function is defined as the natural logarithm of the Likelihood Ratio:

$$\begin{aligned}
\log\left[\Lambda_T(\epsilon_t)\right] =& -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left[\frac{c_1}{T}Y_{t-1}\Delta Y_t - \frac{1}{2}\left(\frac{c_1}{T}\right)^2 Y_{t-1}^2\right] \\
=& -\frac{c_1}{\sigma^2 T}\sum_{t=1}^{T}Y_{t-1}\Delta Y_t + \frac{c^2}{2T^2\sigma^2}\sum_{t=1}^{T}Y_{t-1}^2.
\end{aligned} \tag{6.1.6}$$

We can simplify (6.1.6) by substituting:

$$A_T = \frac{1}{T\sigma^2}\sum_{t=2}^{T}Y_{t-1}\Delta Y_t,$$

$$B_T = \frac{1}{T^2\sigma^2}\sum_{t=2}^{T}Y_{t-1}^2,$$

and we obtain the test statistic of the Likelihood Ratio test:

$$\log\left[\Lambda_T(\epsilon_t)\right] = -c_1 A_T + \frac{1}{2}c_1^2 B_T. \tag{6.1.7}$$

As a result we have obtained the test statistic for the log Likelihood Ratio test. The test rejects the null hypothesis in favor of the process being stationary, for small values of (6.1.7). Since the alternative hypothesis is dependent on the fixed alternative c_1 such that each local alternative corresponds to its own critical value $l_T^\alpha(c_1)$ at significance level α. Let us define the set of $M \in \mathbb{N}$ negative local alternatives

$$C = \{c_1, c_2, \ldots, c_M | c_i \in \mathbb{Z}_{<0} \ \&\ c_i > c_{i+1}\ \forall i \leq M\}.$$

Such that each critical value $l_T^\alpha(c_i)$ corresponds to the fixed local alternative $c_i \in C$.

6.2 Asymptotic Critical Values of the Likelihood Ratio Test

The previous section resulted in the find of the log likelihood ratio test statistic. Since each critical value $l_T^\alpha(c_i)$ corresponds to a fixed local alternative $c_i \in C$ being tested, we have to determine the critical values corresponding to the M fixed alternatives. The likelihood ratio test rejects the null hypothesis in favor of the fixed alternative c_i for small values of the test statistic $\log[\Lambda_T(Y_t)]$, the critical values satisfy

$$\Pr(\log[\Lambda_T(Y_t)] \leq l_T^\alpha(c_i)) = \alpha \tag{6.2.1}$$

where we fix the significance level at $\alpha = 0.05$. Since we are interested in the asymptotics of the unit root process, we do not calculate the exact critical value $l_T^\alpha(c_i)$, but we approximate the asymptotic critical value $l_\infty^\alpha(c_i)$, which is the approximation for $l_T^\alpha(c_i)$ when T becomes large. We simulate $N = 50,000$ times from the first order auto regressive process under the null hypothesis ($\phi_1 = 1$) and for each replication we will calculate the test statistic and after we have sorted the vector filled with the test statistics, we obtain the critical value corresponding to the fixed alternative. This procedure will be repeated for the set of fixed alternatives C to obtain the critical values we need to perform the power analysis.

6.3 The Power Envelope

This section discusses the power envelope for unit root tests. Elliot, Stock and Rothenberg (1996) [5] did a study on the power envelope for unit root testing and produced a straightforward analysis to derive the power envelope for the class of unit root tests. The following section will give the outline of the analysis introduced by Elliot et al. (1996) to calculate the power envelope for unit root tests. The power envelope is a useful tool for examining power properties of a statistical test. In unit root testing, the power envelope returns the attainable upperbound on the power of the class of unit root tests at significance level α.

If we define the set of functions $R = \{\rho_T : \rho_T \text{ is a unit root test}\}$ as the class of unit root tests at significance level $\alpha = 0.05$, the Dickey-Fuller test and the Likelihood Ratio test are both members of this class. These unit root tests, such as the Dickey-Fuller test and the Likelihood Ratio test, can be defined in the form of a function

$$\rho_T : \overline{Y} = (Y_1, \ldots, Y_T) \to [0, 1],$$

which returns the probability that the test $\rho_T(\overline{Y})$ rejects the null hypothesis, subject to the test statistic calculated from the observations $\overline{Y} = (Y_1, \ldots, Y_T)$. The power of a unit root test ρ_T is the expectation under the alternative hypothesis of $\rho_T(\overline{Y})$, $\mathrm{E}_{\phi_i = 1 - \frac{c_i}{T}}\left[\rho_T(\overline{Y})\right]$. We define the power envelope as the function $\Pi_T^\alpha(\cdot)$ which is the maximum attainable power of the class of unit root tests at significance level $\alpha = 0.05$, subject to the alternative hypothesis $c = c_i \in C$. Since the power envelope calculates the maximum attainable power, it gives a good standard to which power properties of several tests can be compared. Earlier we computed the power of the Dickey-Fuller test for several sample sizes T and for several (local-to-unity) fixed alternatives $c_i \in C$ such that $\phi_i = 1 + \frac{c_i}{T}$. The power envelope is defined by:

$$\Pi_T^\alpha(\phi) = \max_{\rho_T : \mathrm{E}_1(\rho_T(\overline{Y}))=\alpha} \mathrm{E}_{\phi_i}(\rho_T(\overline{Y})) \tag{6.3.1}$$

where ρ_T is a unit root test from the class, $\overline{Y}$ defines the time series of length T and ϕ_i the fixed alternative. With help of the Neyman-Pearson lemma (explained in Appendix A.4), there is a simple way to derive the power envelope. The Neyman-Pearson lemma states that the Likelihood Ratio test is the most powerful test for two simple hypotheses $H_0 : \phi = 1$ and $H_1 : \phi = \phi_i < 1$. In section 6.1 we have calculated the Likelihood Ratio test statistic and by the lemma we conclude that the point optimal unit root test rejects the null hypothesis for small values of

$$\log\left[\Lambda_T(\epsilon_t)\right] = -T(\phi_i - 1)A_T + \frac{1}{2}[T(\phi_i - 1)]^2 B_T. \tag{6.3.2}$$

This results in an explicit formula for the power envelope $\Pi_T^\alpha(\phi_i)$

$$\begin{aligned}\Pi_T^\alpha(\phi_i) =& \max_{\rho_T : \mathrm{E}_1(\rho_T(\overline{Y}))=\alpha} \mathrm{E}_{\phi_i}(\rho_T(\overline{Y})) \\ =& \Pr_{\phi_i}\left[-T(\phi_i - 1)A_T + \frac{1}{2}[T(\phi_i - 1)]^2 B_T \le l_T^\alpha(\phi_i)\right]\end{aligned} \tag{6.3.3}$$

with A_T and B_T defined in (6.1.7), $l_T^\alpha(\phi_i)$ the critical value corresponding to the fixed alternative $\phi_i = 1 + \frac{c}{T}$ at significance level $\alpha = 0.05$, which satisfies

$$\Pr_{\phi=1}\left[-T(\phi_i - 1)A_T + \frac{1}{2}[T(\phi_i - 1)]^2 B_T \le l_T^\alpha(\phi_i)\right] = \alpha$$

For every fixed alternative ϕ_i we can compute the corresponding $l_T^\alpha(\phi_i)$ and calculate the maximum power $\Pi_T^\alpha(\phi_i)$, such that the power envelope is pointwise obtainable. The result is a sequence of most powerful test depending on the alternative being considered for test size $\alpha = 0.05$. The optimal test against the alternative $\phi = \phi_i < 1$ is dependent on the fixed alternative ϕ_i and as as result there does not exist a uniformly most powerful test at significance level α. As we shall see later on this also holds for the asymptotic power envelope.

6.4 The Asymptotic Power Envelope

In the power envelope for the finite sample process, we concluded that there does not exist a uniform most powerful test for size α since the power envelope depends on the local-to-unity fixed alternative $\phi_i = 1 + \frac{c_i}{T}$ being evaluated. In this section we will discuss and derive the asymptotic power envelope for unit root tests. The asymptotic power envelope gives an asymptotically attainable upperbound on the local-to-unity asymptotic power of a sequence of unit root tests and is defined by

$$\lim_{T\to\infty} \Pi_T^\alpha(c_i) = \Pi_\infty^\alpha(c_i). \tag{6.4.1}$$

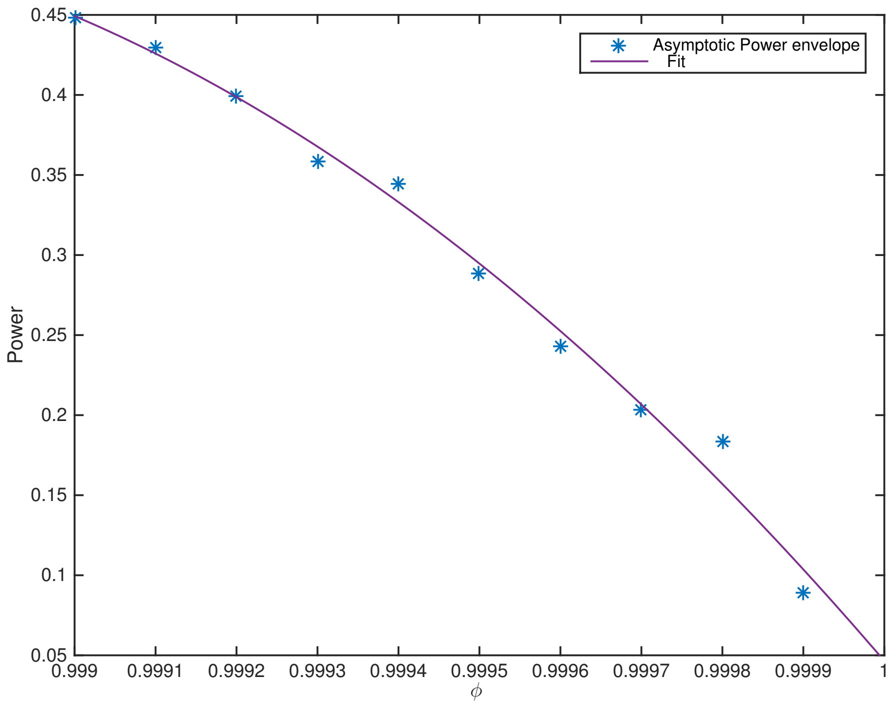


Figure 6.1: The power of the test close to unity corresponds to the asymptotic size $\alpha = 0.05$.

Figure 6.1 shows the asymptotic power envelope for local alternatives extremely close to unity. The figure shows that the size is just achieved asymptotically, such that we can no longer speak of the exact test size, but of the *asymptotic size* of the test. The power of the test for local alternatives close to unity approaches the value 0.05 which corresponds with the test size $\alpha = 0.05$[1]. If we assume that the limit (6.4.1) exists, there is an explicit formula for the asymptotic power envelope:

$$\Pi_{\infty}^{\alpha}(c_i) = \lim_{T\to\infty}\left(\Pr_{c_i}\left[-c_iA_T + \frac{1}{2}c_i^2B_T \leq l_{\infty}^{\alpha}(c_i)\right]\right). \tag{6.4.2}$$

Elliot et al. have proven that the asymptotic power envelope equals:

$$\begin{aligned}\Pi_{\infty}^{\alpha}(c_i) =& \lim_{T\to\infty}\left(\Pr_{c_i}\left[-c_iA_T + \frac{1}{2}c_i^2B_T \leq l_{\infty}^{\alpha}(c_i)\right]\right)\\ =&\Pr\left[-c_i\int_0^1 W_{c_i}(r)dW(r) - \frac{1}{2}c_i^2\int_0^1 W_{c_i}(r)dr \leq l_{\infty}^{\alpha}(c_i)\right].\end{aligned} \tag{6.4.3}$$

Equation (6.4.3) is obtained as follows:

[1]Since we are not able to examine the power of the test for $c = 0$, we extrapolated a quadratic fit in the figure to show the power of the test for $c_i \to 0$. Figure 6.1 illustrates that the power approaches 0.05 when $c_i \to 0$.

$$
\begin{aligned}
\Pi_T^\alpha(\phi_i) =&\mathrm{Pr}_{\phi_i}\left[-T(\phi_i-1)A_T+\frac{1}{2}[T(\phi_i-1)]^2B_T\le l_T^\alpha(\phi_i)\right]\\
=&\mathrm{Pr}_{c_i}\left[-c_iA_T+\frac{1}{2}c_i^2B_T\le l_T^\alpha(c_i)\right]\\
=&\mathrm{Pr}_{c_i}\left[-c_iA_T+c_i^2B_T-\frac{1}{2}c_i^2B_T\le l_T^\alpha(c_i)\right]\\
=&\mathrm{Pr}_{c_i}\left[-c_i\left(A_T-c_iB_T\right)-\frac{1}{2}c_i^2B_T\le l_T^\alpha(c_i)\right].
\end{aligned}
\tag{6.4.4}
$$

Elliot et al. provided us with the asymptotic distribution of the expressions for A_T and B_T:

- $A_T - c_iB_T \xrightarrow{d} \int_0^1 W_{c_i}(t)dW(t)$;
- $B_T \xrightarrow{d} \int_0^1 W_{c_i}(t)^2dt$.

Such that the asymptotic power envelope (6.4.3) yields

$$
\begin{aligned}
\Pi_T^\alpha(c_i) =&\mathrm{Pr}_{c_i}\left[-c_i\left(A_T-c_iB_T\right)-\frac{1}{2}c_i^2B_T\le l_T^\alpha(c_i)\right]\\
&\xrightarrow{L}\\
\Pi_\infty^\alpha(c_i) =&\mathrm{Pr}_{c_i}\left[-c_i\int_0^1 W_{c_i}(t)dW(t)-\frac{1}{2}c_i^2\int_0^1 W_{c_i}(t)^2dt\le l_\infty^\alpha(c_i)\right].
\end{aligned}
\tag{6.4.5}
$$

Where $W(t)$ indicates a Wiener Process and $W_{c_i}(t)$ denotes an Ornstein-Uhlenbeck Process which satisfies the following stochastic differential equation:

$$
dW_{c_i}(t) = -c_iW_{c_i}(t)dt + dW(t) \tag{6.4.6}
$$

with initial value $W_{c_i}(0) = 0$.

6.5 Analytic Solution to the Ornstein-Uhlenbeck Process

An Ornstein-Uhlenbeck process is a stochastic continuous time process. In order to construct the asymptotic power envelope, we need to simulate a number of Ornstein Uhlenbeck processes to approximate the asymptotic power for each fixed alternative $c_i \in C$. To simplify the Matlab simulation, we shall solve the stochastic differential equation (6.4.6) to obtain the analytic solution and we will simulate an Ornstein-Uhlenbeck process by the analytic solution $W_{c_i}(t)$. The general Ornstein-Uhlenbeck process satisfies the following stochastic differential equation[2]:

[2]Since we defined the fixed local alternative as $\phi_i = 1 + \frac{c_i}{T}$, $c_1 < 0$ we could substitute k with $-c_i$ in the solution of the Ornstein-Uhlenbeck process.

$$dW_{c_i}(t) = k(\mu - W_{c_i}(t))dt + \sigma dW(t). \tag{6.5.1}$$

In (6.5.1) $W_{c_i}(\cdot)$ defines an Ornstein-Uhlenbeck process, $W(\cdot)$ a Wiener process, μ is the long term mean of the Ornstein-Uhlenbeck process, $\sigma = 1$ and $c_i \in C$ is the fixed alternative after substituting it by $k = -c_i$. μ is the long term mean, which means that over time the process will tend to drift towards this value μ. If $\mu \neq 0$ the data is not centered, therefore we will substitute $Y(t) = W_{c_i}(t) - \mu$. The process Y_t satisfies the stochastic differential equation

$$dY(t) = dW_{c_i}(t) = -kY(t)dt + \sigma dW(t). \tag{6.5.2}$$

From equation (6.5.2) we can conclude that $Y(t)$ will converge to zero at exponential rate, thus it seems a good idea to perform the following variable transformation:

$$Z(t) = e^{-kt}Y(t) \Leftrightarrow Y(t) = e^{kt}Z(t) \tag{6.5.3}$$

such that the stochastic differential equation for $Z(t)$ satisfies[3].

$$\begin{aligned} dZ(t) =& ke^{kt}Y(t)dt + e^{kt}dY(t) \\ =& ce^{kt}Y(t)dt + e^{kt}\left[-kY(t)dt + \sigma dW(t)\right] \\ =& e^{kt}\sigma dW(t). \end{aligned} \tag{6.5.4}$$

By Itô integration of $Z(t)$ we obtain the following solution for $Z(t)$

$$Z(t) = Z(s) + \int_s^t \sigma e^{ku} dW(u). \tag{6.5.5}$$

If we now substitute $Z(t) = e^{kt}Y(t)$ we obtain

$$\begin{aligned} e^{kt}Y(t) =& e^{-ks}Y(s) + \int_s^t \sigma e^{ku} dW(u) \\ &\Leftrightarrow \\ Y(t) =& e^{-k(t-s)}Y(s) + \int_s^t \sigma e^{-k(t-u)} dW(u). \end{aligned} \tag{6.5.6}$$

As a result, after back substitution of $Y(t) = W_{c_i}(t) - \mu$ and $k = -c_i$ we have found the analytic solution to the Ornstein-Uhlenbeck stochastic differential eqaution:

$$W_{c_i}(t) = \mu + e^{c_i(t-s)}(W_{c_i}(t) - \mu) + \int_s^t \sigma e^{c_i(t-u)} dW(u). \tag{6.5.7}$$

Note that (6.4.6) is a simplified version of the general Ornstein-Uhlenbeck stochastic differential equation (6.5.1) with $\mu = 0$ and $\sigma = 1$, by means of this solution the simulation of an Ornstein-Uhlenbeck process in Matlab is straightforward. With the solution to this process we can calculate the asymptotic power envelope. Figure 6.2 illustrates the asymptotic power envelope for fixed local alternatives $c_i \in C$. The graph shows for each alternative c_i the maximum asymptotically attainable power of the class of unit root tests.

[3]By the product rule for Itô integrals [14].

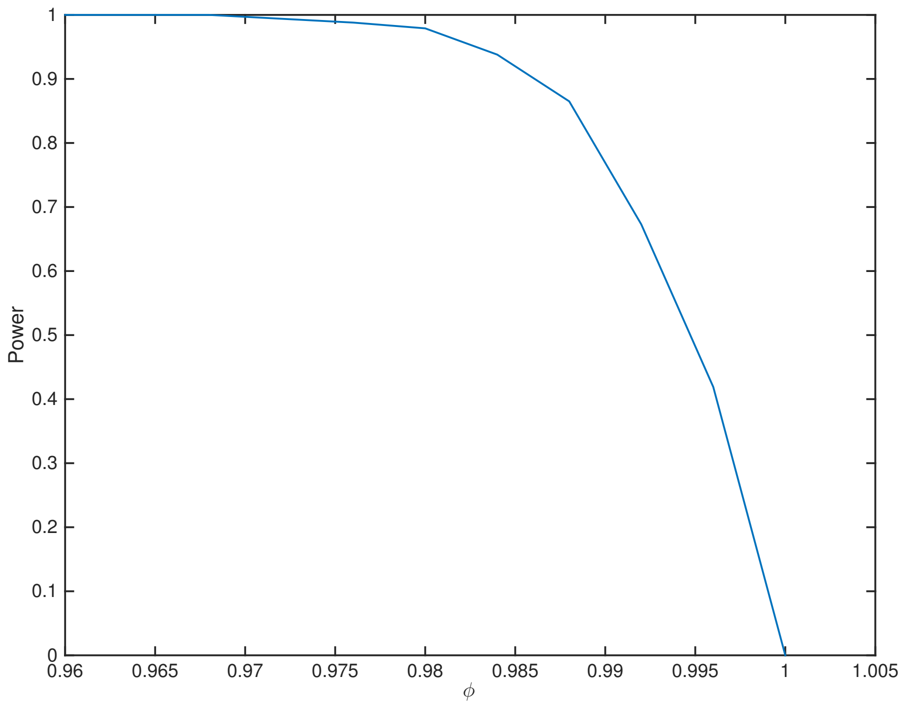


Figure 6.2: The asymptotic power envelope.

6.6 Asymptotically Point Optimal Unit Root Test

Since the asymptotic power envelope is pointwise attainable and there does not exist a uniform most powerful unit root test, there is not a most optimal test and therefore point optimal (against the alternative $c = c_i$) unit root tests have our interest. The asymptotic power envelope can be used to gain a test based on the use of local alternatives with higher power than standard tests based on least square regression. However, due to the fact that each fixed alternative c_i results in a different unit root test, we will provide the analysis on how to obtain the alternative $c_i \in C$ that results in the overall best performing unit root test. The closeness of the asymptotic power curve to the asymptotic power envelope is a good criterion for measuring the performance of the tests based on the alternatives. In order to accomplish that, for each local alternative we will compute the asymptotic powercurve and we will determine which of these powercurves is the closest to the asymptotic power envelope by the method of least squares. Recall the definition of the asymptotic power envelope Π_∞^α:

$$\Pi_\infty^\alpha(c_i) = \Pr_{c_i}\left[-c_i \int_0^1 W_{c_i}(t)dW(t) - \frac{1}{2}c_i^2 \int_0^1 W_{c_i}(t)^2 dt \leq l_\infty^\alpha(c_i)\right]. \tag{6.6.1}$$

The asymptotic power envelope is the maximum attainable asymptotic power of unit root tests against the fixed alternative c_i. Another way to define this maximum attainable power is in terms of the powercurve $\Pi_\infty^\alpha(c, c_i)$, which is the power of the unit root test corresponding to the critical value $l^{\alpha}_{\infty}(c_i)$ subject to the set of alternatives $C = \{c_1, c_2, \ldots, c_M | c_i \in \mathbb{Z}_{<0} \ \& \ c_i > c_{i+1} \ \forall i \leq M\}$:

$$\Pi^{\alpha}_{\infty}(c, c_i) = \Pr\left[-c_i \int_0^1 W_c(t)dW(t) - \frac{1}{2}c_i^2 \int_0^1 W_c(t)^2 dt \leq l^{\alpha}_{\infty}(c_i)\right] \tag{6.6.2}$$

where c_i denotes the alternative of which the powercurve is computed and c defines the actual value of the AR(1) process. If we maximize this powercurve (6.6.2) $\Pi^{\alpha}_{\infty}(c_i)$ over the range of local alternatives in the small neighborhood around unity, we will obtain the asymptotic power envelope for the fixed alternative c_i:

$$\begin{aligned} \max_c \Pi^{\alpha}_{\infty}(c, c_i) =& \max_c \Pr\left[-c_i \int_0^1 W_c(t)dW(t) - \frac{1}{2}c_i^2 \int_0^1 W_c(t)^2 dt \leq l^{\alpha}_{\infty}(c_i)\right] \\ =& \Pr\left[-c_i \int_0^1 W_{c_i}(t)dW(t) - \frac{1}{2}c_i^2 \int_0^1 W_{c_i}(t)^2 dt \leq l^{\alpha}_{\infty}(c_i\right] \\ =& \Pi^{\alpha}_{\infty}(c_i). \end{aligned} \tag{6.6.3}$$

To select the overall best performing unit root test, we will compare each powercurve to the asymptotic power envelope and the method of least squares will result in the best overall performing test. The method of least squares minimizes the following function:

$$\min_{c_i} \sum_{i=1}^{M} (\Pi^{\alpha}_{\infty}(c_i) - \Pi^{\alpha}_{\infty}(c, c_i))^2$$

This method returns the fixed alternative c_i such that the test obtained with this fixed alternative is the best performing one. As a result we have found a unit root test with the best overall performance.

Chapter 7

Summary of Results

This thesis has focused the unit root process. In the first order auto regressive process $Y_t = \phi Y_{t-1} + \epsilon_t$, a unit root is present if $\phi = 1$ and we investigated the ordinary least square estimator to construct a standard t-test statistic. The test we examined is known as the Dickey-Fuller test which rejects the null hypothesis for small values of $t_T = \frac{\hat{\phi}-1}{\hat{\sigma}_{\hat{\phi}}}$ and we concluded that the limiting distribution corresponds to the Dickey-Fuller distribution (4.1.4). The distributions of the innovations is important for the type of unit root test being considered. If we assumed $\epsilon_t \sim WN(0, \sigma^2)$ the Dickey-Fuller test is an appropriate unit root test, instead if we assumed $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ we could test the null hypothesis with the Likelihood Ratio test, which has the benefit of performing further power analysis on the unit root process.

The Likelihood Ratio test rejects the null hypothesis of the presence of a unit root for small values of

$$\log\left[\Lambda_T(\epsilon_t)\right] = -c_1 A_T + \frac{1}{2} c_1^2 B_T$$

and with help of the Neyman-Pearson lemma we obtained that this test provided for each fixed alternative the point optimal test. The Likelihood Ratio gave the opportunity to compute the asymptotic power envelope. The analysis on the asymptotic power envelope done by Elliot, Stock and Rothenberg did prove that there exists an explicit formula for the asymptotic power envelope:

$$\Pi_\infty^\alpha(c_i) = \Pr_{c_i}\left[-c_i \int_0^1 W_{c_i}(t) dW(t) - \frac{1}{2} c_i^2 \int_0^1 W_{c_i}(t)^2 dt \leq l_\infty^\alpha(c_i)\right].$$

The asymptotic power envelope consists of an Ornstein-Uhlenbeck process, such that we have computed the analytic solution by means of stochastic Itô integration. With the help of Monte Carlo simulation we obtained the asymptotic power envelope for fixed alternatives $c_i \in C$. Since each fixed alternative resulted in a different Likelihood Ratio test, by minimizing the distance between the asymptotic powercurve and the asymptotic power envelope, we could select the local alternative corresponding to the overall asymptotically best performing test.

Chapter 8

Discussion

This thesis discussed the Dickey-Fuller test, the Likelihood Ratio test and asymptotic theory of unit root processes and power envelopes for the AR(1) process without intercept and deterministic trend. AR(1) processes are not always suitable in empirical studies on time series and, often, higher order auto regressive processes, AR(p) with $p > 1 \in \mathbb{Z}$, with intercept and/or a deterministic trend are considered. Therefore we recommend to do further research on the asymptotic distribution of the AR(p) process when Assumption 1 is not valid. We will give an outline of the extra simulation and transformations that need to be done in the case of several other auto regressive processes. Since the unit root problem is widely studied in econometrics, there are quite some papers which shed light on the less simplified auto regressive process with a unit root. In the following paragraphs we will give an insight on the way one could obtain these different limiting distributions.

The first order auto regressive process under Assumption 1, is a simplified version of the general first order auto regressive process. In Assumption 1 there is no intercept present and no deterministic trend. The Dickey Fuller test we examined is only valid when we work with the AR(1) process (2.1.1) under Assumption 1. If we add an intercept m or a deterministic trend γt to the process' equation it will result in a different test statistic and asymptotic distribution compared to the one we examined in this thesis. If in the AR(1) process an intercept $m \neq 0$ and deterministic trend are present it yields the following auto regressive process:

$$Y_t = m + \phi Y_{t-1} + \gamma_t + \epsilon_t \tag{8.0.1}$$

where ϵ_t represents IID White Noise with zero mean and constant variance. David Dickey and Wayne Fuller computed alternative ways to obtain the limiting distribution of 8.0.1. If one is interested in computing the test statistic and the power of the Dickey-Fuller test when an intercept $m \neq 0$ is present, it is possible to center the data by examining $Y_t - m$ opposed to Y_t. In this case under the null hypothesis the increments are equal to the innovations and the process $Y_t - m$ behaves as a random walk.

Most often it is not evident if a time series contains a deterministic trend, thus one has to examine the presence of a trend before choosing the appropriate Dickey-Fuller test. Harvey, Leybourne and Taylor (2009) [9] have written a paper in which they discuss the uncertainty about the trend and initial value Y_0. They concluded that it is possible to detrend the data in order to perform a Dickey-Fuller type unit root test. As a result they could argue that Dickey-Fuller type unit root tests are almost asymptotically efficient when a deterministic trend is present in the data. We can conclude that there are ways to use the Dickey-Fuller unit root test for data consisting a deterministic trend and we would recommend to consider several detrending methods in order to find the test with locally optimal power.

In reality the first order auto regressive process is not often being considered to fit a time series. Often higher orders $p \in \mathbb{N}$ of regression are fitted to a time series. Since the characteristic equation of that series has several roots, testing for the presence of a unit root becomes more difficult. Instead of using the standard Dickey-Fuller test, the Augmented Dickey-Fuller test was introduced by Wayne Fuller [6]. An important implication of testing for a unit root in higher order auto regressive processes is the proper selection of the lag p.

Another point of criticism we should mention is the fact that the least square estimator is biased. By definition of the critical values, the probability of a type I error must be smaller then the significance level $\alpha = 0.05$. Table 5.4 shows us that these results do not match the significance level α, thus we can conclude that the results are biased. There are several methods available to reduce bias in a Monte Carlo simulation, but as this was not the aim of the thesis, we did not implement and test these bias reduction methods. To perform a more accurate power analysis of the Dickey-Fuller test, we recommend to perform a bias reduction method.

Bibliography

[1] David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

[2] David A Dickey and Wayne A Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, pages 1057–1072, 1981.

[3] Joseph L. Doob. *Stochastic processes*, volume 101. New York Wiley, 1953.

[4] Graham Elliott, Thomas J. Rothenberg, and James H. Stock. Efficient tests for an autoregressive unit root, 1992.

[5] Graham Elliott, Thomas J. Rothenberg, and James H. Stock. Efficient tests for an autoregressive unit root, 1996.

[6] Wayne A. Fuller. *Introduction to statistical time series*, volume 428. John Wiley & Sons, 2009.

[7] Niels Haldrup and Michael Jansson. Improving size and power in unit root testing. *Palgrave handbook of econometrics*, 1:252–277, 2006.

[8] James D. Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

[9] David I. Harvey, Stephen J. Leybourne, and Robert Taylor. Unit root testing in practice: dealing with uncertainty over the trend and initial condition. *Econometric Theory*, 25(03):587–636, 2009.

[10] Robin High. Important factors in designing statistical power analysis studies. *Computing News, Summer,(14-15)*, 2000.

[11] Jerzy Neyman and Egon S. Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Royal Society of London Philosophical Transactions Series A*, 231:289–337, 1933.

[12] Peter C.B. Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988.

[13] G. William Schwert. Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, 20(1):5–17, 2002.

[14] Steven E. Shreve. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer, 2004.

[15] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Phys. Rev.*, 36:823–841, Sep 1930.

Appendices

Appendix A

Auxiliary results

A.1 Central Limit Theorem

Theorem A.1.1 (Central Limit Theoreom). *Let $X_1, X_2, \ldots, X_T$ be a set of T IID random variables and each X_i have an arbitrary probability distribution $P(x_1, ..., x_T)$ with mean μ_i and a finite variance σ_i^2. Then it follows that the normal variable $\overline{X}$*

$$\overline{X} = \frac{\sum_{i=1}^{T} x_i - \sum_{i=1}^{T} \mu_i}{\sqrt{\sum_{i=1}^{T} \sigma_i^2}}$$

has a limiting cumulative distribution function which approaches a normal distribution.

A.2 Functional Central Limit Theorem

Theorem A.2.1 (Functional Central limit Theorem). *Let $X_1, X_2, \ldots$ be IID random variables with mean zero and variance 1* [1] *and define:*

$$Y_n(t) = \frac{1}{\sqrt{(n)}} \sum_{k \leq nt} X_k$$

with $t \in [0,1]$ and $n \in \mathcal{N}$. Consider a Brownian Motion W on $[0,1]$ and let $g : D[0,1] \to \mathcal{R}$ measurable and continuous at W. Then $g(Y_n) \to g(W)$ in distribution.

A.3 Continuous Mapping Theorem

[1] Without loss of generality we can take the variance $\sigma^2 = 1$

Theorem A.3.1 (Continuous Mapping Theorem). *The continuous mapping theorem states that if $S_T(\cdot) \to S(\cdot)$ and $g(\cdot)$ is a continuous functional, with $S(\cdot)$ a continuous-time stochastic process and with $S(r)$ representing its value at some time $r \in [0, 1]$, then $g(S_T(\cdot)) \to g(S(\cdot))$.*

A.4 Neyman-Pearson Lemma

Theorem A.4.1 (Neyman-Pearson Lemma). *If there exists a critical region K of size α and $k \in \mathbb{R} > 0$ such that*

$$\frac{\prod_{i=1}^{n} f(x_i|\theta_1)}{\prod_{i=1}^{n} f(x_i|\theta_0)} \geq k$$

for points in K and

$$\frac{\prod_{i=1}^{n} f(x_i|\theta_1)}{\prod_{i=1}^{n} f(x_i|\theta_0)} \leq k$$

for points not in K, $\overline{K}$, then K is the best critical region of size α

Appendix B

Matlab programs of the Monte Carlo simulation

B.1 Critical values of the Dickey-Fuller test.

B.1.1 Finite sample critical values

```
% Emma Ingeleiv Wagner - 4004949


% Dickey Fuller Asymptotic Critical values simulation
% NO trend and NO constant

clear all;
%randn('seed',1234);

%%%%%%%%%%%%%%%%%%%%%%%% Declaration of variables %%%%%%%%%%%%%%%%%%%%%%
T=25; %sample size
N=100000; %runs
dt=T/N;%intervals
phi=1;

hatsigma=zeros(1,N);
hatphi=zeros(1,N);


%denominator of the DF distr.
ttest=zeros(1,N);

%%%%%%%%%%%%%%%%%%%%% Simulating a Brownian path & calculating critical values %%%%%%

for k=1:N

X=zeros(1,T);
Z1=zeros(1,T);
Z2=zeros(1,T);
Y1=zeros(1,T);
```

```
for j = 2:T

    X(j)= phi*X(j-1)+rand-0.5;
    Z1(j)=X(j-1)*X(j);
    Z2(j)=X(j-1)^2;

end
Z1=cumsum(Z1);
Z2=cumsum(Z2);
hatphi(k)=Z1(T)/Z2(T);

for j=2:T
    Y1(j)=(X(j)-hatphi(k)*X(j-1))^2;
end

Y1 = cumsum(Y1);

ttest(k)=(hatphi(k)-1)/sqrt(((Y1(T)/(T-1))/Z2(T))); %DF distr

end


% Compute the critical value for the test statistic

dfcritical=sort(ttest);
plot(dfcritical);
nbins=100;
hist(dfcritical,nbins)

% Critical values

alpha1=dfcritical(0.01*N); % 1% significance level
alpha2=dfcritical(0.025*N); % 2.5% significance level
alpha3=dfcritical(0.05*N); % 5% significance level

alpha=[alpha1,alpha2,alpha3]
```

B.1.2 Asymptotic critical values

```
% Emma Ingeleiv Wagner - 4004949


% Dickey Fuller Asymptotic Critical values simulation
% NO trend and NO constant

clear all;
%randn('seed',1234);

%%%%%%%%%%%%%%%%%%%%%%%% Declaration of variables %%%%%%%%%%%%%%%%%%%%%%
T=1;
N=10000; %number of intervals
N2=50000; %runs
dt=T/N;%intervals
```

```
num=zeros(1,N2); %numerator of the DF distr.
den=zeros(1,N2); %denominator of the DF distr.
df=zeros(1,N2);

%%%%%%%%%%%%%%%%%%%%% Simulating a Brownian path & calculating critical values %%%%%%

for k=1:N2

dW = zeros(1,N);
W = zeros(1,N);
Z=zeros(1,N);
Y= zeros(1,N);
dW(1) = sqrt(dt)*randn;
W(1) = dW(1);

for j = 2:N
dW(j) = sqrt(dt)*randn; % Wiener process simulation
W(j) = W(j-1)+dW(j);
Z(j) = (W(j-1)^2);
Y(j) = W(j-1) * dW(j);
end

Z = cumsum(Z);
Y = cumsum(Y);
den(k)=sqrt(dt*Z(N)); %denominator of the DF distr
num(k)= Y(N); %numerator of the DF distr

df(k)=num(k)/den(k); %DF distr

end


% Compute the critical value for the test statistic

dfcritical=sort(df);
plot(dfcritical);
nbins=100;
hist(dfcritical,nbins)

% Critical values

alpha1=dfcritical(0.01*N2); % 1% significance level
alpha2=dfcritical(0.025*N2); % 2.5% significance level
alpha3=dfcritical(0.05*N2); % 5% significance level

alpha=[alpha1,alpha2,alpha3]
```

B.2 Dickey-Fuller Test - White Noise innovations

```
% Emma Ingeleiv Wagner - 4004949
```

```
% Dickey Fuller test
% White Noise errors (rand-0.5)

% Monte Carlo Simulatie
% AR(1) no intercept no trend

%%%%%%%%%% Declaration of variables %%%%%%%%%%%%%%
clear all;

N=1000;
c=[-10:1:0]; %fixed alternatives
%phi1=0.5:0.05:1;
M=length(c);
type2=zeros(1,M);
power=zeros(1,M);

for i=1:M

T=250; % aantal observaties
rand('seed',1234);

% Generate one unit root process of length T
% Calculate type II error

phi=1+c(i)/T;


Y=zeros(T,1);
hatphi=zeros(1,N);
teller=zeros(T,N);
noemer=zeros(T,N);
teller2=zeros(T,N);
hatsigma=zeros(1,N);
testvalue=zeros(1,N);

criticalvalue=-1.941;
count=0;

for k=1:N

Y(1)=0;


for t=2:T
    Y(t)=phi*Y(t-1)+randn;

    % constructie hatphi
    teller(t,k)=Y(t-1)*Y(t);
    noemer(t,k)=Y(t-1)^2;
end

% constructie hatphi
teller=cumsum(teller);
noemer=cumsum(noemer);
hatphi(k)=teller(T,k)/noemer(T,k);
```

```
% constructie hatsigma

for t=2:T

    teller2(t,k)=(Y(t)-hatphi(k)*Y(t-1))^2;
end

teller2=cumsum(teller2);
hatsigma(k)=sqrt(((teller2(T,k))/(T-1))/noemer(T,k));

testvalue(k)=(hatphi(k)-1)/hatsigma(k);

 if(testvalue(k)> criticalvalue)
    count=count+1;
end

end

%nbins=100;
%hist(testvalue,nbins);
type2(i)=count/N;
power(i)=1-type2(i);

end

plot(1+c/T, power)
```

B.3 Dickey-Fuller Test - Gaussian innovations

```
% Emma Ingeleiv Wagner - 4004949


% Dickey Fuller test
% Gaussian errors
% AR(1) no intercept no trend

clearvars -except maxpower;

%%%%%%%%%% Declaration of variables %%%%%%%%%%%%%

c=[-10:1:0]; %fixed alternatives
N=10000;

M=length(c);
type2=zeros(1,M);
power5=zeros(1,M);
power6=zeros(1,M);
power4=zeros(1,M);

for i=1:M

T=500; % aantal observaties
```

```
% Generate one unit root process of length T
% Calculate type II error

phi=1+(c(i)/T);

Y=zeros(T,1);
hatphi=zeros(1,N);
teller=zeros(T,N);
noemer=zeros(T,N);
teller2=zeros(T,N);
hatsigma=zeros(1,N);
testvalue=zeros(1,N);

criticalvalue=-1.9522;
count=0;

for k=1:N

Y(1)=0;

for t=2:T
    Y(t)=phi*Y(t-1)+randn;

    % constructie hatphi
    teller(t,k)=Y(t-1)*Y(t);
    noemer(t,k)=Y(t-1)^2;
end

% constructie hatphi
teller=cumsum(teller);
noemer=cumsum(noemer);
hatphi(k)=teller(T,k)/noemer(T,k);

% constructie hatsigma

for t=2:T

    teller2(t,k)=(Y(t)-hatphi(k)*Y(t-1))^2;
end

teller2=cumsum(teller2);
hatsigma(k)=sqrt(((teller2(T,k))/(T-1))/noemer(T,k));

testvalue(k)=(hatphi(k)-1)/hatsigma(k);

 if(testvalue(k)> criticalvalue)
    count=count+1;
end

end

type2(i)=count/N;
power5(i)=1-type2(i);
```

```
end

plot(1+c/T,power5)
hold on;

for i=1:M

T=100; % aantal observaties


% Generate one unit root process of length T
% Calculate type II error

phi=1+(c(i)/T);

Y=zeros(T,1);
hatphi=zeros(1,N);
teller=zeros(T,N);
noemer=zeros(T,N);
teller2=zeros(T,N);
hatsigma=zeros(1,N);
testvalue=zeros(1,N);

criticalvalue=-1.9645;
count=0;

for k=1:N

Y(1)=0;

for t=2:T
    Y(t)=phi*Y(t-1)+2*randn;

    % constructie hatphi
    teller(t,k)=Y(t-1)*Y(t);
    noemer(t,k)=Y(t-1)^2;
end

% constructie hatphi
teller=cumsum(teller);
noemer=cumsum(noemer);
hatphi(k)=teller(T,k)/noemer(T,k);

% constructie hatsigma

for t=2:T

    teller2(t,k)=(Y(t)-hatphi(k)*Y(t-1))^2;
end

teller2=cumsum(teller2);
hatsigma(k)=sqrt(((teller2(T,k))/(T-1))/noemer(T,k));

testvalue(k)=(hatphi(k)-1)/hatsigma(k);
```

```
 if(testvalue(k)> criticalvalue)
    count=count+1;
end

end

type2(i)=count/N;
power6(i)=1-type2(i);

end

for i=1:M

T=50; % aantal observaties
rand('seed',1234);

% Generate one unit root process of length T
% Calculate type II error

phi=1+c(i)/T;

Y=zeros(T,1);
hatphi=zeros(1,N);
teller=zeros(T,N);
noemer=zeros(T,N);
teller2=zeros(T,N);
hatsigma=zeros(1,N);
testvalue=zeros(1,N);

criticalvalue=-1.9682;
count=0;

for k=1:N

Y(1)=0;

for t=2:T
    Y(t)=phi*Y(t-1)+2*randn;

    % constructie hatphi
    teller(t,k)=Y(t-1)*Y(t);
    noemer(t,k)=Y(t-1)^2;
end

% constructie hatphi
teller=cumsum(teller);
noemer=cumsum(noemer);
hatphi(k)=teller(T,k)/noemer(T,k);

% constructie hatsigma

for t=2:T

    teller2(t,k)=(Y(t)-hatphi(k)*Y(t-1))^2;
```

```
end

teller2=cumsum(teller2);
hatsigma(k)=sqrt(((teller2(T,k))/(T-1))/noemer(T,k));

testvalue(k)=(hatphi(k)-1)/hatsigma(k);

 if(testvalue(k)> criticalvalue)
    count=count+1;
end

end

type2(i)=count/N;
power4(i)=1-type2(i);

end

plot(-c, power4, -c, power5, -c, power6)
```

B.4 Critical values of the Likelihood ratio test.

```
% Emma Ingeleiv Wagner - 4004949


% Likelihood Ratio Test Critical values simulation

% critical values for the likelihood ratio test are dependent on the fixed
% alternative which is being used. Therefore we construct a vector cv with
% a critical value for each fixed alternative.

clear all

%%%%%%%%%%%%%%%%%%%%%%%%% Declaration of variables %%%%%%%%%%%%%%%%%%%%%%
T=1;
N=1000; %number of intervals for the integral
N2=1000; %runs
dt=T/N; %interval size

B=zeros(1,N2); %numerator of the DF distr.
A=zeros(1,N2); %denominator of the DF distr.
lr=zeros(1,N2);
lr2=zeros(1,N2);
c=-10:0.5:0; %fixed local alternative vector
M=length(c);
cv=zeros(1,M);

%%%%%%%%%%%%%%%%%%%%%% Simulating a Brownian path & calculating critical values %%%%%%

for i=1:M

for k=1:N2
```

```
dW = zeros(1,N);
W = zeros(1,N);
Z=zeros(1,N);
Y= zeros(1,N);
X=zeros(1,N);


dW(1) = sqrt(dt)*randn;
W(1) = dW(1);

for j = 2:N %Wiener Process simulation

dW(j) = sqrt(dt)*randn;
W(j) = W(j-1)+dW(j);
Z(j) = (W(j-1)^2);
Y(j) = W(j-1) * dW(j);

end

Z = cumsum(Z);
Y = cumsum(Y);
A(k)=dt*Z(N);
B(k)= Y(N);


lr(k)=-c(i)*B(k)+0.5*c(i)^2*A(k);

end

% Compute the critical value for the test statistic
lrcritical=sort(lr);
cv(i)=lrcritical(0.05*N2);


end
```

B.5 Likelihood Ratio Test

```
% Emma Ingeleiv Wagner - 4004949

% Likelihood Ratio test

randn('seed',1234);
clearvars -except cv; %delete all variables except cv
%%%%%%%%%% Declaration of variables %%%%%%%%%%%%%%

c=-10:1:0; %fixed alternatives
N=1000; %reps
M=length(c);
powerlik=zeros(1,M); %powervector

% Generate an AR(1) process of length T
% Calculate type II error
```

```
for i=1:M

T=25; %sample size

falt=c(i); %fixed alternative
Y=zeros(T,1); %AR(1) process
lrtest=zeros(1,N); %LR test statistic
A=zeros(T,N);
B=zeros(T,N);
count=0;


for k=1:N

Y(1)=0;

for t=2:T
    Y(t)=(1+(falt/T))*Y(t-1)+randn; %AR(1) process simulation
    A(t,k)=Y(t-1)*(Y(t)-Y(t-1));
    B(t,k)=Y(t-1)^2;
end

% constructie LR test statistic
A=cumsum(A);
B=cumsum(B);
lrtest(k)=-(falt/T)*A(T,k)+0.5*(falt^2)/(T^2)*B(T,k); %test statistic

if(lrtest(k)< cv(i))
    count=count+1;
end

end

%computing the power for each fixed alternative
powerlik(i)=count/N;

end

plot(-c,powerlik);
hold on;
```

B.6 Asymptotic Power Envelope

```
% Emma Ingeleiv Wagner - 4004949


% Asymptotic Power Envelop (as in the paper)
% Simulation of a Wiener Process
% Simulation of an Ornstein-Uhlenbeck process
clearvars -except cv

%%%%%%%%%%%%%%%%%% Declaration of variables %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
N=1000; %number of intervals of the integral
N2=1000; %reps
T=1;
dt=T/N;
c=-10:0.5:0; %fixed alternative
critval=cv;
powerenv=zeros(1,length(cv));
lrtest=zeros(1,N2);

for k=1:length(c);

count=0;

for i=1:N2

W = zeros(1,N); %Wiener process
X=zeros(1,N); %Ornstein-Uhlenbeck process
Y=zeros(1,N);
Z=zeros(1,N);
dW=zeros(1,N);
dX=zeros(1,N);

dW(1) = sqrt(dt)*randn;
W(1) = dW(1);

for j=2:N

dW(j) = sqrt(dt)*randn; %Wiener process simulation
W(j) = W(j-1)+dW(j);

dX(j)=exp(-c(k)*j/N)*exp(c(k)*(j-1)/N)*(W(j)-W(j-1)); %Ornstein-Uhlenbeck process simu
X(j)=X(j-1)+dX(j);

Z(j)=X(j-1)*(W(j)-W(j-1)); %terms integral 0 to 1 XdW
Y(j)=X(j)^2; %terms integral 0 to 1 Xdt
end

Z = cumsum(Z);
Y = dt*cumsum(Y);

lrtest(i)=-c(k)*Z(N)-0.5*c(k)^2*Y(N);

    if lrtest(i)<cv(k)
        count=count+1;
    end

end

powerenv(k)=count/N2;

end

plot(-c,powerenv)
```