Anders Rahbek
Rasmus Søndergaard Pedersen
University of Copenhagen

September 2024

Part III
Autoregressive modelling and unit roots

III.1 Introduction

This is an introduction to the analysis of autoregressive (AR) models in the case where the characteristic polynomial is allowed to have a root at one, a so-called unit root, rather than as has been the case until now, all roots outside the unit circle. The focus here is the univariate case, whereas the multivariate analysis, or in other words, cointegration analysis, is treated separately. Apart from reparametrizations of the univariate AR models, estimation in the presence of unit roots is based on linear regression analysis. What is changed is the asymptotic inference which is based on non classic limit distributions.

The explicit underlying assumption of the previous analyses of autoregressive models has been that of geometric ergodicity, which in terms of the simple AR(1) model with x_0 fixed and ε_t i.i.d.N$(0, \sigma^2)$,

$$x_t = \rho x_{t-1} + \varepsilon_t,\ t = 1, ..., T \tag{III.1}$$

can be stated as the parameter restriction $|\rho| < 1$. When $\rho = 1$,

$$x_t = x_{t-1} + \varepsilon_t = \sum_{t=1}^{t} \varepsilon_i + x_0, \tag{III.2}$$

that is, x_t is the sum of a random walk $\sum_{t=1}^{t} \varepsilon_i$ and the initial value x_0. Clearly when $\rho = 1$, x_t is not stationary, not even asymptotically, as for example the variance conditional on x_0, $\mathbb{V}[x_t] = t\sigma^2$, which is increasing in t. This is the kind of non-stationarity that is commonly referred to when discussing 'non-stationary' in the context of unit root analysis. It is easy to see that,

$$\Delta x_t = x_t - x_{t-1} = \varepsilon_t$$

i.e. the differenced process is stationary.

In short, for $\rho = 1$, x_t is a simple example of a so called I(1) process, in the sense that x_t is a non-stationary process with a random walk component, while Δx_t is stationary and also geometrically ergodic. Unit root analysis provides a framework to discriminate these two situations: the geometrically ergodic (or, simply stationary) case, and the non-stationary random walk type behavior. As is common in the literature, this is sometimes referred to as the hypothesis of stationarity and non-stationarity respectively, which should not cause any confusion.

The assumption of $|\rho| < 1$ implies in particular that inference on the parameters is based on well-known asymptotic Gaussian and χ^2 distributions. However, many if not most economic time series do not show stationary behavior and inference with such variables is not standard. To avoid this, the data series are often transformed by for example differencing (Δx_t) and taking the logarithm, or a combination thereof, to obtain approximately stationary series, which may then be analyzed using autoregressive models under the assumption of geometric ergodicity.

Univariate unit root analysis is a first step towards cointegration analysis where relations between non-stationary key economic variables can be analyzed.

In most analyses of economic time series it is not easy to distinguish if the series analyzed behave as stationary processes with possibly a linear trend or, as process with a random walk component with or without a linear trend. Unit root analysis in AR(k) models with deterministic terms help to do so and will be discussed.

As a final remark before turning to the analysis of unit roots it is important to stress that many other kinds of non-stationarity appears in the literature. For example, fractionally integrated processes where $\Delta^d x_t = (1-L)^d x_t$ is stationary for some $0 < d < 1$, and processes with breaks in, say, the mean. The former kind is not treated here and with regard to the latter these can be analyzed by inclusion of dummy variables as in standard AR models.

III.2 The AR(1) model

In order to introduce the kind of new inference due to non-stationary processes consider again the simple AR(1) model in (III.1),

$$x_t = \rho x_{t-1} + \varepsilon_t.$$

Now with $\rho \in \mathbb{R}$ the maximum likelihood estimator is given by,

$$(\hat{\rho} - \rho_0) = S_{\varepsilon z} S_{zz}^{-1} = \frac{\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2} \tag{III.3}$$

From previous analysis it follows that if $|\rho_0| < 1$ then $x_{t-1}\varepsilon_t$ is a martingale difference sequence satisfying the assumptions of the Central Limit Theorem (CLT) for martingale differences. As a result the estimator is consistent, $\hat{\rho} \xrightarrow{P} \rho_0$ and asymptotically normally distributed,

$$\sqrt{T}(\hat{\rho} - \rho_0) = \sqrt{T} S_{\varepsilon z} S_{zz}^{-1} \xrightarrow{D} N(0, 1 - \rho_0^2).$$

In the case of $\rho_0 = 1$ and with $x_0 = 0$,

$$(\hat{\rho} - 1) = S_{\varepsilon z} S_{zz}^{-1} = \frac{\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t x_{t-1}}{\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2} = \frac{\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t (\sum_{i=1}^{t-1} \varepsilon_i)}{\frac{1}{T}\sum_{t=1}^{T} (\sum_{i=1}^{t-1} \varepsilon_i)^2}$$

Now first of all $x_t = \sum_{i=1}^{t} \varepsilon_i$ is not (geometrically) ergodic, let alone i.i.d., so that the usual class of laws of large numbers (LLN) do not apply. Second, while $\varepsilon_t x_{t-1}$ is a martingale difference (MGD), it does not satisfy the assumptions of a CLT. In particular with $\mathcal{F}_t = (x_t, x_{t-1}, ..)$ (or, $\sigma(x_t, ...)$) we find

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\varepsilon_t^2 x_{t-1}^2 | \mathcal{F}_{t-1}\right] = \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2\right)\sigma_0^2, \quad x_t = \sum_{i=1}^{t} \varepsilon_i.$$

This does not converge in probability to a constant as required for the CLT to hold. In fact it holds that $T^{1/2}(\hat{\rho} - 1) \xrightarrow{P} 0$, $\hat{\rho}$ is said to be 'super-consistent', and furthermore, as we will see

$$T(\hat{\rho} - 1) \xrightarrow{D} \frac{\int_0^1 \mathcal{W}_u d\mathcal{W}_u}{\int_0^1 \mathcal{W}_u^2 du} = \frac{\int_0^1 \mathcal{W} d\mathcal{W}}{\int_0^1 \mathcal{W}_u^2 du}$$

where $\mathcal{W}$ is a standard Brownian motion – which is introduced in the next section – on the interval $u \in [0, 1]$. The distribution is a non-standard asymmetric distribution and was tabulated originally by Dickey and Fuller (1979). This implies that reporting the usual $t-$statistic for $\hat{\rho}$ has no meaning if x_t is indeed a random walk.

Recall that the likelihood ratio test statistic of the hypothesis $H : \rho = \rho_0$ is given by

$$\mathrm{LR}(\rho = \rho_0) = T \log(1 + W_T), \text{ where } W_T = (\hat{\rho} - \rho_0)^2 S_{zz}/\hat{\sigma}^2. \tag{III.4}$$

When $|\rho_0| < 1$,

$$W = TW_T \xrightarrow{D} \chi^2, \tag{III.5}$$

while for the case of a unit root, $\rho_0 = 1$,

$$W = TW_T \xrightarrow{D} \frac{(\int_0^1 \mathcal{W} d\mathcal{W})^2}{\int_0^1 \mathcal{W}_u^2 du} \neq \chi^2, \tag{III.6}$$

which is the (square) of the so-called Dickey-Fuller distribution. It has broader tails than the χ^2 distribution. For example the 95% quantile is approximately 4.2 which should be compared with 95% quantile of the χ^2 distribution, 3.84.

These results are explained in detail in the next sections and it is briefly commented upon that the results are similar for AR processes with more lags.

III.3 Brownian motion

In this section it is discussed in what sense LLN and CLTs hold for functions of the random walk.

III.3.1 Brownian motion

Consider again the random walk,

$$x_0 = 0 \tag{III.7}$$

$$x_t = \sum_{i=1}^{t} \varepsilon_i \text{ for } t = 1, \ldots, T \tag{III.8}$$

The main features of the random walk defined this way are that x_t is $\mathrm{N}(0, t\sigma^2)$ distributed and that x_t has independent increments, i.e. Δx_t and Δx_{t+k} are independent for $k \neq 0$. That $x_0 = 0$ is merely a convenient convention. In the analyses of AR models the initial value x_0 is fixed, but can be ignored in the asymptotic analysis.

The continuous time equivalent of the random walk relevant here is the Brownian motion, $\mathcal{B}$ defined on the unit-interval $[0, 1]$. Now $\mathcal{B}$ is a function of time $u \in [0, 1]$, it is a stochastic process, and a realization of $\mathcal{B}$ is a continuous function on $[0, 1]$ with the following properties:

Definition III.3.1 *The Brownian motion $\mathcal{B}$ with variance σ^2 defined on $[0, 1]$ is a stochastic process with the properties:*

1. $\mathcal{B}_0 = 0$

2. $\mathcal{B}_u$ *is* $\mathrm{N}(0, u\sigma^2)$ *distributed for all* $u \in [0,1]$

3. *For any* $0 \leq u_1 < \ldots < u_k \leq 1$ *the increments* $(\mathcal{B}(u_2) - \mathcal{B}(u_1)), \ldots, (\mathcal{B}(u_k) - \mathcal{B}(u_{k-1}))$ *are independent*

4. $\mathcal{B}_u$ *is continuous as a function of* u.

If $\sigma^2 = 1$, $\mathcal{B} \equiv \mathcal{W}$ *is a standard Brownian motion.*

The main difference between the random walk and the Brownian motion is the continuity. The class of functions on the unit interval which are continuous will be referred to as $C(0,1)$.

In order to change the time scaling of the random walk to be on the unit interval define for $u \in [0,1]$ the process,

$$x_0^T = 0 \tag{III.9}$$

$$x_u^T = \frac{1}{\sqrt{T}} \sum_{i=1}^{[Tu]} \varepsilon_i \tag{III.10}$$

where $[Tu]$ is the integer value of Tu, $u \in [0,1]$. For $u = 0, 1/T, 2/T, .., 1 = T/T$ clearly x_u^T is just the random walk divided by $\sqrt{T}$. In between these points x_u^T is constant, see figure 1.

Thus x_u^T is an example of a process on $[0,1]$, which is right-continuous and has limits from the left, a so-called càdlàg process. I.e. it is on the right scale but not continuous. It is no problem to connect the points of the random walk to make it a $C(0,1)$ function, but this makes the derivations complicated. In fact, that x_u^T is not continuous can be ignored in the following, since the limit is indeed continuous as demonstrated in the outline of the proof of Theorem III.3.2 below. The main theorem (the invariance principle) states that

$$x_\cdot^T = \frac{1}{\sqrt{T}} \sum_{i=1}^{[T\cdot]} \varepsilon_i \xrightarrow{D} \mathcal{B}. \tag{III.11}$$

uniformly on $[0,1]$ as $T \to \infty$, where $\mathcal{B}$ is a Brownian motion with variance σ^2. This is in accordance with Figure 1.

III.3.2 Invariance principle

The limit theorems so far regarding convergence of stationary processes have treated convergence in distribution and probability of random variables defined on $\mathbb{R}$ or $\mathbb{R}^p$. The statement in (III.11) is a statement about convergence in distribution of a random variable on $C(0,1)$ instead, i.e. a functional limit theorem.

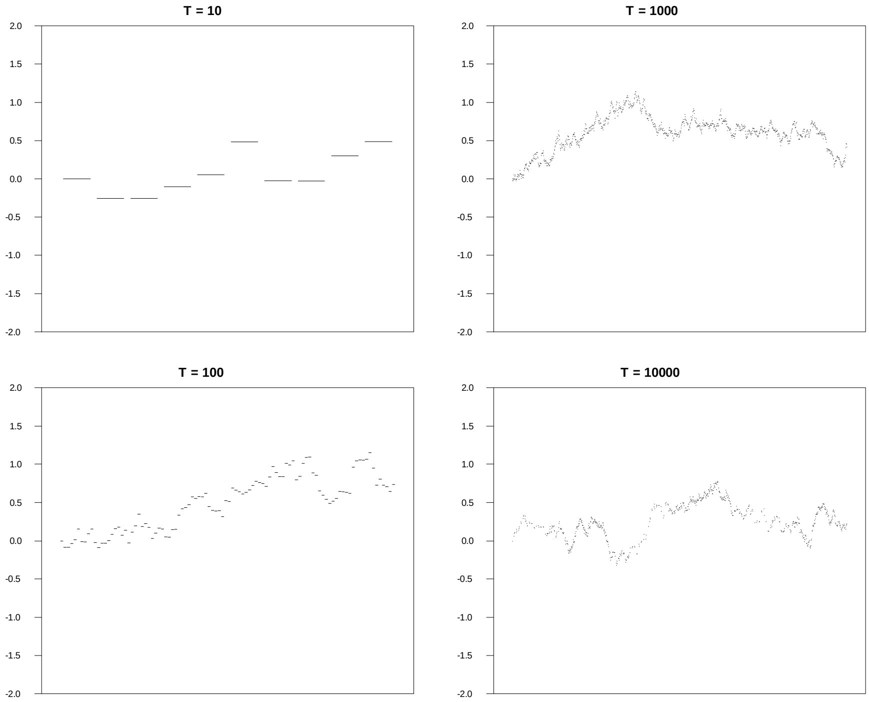


Figure 1: Simulations of $X_{[Tu]}$ based on $\epsilon_t's$ drawn from the $N(0,1)$ distribution.

In order to understand the convergence, introduce the metric on $C(0,1)$ which is given by the supremum, i.e. with x in $C(0,1)$,

$$x^T \to x \text{ if } \sup_u |x_u^T - x_u| \to 0,\, T \to \infty \tag{III.12}$$

Thus for example if x_u^T, or simply x^T, is a sequence of random variables on $C(0,1)$ then,

$$x_u^T \xrightarrow{P} 0, \tag{III.13}$$

where P is defined on $C(0,1)$, means that $\sup_{u\in[0,1]} |x_u^T| \xrightarrow{P} 0$ on $\mathbb{R}$.

To prove convergence in distribution on $C(0,1)$, as in (III.11), two things are needed. First the finite-dimensional distributions of x_u^T need to converge, i.e. for any $0 \leq u_1 < u_2 < \ldots < u_k \leq 1$,

$$(x_{u_1}^T, \ldots, x_{u_k}^T) \xrightarrow{D} (\mathcal{B}_{u_1}, \ldots, \mathcal{B}_{u_k}) \tag{III.14}$$

Thus e.g. for any fixed u, $x_u^T \xrightarrow{D} \mathcal{B}_u$ which is simply the $N(0, \sigma^2 u)$ distribution. For the case here this is trivial since any CLT gives for i.i.d. processes gives,

$$x_u^T = \frac{1}{\sqrt{T}} \sum_{i=1}^{[Tu]} \varepsilon_i = \sqrt{\frac{[Tu]}{T}} \frac{1}{\sqrt{[Tu]}} \sum_{i=1}^{[Tu]} \varepsilon_i \xrightarrow{D} \sqrt{u} N(0, \sigma^2) = N(0, \sigma^2 u). \tag{III.15}$$

What is needed ,in addition to prove convergence on $C(0,1)$, is 'tightness'. When discussing convergence on $\mathbb{R}$ for fixed u this is equivalent to $x_u^T = O_P(1)$. On $C(0,1)$ the equivalent definition of tightness is that there exists a compact set $K \subset C(0,1)$, such that

$$P(x^T \in K) \geq 1 - \delta, \ \delta > 0 \ \text{ for all } T. \tag{III.16}$$

Thus the probability mass cannot 'escape to infinity'. Tightness is not discussed further. Instead an excellent reference is Billingsley (1968), where also a detailed proof of Theorem III.3.1 below is given.

Theorem III.3.1 *(Donsker's Theorem or Invariance Principle)*
Let $\varepsilon_t, t = 1, .., T$ be i.i.d. $\mathrm{N}(0, \sigma^2)$ *then*

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{[T\cdot]} \varepsilon_i \xrightarrow{D} \mathcal{B}. \quad , \tag{III.17}$$

on $C(0,1)$, with $\mathcal{B}$ a Brownian motion on $[0,1]$ with variance σ^2.

III.3.3 Invariance principle for martingale differences

In fact, Theorem III.3.1 is a corollary to the general invariance principle for martingales, the functional central limit theorem (FCLT). This we stated for $u = 1$ as the CLT in Theorem I.4.4 (stated as a corollary to Brown, 1971).

Theorem III.3.2 *Let $(Y_t)_{t=1,2,\ldots}$, with $\mathbb{E}[Y_t^2] < \infty$, be a martingale difference sequence with respect to the increasing sequence $\mathcal{F}_t$. Assume further that, (i) and (ii), or (i) and (ii') hold for some $\delta > 0$ and as $T \to \infty$,*

$$\textit{(i)}: \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[Y_t^2 | \mathcal{F}_{t-1}\right] \xrightarrow{P} \sigma_y^2 > 0 \quad \textit{and} \tag{III.18}$$

$$\textit{(ii)}: \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[Y_t^2 \mathbb{I}(|Y_t| > \delta\sqrt{T})\right] \to 0, \ \textit{or} \tag{III.19}$$

$$\textit{(ii)'}: \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[Y_t^2 \mathbb{I}(|Y_t| > \delta\sqrt{T}) | \mathcal{F}_{t-1}\right] \xrightarrow{P} 0. \tag{III.20}$$

Then, as $T \to \infty$,

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{[T\cdot]} Y_i \xrightarrow{D} \mathcal{B}.$$

where $\mathcal{B}$ is a Brownian motion with variance σ^2.

III.3.4 Convergence to integrals

Now turning to the estimator $\hat{\rho}$ and the likelihood ratio test statistic of the previous section for the hypothesis that $\rho = 1$ it follows that what is interesting is the asymptotic behavior of terms such as,

$$\sum_{t=1}^{T} x_{t-1}^2 = \sum_{t=1}^{T} (\sum_{i=1}^{t-1} \varepsilon_i)^2 \text{ and } \sum_{t=1}^{T} x_{t-1}\varepsilon_t = \sum_{t=1}^{T} (\sum_{i=1}^{t-1} \varepsilon_i)\varepsilon_t.$$

The latter involves the definition of the stochastic integral $\int_0^1 \mathcal{B}_u d\mathcal{B}_u$ (or, simply $\int \mathcal{B} d\mathcal{B}$) whereas the former involves $\int_0^1 \mathcal{B}_u^2 du$.

Recall that $x_T \xrightarrow{D} x$ on $\mathbb{R}$ means that $f(x_T) \xrightarrow{D} f(x)$ for any continuous function $f : \mathbb{R} \mapsto \mathbb{R}$ This part of the definition of convergence in distribution is important, and is commonly referred to as the 'continuous mapping theorem'. By definition of convergence in distribution, this holds for any metric space, in particular $C(0,1)$. That is, for continuos functions, or mappings, as for example, $f : C(0,1) \to \mathbb{R}$ or $f : C(0,1) \to C(0,1)$.

Rewrite next, $\sum_{t=1}^{T} x_{t-1}^2$ for $\rho = 1$ as,

$$T^{-2} \sum_{t=1}^{T} (\sum_{i=1}^{t} \varepsilon_i)^2 = T^{-1} \sum_{u=1/T}^{1} (\frac{1}{\sqrt{T}} \sum_{i=1}^{[Tu]} \varepsilon_i)^2 = \int_0^1 (\frac{1}{\sqrt{T}} \sum_{i=1}^{[Tu]} \varepsilon_i)^2 du, \quad \text{(III.21)}$$

using the piecewise constancy of $\sum_{i=1}^{[Tu]} \varepsilon_i$. Now the mapping $x \mapsto f(x) = \int_0^1 x_u du$ from $C(0,1)$ into $\mathbb{R}$ is continuous. To see this let $x^T \to x$ on $C(0,1)$ and evaluate

$$|\int_0^1 (x_u^T - x_u) du| \leq \int_0^1 |x_u^T - x_u| du \leq \sup_{u \in [0,1]} |x_u^T - x_u| \quad \text{(III.22)}$$

which tends to zero by definition of convergence on $C(0,1)$. Similarly, $f(x) = \int_0^1 x_u^2 du$ is continuous and hence by applying Donsker's theorem and the definition of convergence in distribution,

$$T^{-2} \sum_{t=1}^{T} (\sum_{i=1}^{t} \varepsilon_i)^2 \xrightarrow{D} \int_0^1 \mathcal{B}_u^2 du \quad \text{(III.23)}$$

Unfortunately, the continuous mapping argument cannot be applied for the convergence of $\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\varepsilon_i)\varepsilon_t$ and instead a heuristic argument is given. A proof of the result is found in Chan & Wei (1988). Rewrite the term as follows

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\varepsilon_i)\varepsilon_t &= \sum_{u=1/T}^{1}(\frac{1}{\sqrt{T}}\sum_{i=1}^{[Tu]-1}\varepsilon_i)\Delta(\frac{1}{\sqrt{T}}\sum_{i=1}^{[Tu]}\varepsilon_i) \\ &\simeq \sum_{u=1/T}^{1}\mathcal{B}_u d\mathcal{B}_u \overset{"D"}{\rightarrow} \int_0^1 \mathcal{B}_u d\mathcal{B}_u \text{ as T} \rightarrow \infty. \qquad \text{(III.24)}\end{aligned}$$

In both cases the important thing is not the exact form of the limit, but that in fact the terms do converge in distribution.

What it means is that for example the stochastic variable $\int_0^1 \mathcal{B}d\mathcal{B}$ can be simulated by $T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\varepsilon_i)\varepsilon_t$ for large T. In fact all the quoted limit distributions have been tabulated that way.

Collecting the results gives:

Theorem III.3.3 *Under the assumption that $\varepsilon_t, t = 1, .., T$, are $iidN(0, \sigma^2)$ then*

$$(\frac{1}{\sqrt{T}}\sum_{i=1}^{[Tu]}\varepsilon_t, T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\varepsilon_i)\varepsilon_t, T^{-2}\sum_{t=1}^{T}(\sum_{i=1}^{t}\varepsilon_i)^2) \overset{D}{\rightarrow} (\mathcal{B}_u, \int_0^1 \mathcal{B}_u d\mathcal{B}_u, \int_0^1 \mathcal{B}_u^2 du)$$

where $\mathcal{B}$ is a Brownian motion with variance σ^2.

As for the FCLT the result can be further generalized to martingale difference sequences by Hansen (1992, Theorem 2.1):

Theorem III.3.4 *Let Y_t be a martingale sequence with respect to $\mathcal{F}_t$, for which as $T \rightarrow \infty$,*

$$\begin{aligned}&(i): \quad \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tu]}Y_t \overset{D}{\rightarrow} \mathcal{B}_u \textit{ with variance } \sigma^2 \textit{ and} \\ &(ii): \quad \sup_T \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[Y_t^2\right] < \infty\end{aligned}$$

Then, as $T \rightarrow \infty$,

$$(\frac{1}{\sqrt{T}}\sum_{i=1}^{[Tu]}Y_t, T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}Y_i)Y_t, T^{-2}\sum_{t=1}^{T}(\sum_{i=1}^{t}Y_i)^2) \overset{D}{\rightarrow} (\mathcal{B}_u, \int_0^1 \mathcal{B}d\mathcal{B}, \int_0^1 \mathcal{B}_u^2 du).$$

III.3.5 The AR(1) model reconsidered

Given the above presented theory, the results for the AR(1) model in (III.1) can be collected in the following theorem:

Theorem III.3.5 *Consider the AR(1) model in (III.1) with ML estimators given by,*

$$\hat{\rho} = S_{yz}S_{zz}^{-1} \text{ and } \hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{\rho}x_{t-1})^2 = S_{yy} - S_{yz}S_{zz}^{-1}S_{zy}, \tag{III.25}$$

where y, z in the product moments refer to x_t and x_{t-1} respectively. When $\rho = 1$, they are consistent, that is $\hat{\rho} \xrightarrow{P} 1$ and $\hat{\sigma}^2 \xrightarrow{P} \sigma_0^2$ as $T \to \infty$. Moreover,

$$T(\hat{\rho} - 1) \xrightarrow{D} \int_0^1 \mathcal{W}d\mathcal{W} / \int_0^1 \mathcal{W}_u^2 du, \tag{III.26}$$

where $\mathcal{W}$ is a standard Brownian motion.

The LR test statistic for the hypothesis that $\rho = 1$ is given by,

$$LR(\rho = 1) = T\log(1 + W_T), \qquad W_T = (\hat{\rho} - 1)^2 \frac{S_{zz}}{\hat{\sigma}^2}. \tag{III.27}$$

For $\rho = 1$, the LR statistic is asymptotically Dickey-Fuller type distributed,

$$LR(\rho = 1) \xrightarrow{D} (\int_0^1 \mathcal{W}d\mathcal{W})^2 / \int_0^1 \mathcal{W}_u^2 du \quad \text{as } T \to \infty. \tag{III.28}$$

Proof:

The expressions for the ML estimators and the LR statistic in (III.25) and (III.27) follow by the results for the AR(1) model established in Part 2 of the lecture notes. The result in (III.26), holds as

$$T(\hat{\rho} - 1) = S_{\varepsilon z}S_{zz}^{-1} = \frac{\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t x_{t-1}}{\frac{1}{T^2}\sum_{t=1}^{T}x_{t-1}^2}.$$

Consider first the denominator,

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t x_{t-1} = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{i=1}^{t-1}\varepsilon_i + x_0\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t.$$

The first term converges in distribution to $\int_0^1 \mathcal{B}d\mathcal{B}$ by Theorem III.3.4 as $\Delta x_t = \varepsilon_t$ are i.i.d.N$(0, \sigma^2)$. The second term converges in probability to

$\mathbb{E}[\varepsilon_t] = 0$ by LLN for i.i.d. processes. For the numerator, applying Theorem III.3.4 gives the desired, as $\Delta x_t = \varepsilon_t$ is a martingale difference.

Finally, turn to the LR statistic, where

$$TW_T = TS_{\varepsilon z}S_{zz}^{-1}S_{\varepsilon z}/\hat{\sigma}^2 \xrightarrow{D} \int_0^1 \mathcal{B}d\mathcal{B}(\int_0^1 \mathcal{B}_u^2 du)^{-1}\int_0^1 \mathcal{B}d\mathcal{B}/\sigma_0^2, \tag{III.29}$$

using the same arguments as before, and that $\hat{\sigma}^2$ is consistent. The result in (III.28) follows by setting $\mathcal{W} = \frac{1}{\sqrt{\sigma^2}}\mathcal{B}$. The consistency of $\hat{\sigma}^2$ can be seen by rewriting,

$$\hat{\sigma}^2 = S_{\varepsilon\varepsilon} - (\hat{\rho}-1)^2 S_{zz} = S_{\varepsilon\varepsilon} + o_P(1) \xrightarrow{P} \sigma^2. \tag{III.30}$$

That $(\hat{\rho}-1)^2 S_{zz} = o_P(1)$, holds as $(\hat{\rho}-1) S_{zz} = O_P(1)$, while $\hat{\rho} \xrightarrow{P} 1$. □

III.4 Testing for unit-roots in AR(k) models

It is discussed here what it means to allow for a unit root in the AR(k) model and how to test for it.

III.4.1 I(1) and I(0) Processes

First a definition of 'non-stationarity' and 'stationarity' suitable for the classification of AR processes is needed. Recall that the AR(1) process x_t is geometrically ergodic with stationary version $x_t^* = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$ if $|\rho| < 1$, while if $\rho = 1$, $x_t - x_0$ is a random walk.

Definition III.4.1 *A geometrically ergodic process x_t is called I(0), if the stationary version x_t^* is a linear proces which satisfies $x_t^* = \phi(L)\varepsilon_t = \sum_{i=0}^{\infty} \phi_i \varepsilon_{t-i}$ with $\phi(1) = \sum_{i=0}^{\infty} \phi_i \neq 0$.*

Example III.4.1 *For $|\rho| < 1$, the AR(1) process $x_t^* = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$ is I(0), because $\sum_{i=0}^{\infty} \rho^i = (1-\rho)^{-1} \neq 0$.*

Example III.4.2 *Consider the AR(1) process with a constant μ and $|\rho| < 1$,*

$$x_t = \rho x_{t-1} + \mu + \varepsilon_t, \tag{III.31}$$

which is geometrically ergodic with stationary solution x_t^* for which,

$$x_t^* - \mathbb{E}[x_t^*] = x_t^* - \frac{\mu}{1-\rho} = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}, \quad \text{and} \ \sum_{i=0}^{\infty} \rho^i = (1-\rho)^{-1}. \tag{III.32}$$

Hence, $x_t - \frac{\mu}{1-\rho}$ is an I(0) process. Similarly, for $|\rho| < 1$ the AR(1) process with a linear trend,

$$x_t = \rho x_{t-1} + \mu_0 + \mu_1 t + \varepsilon_t, \tag{III.33}$$

can be written as $x_t = x_t + \frac{\mu_1}{1-\rho}t$ where

$$x_t = \rho x_{t-1} + \tilde{\mu} + \varepsilon_t, \tag{III.34}$$

and $\tilde{\mu} = \mu_0 - \frac{\mu_1}{1-\rho}$. Hence, $x_t - \frac{\tilde{\mu}}{1-\rho}t$ is an I(0) process, with

$$x_t^* = \frac{\tilde{\mu}}{1-\rho} + \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}. \tag{III.35}$$

That is, x_t with its linear trend subtracted, $x_t - \frac{\mu_1}{1-\rho}t$ is an I(0) process and has a stationary version. Thus x_t is trend-I(0), and x_t^* trend-stationary.

As an example of a stationary process, which is not I(0) consider,

$$x_t = \Delta\varepsilon_t = \varepsilon_t - \varepsilon_{t-1} \tag{III.36}$$

Clearly, x_t is stationary and linear, but it is not I(0). The reason for not including $\Delta\varepsilon_t$ in the I(0) processes is that, when accumulating it, $\sum_{i=1}^{t} \Delta\varepsilon_i = \varepsilon_t - \varepsilon_0$, it is still stationary and no CLT, and hence FCLT, applies to the accumulated process. This is contrary to $x_t = \varepsilon_t$, where both a FCLT and CLT applies.

Next define I(1) and I(2) processes.

Definition III.4.2 *A stochastic process x_t is called integrated of order $d = 1, 2$, I(d), if $\Delta^d x_t$ is I(0).*

Example III.4.3 *A random walk, $x_t = \sum_{i=1}^{t} \varepsilon_i$, is indeed an I(1) process. For the random walk with drift,*

$$x_t = \sum_{i=1}^{t} \varepsilon_t + \mu t,$$

$x_t - \mu t$ is an I(1) processes.

Example III.4.4 *An example of an I(2) process is $\Delta^2 x_t = \varepsilon_t$. While processes integrated of order 2 are relevant for empirical applications, processes integrated of order higher than 2 have so far no practical applications.*

III.4.2 The AR(2) model

Consider the AR(2) model as given by

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t, \quad t = 1, 2, ..., T \tag{III.37}$$

with x_0 and x_{-1} fixed, ε_t i.i.d.N(0,σ^2) and parameters $(\rho_1, \rho_2, \sigma^2) \in \mathbb{R}^2 \times \mathbb{R}_+$.

The characteristic polynomial evaluated at $z = 1$,

$$A(1) = 1 - \rho_1 - \rho_2,$$

is zero if and only if $\rho_1 + \rho_2 = 1$. To make this restriction even simpler, reparametrize the AR(2) model as

$$\Delta x_t = \pi x_{t-1} + \gamma \Delta x_{t-1} + \varepsilon_t, \tag{III.38}$$

where $\pi = \rho_1 + \rho_2 - 1$, and $\gamma = -\rho_2$ are freely varying parameters, and $(x_0, \Delta x_0)$ fixed. In this parametrization, the characteristic polynomial is given by

$$A(z) = (1 - z) - \pi z - \gamma z (1 - z), \tag{III.39}$$

which is zero at $z = 1$ if and only if $\pi = 0$. Hence the hypothesis of a unit root is equivalent to,

$$H_0 : \pi = 0.$$

III.4.2.1 Properties of x_t when $\pi = 0$.

Under H_0, and if $|\gamma| < 1$, then $S_t \equiv \Delta x_t$ is geometrically ergodic with stationary solution,

$$S_t^* = \sum_{i=0}^{\infty} \gamma^i \varepsilon_{t-i}.$$

Next, use that by definition under H_0, $\gamma(z) = 1 - \gamma z = \gamma(1) + \gamma(1 - z)$, with $\gamma(L) \Delta x_t = \varepsilon_t$, such that

$$\gamma(L) \Delta x_t = \gamma(1) \Delta x_t + \gamma \Delta^2 x_t = \varepsilon_t.$$

Hence, with $|\gamma| < 1$,

$$\Delta x_t = \frac{1}{\gamma(1)}(-\gamma \Delta^2 x_t + \varepsilon_t),$$

and summation over t gives,

$$x_t = \frac{1}{\gamma(1)} \sum_{i=1}^{t} \varepsilon_i + \frac{\gamma}{\gamma(1)} S_t + (x_0 - \frac{\gamma}{\gamma(1)} \Delta x_0). \tag{III.40}$$

In other words, under the hypothesis of a unit root, or $\pi = 0$, and if $|\gamma| < 1$, then x_t can be represented as the sum of a random walk, a geometrically ergodic process S_t and initial values as a function of x_0 and x_{-1}. Thus x_t is non-stationary as an I(1) process.

Furthermore, with $u \in [0, 1]$,

$$\frac{1}{\sqrt{T}} x_{[Tu]} = 1/\gamma(1) \frac{1}{\sqrt{T}} \sum_{i=1}^{[Tu]} \varepsilon_i + o_P(1) \xrightarrow{D} 1/\gamma(1) \mathcal{B}_u = \frac{1}{1-\gamma} \mathcal{B}_u, \quad \text{(III.41)}$$

where $\mathcal{B}$ is a Brownian motion with variance σ^2. That is, a FCLT applies to x_t, and the limit is the same Brownian motion as in the AR(1) case but multiplied by a constant which reflects the further lag.

That (III.41) holds, follows by applying Theorem III.3.2 to ε_t, which shows that $\sum_{i=1}^{[T\cdot]} \varepsilon_i / \sqrt{T}$ converge in distribution to $\mathcal{B}$ on $C(0, 1)$. Next, note that $\frac{1}{\sqrt{T}}(x_0 - \frac{\gamma}{\gamma-1} \Delta x_0) \xrightarrow{P} 0$ as x_0 and Δx_0 are fixed initial values. Finally, for the $o_P(1)$ term one needs to show $\frac{1}{\sqrt{T}} S_{[Tu]}$ converge to zero in probability on $C[0, 1]$.

Some further results are needed for the product moments of x_t and $\Delta x_t = S_t$. These, as well as the previous considerations, are stated in the proposition:

Proposition III.4.1 *Consider x_t given by (III.38). If $\pi = 0$, then $\gamma(L) \Delta x_t = \varepsilon_t$, where $\gamma(z) = 1 - \gamma z$. If furthermore, $|\gamma| < 1$, $S_t \equiv \Delta x_t$ is geometrically ergodic, and x_t has the representation in (III.40). Moreover, as $T \to \infty$, it holds that with $u \in [0, 1]$ and $\mathcal{B}_u$ a Brownian motion with variance σ^2,*

$$\begin{aligned} &\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\cdot]} x_t, \frac{1}{T^2} \sum_{t=1}^{T} x_{t-1}^2, \frac{1}{T} \sum_{t=1}^{T} x_{t-1} S_t\right) \xrightarrow{D} \quad \text{(III.42)} \\ &\left(\frac{1}{\gamma(1)} \mathcal{B}_\cdot, \frac{1}{\gamma(1)^2} \int_0^1 \mathcal{B}_u^2 du, \frac{1}{\gamma(1)^2} \int_0^1 \mathcal{B} d\mathcal{B} + \omega\right), \end{aligned}$$

where $\omega = \sum_{h=1}^{\infty} Cov(S_t^*, S_{t+h}^*)$. Also, jointly,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} S_{t-1} \varepsilon_t \xrightarrow{D} N\left(0, \sigma^2 \mathbb{V}(S_t^*)\right). \quad \text{(III.43)}$$

Proof: The convergence to $\mathcal{B}$ and $\int \mathcal{B}^2 du$ hold by (III.41) and the continuity of $f(x) = \int_0^1 x_u^2 du$, $f : C(0, 1) \to \mathbb{R}$. The convergence towards the stochastic integral holds by Theorem 4.1 in Hansen (1992), in combination with the LLN. The Gaussian limit holds by the CLT for martingale differences. $\square$

The following result is used repeatedly:

Lemma III.4.1 *Let S_t be a geometrically ergodic process, with $\mathbb{E}\left[|S_t^*|^3\right] < \infty$. Then*

$$\frac{1}{\sqrt{T}} S_{[T\cdot]} \xrightarrow{P} 0$$

as $T \to \infty$.

Proof: By definition of convergence in probability on $C(0,1)$ it has to be shown that

$$P(\sup_{u\in[0,1]} \frac{1}{\sqrt{T}}|S_{[T\cdot]}| > \delta) \to 0.$$

As before, this follows by,

$$P(\sup_{u\in[0,1]} \frac{1}{\sqrt{T}}|S_{[T\cdot]}| > \delta) = P(\max_{t=1,..,T} |S_t| > \delta\sqrt{T}) \tag{III.44}$$

$$\leq \frac{1}{T^{3/2}\delta^3} \sum_{t=1}^{T} E(|S_t|^3 \mathbb{I}(|S_t| > \delta\sqrt{T})) = O(T^{-1/2})$$

$\square$

III.4.2.2 Testing $\pi = 0$

With $Y_t \equiv \Delta x_t$, $\beta' = (\pi, \gamma)$ and $Z_t = (x_{t-1}, \Delta x_{t-1})'$, the AR(2) model in (III.38) can be analyzed by OLS in the linear regression, $Y_t = \beta' Z_t + \varepsilon_t$ and the following theorem holds:

Theorem III.4.1 *Consider the AR(2) model in (III.38) with ML estimators given by,*

$$\hat{\beta}' = (\hat{\pi}, \hat{\gamma}) = S_{yz} S_{zz}^{-1}, \quad \text{and } \hat{\sigma}^2 = S_{yy\cdot z}. \tag{III.45}$$

where y, z in the product moments refer to $Y_t = \Delta x_t$ and $Z_t = (x_{t-1}, \Delta x_{t-1})'$ respectively.

When $\pi = 0$, and $|\gamma| < 1$, they are consistent, that is $\hat{\pi} \xrightarrow{P} 1, \hat{\gamma} \xrightarrow{P} \gamma$ and $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ as $T \to \infty$. Moreover,

$$(T\hat{\pi}, \sqrt{T}(\hat{\gamma} - \gamma)) \xrightarrow{D} ((1-\gamma) \int_0^1 \mathcal{W} d\mathcal{W} / \int_0^1 \mathcal{W}_u^2 du,\ N\left(0, 1-\gamma^2\right)) \tag{III.46}$$

where $\mathcal{W}$ is a standard Brownian motion.

The LR statistic for the hypothesis $H_0 : \pi = 0$, *is given by,*

$$LR\,(\pi = 0) = T \log\left(1 + W_T\right), \tag{III.47}$$
$$W = TW_T = TS_{yz}S_{zz}^{-1}R(R'S_{zz}^{-1}R)^{-1}R'S_{zz}^{-1}S_{zy}/\hat{\sigma}^2, \tag{III.48}$$

where $R' = (1, 0)$. *When* $\pi = 0$, *and* $|\gamma| < 1$, *the LR statistic is asymptotically Dickey-Fuller type distributed,*

$$LR\,(\pi = 0) \xrightarrow{D} (\int_0^1 \mathcal{W}d\mathcal{W})^2 / \int_0^1 \mathcal{W}_u^2 du \quad \textit{as } T \to \infty. \tag{III.49}$$

Note that the limiting distributions are not affected by the presence of the additional lag as can be seen by comparing with for example (III.29) and Theorem III.3.5. However, the small sample distributions may be severely affected by (estimation of) γ.

Observe that $\hat{\pi}$ is super-consistent, while $\hat{\gamma}$ is the usual $\sqrt{T}$ consistent.

Note also that the LR statistic can be written in many ways, the one in (III.48) has been chosen to emphasize the role of the restriction matrix $R' = (0, 1)$. With $Z_{1t} = x_{t-1}, Z_{2t} = \Delta x_{t-1}$, the formula may also be written as

$$W = TW_T = TS_{y1\cdot 2}S_{11\cdot 2}^{-1}S_{1y\cdot 2}/\hat{\sigma}^2, \tag{III.50}$$

where, for example, $S_{y1\cdot 2} = S_{yz_1} - S_{yz_2}S_{z_2z_2}^{-1}S_{z_2z_1}$.

This corresponds to the usual regression interpretation of $\hat{\pi}$, that is, $\hat{\pi}$ can be found in two steps. First regress Δx_t and x_{t-1} on Δx_{t-1} and obtain the residuals,

$$R(\Delta x_t|\Delta x_{t-1}) = \Delta x_t - S_{yz_2}S_{z_2z_2}^{-1}\Delta x_{t-1},$$
$$R(x_{t-1}|\Delta x_{t-1}) = x_{t-1} - S_{z_1z_2}S_{z_2z_2}^{-1}\Delta x_{t-1}.$$

Next, $\hat{\pi}$ is obtained by OLS regression of $R(\Delta x_t|\Delta x_{t-1})$ on $R(x_{t-1}|\Delta x_{t-1})$, giving $\hat{\pi} = S_{y1\cdot 2}S_{11\cdot 2}^{-1}$.

Proof of Theorem III.4.1:

Using Theorems III.6 and III.9, the ML estimators are given by (III.45). Likewise the LR statistic of $\pi = 0$ can be written as in (III.47) using for example equation (III.43) and the proof of Theorem III.9.

Corresponding to the different rates of convergence, define the normalization matrix N_T as,

$$N_T \equiv \begin{pmatrix} \frac{1}{\sqrt{T}} & 0 \\ 0 & 1 \end{pmatrix}.$$

By definition, $\hat{\beta} - \beta = S_{zz}^{-1} S_{z\varepsilon}$ and hence,

$$\begin{pmatrix} T\hat{\pi} \\ \sqrt{T}(\hat{\gamma} - \gamma) \end{pmatrix} = (N_T S_{zz} N_T)^{-1} \sqrt{T} N_T S_{z\varepsilon}$$

Note that,

$$\begin{aligned} N_T S_{zz} N_T = & \begin{pmatrix} \frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2 & \frac{1}{T^{3/2}} \sum_{t=1}^T x_{t-1} \Delta x_{t-1} \\ \frac{1}{T^{3/2}} \sum_{t=1}^T x_{t-1} \Delta x_{t-1} & \frac{1}{T} \sum_{t=1}^T \Delta x_{t-1}^2 \end{pmatrix} \\ & \xrightarrow{D} \begin{pmatrix} (\frac{1}{1-\gamma})^2 \int \mathcal{B}_u^2 du & 0 \\ 0 & \frac{\sigma^2}{(1-\gamma^2)} \end{pmatrix}, \end{aligned}$$

where $\mathcal{B}$ is a Brownian motion with variance σ^2. This follows by Proposition III.4.1. Likewise,

$$\sqrt{T} N_T S_{z\varepsilon} = \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T x_{t-1} \varepsilon_t \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T \Delta x_{t-1} \varepsilon_t \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \frac{1}{1-\gamma} \int \mathcal{B} d\mathcal{B} \\ N(0, \frac{\sigma^4}{(1-\gamma^2)}) \end{pmatrix},$$

Collecting terms,

$$\begin{pmatrix} T\hat{\pi} \\ \sqrt{T}(\hat{\gamma} - \gamma) \end{pmatrix} \xrightarrow{D} \begin{pmatrix} (1-\gamma) \left(\int \mathcal{B}_u^2 du \right)^{-1} \int \mathcal{B} d\mathcal{B} \\ N\left(0, 1 - \gamma^2\right) \end{pmatrix}$$

and (III.46) follows. For the LR statistic, consider TW_T,

$$TW_T = T S_{\varepsilon z} S_{zz}^{-1} R (R' S_{zz}^{-1} R)^{-1} R' S_{zz}^{-1} S_{z\varepsilon} / \hat{\sigma}^2.$$

Write the denominator as,

$$\begin{aligned} & \sqrt{T} S_{\varepsilon z} N_T (N_T S_{zz} N_T)^{-1} N_T R (R' N_T (N_T S_{zz} N_T)^{-1} N_T R)^{-1} \times \\ & \qquad R' N_T (N_T S_{zz} N_T)^{-1} N_T S_{z\varepsilon} \sqrt{T} \\ = & \sqrt{T} S_{\varepsilon z} N_T (N_T S_{zz} N_T)^{-1} R (R' (N_T S_{zz} N_T)^{-1} R)^{-1} R' (N_T S_{zz} N_T)^{-1} N_T S_{z\varepsilon} \sqrt{T} \end{aligned}$$

where it has been used that $N_T' R = \frac{1}{\sqrt{T}} R$ as $R' = (1, 0)$. The just applied arguments gives that this converge in distribution to,

$$\int \mathcal{B} d\mathcal{B} \left(\int \mathcal{B}^2 du \right)^{-1} \int \mathcal{B} d\mathcal{B} \overset{D}{=} \sigma^2 \int \mathcal{W} d\mathcal{W} \left(\int \mathcal{W}^2 du \right)^{-1} \int \mathcal{W} d\mathcal{W}$$

with $\mathcal{W} = \mathcal{B}/\sigma$. Using that $\hat{\sigma}^2$ is consistent, the result in (III.49) follows as $W_T = O_P(T^{-1})$ and the usual Taylor expansion of $\log(1 + w)$ can be applied. $\square$

III.4.3 The AR(k) model

Next turn to the AR(k) model which will only briefly be discussed since the main arguments have been given in connection with the discussion of the AR(2) model. The AR(k) model is given by,

$$x_t = \rho_1 x_{t-1} + \ldots + \rho_k x_{t-k} + \varepsilon_t, \ t = 1, ..., T \tag{III.51}$$

ε_t i.i.d.N$(0, \sigma^2)$ and $x_0, ..., x_{-k}$ fixed. To anticipate the unit root analysis reparametrize (III.51) as

$$\Delta x_t = \pi x_{t-1} + \gamma_1 \Delta x_{t-1} + \ldots + \gamma_{k-1} \Delta x_{t-k+1} + \varepsilon_t$$

where $\pi = -A(1) = \sum_{i=1}^{k} \rho_i - 1$ and $\gamma_i = -\sum_{j=i+1}^{k} \rho_j$. Thus the characteristic polynomial takes the form

$$A(z) = (1 - z) - \pi z - \gamma_1 (1 - z) z - \ldots - \gamma_{k-1} (1 - z) z^{k-1}$$

which has a unit root at $z = 1$ if and only if $\pi = 0$.

III.4.3.1 Representation for the AR(k)

Suppose that $\pi = 0$, or equivalently,

$$\Delta x_t = \gamma_1 \Delta x_{t-1} + .. + \gamma_{k-1} \Delta x_{t-k+1} + \varepsilon_t$$

Set $\gamma(z)$ equal to the characteristic polynomial of Δx_t, that is

$$\gamma(z) \equiv 1 - \gamma_1 z - ... - \gamma_{k-1} z^{k-1}, \tag{III.52}$$

and $A(z) = (1 - z)\gamma(z)$ when $\pi = 0$. Then if $|\gamma(z)| = 0$ implies $|z| > 1$, $(\Delta x_t, ..., \Delta x_{t-k+1})$ is geometrically ergodic. In particular, Δx_t admits a stationary representation of the form,

$$\Delta x_t^* = \theta(L) \varepsilon_t \tag{III.53}$$

where $\theta(z) = \gamma(z)^{-1} = \sum_{i=0}^{\infty} \theta_i z^i$, with θ_i given in Theorem II.9.

Similar to the AR(2) case the following result holds:

Theorem III.4.2 *Let x_t be an AR(k) process given by (III.51). Assume that $A(z)$ has one, and only one, unit-root, while the remaining roots, that is the roots of $\gamma(z)$ in (III.52), are larger than one in absolute value. Then x_t is an I(1) process, with representation,*

$$x_t = \phi \sum_{i=1}^{t} \varepsilon_t + \lambda' S_t + \lambda_0, \tag{III.54}$$

where $\phi = \gamma(1)^{-1}$, $S_t = (\Delta x_t, ..., \Delta x_{t-k+1})'$ is geometrically ergodic, and $\lambda' S_t$ is a linear combination of S_t. The constant λ_0 depends on initial values of the AR(k) process and is given by $\lambda_0 = \lambda' S_0 + x_0$.

The vector $\lambda \in \mathbb{R}^{k-1}$ is given by $\lambda' = \phi\left(\gamma_0^*, ..., \gamma_{k-2}^*\right)$, where $\gamma_0^* = 1 - \gamma(1)$, and $\gamma_i^* = \gamma_{i-1}^* - \gamma_i$ for $i = 1, .., k-2$.

Thus under the assumption of a unit root, and the additional assumption that the remaining roots are outside the unit circle, x_t has a representation as a random walk plus a linear combination of a geometrically ergodic process. The LLN applies to the $\gamma' S_t$ term, see Lemma III.4.1, and the FCLT apply to the random walk part such that,

$$\frac{1}{\sqrt{T}} x_{[T\cdot]} \xrightarrow{D} \phi \mathcal{B}, \tag{III.55}$$

where $\mathcal{B}$ is a Brownian motion on [0,1] with variance σ^2.

Hence testing for a unit root in the AR(k) model, that is $\pi = 0$, is equivalent to testing if the autoregressive process is an I(1) process provided that the remaining roots are larger than one in absolute value.

Proof:

The arguments are analogous to the AR(2) case. Expand $\gamma(z)$ as follows,

$$\gamma(z) = \gamma(1) + \gamma^*(z)(1-z), \tag{III.56}$$

where $\gamma^*(z) = (\gamma(z) - \gamma(1))/(1-z)$ is a polynomial of order $k-2$, $\gamma^*(z) = \gamma_0^* + \gamma_1^* z + ... + \gamma_{k-2}^* z^{k-2}$. Simple identification of coefficients show that γ_i^* and γ_i are related by

$$\gamma_0^* = 1 - \gamma(1), \quad \gamma_i^* = \gamma_{i-1}^* - \gamma_i \text{ for } i = 1, 2, ..., k-2. \tag{III.57}$$

Use this to rewrite $\gamma(L)\Delta x_t = \varepsilon_t$ as,

$$\gamma(L)\Delta x_t = \gamma(1)\Delta x_t + \gamma^*(L)\Delta^2 x_t = \varepsilon_t. \tag{III.58}$$

Next, divide by $\gamma(1)$, and consider $\sum_{i=1}^t \Delta x_i$ which, with $\phi = 1/\gamma(1)$, equals,

$$x_t = x_0 + \phi\left(\sum_{i=1}^t \varepsilon_i + \gamma^*(L)(\Delta x_t - \Delta x_0)\right). \tag{III.59}$$

Now, $\gamma^*(L)\Delta x_t = \gamma_0^* \Delta x_t + ... + \gamma_{k-2}^* \Delta x_{t-k+2}$, and hence $S_t \equiv \phi\gamma^*(L)\Delta x_t$ is a linear combination of the geometrically ergodic process $(\Delta x_t, ..., \Delta x_{t-k+2})'$ such that the LLN applies to S_t. Likewise, S_0 is a linear combination of the initial values $(\Delta x_0, ..., \Delta x_{-k+2})'$. □

III.4.3.2 Testing for a unit-root

The assumption of a unit root in the AR(k) model in (III.51) is equivalent to the assumption that

$$H_0: \ \pi = 0. \tag{III.60}$$

The analysis of the AR(k) model is identical to the analysis of the AR(2) model, with $Y_t \equiv \Delta x_t$, $\beta' = (\pi, \gamma')$, where $\gamma' \equiv (\gamma_1, .., \gamma_{k-1})$ and $Z_t = (x_{t-1}, \Delta x_{t-1}, ..., \Delta x_{t-k+1})'$. Using Theorem III.4.2, mimicking the proof for the AR(2) case of Theorem III.4.1 and noting that Proposition III.4.1 generalizes immediately to the AR(k) case, the following holds:

Theorem III.4.3 *Consider the AR(k) model in (III.51) with ML estimators given by,*

$$\hat{\beta}' = (\hat{\pi}, \hat{\gamma}') = S_{yz}S_{zz}^{-1}, \quad \text{and } \hat{\sigma}^2 = S_{yy\cdot z}. \tag{III.61}$$

where z, y in the product moments refer to $Z_t = (x_{t-1}, \Delta x_{t-1}, ..., \Delta x_{t-k+1})'$ and $Y_t = \Delta x_t$ respectively.

When $\pi = 0$, and $|\gamma(z)| = 0$ implies $|z| > 1$, $\hat{\beta}$ and $\hat{\sigma}^2$ are consistent, that is $\hat{\pi} \xrightarrow{P} 1, \hat{\gamma} \xrightarrow{P} \gamma$ and $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ as $T \to \infty$. Moreover,

$$(T\hat{\pi}, \sqrt{T}(\hat{\gamma} - \gamma)') \xrightarrow{D} (\phi^{-1} \int_0^1 \mathcal{W}d\mathcal{W} / \int_0^1 \mathcal{W}_u^2 du, \ N_{k-1}(0, V)) \tag{III.62}$$

where $\mathcal{W}$ is a standard Brownian motion and $V = V(S_t^*)$, see Theorem III.4.2.

The LR statistic for the hypothesis $H_0 : \pi = 0$, or with $R'\beta = (1, 0)\beta = \pi = 0$, is given by,

$$LR(\pi = 0) = T \log (1 + W_T), \ \text{where} \tag{III.63}$$

$$W = TW_T = TS_{yz}S_{zz}^{-1}R(R'S_{zz}^{-1}R)^{-1}R'S_{zz}^{-1}S_{zy}/\hat{\sigma}^2, \tag{III.64}$$

When $\pi = 0$, and $|\gamma(z)| = 0$ implies $|z| > 1$, the LR statistic is asymptotically Dickey-Fuller type distributed,

$$LR(\pi = 0) \xrightarrow{D} (\int_0^1 \mathcal{W}d\mathcal{W})^2 / \int_0^1 \mathcal{W}_u^2 du \quad \text{as } T \to \infty. \tag{III.65}$$

III.5 The role of deterministic terms

Consider the simple AR(1) model with a constant regressor

$$\Delta x_t = \pi x_{t-1} + \mu + \varepsilon_t$$

Under the assumption $\pi = 0$,

$$x_t = x_0 + \sum_{i=1}^{t} \varepsilon_i + \mu t, \tag{III.66}$$

whereas, if $\pi \in (-2, 0)$ or equivalently $|\rho| < 1$,

$$x_t^* = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} + \frac{\mu}{1-\rho}. \tag{III.67}$$

Thus testing the assumption of $\pi = 0$ against the alternative $\pi \in (-2, 0)$ essentially tests if x_t is a random walk with drift against it being a stationary process with a constant mean. This lack of deterministic 'balance' shows up in the limit distribution of the likelihood ratio test which will have nuisance parameters in the sense that there are two different limit distributions depending on whether or not $\mu = 0$. Apart from this in most practical situations such as the analysis of the US-GNP below it is of interest to test if the process is trend-stationary against it being a random walk with 'stationary noise'. This can be accomplished by first testing whether the process is I(1) in a model which allows a deterministic trend both under and outside the alternative. Next, the determination of whether there is a linear trend or not, can be done by the standard χ^2 distributed likelihood ratio test statistics.

Similar considerations lead as in Dickey and Fuller (1979) to consider the following AR(k) model when analyzing the I(1) hypothesis in the case of deterministic linear trends,

$$\Delta x_t = \pi_1 x_{t-1} + \pi_2 t + \gamma_1 \Delta x_{t-1} + \ldots + \gamma_{k-1} \Delta x_{t-k-1} + \mu + \varepsilon_t$$

where the hypothesis of interest is,

$$H: \ \pi_1 = \pi_2 = 0.$$

Under this hypothesis it follows, under the assumption of Theorem III.4.2, that

$$x_t = \phi \sum_{i=1}^{t} \varepsilon_i + \phi \mu t + \lambda' S_t + a + c,$$

where c is a constant. That is,x_t is an I(1) process with a linear trend. Under the alternative hypothesis that all roots of the characteristic polynomial are outside the unit circle, x_t is trend-I(0) and has a representation as,

$$x_t^* = \theta(L)\varepsilon_t + \theta(1)\mu t + \text{constant}$$

i.e. trend-stationary or a stationary process with a linear trend. Having determined whether or not $\pi_1 = \pi_2 = 0$ a successive test of the presence of a linear trend is as mentioned the usual χ^2 test. Note that if x_t is I(1) this is a test of $\mu = 0$, whereas if x_t is trend-stationary it is a test of $\pi_2 = 0$.

The likelihood ratio test statistic of the hypothesis, $\pi_1 = \pi_2 = 0$ has the same form as before, and is based on OLS regression of Δx_t on $(x_{t-1}, t)'$ corrected for a constant and the lagged differences, Δx_{t-i}. The limit distribution of the likelihood ratio test statistic is in this case given by

$$\mathrm{LR}(\pi_1 = \pi_2 = 0) \xrightarrow{D} (\int_0^1 F d\mathcal{W})'(\int_0^1 F_u F_u')^{-1} \int_0^1 F d\mathcal{W},$$

where $\mathcal{W}$ is a standard Brownian motion and the two-dimensional process F is given by

$$F_u = \begin{pmatrix} \mathcal{W}_u - \int_0^1 \mathcal{W}_s ds \\ u - \int_0^1 s ds \end{pmatrix}.$$

This distribution is tabulated in Johansen (1996) and some quantiles are reported below.

Mimicking the ideas above, the model which allows for a constant level is given by

$$\Delta x_t = \pi_1 x_{t-1} + \pi_2 + \gamma_1 \Delta x_{t-1} + \ldots + \gamma_{k-1} \Delta x_{t-k-1} + \varepsilon_t.$$

In this case, under the hypothesis of $\pi_1 = \pi_2 = 0$, the likelihood ratio test statistic has a limit distribution as above, but with $F_u = (\mathcal{W}_u, 1)$.

III.5.1 Quantiles for LR testing

Summarizing the discussion above three different models were of interest. With the notation $V_t = (\Delta x_{t-1}, .., \Delta x_{t-k+1})'$ and $\gamma' = (\gamma_1, .., \gamma_{k-1})$ these can be rewritten as,

$$\begin{aligned} H_0 &: \ \Delta x_t = \pi x_{t-1} + \gamma' V_t + \varepsilon_t \\ H_1 &: \ \Delta x_t = (\pi_1, \pi_2)(x_{t-1}, 1)' + \gamma' V_t + \varepsilon_t \\ H_2 &: \ \Delta x_t = (\pi_1, \pi_2)'(x_{t-1}, t)' + \gamma' V_t + \mu + \varepsilon_t \end{aligned}$$

The hypotheses of interest are $H_0^* : \pi = 0$, $H_1^* : \pi_1 = \pi_2 = 0$ and $H_2^* : \pi_1 = \pi_2 = 0$ respectively.

For $i = 0, 1, 2$, the limit distributions of the likelihood ratio test statistics of H_i^* against H_i are given by

$$\int d\mathcal{W} F (\int F F du)^{-1} \int F d\mathcal{W} \qquad \text{(III.68)}$$

under the assumptions of Theorem 4.3. Here $\mathcal{W}$ is a standard Brownian motion and F takes the forms,

$$H_0^* : F_u = \mathcal{W}_u$$
$$H_1^* : F_u = (\mathcal{W}_u, 1)'$$
$$H_2^* : F_u = (\mathcal{W}_u - \int \mathcal{W}_s ds, u - 1/2)'.$$

The quantiles of $(III.68)$ given below are from Johansen (1996). For comparison also the quantiles of the χ^2 distributions with 1 and 2 degrees of freedom are given.

Quantiles of the Likelihood Ratio Tests for Unit Roots

Hypothesis	95% quantile	97.5 % quantile	99 % quantile
H_0^*	4.2	5.3	7.0
H_1^*	9.1	10.7	12.7
H_2^*	12.4	14.1	16.4
χ_1^2	3.84	5.02	6.64
χ_2^2	6.0	7.4	9.21

III.5.2 The US-GNP example

A preliminary analysis of log(US-GNP) quarterly data from the period 1959:3 – 1996:4 indicates that an AR(6) model with a linear trend describe well the dynamics. This is in accordance with the economic literature on growth. But it is also found that one of the roots in the characteristic polynomial is close to one, $z = 0.916$.

The interpretation of μ changes dramatically depending on whether or not there is a unit root. Write the AR(6) model to anticipate the I(1) analysis as

$$\Delta x_t = \pi_1 x_{t-1} + \pi_2 t + \gamma_1 \Delta x_{t-1} + .. + \gamma_5 \Delta x_{t-5} + \mu + \varepsilon_t \tag{III.69}$$

The reported values are

$$\begin{array}{rrr} \text{Parameter} & \text{Estimate} & t\text{-value.} \\ \pi_1 & -0.055 & -3.05 \\ \pi_2 & 0.004 & 2.74 \end{array} \tag{III.70}$$

Clearly, based on the discussion so far, it is not clear how the reported t-values should be interpreted if indeed "0.916 = 1" and $\pi_2 = 0$.

The likelihood ratio test statistic of the hypothesis $\pi_1 = \pi_2 = 0$ equals,

$$\text{LR}(\pi_1 = \pi_2 = 0) = 14.3 \tag{III.71}$$

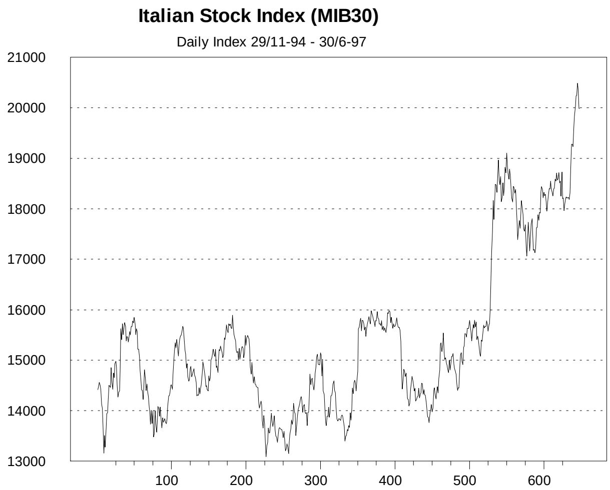
which according to the χ^2_2 distribution is clearly rejected. However, it is not the χ^2 distribution but the Dickey-Fuller type distribution that should be used with a 95% quantile of 12.4 and a 97.5% quantile of 14.13. Hence it is not clear if the hypothesis should be rejected. Thus a model for log(US-GNP) is a random walk with drift.

To test whether indeed the drift is present is very simple, since this is a simple hypothesis in the stationary model of $\Delta \log$(US-GNP). The reported t-value is 2.6 and, based on the normal distribution, the drift is present.

On the other hand if one rejects the unit root then it is of interest to test if the linear trend is present in the stationary model. Clearly the trend plays a significant role since the $t-$value is 2.7.

III.5.3 Stock market index

Consider the log of Italian stock market index, x_t in figure below and compare with the simulated realization of the random walk. It appears that there is a random walk part in the series.



Indeed an overly simplified implication of the efficient market hypothesis is that the log of stock prices follow a random walk with a drift. In particular, in the basic Black-Scholes set-up,

$$dx_u = \mu du + d\mathcal{B}_u \tag{III.72}$$

with $\mathcal{B}_u$ a Brownian motion with variance σ^2, which in discrete time can be represented as

$$\Delta x_t = \mu + \varepsilon_t \tag{III.73}$$

To see if this describes the variation in the data an AR(4) model was fitted ,

$$\Delta x_t = (\pi_1, \pi_2)(x_{t-1}, t)' + \gamma_1 \Delta x_{t-1} + \ldots + \gamma_3 \Delta x_{t-3} + \mu + \varepsilon_t \qquad \text{(III.74)}$$

to the log of the Italian MIB30 index.

One finds that indeed there is a root in the characteristic polynomial of 0.98 (and the remaining roots outside the unit circle) and that

$$\mathrm{LR}(\pi_1 = \pi_2 = 0) = 5.6 \qquad \text{(III.75)}$$

and hence based on the 95 % quantile of the Dickey-Fuller type distribution this is accepted. Hence, we maintain the hypothesis that x_t has a drift and a random walk part.

Note in particular that the misspecification test for ARCH is significant which will be explored further when discussing ARCH models. Thus the AR model does not describe fully the variation in the data and a different model is needed, see later. Also the above analysis is based on the assumption of no ARCH and hence is by no means a "valid" analysis.

References

[1] Chan & Wei, 1988, ”Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes”, Annals of Math. Statistics

[2] Billingsley, 1968, ” Convergence of Probability Measures”, Wiley.

[3] Brown, B.M. (1971), Martingale Central Limit Theorems, *The Annals of Mathematical Statistics* 42:59-66.

[4] Dickey and Fuller, 1979, ”Distributions of the Estimators for Autoregressive Time Series with a Unit Root”, JASA.

[5] Hansen, B., 1992, ”Convergence to Stochastic Integrals for Dependent Heterogenous Processes”, Econometric Theory, 8:489-500

[6] Johansen, S., 1996, "Likelihood-Based Inference in Cointegrated Vector Autoregressive Models", Oxford University Press