# Pset2 - Econometrics II

Estevão Cardoso

March 2025

## 1 Question 1 (Dynamic Multipliers of Nonstationary Processes - 40 points)

For any $s \in \{0\} \cup \mathbb{N}$, the dynamic multiplier $s$-periods ahead for a stochastic process $\{Y_t\}$ is given by $\frac{\partial Y_{t+s}}{\partial \varepsilon_t}$ and it captures the consequences for $Y_{t+s}$ if $\varepsilon_t$ were to increase by one unit with $\varepsilon$'s for all other dates unaffected.

**a)** Let $\{Y_t\}$ be a MA(1) process with a deterministic time trend, i.e., $Y_t = \alpha + \delta \cdot t + \varepsilon_t + \theta \cdot \varepsilon_{t-1}$, where $\{\varepsilon_t\}$ is a white noise process.
What is the dynamic multiplier $s$-periods ahead for this stochastic process and any $s \in \{0\} \cup \mathbb{N}$? (5 points for $s = 0$; 5 points for $s = 1$; and 5 points for $s > 1$ — 15 points in total).

As stated in Hamilton (2020), if one's want to look to what happens to $Y_{i+s}$ if $\varepsilon_t$ were to increase by one unit with $\varepsilon_t$'s for all other dates unaffected, then the more effective way to do that is to use the multiplier $s$-periods ahead holding all other $\varepsilon$'s fixed.

Now consider the model,

$$Y_t = \alpha + \delta t + \varepsilon_t + \theta \varepsilon_{t-1}.$$

We need to check what happens when we change the horizont for each potential value of $s$. First consider the case for $s = 0$. We want $\frac{\partial Y_t}{\partial \varepsilon_t}$.

From the equation for $Y_t$,

$$Y_t = \alpha + \delta t + \varepsilon_t + \theta \varepsilon_{t-1},$$

given the term $\varepsilon_t$ depends on $\varepsilon_t$ directly, so

$$\frac{\partial Y_t}{\partial \varepsilon_t} = 1.$$

Now the case when $s = 1$. We want $\frac{\partial Y_{t+1}}{\partial \varepsilon_t}$.

From the equation for $Y_t$,

$$Y_{t+1} = \alpha + \delta(t+1) + \varepsilon_{t+1} + \theta\varepsilon_t.$$

Here, $\varepsilon_t$ affects $Y_{t+1}$ through the MA(1) term $\theta\varepsilon_t$, thus

$$\frac{\partial Y_{t+1}}{\partial \varepsilon_t} = \theta.$$

Last, for the case when $s > 1$, we want $\frac{\partial Y_{t+s}}{\partial \varepsilon_t}$ for $s > 1$. But note that for $s > 1$,

$$Y_{t+s} = \alpha + \delta(t+s) + \varepsilon_{t+s} + \theta\varepsilon_{t+s-1}.$$

Since $\varepsilon_t$ does not appear in $Y_{t+s}$ for $s > 1$ (because the MA(1) only depends on current and one lagged $\varepsilon$), we have

$$\frac{\partial Y_{t+s}}{\partial \varepsilon_t} = 0.$$

Therefore we can observe that the result for $s = 2$ also holds for all $s > 2$.

**b)** Let $\{Y_t\}$ be an I(1) process with a drift, i.e.,

$$Y_t = \delta + Y_{t-1} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a white noise process. What is the dynamic multiplier $s$-periods ahead for this stochastic process and any $s \in \{0\} \cup \mathbb{N}$? (15 points for $s \in \{0\} \cup \mathbb{N}$)

Recall that,

$$\begin{aligned}
Y_{t+s} &= \delta + Y_[(t+s) - 1] + \varepsilon_{t+s} \\
&= \delta + \big(\delta + Y_{[(t+s)-2]}\big) + \varepsilon_{t+s} \\
&= \cdots = (s+1)\cdot\delta + Y_{t-1} + \sum_{i}^{s} \varepsilon_{[(t+s)-i]}
\end{aligned}$$

Therefore, we can see that $\forall s \in \{0\} \cup \mathbb{N}$.

$$\frac{\partial Y_{t+s}}{\partial \varepsilon_t} = \frac{\partial}{\partial \varepsilon_t}\left[(s+1)\cdot\delta + Y_{t-1} + \sum_{i=1}^{s} \varepsilon_{[(t+s)-i]}\right] = 1$$

**c)** Given the results above, explain why unit root processes are considered to have infinite memory. (10 points)

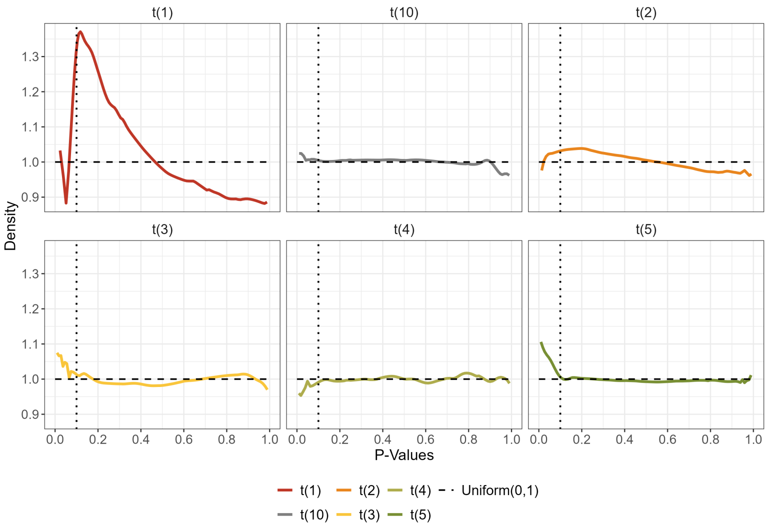
For a unit root process the inovation given by $\varepsilon_t$ have a permanent effect on $Y_t$ that is captured by

$$\lim_{s \to \infty} \frac{\partial y_{t+s}}{\varepsilon_t} = 1 + \psi_1 + \psi_2 + \cdots + = \psi(1)$$

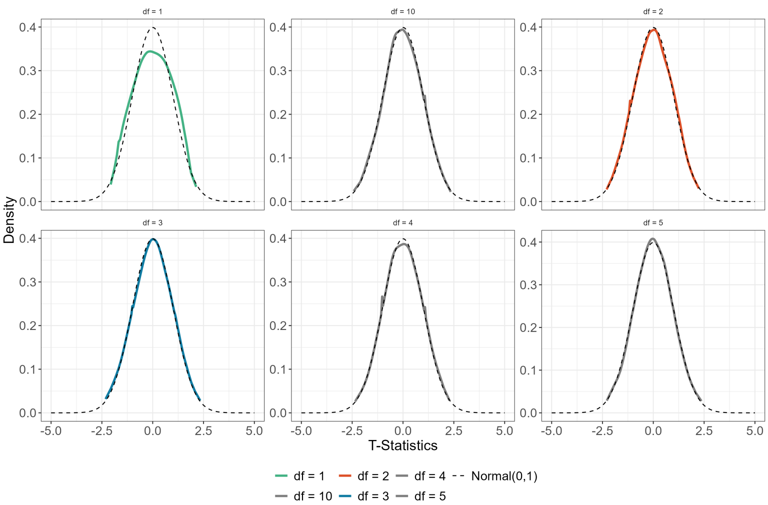
Hence, the inovation don't "forget" the past choks because their multiplier is always equal to one, and the autocorrelations of the process decay very slowly (if at all), the time series does not revert to a fixed mean. Therefore, in an I(1) process such as $Y_t = \delta + Y_{t-1} + \varepsilon_t$, each shock $\varepsilon_t$ is carried forward indefinitely. As a result, the current value $Y_t$ depends not only on recent values and shocks, but on the entire history of past shocks.

## 2 Deterministic Time Trends

**a) and b)** Impose that $\varepsilon_t$ follows a t-distribution with 5 degrees of freedom. What is the test's rejection rate? (15 points). b) Impose that $\varepsilon_t$ follows a t-distribution with 1 degree of freedom. What is the test's rejection rate? (15 points)?




**(a)** P-valued Estimated for $\varepsilon_t \sim t_{1,2,3,4,5,10}$




**(b)** T-statistic estim. density for for $\varepsilon_t \sim t_{1,2,3,4,5,10}$

**Figure 1:** t-distributions and all the rejection rates

In figure 1a, it becomes evident that as we increase the degrees of freedom (DF), the t-distribution in our case more closely resembles a normal distribution. The dotted line indicates the actual p-value we are interested in, and the dashed line indicates the density of an uniform distribution which we use to compare our rejection rate. In figure 1a, for 1 and 2 DF, the rejection rate is very noisy, and we cannot confidently say whether we are correctly rejecting the null hypothesis when we should, or if the test has enough power to accept it when appropriate. We can observe in table 1 that the rejection rate for 1 DF is below the 10% level we set. As we increase the DF, the test size behaves more accurately, reaching a rejection rate of 10.54% for 5 DF. When we reach 10 DF, it is almost certain that we are making nearly all correct decisions regarding rejection or acceptance of the null hypothesis. Overall with 1 DF we expect to unde reject the null due to its lower rejection rate. In figure 1b, the first thing we notice is that the distribution of the $t$-statistic with 1 DF is wider and seems to deviate somewhat from a normal distribution havin a higher variance. In contrast, when we set 5 DF and 10 DF, it more closely resembles a normal distribution.

| DF | $t$-stat Decision Rate (%) |
|---|---|
| 1 | 8.54 |
| 2 | 9.79 |
| 3 | 10.56 |
| 4 | 9.83 |
| 5 | 10.52 |
| 10 | 10.20 |

**Table 1:** Rejection rates by degrees of freedom

**c)** Are the rejection rates in items 1 and 2 very different from each other? If so, what features of our model can explain this difference? If not, why are they similar? (10 points)

The two rejection rates are a bit different. When we compare the rejection rates for 1 degree of freedom (DF) and 5 DF, we observe that the null hypothesis is under-rejected more frequently when using 1 DF. This is mainly due to violations of the assumptions underlying the OLS model, particularly the distribution of the error terms. Specifically, the OLS estimator lacks a finite fourth moment when errors follow a heavy-tailed distribution such as the $t$-distribution with low DF. This issue is also related to the mathematical properties of the $t$-distribution: a $t$-distribution with $n$ degrees of freedom does not have well-defined moments of order $n$ or higher. In our case, the $t_1$ distribution does not even have a defined mean. When the error terms follow a $t$-distribution with low DF, such as 1, the heavier tails increase the variability of the test statistic, which in turn affects the rejection rates. This often results in more noise and less accurate control of the type I error rate. As we increase the DF, the $t$-distribution approaches the normal distribution, improving the approximation on which the test relies. Consequently, the test yields rejection rates that are closer to the nominal significance level.

## 3 (Asymptotic Properties of an AR(1) with a unit root

**Item a)** No Constant Term or Time Trend in the Regression, True Process is a Random Walk.

### Item a.1)

Our proof is based on heavily in Hamilton (2020) chapther 17. Now, assume that the true process is $Y_t = \rho Y_{t-1} + \varepsilon_t$, where $\rho = 1$ and $\varepsilon_t$ is a white noise with zero mean and variance $\sigma^2$. We want to show that,

$$T \cdot (\hat{\rho}_T - 1) \xrightarrow{d} \frac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\int_0^1 [W(r)]^2 \, dr}$$

Since our model has no constant, the OLS we estimate for $\hat{\rho_T}$ is,

$$\hat{\rho_T} = \frac{\sum_t X_t Y_t}{\sum_t X^2} = \frac{\sum_t Y_{t-1} Yt}{\sum_t Y_{t-1}^2}$$

We can substitute $Y_t = \rho Y_{t-1} + \varepsilon_t$ into the previous equation to get,

$$\hat{\rho_T} = \rho + \frac{\sum_t Y_{t-1}\varepsilon_t}{\sum_t Y_{t-1}^2} \Rightarrow T(\hat{\rho_T} - \rho) = \frac{T^{-1}\sum_t Y_{t-1}\varepsilon_t}{T^{-2}\sum_t Y_{t-1}^2}$$

Hamilton (2020) mention two useful results we can use, the first is the *useful results 2*,

$$T^{-1}\sum_t Y_{t-1}\varepsilon_t \xrightarrow{d} \frac{1}{2}\sigma^2 \cdot \{[W(1)]^2 - 1\}$$

and the second is the *useful results 5*

$$T^{-2}\sum_t Y_{t-1}^2 \xrightarrow{d} \sigma^2 \cdot \int_0^1 [W(r)]^2 \, dr$$

By Slutsky's theorem, the assymptotic distribution of the OLS for the estimate of $\hat{\rho_T}$ is,

$$T(\hat{\rho} - \rho) = T(\hat{\rho_T} - \rho) = \frac{T^{-1}\sum_t Y_{t-1}\varepsilon_t}{T^{-2}\sum_t Y_{t-1}^2} \xrightarrow{d} \frac{\frac{1}{2} \cdot \{[W(1)]^2 - 1\}}{\cdot \int_0^1 [W(r)]^2 \, dr}$$

### Item a.2)

Our goal is to prove that,

$$t_T = \frac{\hat{\rho} - 1}{\hat{\sigma}_{\hat{\rho}T}} \xrightarrow{d} \frac{\frac{1}{2}\{[W(1) - 1]\}}{\{\int_0^1 [W(r)]^2 \, dr\}^{\frac{1}{2}}}$$

We recognize that the above equation is the OLS $t$-statistic. Hence, the standard error of $\hat{\rho}$ is

$$\hat{\sigma}^2_{\hat{\rho}T} = \left( \frac{S_T^2}{\sum_t Y_{t-1}^2} \right) \quad \text{where } S_T^2 = \frac{\sum_t \hat{\varepsilon}_t^2}{T - 1}$$

From stabilished results we have seen that the estimator for the error variance $S_T^2$ is consistent for the variance error $\hat{\sigma}^2_{\hat{\rho}T}$. Now note that wince we are under the white noise assumption, we can rely on the homokedasticity assumption. Then replacing $\hat{\sigma}^2_{\hat{\rho}T}$ we can define $t_T$ as,

$$\frac{\hat{\rho_T} - 1}{\hat{\sigma}_{\hat{\rho}T}} = \frac{\hat{\rho_T} - 1}{\left( \frac{S_T^2}{\sum_t Y_{t-1}^2} \right)^{\frac{1}{2}}} = \frac{T(\hat{\rho_T - 1})(T^{-2}\sum_t Y_{t-1}^2)^{\frac{1}{2}}}{(s_T^2)^{\frac{1}{2}}}$$

From item 1.a we know that $T(\hat{\rho_T} - 1) = \frac{T^{-1}\sum_t Y_{t-1}\varepsilon_t}{T^{-2}\sum_t Y_{t-1}^2}$. Hence, we can substitute it to get,

$$\frac{\hat{\rho_T} - 1}{\hat{\sigma}_{\hat{\rho}T}} = \frac{T^{-1}\sum_t Y_{t-1}\varepsilon_t}{(T^{-2}\sum_t Y_{t-1}^2)^{\frac{1}{2}}(s_T^2)^{\frac{1}{2}}} = \frac{T^{-1}\sum_t Y_{t-1}\varepsilon_t}{((s_T^2)T^{-2}\sum_t Y_{t-1}^2)^{\frac{1}{2}}}$$

By the same results we incorporated in item 1.a we again make use of them to show that,

$$\frac{\hat{\rho_T} - 1}{\hat{\sigma}_{\hat{\rho}T}} == \frac{T^{-1}\sum_t Y_{t-1}\varepsilon_t}{((s_T^2)T^{-2}\sum_t Y_{t-1}^2)^{\frac{1}{2}}} \xrightarrow{d} \frac{\frac{1}{2}\sigma^2\{[W(1)^2 - 1]\}}{\{\sigma^2\sigma^2\int_0^1 [W(r)]^2 \, dr\}^{\frac{1}{2}}} = \frac{\frac{1}{2}\{[W(1)^2 - 1]\}}{\{\int_0^1 [W(r)]^2 \, dr\}^{\frac{1}{2}}}$$

As we have proof, the $t$-statistic is calculated as we used to do, but it is assymptotic distribution is even close to a normal when the true process we assume is a random walk because the sums of $Y_{T-1}$ grow faster with order $T^2$.

Item **2** Constant Term but No Time Trend in the Regression, True Process is a Random Walk.

Now the true process is $Y_t = \alpha + \rho \cdot Y_{t-1} + \varepsilon_t$ , where $\alpha = 0,\ \rho = 1$ and $\{\varepsilon_t\}$ is a white noise with zero mean and variance $\sigma^2$.

Our goal is to prove that

$$T \cdot (\hat{\rho_T} - 1) \xrightarrow{d} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int_0^1 W(r)\ dr}{\int_0^1 [W(r)]^2 dr - [\int_0^1 W(r)\ dr]^2}$$

The OLS estimates $\hat{\alpha}_T$ and $\hat{\rho}_T$ are calculated in the usual way:

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T \end{bmatrix} = \left( \sum_t X_t X_t' \right)^{-1} \left( \sum_t X_t Y_t \right), \quad \text{where } X_t = \begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix}$$

To make things easier we can define,

$$b_T = \begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T \end{bmatrix}, \quad \beta = \begin{bmatrix} \alpha \\ \rho \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

We now substitute the true DGP: $Y_t = \rho Y_{t-1} + \varepsilon_t = X_t'\beta + \varepsilon_t$, we get:

$$b_T = \left( \sum_t X_t X_t' \right)^{-1} \left( \sum_t X_t (X_t'\beta + \varepsilon_t) \right)$$

$$\Rightarrow b_T - \beta = \left( \sum_t X_t X_t' \right)^{-1} \left( \sum_t X_t \varepsilon_t \right)$$

Now, expand the terms using the form of $X_t = \begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix}$:

$$X_t X_t' = \begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix} \begin{bmatrix} 1 & y_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & y_{t-1} \\ y_{t-1} & y_{t-1}^2 \end{bmatrix}$$

Summing over $t$, we get the design matrix:

$$\sum_t X_t X_t' = \begin{bmatrix} \sum_t 1 & \sum_t y_{t-1} \\ \sum_t y_{t-1} & \sum_t y_{t-1}^2 \end{bmatrix}$$

and similarly:

$$\sum_t X_t \varepsilon_t = \sum_t \begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix} \varepsilon_t = \begin{bmatrix} \sum_t \varepsilon_t \\ \sum_t y_{t-1}\varepsilon_t \end{bmatrix}$$

Thus:

$$b_T - \beta = \left( \begin{bmatrix} \sum_t 1 & \sum_t y_{t-1} \\ \sum_t y_{t-1} & \sum_t y_{t-1}^2 \end{bmatrix} \right)^{-1} \begin{bmatrix} \sum_t \varepsilon_t \\ \sum_t y_{t-1}\varepsilon_t \end{bmatrix}$$

Next, define the scaling matrix $\Upsilon_T$ as we saw in class:

$$\Upsilon_T = \begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix} \quad \Rightarrow \quad \Upsilon_T^{-1} = \begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{-1} \end{bmatrix}$$

Multiplying both sides of the expression for $b_T - \beta$ by $\Upsilon_T$, we obtain the scaled OLS deviation:

$$\Upsilon_T(b_T - \beta) = \begin{bmatrix} T^{1/2}(\hat{\alpha}_T - 0) \\ T(\hat{\rho}_T - 1) \end{bmatrix} = \Upsilon_T \left( \sum_t X_t X_t' \right)^{-1} \left( \sum_t X_t \varepsilon_t \right)$$

Now we expand step by step and scale the design matrix:

$$\Upsilon_T \left( \sum_t X_t X_t' \right)^{-1} = \Upsilon_T^{-1} \begin{bmatrix} 1 & T^{-1} \sum_t y_{t-1} \\ T^{-1} \sum_t y_{t-1} & T^{-1} \sum_t y_{t-1}^2 \end{bmatrix}^{-1}$$

And we Combine with the scaled residual vector to get

$$\begin{aligned} \Upsilon_T(b_T - \beta) &= \left[ \begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{-1} \end{bmatrix} \begin{bmatrix} \sum_t 1 & \sum_t y_{t-1} \\ \sum_t y_{t-1} & \sum_t y_{t-1}^2 \end{bmatrix} \right]^{-1} \begin{bmatrix} T^{-1/2} \sum_t \varepsilon_t \\ T^{-1} \sum_t y_{t-1} \varepsilon_t \end{bmatrix} \\ &= \begin{bmatrix} T^{-3/2} & -T^{-3/2} \sum_t y_{t-1} \\ -T^{-3/2} \sum_t y_{t-1} & T^{-2} \sum_t y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2} \sum_t \varepsilon_t \\ T^{-1} \sum_t y_{t-1} \varepsilon_t \end{bmatrix} \end{aligned}$$

We highlight the selected term in purple because it will be useful later in part *(b)*. By *Useful Results 1, 2, 4, and 5*, we have the following convergence results:

$$\begin{aligned} T^{-\frac{1}{2}} \sum_t \varepsilon_t &\xrightarrow{d} \sigma \cdot W(1) \\ T^{-1} \sum_t Y_{t-1} \varepsilon_t &\xrightarrow{d} \frac{1}{2} \sigma^2 \cdot \left( [W(1)]^2 - 1 \right) \\ T^{-\frac{3}{2}} \sum_t Y_{t-1} &\xrightarrow{d} \sigma \cdot \int_0^1 W(r)\, dr \\ T^{-2} \sum_t Y_{t-1}^2 &\xrightarrow{d} \sigma^2 \cdot \int_0^1 [W(r)]^2\, dr \end{aligned}$$

Then, applying asymptotic theorems and doing some algebra, we obtain:

$$\begin{aligned}
\Upsilon_T(b_T-\beta) &= \begin{bmatrix} 1 & T^{-\frac{3}{2}}\cdot\sum_t Y_{t-1} \\ T^{-\frac{3}{2}}\cdot\sum_t Y_{t-1} & T^{-2}\cdot\sum_t Y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}}\sum_t \varepsilon_t \\ T^{-1}\sum_t Y_{t-1}\varepsilon_t \end{bmatrix} \\
&\xrightarrow{d} \begin{bmatrix} 1 & \sigma\int_0^1 W(r)\,dr \\ \sigma\int_0^1 W(r)\,dr & \sigma^2\int_0^1 [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} \sigma\cdot W(1) \\ \frac{1}{2}\sigma^2\cdot\{[W(1)]^2-1\} \end{bmatrix} \\
&= \sigma\cdot\left\{ \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} 1 & \int_0^1 W(r)\,dr \\ \int_0^1 W(r)\,dr & \int_0^1 [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix} \right\} \\
&\quad\cdot\sigma \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2-1\} \end{bmatrix} \\
&= \sigma\cdot \begin{bmatrix} 1 & 0 \\ 0 & \sigma^{-1} \end{bmatrix} \begin{bmatrix} 1 & \int_0^1 W(r)\,dr \\ \int_0^1 W(r)\,dr & \int_0^1 [W(r)]^2\,dr \end{bmatrix}^{-1} \\
&\quad\cdot \begin{bmatrix} 1 & 0 \\ 0 & \sigma^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2-1\} \end{bmatrix} \\
&= \begin{bmatrix} \sigma & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \int_0^1 [W(r)]^2\,dr & \int_0^1 W(r)\,dr \\ \int_0^1 W(r)\,dr & 1 \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2-1\} \end{bmatrix}
\end{aligned}$$

To invert the middle matrix:

$$\begin{bmatrix} 1 & \int_0^1 W(r)\,dr \\ \int_0^1 W(r)\,dr & \int_0^1 [W(r)]^2\,dr \end{bmatrix}^{-1} = \frac{1}{\Delta}\begin{bmatrix} 1 & -\int_0^1 W(r)\,dr \\ -\int_0^1 W(r)\,dr & \int_0^1 [W(r)]^2\,dr \end{bmatrix}$$

$$\text{where } \Delta = \int_0^1 [W(r)]^2\,dr - \left(\int_0^1 W(r)\,dr\right)^2$$

Substituting the inverse in $\Upsilon_T(b_T-\beta)$:

$$\begin{aligned}
\Upsilon_T(b_T-\beta) &\xrightarrow{d} \begin{bmatrix} \sigma & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{\Delta} \begin{bmatrix} \int_0^1 [W(r)]^2\,dr & \int_0^1 W(r)\,dr \\ \int_0^1 W(r)\,dr & 1 \end{bmatrix} \cdot \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)^2-1]\} \end{bmatrix} \\
&= \frac{1}{\Delta} \begin{bmatrix} \sigma\int_0^1 [W(r)]^2\,dr & -\sigma\int_0^1 W(r)\,dr \\ -\int_0^1 W(r)\,dr & 1 \end{bmatrix} \cdot \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)^2-1]\} \end{bmatrix}
\end{aligned}$$

As we are interested solely in the asymptotic distribution of $T(\hat{\rho}_T-1)$, we focus on the element in the **second row and first column** of the matrix product derived above. Thus, the asymptotic distribution of $T(\hat{\rho}_T-1)$ is **not normal** when the process contains a unit root. Moreover, it differs from the asymptotic distribution obtained in part *(a)*, where no constant was included in the estimated model. Then, the desired result is:

$$T(\hat{\rho}_T - 1) \xrightarrow{d} \frac{1}{\Delta} \cdot \frac{1}{2} \cdot \left\{[W(1)]^2 - 1\right\} - \left(W(1) \int_0^1 W(r)\, dr\right)$$
$$= \frac{1}{2} \cdot \frac{\{[W(1)]^2 - 1\} - \left(W(1) \int_0^1 W(r)\, dr\right)}{\int_0^1 [W(r)]^2\, dr - \left(\int_0^1 W(r)\, dr\right)^2}$$

**Item 2.b**

**(b)** We aim to show that:

$$t_T = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} \xrightarrow{d} \frac{1}{\left\{\int_0^1 [W(r)]^2 dr - \left(\int_0^1 W(r) dr\right)^2\right\}^{1/2}} \cdot \frac{1}{2} \left\{[W(1)]^2 - 1\right\} - W(1) \int_0^1 W(r) dr$$

where $\hat{\sigma}_{\hat{\rho}_T}$ is the usual OLS standard error, defined as:

$$s_T^2 = \frac{\hat{\varepsilon}_t^2}{T-2}, \qquad \hat{\sigma}_{\hat{\rho}_T} = \frac{s_T^2}{\sum_t Y_{t-1}^2}$$

We simplify a bit this step to define $\hat{\sigma}_{\hat{\rho}_T}$ as:

$$\hat{\sigma}_{\hat{\rho}_T} = \frac{s_T^2}{\sum_t Y_{t-1}^2} = s_T^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \left[\sum_t \begin{bmatrix} 1 \\ Y_{t-1} \end{bmatrix} \begin{bmatrix} 1 & Y_{t-1} \end{bmatrix}\right]^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Define the scaling matrix $\Upsilon_T$ as in part (a). If we multiply $\hat{\sigma}_{\hat{\rho}_T}$ by $T^2$, we get:

$$T^2 \hat{\sigma}_{\hat{\rho}_T} = s_T^2 \begin{bmatrix} 0 & 1 \end{bmatrix} T \left[\sum_t \begin{bmatrix} 1 \\ Y_{t-1} \end{bmatrix} \begin{bmatrix} 1 & Y_{t-1} \end{bmatrix}\right]^{-1} T \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$= s_T^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \Upsilon_T \left[\Upsilon_T^{-1} \sum_t \begin{bmatrix} 1 \\ Y_{t-1} \end{bmatrix} \begin{bmatrix} 1 & Y_{t-1} \end{bmatrix} \Upsilon_T^{-1}\right]^{-1} \Upsilon_T \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

From part (a), we have already shown that:

$$\left\{\Upsilon_T^{-1} \left[\sum_t \begin{bmatrix} 1 \\ Y_{t-1} \end{bmatrix} \begin{bmatrix} 1 & Y_{t-1} \end{bmatrix}\right] \Upsilon_T^{-1}\right\}^{-1} \xrightarrow{d} \begin{bmatrix} 1 & 0 \\ 0 & \sigma^{-1} \end{bmatrix} \begin{bmatrix} 1 & \int_0^1 W(r) dr \\ \int_0^1 W(r) dr & \int_0^1 [W(r)]^2 dr \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & \sigma^{-1} \end{bmatrix}$$

We know from previous econometrics courses that the estimator of the error variance, $s_T^2$, is consistent for the true variance of the error, $\sigma^2$. This allows us to invoke the Continuous Mapping Theorem on the expression involving $T^2 \hat{\sigma}_{\hat{\rho}_T}$.

$$\xrightarrow{d} \sigma^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \left\{\begin{bmatrix} 1 & 0 \\ 0 & \sigma^{-1} \end{bmatrix} \begin{bmatrix} \int_0^1 1\, dr & \int_0^1 W(r)\, dr \\ \int_0^1 W(r)\, dr & \int_0^1 W(r)^2\, dr \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sigma^{-1} \end{bmatrix}\right\}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} \int_0^1 1\,dr & \int_0^1 W(r)\,dr \\ \int_0^1 W(r)\,dr & \int_0^1 W(r)^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Recall that, from *(a)*,

$$\begin{bmatrix} \int_0^1 1\,dr & \int_0^1 W(r)\,dr \\ \int_0^1 W(r)\,dr & \int_0^1 W(r)^2\,dr \end{bmatrix}^{-1} = \frac{1}{\Delta} \begin{bmatrix} \int_0^1 W(r)^2\,dr & -\int_0^1 W(r)\,dr \\ -\int_0^1 W(r)\,dr & 1 \end{bmatrix},$$

where

$$\Delta = \left( \int_0^1 W(r)^2\,dr - \left[ \int_0^1 W(r)\,dr \right]^2 \right),$$

which corresponds to the variance of the Brownian motion over $[0,1]$.

Substituting back into the earlier expression we can get:

$$T^2 \hat{\sigma}_{\hat{\rho}_T} = [0 \quad 1] \frac{1}{\Delta} \begin{bmatrix} \int_0^1 W(r)^2\,dr & -\int_0^1 W(r)\,dr \\ -\int_0^1 W(r)\,dr & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \frac{1}{\int_0^1 W(r)^2\,dr - \left( \int_0^1 W(r)\,dr \right)^2}$$

Now we are interested in the asymptotic distribution of the $t$-statistic associated with the unit root test:

$$\frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} = \frac{T(\hat{\rho}_T - 1)}{(T^2 \hat{\sigma}^2_{\hat{\rho}_T})^{1/2}}$$

From result $(a)$ and the Continuous Mapping Theorem we end up with:

$$\xrightarrow{d} \frac{\frac{1}{2} \{ [W(1)]^2 - 1 \} - W(1) \int_0^1 W(r)\,dr}{\left\{ \int_0^1 W(r)^2\,dr - \left( \int_0^1 W(r)\,dr \right)^2 \right\}^{1/2}}$$

$$\xrightarrow{d} \frac{1}{2} \cdot \frac{\{ [W(1)]^2 - 1 \} - W(1) \int_0^1 W(r)\,dr}{\left( \int_0^1 W(r)^2\,dr - \left[ \int_0^1 W(r)\,dr \right]^2 \right)} \cdot \left( \int_0^1 W(r)^2\,dr - \left[ \int_0^1 W(r)\,dr \right]^2 \right)^{1/2}$$

$$= \frac{1}{2} \cdot \frac{\{ [W(1)]^2 - 1 \} - W(1) \int_0^1 W(r)\,dr}{\left( \int_0^1 W(r)^2\,dr - \left[ \int_0^1 W(r)\,dr \right]^2 \right)^{1/2}}$$

**Item 3** Constant Term but No Time Trend in the Regression, True Process is a Random Walk with Drift.

## Model Specification

Assume the true data generating process is:

$$Y_t = \alpha + \rho \cdot Y_{t-1} + \varepsilon_t$$

where $\alpha \neq 0,\ \rho = 1,\ \{\varepsilon_t\}$ is white noise with zero mean and $\text{Var}(\varepsilon_t) = \sigma^2$. We are going to estimate $Y_t = \alpha + \rho \cdot Y_{t-1} + \varepsilon_t$.
We want to prove that:

$$\begin{bmatrix} T^{\frac{1}{2}}(\hat{\alpha}_T - \alpha) \\ T^{\frac{3}{2}}(\hat{\rho}_T - 1) \end{bmatrix} \xrightarrow{d} N\left(0, \sigma^2 \cdot \begin{bmatrix} 1 & \frac{\alpha}{2} \\ \frac{\alpha}{2} & \frac{\alpha^2}{3} \end{bmatrix}^{-1}\right)$$

The OLS estimates $\hat{\alpha}_T$ and $\hat{\rho}_T$ are calculated in the standard way:

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T \end{bmatrix} = \left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} \left(\sum_{t=1}^{T} X_t Y_t\right), \quad \text{where } X_t = \begin{bmatrix} 1 & Y_{t-1} \end{bmatrix}'$$

For notational convenience, define:

$$\boldsymbol{b}_T = \begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T \end{bmatrix},$$
$$\boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \rho \end{bmatrix} = \begin{bmatrix} \alpha \\ 1 \end{bmatrix}$$

We can plug in the true process $Y_t = X_t'\boldsymbol{\beta} + \varepsilon_t$ into the estimator:

$$\boldsymbol{b}_T = \left(\sum_t X_t X_t'\right)^{-1} \left[\sum_t X_t(X_t'\boldsymbol{\beta} + \varepsilon_t)\right]$$
$$\Rightarrow \boldsymbol{b}_T - \boldsymbol{\beta} = \left(\sum_t X_t X_t'\right)^{-1} \left(\sum_t X_t \varepsilon_t\right)$$

If we expand the matrix terms carefully we obtain:

$$\boldsymbol{b}_T - \boldsymbol{\beta} = \begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\rho}_T - 1 \end{bmatrix} = \begin{bmatrix} T & \sum_t Y_{t-1} \\ \sum_t Y_{t-1} & \sum_t Y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_t \varepsilon_t \\ \sum_t Y_{t-1}\varepsilon_t \end{bmatrix}$$

And under the null ($\rho = 1$), we can express $Y_t$ as:

$$\begin{aligned} Y_t &= \alpha + Y_{t-1} + \varepsilon_t \\ &= \alpha + (\alpha + Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \cdots \\ &= Y_0 + \alpha \cdot t + \sum_{s=1}^{t} \varepsilon_s \end{aligned}$$

Without loss of generality for asymptotic analysis (as $Y_0$ is a constant), assume $Y_0 = 0$. The term $\sum_{s=1}^{t} \varepsilon_s$ represents a random walk without drift:

$$\begin{aligned} \xi_t &= \xi_{t-1} + \varepsilon_t \\ &= (\xi_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \cdots \\ &= \xi_0 + \sum_{s=1}^{t} \varepsilon_s \end{aligned}$$

Assuming $\xi_0 = 0$, we examine the behavior of the sum $\sum_{t=1}^{T} Y_{t-1}$ that appears in the OLS estimator matrices:

$$\begin{aligned} \sum_{t=1}^{T} Y_{t-1} &= \sum_{t=1}^{T} \alpha \cdot (t-1) + \sum_{t=1}^{T} (\varepsilon_1 + \cdots + \varepsilon_{t-1}) \\ &= \sum_{t=1}^{T} \alpha \cdot (t-1) + \sum_{t=1}^{T} \xi_{t-1} \end{aligned}$$

Making use of Hamilton (2020) *Useful Results 4 and 8* we can show that,

$$\begin{gathered} T^{-\frac{3}{2}} \sum_{t=1}^{T} \xi_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r)dr \\ T^{-2} \sum_{t=1}^{T} \alpha \cdot (t-1) = \alpha \sum_{t=1}^{T} \frac{(t-1)^1}{T^2} \to \frac{\alpha}{2} \end{gathered}$$

The are some aspects of the magnitude of each term: the deterministic trend component $\sum_{t=1}^{T} \alpha \cdot (t-1)$ is, $O(T^2)$ in exact terms $O_p(T^2)$ in probability (forms a degenerate distribution); the stochastic component $\sum_{t=1}^{T} \xi_{t-1}$ is $O_p(T^{\frac{3}{2}})$ (must be scaled by $T^{-\frac{3}{2}}$ to become $O_p(1)$). This implies that the deterministic trend $\alpha \cdot (t-1)$ asymptotically dominates the stochastic components. Thus, the scaled sum converges as follows:

$$
\begin{aligned}
T^{-2}\sum_{t=1}^{T} Y_{t-1} &= T^{-2}\sum_{t=1}^{T} \alpha\cdot(t-1) + T^{-2}\sum_{t=1}^{T}\xi_{t-1} \\
&= T^{-2}\sum_{t=1}^{T} \alpha\cdot(t-1) + \frac{1}{T}\left(T^{-\frac{3}{2}}\sum_{t=1}^{T}\xi_{t-1}\right) \\
&\xrightarrow{p} \frac{\alpha}{2} + 0
\end{aligned}
$$

where, the first term converges to $\alpha/2$; The second term vanishes because $T^{-\frac{3}{2}}\sum \xi_{t-1}$ is $O_p(1)$ by the additional $1/T$ scaling drives it to zero. Next it is easy to decompose the sum of squared lagged variables as,

$$
\begin{aligned}
\sum_{t=1}^{T} Y_{t-1}^2 &= \sum_{t=1}^{T}\left[\alpha\cdot(t-1)+\xi_{t-1}\right]^2 \\
&= \sum_{t=1}^{T}\alpha^2\cdot(t-1)^2 + 2\sum_{t=1}^{T}\alpha\cdot(t-1)\cdot\xi_{t-1} + \sum_{t=1}^{T}\xi_{t-1}^2 \\
&= O_p(T^3) + O_p(T^2\cdot T^{\frac{3}{2}}) + O_p(T^2)
\end{aligned}
$$

the first terms will converge to,

$$
\begin{aligned}
\sum_{t=1}^{T}\alpha^2\cdot(t-1)^2 &= O_p(T^3) \\
T^{-3}\sum_{t=1}^{T}\alpha^2\cdot(t-1)^2 &\to \frac{\alpha^2}{3}
\end{aligned}
$$

the second term in its turn will be,

$$
\begin{aligned}
2\sum_{t=1}^{T}\alpha\cdot(t-1)\cdot\xi_{t-1} &= O_p(T^{\frac{7}{2}}) \\
T^{-3}\left(2\sum_{t=1}^{T}\alpha\cdot(t-1)\cdot\xi_{t-1}\right) &= O_p(T^{\frac{1}{2}}) \to 0
\end{aligned}
$$

and using *Useful Result 5* from Hamilton (2020) it yields,

$$
\begin{aligned}
\sum_{t=1}^{T}\xi_{t-1}^2 &= O_p(T^2) \\
T^{-3}\sum_{t=1}^{T}\xi_{t-1}^2 &\to 0
\end{aligned}
$$

And the total sum converges to,

$$
\begin{aligned}
T^{-3}\sum_{t=1}^{T} Y_{t-1}^2 = T^{-3}\sum_{t=1}^{T}\alpha^2\cdot(t-1)^2 \\
&+T^{-3}\left(2\sum_{t=1}^{T}\alpha\cdot(t-1)\cdot\xi_{t-1}\right) \\
&+T^{-3}\sum_{t=1}^{T}\xi_{t-1}^2 \\
&=\alpha^2\sum_{t=1}^{T}\frac{(t-1)^2}{T^3}+O_p(T^{-\frac{1}{2}})+O_p(T^{-1}) \\
&\xrightarrow{p}\frac{\alpha^2}{3}
\end{aligned}
$$

Additionaly, by the Central Limit Theorem, $\sum_{t=1}^{T}\varepsilon_t = O_p(\sqrt{T})$, and we can show that,

$$
\begin{aligned}
\sum_{t=1}^{T} Y_{t-1}\varepsilon_t &= \sum_{t=1}^{T}[\alpha\cdot(t-1)+\xi_{t-1}]\varepsilon_t \\
&= \underbrace{\sum_{t=1}^{T}\alpha\cdot(t-1)\cdot\varepsilon_t}_{O_p(T^{3/2})}+\underbrace{\sum_{t=1}^{T}\xi_{t-1}\varepsilon_t}_{O_p(T)} \\
&= O_p(T^{3/2})+O_p(T)
\end{aligned}
$$

where the orders follow from *Useful Results 2 and 3* from Hamilton (2020). Thus, the dominant term is $O_p(T^{3/2})$, and the estimation error can be expressed as:

$$
\begin{bmatrix}\hat{\alpha}_T-\alpha\\ \hat{\rho}_T-1\end{bmatrix}=\begin{bmatrix}T & \sum_t Y_{t-1}\\ \sum_t Y_{t-1} & \sum_t Y_{t-1}^2\end{bmatrix}^{-1}\begin{bmatrix}\sum_t\varepsilon_t\\ \sum_t Y_{t-1}\varepsilon_t\end{bmatrix}=\begin{bmatrix}O_p(T) & O_p(T^2)\\ O_p(T^2) & O_p(T^3)\end{bmatrix}^{-1}\begin{bmatrix}O_p(T^{1/2})\\ O_p(T^{3/2})\end{bmatrix}
$$

Observe that if we use the appropriate scaling matrix, and apply it, we will get

$$
\Upsilon_T=\begin{bmatrix}T^{1/2} & 0\\ 0 & T^{3/2}\end{bmatrix},\quad \Upsilon_T^{-1}=\begin{bmatrix}T^{-1/2} & 0\\ 0 & T^{-3/2}\end{bmatrix}
$$

And,

$$
\begin{aligned}
\Upsilon_T\begin{bmatrix}\hat{\alpha}_T-\alpha\\ \hat{\rho}_T-1\end{bmatrix} &= \left\{\Upsilon_T^{-1}\begin{bmatrix}T & \sum_t Y_{t-1}\\ \sum_t Y_{t-1} & \sum_t Y_{t-1}^2\end{bmatrix}\Upsilon_T^{-1}\right\}^{-1}\Upsilon_T^{-1}\begin{bmatrix}\sum_t\varepsilon_t\\ \sum_t Y_{t-1}\varepsilon_t\end{bmatrix} \\
&= \begin{bmatrix}1 & T^{-2}\sum_t Y_{t-1}\\ T^{-2}\sum_t Y_{t-1} & T^{-3}\sum_t Y_{t-1}^2\end{bmatrix}^{-1}\begin{bmatrix}T^{-1/2}\sum_t\varepsilon_t\\ T^{-3/2}\sum_t Y_{t-1}\varepsilon_t\end{bmatrix}
\end{aligned}
$$

so far, we have established that:

$$\begin{bmatrix} 1 & T^{-2}\sum_t Y_{t-1} \\ T^{-2}\sum_t Y_{t-1} & T^{-3}\sum_t Y_{t-1}^2 \end{bmatrix} \xrightarrow{p} Q := \begin{bmatrix} 1 & \frac{\alpha}{2} \\ \frac{\alpha}{2} & \frac{\alpha^2}{3} \end{bmatrix}$$

and we have shown that the scaled system converges to the same form as a deterministic time trend model. By Theorem 1 from the lecture slides, we obtain joint normality where the asymptotic variance is scaled by $\alpha$ in the off-diagonal elements, so we get,

$$\begin{aligned}
\begin{bmatrix} T^{-\frac{1}{2}}\sum_t \varepsilon_t \\ T^{-\frac{3}{2}}\sum_t Y_{t-1}\varepsilon_t \end{bmatrix} &\xrightarrow{P} \begin{bmatrix} T^{-\frac{1}{2}}\sum_t \varepsilon_t \\ T^{-\frac{3}{2}}\sum_{t=1}^T \alpha\cdot(t-1)\cdot\varepsilon_t \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} T^{-\frac{1}{2}}\sum_t \varepsilon_t \\ T^{-\frac{3}{2}}\sum_{t=1}^T (t-1)\cdot\varepsilon_t \end{bmatrix}
\end{aligned}$$

The transformed vector converges in distribution to:

$$\begin{aligned}
&\xrightarrow{d} \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}\right) \\
&= N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix}\right) \\
&= N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \frac{\alpha}{2} \\ \frac{\alpha}{2} & \frac{\alpha^2}{3} \end{bmatrix}\right) \\
&= N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 Q\right)
\end{aligned}$$

And pplying the Continuous Mapping Theorem to the scaled estimators we obtain,

$$\begin{aligned}
\Upsilon_T \begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\rho}_T - 1 \end{bmatrix} &= \begin{bmatrix} T^{\frac{1}{2}} & 0 \\ 0 & T^{\frac{3}{2}} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\rho}_T - 1 \end{bmatrix} \\
&= \begin{bmatrix} T^{\frac{1}{2}}(\hat{\alpha}_T - \alpha) \\ T^{\frac{3}{2}}(\hat{\rho}_T - 1) \end{bmatrix} \\
&\xrightarrow{d} Q^{-1}\cdot N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 Q\right) \\
&= N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 Q^{-1}\right)
\end{aligned}$$

Finally, the asymptotic distribution is *normal*, in contrast to some other unit root cases, the properties of $\hat{\alpha}_T$ and $\hat{\rho}_T$ match those of a deterministic time trend model (as discussed in Hamilton (2020)). Moreover, the regressor $Y_{t-1}$ is asymptotically dominated by the time trend component arising from the drift ($\alpha \neq 0$). The scaling factors ($T^{1/2}$ for intercept and $T^{3/2}$ for slope) reflect the different convergence rates.

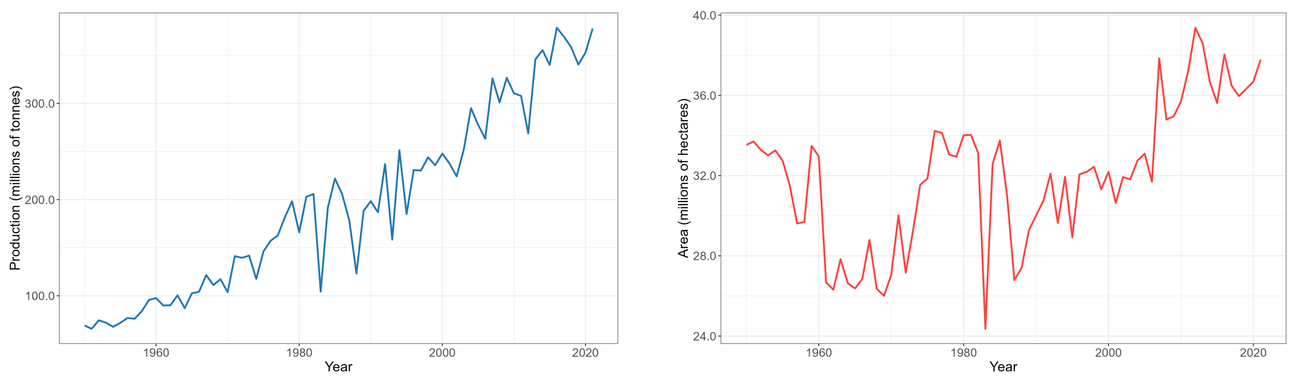
# 4 Augmented Dickey-Fuller Test in Practice - 70 points

In the file *corn_production_land_us.csv*, you can find the time series for corn production (tonnes) in the U.S.. Using the data between 1950 and 2021, test whether this stochastic process has a unit root or not.

We can see from Figure 2 that the production series exhibits a clear positive growth trend, likely due to the incorporation of new technologies over the years. In contrast, the series for land use remains relatively constant over time. Therefore, we assume that the most appropriate model to describe this process is an AR model with both a drift and a deterministic trend for U.S. corn production. The model is specified as follows:

$$\Delta\text{Production}_t = \alpha + \beta t + \gamma \cdot \text{Production}_{t-1} + \sum_{i=1}^{p} \delta_i \Delta\text{Production}_{t-i} + \varepsilon_t$$

where $\alpha$ is the intercept or drift; $\beta$ is the deterministic time trend we are assuming; $\gamma$ is the unit root coefficient we are interested in estimating to test; and $\delta$ is the coefficient for the other lags. We are also testing the null hypothesis that $H_0 : \gamma = 0$ which assumes that the series has a unit root and its not sationary, against $H_1 : \gamma < 0$ so the serie is stationary.




**(a)** Corn production (in millions of tonnes).

**(b)** Corn harvested area (in millions of hectares).

**Figure 2:** Corn production and harvested area in the United States from 1950 to 2021.

The optimal choice of the number of lags was made upon the automatic decision of the R function `ur.df` from the library `urca` is, and it chooses 2 two lags, including the first difference of the outcome variable, so 1 lag. The final form of the estimated model is

$$\begin{aligned}\Delta\text{Production}_t = 2.684 \times 10^7 + 2.982 \times 10^6 \cdot t - 0.6705 \cdot \text{Production}_{t-1} \\ - 0.1808 \cdot \Delta\text{Production}_{t-1} + \varepsilon_t\end{aligned}$$

The results of the test for each coefficient are displayed in table 2. The most important point to note is that we reject the null hypothesis of the presence of a unit root.

| Statistic | Observed Value | 1% | 5% | 10% |
|---|---|---|---|---|
| $\tau_3$ | -4.3468 | -4.04 | -3.45 | -3.15 |
| $\phi_2$ | 7.8636 | 6.50 | 4.88 | 4.16 |
| $\phi_3$ | 9.5984 | 8.73 | 6.49 | 5.47 |

**Table 2:** ADF Test: Test statistics and critical values

## Statement Regarding Artificial Intelligence Usage

This project made use of ChatGPT 4.0 for coding assistance and language correction.

## References

Hamilton, James D (2020). *Time series analysis*. Princeton university press.