

Econometrics PS2 TIME SERIES

Jordi Torres, ETE2

December 2024

Exercise 1.1

1.1.a

Given the observed autocovariance function $\gamma_y(h)$, we propose an **MA(2) model** for the time series (y_t) , as the autocovariance becomes negligible (close to 0) after lag 2. Specifically, the values of $\gamma_y(h)$ decrease from $h = 0$ to $h = 2$ and drop off further at $h = 3$ and beyond.

To confirm this, we compute the asymptotic variance of $\gamma_y(h)$ under the assumption of an MA(q) process and perform hypothesis testing for $h > q$.

The null hypothesis is:

$$H_0 : \gamma_y(q + h) = 0 \quad \text{for } h > q.$$

The test statistic uses the asymptotic variance of $\gamma_y(h)$, given by:

$$\text{Var}(\gamma_y(h)) = \frac{\gamma_y(0) + 2 \sum_{j=1}^q \gamma_y(j)}{T}.$$

For $q = 2$, we calculate:

$$\gamma_y(0) + 2(\gamma_y(1) + \gamma_y(2)) = 2.25 + 2(1.26 + 0.58) = 5.93.$$

The bounds for accepting H_0 are:

$$\pm 1.96 \sqrt{\frac{5.93}{100}} = \pm 0.476.$$

Since $\gamma_y(3) = 0.18$ and $\gamma_y(4) = -0.22$ lie within $[-0.476, 0.476]$, we do not reject H_0 for $h > 2$. This supports the choice of an MA(2) model:

$$y_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \epsilon_t,$$

where ϵ_t is white noise with variance σ^2 .

1.1.b

To estimate the parameters $(\phi_1, \phi_2, \sigma^2)$, we use **Maximum Likelihood Estimation (MLE)**. The likelihood function is:

$$\ell(\mu, \phi_1, \phi_2, \sigma^2) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=3}^T \epsilon_t^2,$$

where:

$$\epsilon_t = y_t - \phi_1 \epsilon_{t-1} - \phi_2 \epsilon_{t-2}.$$

The goal is to maximize the likelihood function to estimate the parameters:

$$(\hat{\mu}, \hat{\phi}_1, \hat{\phi}_2, \hat{\gamma}) = \arg \max_{\mu, \phi_1, \phi_2, \gamma} \ell(\mu, \phi_1, \phi_2, \gamma).$$

This yields the estimates $\hat{\mu}, \hat{\phi}_1, \hat{\phi}_2$, and the variance $\hat{\sigma}^2$.

Exercise 1.2

1.2.a

We know that an AR time series of order p is characterized by its partial autocorrelation function, which is equal to 0 for lags larger or equal to $p + 1$.

We will follow a similar approach as before. Now we want to test that:

$$H_0 : r_y(p + h) = 0 \quad \text{for } h > p.$$

We also know that:

$$\sqrt{T} \begin{pmatrix} r_y(p+1) \\ r_y(p+2) \\ \vdots \\ r_y(p+h) \end{pmatrix} \sim N(0, I_h),$$

where I_h is the $h \times h$ identity matrix.

And that we do not reject the null hypothesis if:

$$\hat{r}_y(p+h) \in \left(-\frac{1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}} \right).$$

Which in our case becomes:

$$\left(-\frac{1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}} \right) = \left(-\frac{1.96}{\sqrt{100}}, \frac{1.96}{\sqrt{100}} \right) = (-0.196, 0.196).$$

Examining the partial autocorrelation function values for $h > 2$:

$$\hat{r}_y(3) = 0.11, \quad \hat{r}_y(4) = 0.13.$$

Both values lie within the interval $(-0.196, 0.196)$. Therefore, we do not reject H_0 for $h > 2$.

This suggests an AR(2) model is appropriate. The model can be written as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t,$$

where ϵ_t is white noise with variance σ^2 .

1.2.b

To estimate the parameters ϕ_1 and ϕ_2 , we can use the Yule-Walker representation. First, however, we need to calculate the values of the autocovariance function from the partial autocorrelation function (PACF).

We use the following relationship:

$$\begin{pmatrix} \gamma_y(0) & \gamma_y(1) \\ \gamma_y(1) & \gamma_y(0) \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} \gamma_y(1) \\ \gamma_y(2) \end{pmatrix}$$

Where U represents the PACF values. We know from the exercise that $\gamma_y(0) = 1.21$. That is, we need to solve the system:

$$\gamma_y(1) = \gamma_y(0)U_1 + \gamma_y(1)U_2,$$

$$\gamma_y(2) = \gamma_y(1)U_1 + \gamma_y(0)U_2.$$

For the given values of $U_1 = 0.54$, $U_2 = 0.21$, and $\gamma_y(0) = 1.21$, we solve these equations:

$$\gamma_y(1) = \frac{\gamma_y(0)U_1}{1 - U_2} \approx 0.827,$$

$$\gamma_y(2) = \gamma_y(1)U_1 + \gamma_y(0)U_2 \approx 0.701.$$

Next, we use the Yule-Walker equations to estimate the parameters ϕ_1 and ϕ_2 . These equations are:

$$\gamma_y(1) = \gamma_y(0)\phi_1 + \gamma_y(1)\phi_2,$$

$$\gamma_y(2) = \gamma_y(1)\phi_1 + \gamma_y(0)\phi_2.$$

Solving these equations, we get:

$$\phi_2 = \frac{\gamma_y(2) - \frac{\gamma_y^2(1)}{\gamma_y(0)}}{\gamma_y(0) - \frac{\gamma_y^2(1)}{\gamma_y(0)}} \approx 0.21,$$

$$\phi_1 = \frac{\gamma_y(1)(1 - \phi_2)}{\gamma_y(0)} \approx 0.54.$$

Thus, the parameter estimates for the AR(2) model are:

$$\phi_1 = 0.54, \quad \phi_2 = 0.21.$$

1.2.c

We use an independent sample because, without it, reliable inference becomes impossible. The estimates would end up being biased. Moreover, if we don't use another sample, we run the risk of overfitting the model: it may describe our specific data well but not the true data-generating process (DGP). This distinction is hard to make without testing hypotheses and predictions using independent data.

To compute the variance of the forecast, let us first define:

$$P_{H_y^T} Y_{t+1} = \hat{\phi}_1 Y_t + \hat{\phi}_2 Y_{t-1}$$

The forecast error is then given by:

$$\epsilon_{T+1} = Y_{t+1} - P_{H_y^T} Y_{t+1}$$

To compute the confidence interval for this forecast, we need to compute the variance of the forecast error:

$$\text{Var}(\epsilon_{T+1}) = \text{Var}(Y_{t+1} - P_{H_y^T} Y_{t+1})$$

Here, the assumption of independence between the sample and the forecasting process simplifies the computation. Under covariance stationarity, this formula becomes:

$$\text{Var}(\epsilon_{T+1}) = \text{Var}(Y_{t+1}) + \text{Var}(P_{H_y^T} Y_{t+1}) = \sigma^2 + \text{Var}(P_{H_y^T} Y_{t+1})$$

Next, we compute $\text{Var}(P_{H_y^T} Y_{t+1})$ using the law of total variance:

$$\text{Var}(P_{H_y^T} Y_{t+1}) = \text{Var}(\hat{\phi}_1 Y_t + \hat{\phi}_2 Y_{t-1})$$

Given that both Y_t and $\hat{\phi}$ are random variables, this variance can be expressed as:

$$\text{Var}(\hat{\phi}_1 Y_t + \hat{\phi}_2 Y_{t-1}) = \mathbb{E}[\text{Var}(\hat{\phi}_1 Y_t + \hat{\phi}_2 Y_{t-1} \mid Y_t, Y_{t-1})] + \text{Var}(\mathbb{E}[\hat{\phi}_1 Y_t + \hat{\phi}_2 Y_{t-1} \mid Y_t, Y_{t-1}])$$

Approximating these terms, we have:

$$\mathbb{E}[Y_t^2] \text{Var}(\hat{\phi}_1) + \mathbb{E}[Y_{t-1}^2] \text{Var}(\hat{\phi}_2) + 2\mathbb{E}[Y_t Y_{t-1}] \text{Cov}(\hat{\phi}_1, \hat{\phi}_2) + \hat{\phi}_1^2 \text{Var}(Y_t) + \hat{\phi}_2^2 \text{Var}(Y_{t-1}) + 2\hat{\phi}_1 \hat{\phi}_2 \text{Cov}(Y_t, Y_{t-1})$$

Here σ^2 is the innovation variance of the AR(2) process, $\text{Var}(\hat{\phi}_1)$, $\text{Var}(\hat{\phi}_2)$, and $\text{Cov}(\hat{\phi}_1, \hat{\phi}_2)$ are the variances and covariance of the estimated coefficients and $\mathbb{E}[Y_t^2]$, $\mathbb{E}[Y_{t-1}^2]$, and $\mathbb{E}[Y_t Y_{t-1}]$ are sample estimates from the data.

Finally, the 95% confidence interval for the forecast at horizon $T + 1$ is:

$$P_{H_y^T} Y_{t+1} \pm 1.96 \sqrt{\text{Var}(\epsilon_{T+1})}$$

This result accounts for the uncertainty in the AR coefficient estimates and the forecast error.

Exercise 1.3

Now we assume that all these models have passed all the pre-specification tests. That is, both AR(3) and the different ARMA(p,q) models have the correct representation and they have been consistently estimated. To decide which model is the best one for prediction, we will follow a three step procedure:

1. **Check if the errors are white noise (Portmanteau Test)**
2. **Check the significance of the coefficients**
3. **Check information criteria and R^2**

If we start by the first method, we know that the Portmanteau tests the following hypothesis:

H_0 : ϵ is a white noise

And to test this we run the following test:

$$T \sum_{h=1}^K \hat{\rho}_{\epsilon}(h)^2$$

The test statistic follows a χ^2 distribution with $(K - p - q)$ degrees of freedom, where K are the number of lags tested, p the order of the AR process and q the order of the MA process. We reject the null hypothesis if the value of the test is above the rejection region (at 5% significance level)

If we go case by case:

1. **ARMA(1,2)**: with 9 degrees of freedom, $16.92 < 16.99$, we reject the null hypothesis.
2. **AR(3)**: with 9 degrees of freedom, $16.92 > 13.73$ in this case it is clear that we fail to reject the null.
3. **ARMA(2,1)**: 9 degrees of freedom, $16.92 > 10.05$, we also fail to reject.
4. **ARMA(2,2)**: 8 degrees of freedom, $15.51 > 10.26$, we also fail to reject.

In this first method we have discarded the ARMA(1,2) as a possible candidate. That the errors are white noise it is an essential requirement for forecasting, as the canonical representation (or causal) can only be achieved if we have errors that behave as white noise.

Following method **2** and we do some simple test statistics, we find the following:

1. **AR(3)**: if we do a t-test on ϕ_3 we get the following:

$$t_3 = \frac{0.05}{0.03} \approx 1.67 < 1.98$$

Thus we can't reject that the last term of the AR process is 0.

2. **ARMA(2,1)**: all t-test reject that the coefficients are 0.
3. **ARMA(2,2)**: we can't reject that the second error term coefficient is 0:

$$t_2 = \frac{0.34}{0.42} = 0.8 < 1.96$$

Finally, if we take into account **step 3** we want to pick, for those models that have surpassed, all the steps, those that have the lowest BIC or AIC. These measures summarize the trade-off between innovation variance and the number of parameters we include in the model.

Using the results we got before and observing the values of the BIC and AIC information criteria, we can conclude that the best model is the **ARMA(2,1)** as this is the model with the lowest info criteria, the highest R^2 and we fail to reject that some of its parameters are 0.

Exercise 2

We have the following model:

$$y_t = \rho y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$$

Where $|\rho| < 1$, $|\theta| < 1$ and ϵ_t is a strong white noise with variance σ^2 .

2.1

If we estimate $\hat{\rho}$ with OLS, the estimate would be biased. The OLS estimator is given by:

$$\hat{\rho} = \frac{\sum_{t=2}^T (y_t y_{t-1})}{\sum_{t=2}^T (y_{t-1}^2)}$$

Expanding the numerator, we have:

$$\hat{\rho} = \frac{\sum_{t=2}^T ((\rho y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}) y_{t-1})}{\sum_{t=2}^T (y_{t-1}^2)}$$

Breaking this down:

$$\hat{\rho} = \rho + \frac{\sum_{t=2}^T (y_{t-1} \epsilon_t)}{\sum_{t=2}^T (y_{t-1}^2)} + \frac{\sum_{t=2}^T (y_{t-1} \theta \epsilon_{t-1})}{\sum_{t=2}^T (y_{t-1}^2)}$$

The second term, converges to $\frac{\mathbb{E}(y_{t-1} \epsilon_t)}{\text{Var}(y_{t-1})}$ in probability, which will be 0 because ϵ_t is white noise and uncorrelated with past values of y_{t-1} .

However, the third term, converges to $\frac{\mathbb{E}(y_{t-1} \theta \epsilon_{t-1})}{\text{Var}(y_{t-1})}$, and this does not converge to zero. To see why:

$$\mathbb{E}(y_{t-1} \epsilon_{t-1}) = \mathbb{E}((\rho y_{t-2} + \epsilon_{t-1} + \theta \epsilon_{t-2}) \epsilon_{t-1}) = \mathbb{E}(\epsilon_{t-1}^2) = \sigma^2$$

This non-zero expectation implies that the third term introduces asymptotic bias into the OLS estimator. Thus, $\hat{\rho}$ does not converge to the true value of ρ as the sample size grows, making the OLS estimator inconsistent.

To address this inconsistency, we can use y_{t-2} as an instrument. Since y_{t-2} is lagged far enough, it is uncorrelated with the error term ϵ_{t-1} but remains correlated with y_{t-1} . This ensures that y_{t-2} is a valid instrument, eliminating the bias caused by the correlation between y_{t-1} and ϵ_{t-1} .

In the next section, we will show that this instrumental variable estimator is consistent under the given assumptions.

2.2

Is this estimator consistent when ρ is different from 1? Let's check:

$$\begin{aligned} \hat{\rho}_{IV} &= \frac{\sum_{t=2}^T y_t y_{t-2}}{\sum_{t=2}^T y_{t-1} y_{t-2}} \\ \hat{\rho}_{IV} &= \frac{\sum_{t=2}^T (\rho y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}) y_{t-2}}{\sum_{t=2}^T y_{t-1} y_{t-2}} \\ \hat{\rho}_{IV} &= \frac{\sum_{t=2}^T \rho (y_{t-1} y_{t-2})}{\sum_{t=2}^T y_{t-1} y_{t-2}} + \frac{\sum_{t=2}^T (\epsilon_t y_{t-2})}{\sum_{t=2}^T y_{t-1} y_{t-2}} + \frac{\sum_{t=2}^T (\theta \epsilon_{t-1} y_{t-2})}{\sum_{t=2}^T y_{t-1} y_{t-2}} \\ \hat{\rho}_{IV} &= \rho + \frac{\sum_{t=2}^T (\epsilon_t y_{t-2})}{\sum_{t=2}^T y_{t-1} y_{t-2}} + \frac{\theta \sum_{t=2}^T (\epsilon_{t-1} y_{t-2})}{\sum_{t=2}^T y_{t-1} y_{t-2}} \end{aligned}$$

As t grows to infinity, the terms $\frac{\sum_{t=2}^T (\epsilon_t y_{t-2})}{\sum_{t=2}^T y_{t-1} y_{t-2}}$ and $\frac{\theta \sum_{t=2}^T (\epsilon_{t-1} y_{t-2})}{\sum_{t=2}^T y_{t-1} y_{t-2}}$ converge to zero. This is because:

By ergodic theorem, which applies here due to the stationarity and ergodicity of the process, the sample averages converge to their expectations. Thus:

$$\frac{1}{T} \sum_{t=2}^T \epsilon_t y_{t-2} \rightarrow \mathbb{E}(\epsilon_t y_{t-2}) = 0 \quad \text{and} \quad \frac{1}{T} \sum_{t=2}^T \epsilon_{t-1} y_{t-2} \rightarrow \mathbb{E}(\epsilon_{t-1} y_{t-2}) = 0.$$

These two equal 0 because:

1. ϵ_t and y_{t-2} are uncorrelated due to ϵ_t being strong white noise.
2. y_{t-2} is uncorrelated with ϵ_{t-1} since y_{t-2} is a function of lags of ϵ_t , which are independent.

Therefore we can conclude that indeed $\hat{\rho}_{IV}$ converges to ρ , proving that the estimator is consistent when $\rho \neq 1$.

2.3

For any given ARMA(p,q), we would need to use y_{t-p-k} for all $k > 0$ as an instrument. These lagged variables are uncorrelated with the error terms but remain correlated with the autoregressive components of the model, ensuring that they are valid instruments

To relate the ARMA(p,q) representation with the Yule-Walker equations, we use ARMA(1,1) as an example. The model is given by:

$$y_t - \phi y_{t-1} = \epsilon_t + \theta \epsilon_{t-1}$$

Multiplying through by y_t and taking expectations:

$$\gamma_y(0) - \phi \gamma_y(1) = \mathbb{E}[(\epsilon_t + \theta \epsilon_{t-1}) y_t].$$

Since ϵ_t is white noise, uncorrelated with y_t , we simplify:

$$\gamma_y(0) - \phi \gamma_y(1) = \sigma^2 + \theta(\mathbb{E}[\epsilon_{t-1} \phi y_{t-1}] + \theta \sigma^2).$$

Using the independence properties of ϵ_t and recursive relationships, this becomes:

$$\gamma_y(0) - \phi \gamma_y(1) = \sigma^2(1 + \theta + \theta^2).$$

Multiplying through by y_{t-1} and taking expectations:

$$\gamma_y(1) - \phi \gamma_y(0) = \mathbb{E}[(\epsilon_t + \theta \epsilon_{t-1}) y_{t-1}].$$

Here, $\mathbb{E}[\epsilon_t y_{t-1}] = 0$, leaving:

$$\gamma_y(1) - \phi \gamma_y(0) = \sigma^2 \theta.$$

Multiplying through by y_{t-2} and taking expectations:

$$\gamma_y(2) - \phi\gamma_y(1) = 0.$$

This implies:

$$\gamma_y(2) = \phi\gamma_y(1).$$

For lags $\ell \geq 2$, the autocovariance functions derived from the Yule-Walker equations are expressed as linear combinations of previous covariances. This is a characteristic of the autoregressive structure. However, for lower lags ($\ell < 2$), the autocovariance depends not only on the AR parameters but also on the variance of the errors (σ^2) and the moving average parameter (θ).

This distinction highlights the dual nature of ARMA processes. The higher-lag relationships are governed by the AR dynamics, while the lower-lag equations include contributions from the error structure, underscoring the role of θ in shaping the short-term behavior of the process.

2.4

We aim to show that under the null hypothesis $H_0 : \rho = 1$:

$$\frac{1}{T} \sum_{t=2}^T y_{t-1} \epsilon_t = \frac{1}{T} \sum_{t=2}^T y_{t-2} \epsilon_t + o_p(1).$$

Under H_0 , the process y_t follows:

$$y_t = y_{t-1} + \epsilon_t + \theta \epsilon_{t-1},$$

which implies:

$$y_{t-1} = y_{t-2} + \epsilon_{t-1} + \theta \epsilon_{t-2}.$$

Substituting y_{t-1} into the expression $\frac{1}{T} \sum_{t=2}^T y_{t-1} \epsilon_t$, we obtain:

$$\frac{1}{T} \sum_{t=2}^T y_{t-1} \epsilon_t = \frac{1}{T} \sum_{t=2}^T (y_{t-2} + \epsilon_{t-1} + \theta \epsilon_{t-2}) \epsilon_t.$$

Expanding the terms, we get:

$$\frac{1}{T} \sum_{t=2}^T y_{t-1} \epsilon_t = \frac{1}{T} \sum_{t=2}^T y_{t-2} \epsilon_t + \frac{1}{T} \sum_{t=2}^T \epsilon_{t-1} \epsilon_t + \theta \frac{1}{T} \sum_{t=2}^T \epsilon_{t-2} \epsilon_t.$$

We analyze the behavior of each term on the right-hand side as $T \rightarrow \infty$:

- First Term: $\frac{1}{T} \sum_{t=2}^T y_{t-2} \epsilon_t$

The term y_{t-2} grows with T , following the unit root behavior. By the ergodic theorem (ET) this term converges in probability to its expectation:

$$\mathbb{E}[y_{t-2} \epsilon_t] = 0,$$

because y_{t-2} is independent of ϵ_t (white noise). However, the variance of this term scales as:

$$\text{Var} \left(\frac{1}{T} \sum_{t=2}^T y_{t-2} \epsilon_t \right) \sim \frac{1}{T^2} \cdot T^3 = T^{-1},$$

meaning it converges to zero at the rate of $T^{-1/2}$. This makes the first term the dominant component asymptotically.

- Second Term: $\frac{1}{T} \sum_{t=2}^T \epsilon_{t-1} \epsilon_t$

The second term involves the product of uncorrelated white noise terms. By the E.T, this term converges in probability to:

$$\mathbb{E}[\epsilon_{t-1} \epsilon_t] = 0.$$

Its variance scales as T^{-1} , meaning this term vanishes faster than the first term.

- Third Term: $\theta \frac{1}{T} \sum_{t=2}^T \epsilon_{t-2} \epsilon_t$

Similar to the second term, this involves products of uncorrelated white noise terms. By the E.T, it converges in probability to:

$$\mathbb{E}[\epsilon_{t-2} \epsilon_t] = 0,$$

with variance scaling as T^{-1} . This term also vanishes faster than the first term.

While all terms converge to zero in probability, the first term, $\frac{1}{T} \sum_{t=2}^T y_{t-2} \epsilon_t$, dominates asymptotically due to the slower rate of convergence. Therefore, combining these results:

$$\frac{1}{T} \sum_{t=2}^T y_{t-1} \epsilon_t = \frac{1}{T} \sum_{t=2}^T y_{t-2} \epsilon_t + o_p(1),$$

as required.

2.5

We now aim to show that under the null hypothesis $H_0 : \rho = 1$:

$$\frac{1}{T^2} \sum_{t=2}^T y_{t-1} y_{t-2} = \frac{1}{T^2} \sum_{t=2}^T y_{t-2}^2 + o_p(1).$$

Under H_0 , we recall:

$$y_{t-1} = y_{t-2} + \epsilon_{t-1} + \theta\epsilon_{t-2}.$$

Substituting y_{t-1} into $\frac{1}{T^2} \sum_{t=2}^T y_{t-1}y_{t-2}$, we have:

$$\frac{1}{T^2} \sum_{t=2}^T y_{t-1}y_{t-2} = \frac{1}{T^2} \sum_{t=2}^T (y_{t-2} + \epsilon_{t-1} + \theta\epsilon_{t-2})y_{t-2}.$$

Expanding the terms, we get:

$$\frac{1}{T^2} \sum_{t=2}^T y_{t-1}y_{t-2} = \frac{1}{T^2} \sum_{t=2}^T y_{t-2}^2 + \frac{1}{T^2} \sum_{t=2}^T \epsilon_{t-1}y_{t-2} + \theta \frac{1}{T^2} \sum_{t=2}^T \epsilon_{t-2}y_{t-2}.$$

- First Term: $\frac{1}{T^2} \sum_{t=2}^T y_{t-2}^2$

This is the dominant term because $y_{t-2}^2 \sim O(T)$. By the E.T, the term converges to:

$$T\mathbb{E}(y_{t-2}^2) \sim O_p(1).$$

- Second Term: $\frac{1}{T^2} \sum_{t=2}^T \epsilon_{t-1}y_{t-2}$

This term has zero mean because ϵ_{t-1} is white noise and uncorrelated with y_{t-2} . By the E.T, it converges in probability to:

$$T\mathbb{E}[\epsilon_{t-1}y_{t-2}] = 0,$$

and its variance scales as T^{-1} , so:

$$\frac{1}{T^2} \sum_{t=2}^T \epsilon_{t-1}y_{t-2} = o_p(1).$$

- Third Term: $\theta \frac{1}{T^2} \sum_{t=2}^T \epsilon_{t-2}y_{t-2}$

Similarly, this term has zero mean by the E.T:

$$T\mathbb{E}[\epsilon_{t-2}y_{t-2}] = 0,$$

and its variance also scales as T^{-1} , so:

$$\frac{1}{T^2} \sum_{t=2}^T \epsilon_{t-2}y_{t-2} = o_p(1).$$

Since the second and third terms vanish faster than the first term, we conclude:

$$\frac{1}{T^2} \sum_{t=2}^T y_{t-1}y_{t-2} = \frac{1}{T^2} \sum_{t=2}^T y_{t-2}^2 + o_p(1),$$

as required.

2.6

We have:

$$\hat{\rho}_{IV} = \frac{\sum_{t=2}^T y_t y_{t-2}}{\sum_{t=2}^T y_{t-1} y_{t-2}}.$$

Then subtracting 1 from both sides:

$$\hat{\rho}_{IV} - 1 = \frac{\sum_{t=2}^T y_t y_{t-2}}{\sum_{t=2}^T y_{t-1} y_{t-2}} - 1.$$

This can be rewritten as:

$$\hat{\rho}_{IV} - 1 = \frac{\sum_{t=2}^T y_t y_{t-2}}{\sum_{t=2}^T y_{t-1} y_{t-2}} - \frac{\sum_{t=2}^T y_{t-1} y_{t-2}}{\sum_{t=2}^T y_{t-1} y_{t-2}}.$$

Simplifying the expression:

$$\hat{\rho}_{IV} - 1 = \frac{\sum_{t=2}^T y_{t-2}(\epsilon_t + \theta \epsilon_{t-1})}{\sum_{t=2}^T y_{t-1} y_{t-2}}.$$

Now, multiplying and dividing by T , we have:

$$T(\hat{\rho}_{IV} - 1) = \frac{\frac{1}{T} \sum_{t=2}^T y_{t-2}(\epsilon_t + \theta \epsilon_{t-1})}{\frac{1}{T^2} \sum_{t=2}^T y_{t-1} y_{t-2}}.$$

Then, if we analyze term by term we know that by UCLT:

$$\frac{1}{T} \sum_{t=2}^T y_{t-2}(\epsilon_t + \theta \epsilon_{t-1}) \xrightarrow{d} (1 + \theta) \sigma^2 \int_0^1 B(s) dB(s)$$

By exercise 2.5 we know that

$$\frac{1}{T^2} \sum_{t=2}^T y_{t-1} y_{t-2} \xrightarrow{p} \frac{1}{T^2} \sum_{t=2}^T y_{t-2}^2$$

And by UCLT we know that¹:

$$\frac{1}{T^2} \sum_{t=2}^T y_{t-2}^2 \xrightarrow{d} \sigma^2 \int_0^1 [B(s)]^2 dB(s)$$

Then by Slutsky, combining the convergence in distribution and the one in probability, we know that

$$\frac{1}{T^2} \sum_{t=2}^T y_{t-1} y_{t-2} \xrightarrow{d} \sigma^2 \int_0^1 [B(s)]^2 dB(s)$$

¹here I follow Hamilton, though I am not sure

Finally, by Continuous Mapping Theorem, we know that:

$$\left(\frac{1}{T^2} \sum_{t=2}^T y_{t-1}y_{t-2}\right)^{-1} \xrightarrow{d} (\sigma^2 \int_0^1 [B(s)]^2 dB(s))^{-1}$$

Thus, combining the results we get:

$$T(\hat{\rho}_{IV} - 1) \xrightarrow{d} \frac{(1 + \theta)\sigma^2 \int_0^1 B(s)dB(s)}{\sigma^2 \int_0^1 [B(s)]^2 dB(s)}$$

Thus the asymptotic variance does depend on the nuisance parameter θ . As this needs also to be estimated, this brings more uncertainty and unreliability to the inference we can make.

2.7

The test statistic is given by:

$$\hat{t} = \frac{\hat{\rho}_{IV} - 1}{\sqrt{V_{\hat{\rho}_{IV}}}}$$

We want to estimate the asymptotic variance under the null hypothesis H_0 . Under H_0 , the key asymptotic results for the summations involving y_t are:

$$\begin{aligned} \sum_{t=2}^T y_{t-1}y_{t-2} &\xrightarrow{d} \sigma^2 T^3 \int_0^1 [B(s)]^2 ds, \\ \sum_{t=2}^T y_{t-2}^2 &\xrightarrow{d} \sigma^2 T^3 \int_0^1 [B(s)]^2 ds. \end{aligned}$$

Using the Continuous Mapping Theorem (CMT), the inverse squared summation asymptotically behaves as:

$$\left(\sum_{t=2}^T y_{t-1}y_{t-2}\right)^{-2} \xrightarrow{d} \left(\sigma^2 T^3 \int_0^1 [B(s)]^2 ds\right)^{-2}.$$

The residuals η_t are defined as:

$$\eta_t = y_t - \rho y_{t-1}.$$

Under $H_0 : \rho = 1$, this simplifies to:

$$\eta_t = \varepsilon_t + \theta \varepsilon_{t-1}.$$

The variance of η_t is:

$$\text{Var}(\eta_t) = \sigma^2(1 + \theta^2).$$

The variance of the IV estimator is given by:

$$V_{\hat{\rho}_{IV}} = \frac{\hat{\sigma}_\eta^2}{\left(\sum_{t=2}^T y_{t-1}y_{t-2}\right)^2}.$$

Substituting the asymptotic results, we obtain:

$$V_{\hat{\rho}_{IV}} \xrightarrow{d} \frac{\sigma^2(1+\theta^2)}{\left(\sigma^2 T^3 \int_0^1 [B(s)]^2 ds\right)^2}.$$

Simplifying further and applying CMT:

$$\sqrt{V_{\hat{\rho}_{IV}}} \xrightarrow{d} \frac{\sqrt{\sigma^2(1+\theta^2)}}{\sigma^2 T^3 \int_0^1 [B(s)]^2 ds}.$$

The numerator of \hat{t} is:

$$T(\hat{\rho}_{IV} - 1) \xrightarrow{d} \frac{\int_0^1 B(s)dB(s)}{\int_0^1 [B(s)]^2 ds}.$$

Combining this with the denominator, the test statistic \hat{t} has the asymptotic distribution:

$$\hat{t} \xrightarrow{d} \frac{\int_0^1 B(s)dB(s)}{\sqrt{1+\theta^2} \int_0^1 [B(s)]^2 ds}.$$

We can see that the t-test will also depend on the nuisance parameter θ

2.8

We assume $\theta + \rho = 0$ and test the null hypothesis $H_0 : \rho = 1$. The IV estimator is given by:

$$\hat{\rho}_{IV} = \rho - \frac{\sum_{t=2}^T (\epsilon_{t-1} - \epsilon_t)y_{t-2}}{\sum_{t=2}^T y_{t-1}y_{t-2}}.$$

First we will check for the consistency of $\hat{\rho}_{IV}$ under the null:

Under $H_0 : \rho = 1$, the process y_t is integrated of order 1. The denominator $\sum_{t=2}^T y_{t-1}y_{t-2}$ scales as:

$$\sum_{t=2}^T y_{t-1}y_{t-2} = \mathcal{O}(T^2),$$

due to the behavior of the random walk y_t . Meanwhile, the numerator $\sum_{t=2}^T (\epsilon_{t-1} - \epsilon_t)y_{t-2}$ is driven by the innovations ϵ_t and scales as:

$$\sum_{t=2}^T (\epsilon_{t-1} - \epsilon_t)y_{t-2} = \mathcal{O}(T).$$

Since the numerator grows slower than the denominator, the ratio converges to zero²:

$$\frac{\sum_{t=2}^T (\epsilon_{t-1} - \epsilon_t) y_{t-2}}{\sum_{t=2}^T y_{t-1} y_{t-2}} \xrightarrow{p} 0.$$

Thus, $\hat{\rho}_{IV}$ is consistent under the null, meaning:

$$\hat{\rho}_{IV} \xrightarrow{p} \rho.$$

Asymptotic Behavior of $\sqrt{T}(\hat{\rho}_{IV} - 1)$

To determine the asymptotic distribution, we scale the expression for $\hat{\rho}_{IV}$ by \sqrt{T} :

$$\sqrt{T}(\hat{\rho}_{IV} - 1) = \frac{\sqrt{T} \cdot \sum_{t=2}^T (\epsilon_{t-1} - \epsilon_t) y_{t-2}}{\sum_{t=2}^T y_{t-1} y_{t-2}}.$$

Let's break it down by parts, The denominator can be approximated using previous results:

$$T^{-2} \sum_{t=2}^T y_{t-1} y_{t-2} \xrightarrow{d} \sigma^2 \int_0^1 [W(r)]^2 dr,$$

where $W(r)$ is a standard Brownian motion.

Then we need to apply the continuous mapping theorem to find:

$$(T^{-2} \sum_{t=2}^T y_{t-1} y_{t-2})^{-1} \xrightarrow{d} (\sigma^2 \int_0^1 [W(r)]^2 dr)^{-1},$$

The numerator, on the other hand, involves:

$$\sum_{t=2}^T (\epsilon_{t-1} - \epsilon_t) y_{t-2}.$$

Using the fact that $y_t \sim \mathcal{O}(T)$ and $(\epsilon_{t-1} - \epsilon_t)$ are independent of y_{t-2} , the numerator scales as:

$$T^{-3/2} \sum_{t=2}^T (\epsilon_{t-1} - \epsilon_t) y_{t-2} \xrightarrow{d} \int_0^1 W(r) dB(r),$$

where $B(r)$ is a standard Brownian motion representing the innovations ϵ_t .

Finally, we combine the two results using the CMT, the scaled difference converges to:

$$\sqrt{T}(\hat{\rho}_{IV} - 1) \xrightarrow{d} \frac{\int_0^1 W(r) dB(r)}{\sigma^2 \int_0^1 [W(r)]^2 dr}.$$

²we can also show that this will go to 0 by applying the E.T and then arguing that the expectation is 0 due to the components not being correlated.

And we can see that **this does not depend on the nuisance parameter θ** . Thus, the conditions that need to hold are:

1. Unit Root Behavior of y_t : The process y_t must follow a random walk under $H_0 : \rho = 1$.
2. Independence of Innovations: The innovations ϵ_t must be independent and identically distributed with zero mean and finite variance.
3. No Explosive Roots: The condition $\theta + \rho = 0$ ensures stationarity of the errors.

Exercise 3

3.1.

First we know that x_t is a function of y_{ts} . Also, we know that y_{ts} is covariance stationary. We will use this fact to show if x_t is also covariance stationary. We check the usual conditions:

1. Mean:

$$\mathbb{E}(x_t) = \sum_{i=0}^{s-1} \mathbb{E}(y_{ts+i}) = s\mu,$$

where μ is the mean of y_t . Since this does not depend on t , the mean of x_t is constant.

2. Variance:

$$\text{Var}(x_t) = \text{Var}\left(\sum_{i=0}^{s-1} y_{ts+i}\right).$$

Using the properties of variance:

$$\text{Var}(x_t) = \sum_{i=0}^{s-1} \text{Var}(y_{ts+i}) + 2 \sum_{0 \leq i < j \leq s-1} \text{Cov}(y_{ts+i}, y_{ts+j}).$$

Since y_t is covariance stationary, $\text{Var}(y_{ts+i}) = \sigma^2$ and $\text{Cov}(y_{ts+i}, y_{ts+j}) = \gamma(|i - j|)$, we have:

$$\text{Var}(x_t) = s\sigma^2 + 2 \sum_{0 \leq i < j \leq s-1} \gamma(|i - j|).$$

This expression depends only on s and the autocovariance function $\gamma(h)$, and not on t , so the variance is constant.

3. Autocovariance: The autocovariance of x_t at lag h is:

$$\text{Cov}(x_t, x_{t+h}) = \text{Cov}\left(\sum_{i=0}^{s-1} y_{ts+i}, \sum_{j=0}^{s-1} y_{(t+h)s+j}\right).$$

Expanding this:

$$\text{Cov}(x_t, x_{t+h}) = \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} \text{Cov}(y_{ts+i}, y_{(t+h)s+j}).$$

Using the covariance stationarity of y_t , the term $\text{Cov}(y_{ts+i}, y_{(t+h)s+j})$ depends only on the time difference $|(ts+i) - ((t+h)s+j)|$. Therefore, $\text{Cov}(x_t, x_{t+h})$ depends only on h , and not on t .

Conclusion: Since x_t satisfies all three conditions, it is covariance stationary.

3.2.

Using the results derived above, we can express the autocovariance of x_t as:

$$\gamma_x(h) = \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} \gamma_y(hs + j - i).$$

Definition: for an ARMA process with covariance stationarity, the autocovariance function of y_t is given by:

$$\gamma_y(h) = \sum_{l=1}^p c_l \rho_l^{|h|}, \quad \forall h \geq 0,$$

where:

- ρ_l are the roots of the AR polynomial $\phi(B) = 0$, with $|\rho_l| < 1$,
- c_l are constants determined by the MA parameters, σ^2 and θ .

Using this representation, we substitute $\gamma_y(hs + j - i)$ into the expression for $\gamma_x(h)$:

$$\gamma_y(hs + j - i) = \sum_{l=1}^p c_l \rho_l^{|hs+j-i|}.$$

Thus, the autocovariance of x_t becomes:

$$\gamma_x(h) = \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} \sum_{l=1}^p c_l \rho_l^{|hs+j-i|}.$$

The behavior of $\gamma_x(h)$ is dominated by the root ρ_1 of the AR polynomial that is closest to 1, since this root corresponds to the slowest decay. Asymptotically:

$$\gamma_y(hs + j - i) \approx c_1 \rho_1^{|hs+j-i|}.$$

Substituting this approximation:

$$\gamma_x(h) \approx \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} c_1 \rho_1^{|hs+j-i|}.$$

For large h , $\rho_1^{|hs+j-i|}$ decays as ρ_1^{hs} , where ρ_1 is the dominant root. Aggregation over i and j introduces a scaling factor A but does not affect the dominant decay rate. Therefore:

$$\gamma_x(h) \approx A \rho_1^{hs},$$

where:

$$A = c_1 \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} 1 = c_1 s^2.$$

The autocovariance function of x_t exhibits exponential decay at a rate determined by the largest root ρ_1 of the AR polynomial of y_t , scaled by the aggregation parameter s . Hence, for large h :

$$\gamma_x(h) \sim A \rho_1^{hs}, \quad \text{with } A = c_1 s^2.$$

3.3.

We are given a zero-mean ARMA(1,1) process y_t :

$$y_t = \phi y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t,$$

where ε_t is white noise with variance σ^2 .

The aggregated process x_t is defined as:

$$x_t = \sum_{i=0}^{s-1} y_{ts+i}.$$

For each y_{ts+i} , substitute from the ARMA(1,1) representation:

$$y_{ts+i} = \phi y_{ts+i-1} + \theta \varepsilon_{ts+i-1} + \varepsilon_{ts+i}.$$

Substituting this into x_t , we get:

$$x_t = \sum_{i=0}^{s-1} (\phi y_{ts+i-1} + \theta \varepsilon_{ts+i-1} + \varepsilon_{ts+i}).$$

Reorganizing the terms, we have:

$$x_t = \phi \sum_{i=0}^{s-1} y_{ts+i-1} + \theta \sum_{i=0}^{s-1} \varepsilon_{ts+i-1} + \sum_{i=0}^{s-1} \varepsilon_{ts+i}.$$

Note that $\sum_{i=0}^{s-1} y_{ts+i-1}$ is itself a lagged version of x_t . Specifically:

$$\sum_{i=0}^{s-1} y_{ts+i-1} = x_{t-1}.$$

Thus, x_t becomes:

$$x_t = \phi x_{t-1} + \theta \sum_{i=0}^{s-1} \varepsilon_{ts+i-1} + \sum_{i=0}^{s-1} \varepsilon_{ts+i}.$$

We define the combined noise term:

$$\eta_t = \theta \sum_{i=0}^{s-1} \varepsilon_{ts+i-1} + \sum_{i=0}^{s-1} \varepsilon_{ts+i}.$$

Here, η_t represents a moving average component of order $s+1$ because it depends on the white noise terms $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-s-1}$.

The process x_t satisfies the equation:

$$x_t = \phi x_{t-1} + \eta_t,$$

where η_t is an $\text{MA}(s+1)$ process. Therefore, x_t follows an $\text{ARMA}(1, s+1)$ representation.

3.4.

The forecast error variance for the direct prediction of x_{t+1} is given by:

$$\text{Var}(x_{t+1} - \mathbb{E}[x_{t+1}]) = \text{Var}(\eta_{t+1}) = s \cdot \sigma^2,$$

where η_{t+1} is the aggregated white noise term from the ARMA process.

On the other hand, the variance obtained by aggregating the individual forecasts for y_t over s periods is:

$$\text{Var}\left(\sum_{i=0}^{s-1} y_{ts+i+1} - \mathbb{E}[y_{ts+i+1}]\right) = \sum_{i=0}^{s-1} \text{Var}(\varepsilon_{ts+i+1}) = s \cdot \sigma^2.$$

At horizon $t+1$, the two methods yield the same forecast error variance (it is just innovation) because the forecast errors ε_{t+1} are uncorrelated at this horizon. Thus, both the direct forecast of x_{t+1} and the aggregated forecasts for y_t capture the same cumulative variance, determined by the white noise variance σ^2 scaled by the aggregation size s .

However, if we try to forecast at $t+k$ with $k > 1$ and less than the order p , then the forecast prediction error changes, due to the effect of MA dependance that we introduce. Then we would have:

The forecast error variance for the direct prediction of x_{t+k} can be expressed by accounting for the dependency introduced by the ARMA process. Specifically:

$$\text{Var}(x_{t+k} - \mathbb{E}[x_{t+k}]) = \sigma^2 s \cdot \left[1 + \frac{2}{s} \sum_{k=1}^s (\phi^k \theta + \phi^{k+1})^2 + \left(\sum_{k=0}^{s-1} \phi^k \theta \right)^2 \right],$$

Which should differ from the variance computed with aggregating forecasts.