Pset1 - Econometrics II

Estevão Cardoso

March 2025

1 Question 1

The goal is to forecast Brazilian Annual GDP Growth between 2021 and 2030 using the data in the file `data_gdp_brazil.csv` and the following models: AR(1), AR(2), MA(1), MA(2), ARMA(1,1), ARMA(2,1), ARMA(1,2), ARMA(2,2). For each model, you will report the following objects: (1) All estimated coefficients (including the intercept), all coefficients' standard errors, and all coefficients' p-values (testing whether each coefficient is zero or not); (2) The Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC); (3) A plot showing the forecast for the next 10 years (2021-2030) with a 95%-confidence interval for the forecast.

Items a) and b)

	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)	ARMA(2,1)	ARMA(1,2)	ARMA(2,2)
Intercept	4.5380	4.5269	4.5393	4.5349	4.4912	4.2591	4.2605	4.2677
Std Error Intercept	0.5383	0.6051	0.4799	0.5292	0.7062	1.0704	1.0722	1.0643
Pval Intercept	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001
AR1	0.2874	0.2505			0.7400	1.1043	0.9440	1.2901
Std Error AR1	0.0905	0.0942			0.2836	0.1416	0.0702	0.4831
Pval AR1	0.0015	0.0078			0.0091	0.0000	0.0000	0.0076
AR2		0.1218				-0.1489		-0.3235
Std Error AR2		0.0947				0.1104		0.4526
Pval AR2		0.1984				0.1776		0.4747
MA1			0.2383	0.2221	-0.5160	-0.8700	-0.7186	-1.0586
Std Error MA1			0.0855	0.0942	0.3747	0.1016	0.1186	0.4947
Pval MA1			0.0053	0.0184	0.1684	0.0000	0.0000	0.0324
MA2				0.1586			-0.1185	0.1559
Std Error MA2				0.1034			0.0966	0.4100
Pval MA2				0.1250			0.2199	0.7038
AIC	691.9898	692.3487	693.9789	693.7178	691.8091	692.5475	692.7290	694.4245
BIC	700.3523	703.4986	702.3413	704.8678	702.9591	706.4850	706.6664	711.1494

Table 1: Overall Results (AR-MA-ARMA)

Item c)

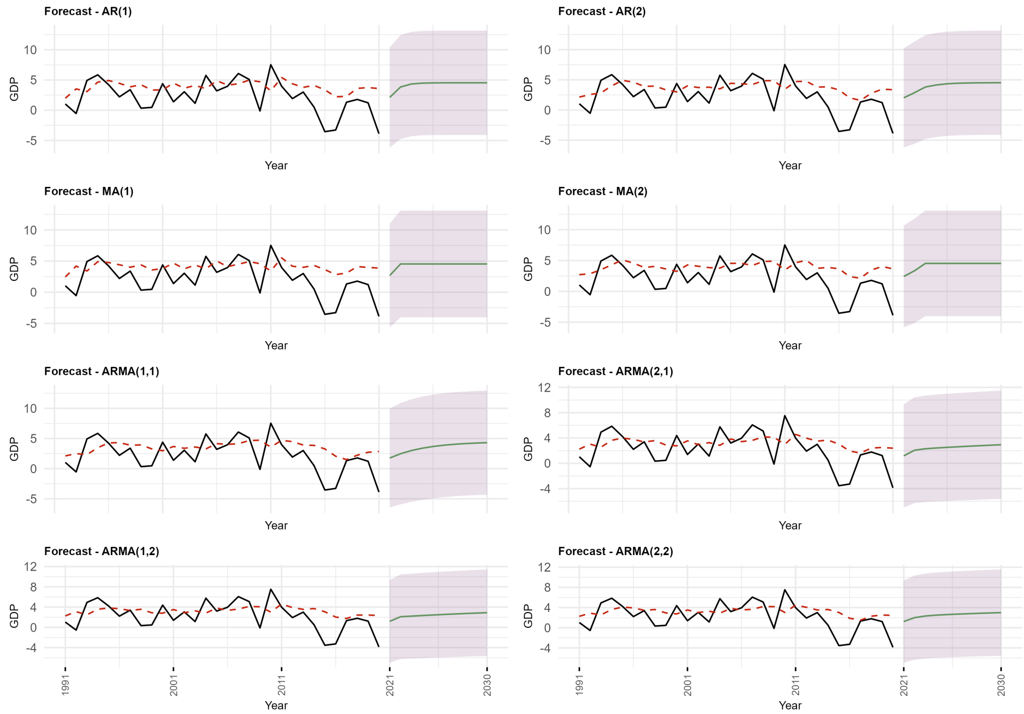


Figure 1: GDP forecast for each model with a 95% confidence interval

Item d)

In order to evaluate each of the eight models, we first examined the BIC and AIC criteria. Broadly speaking, the BIC and AIC criteria both penalize model complexity, though they do so in slightly different ways. The BIC penalizes complexity more heavily than the AIC, making it more conservative in terms of selecting a more complex model. Therefore, a model that is ”overfitted” may be penalized more by the BIC to avoid overfitting. On the other hand, the AIC prioritizes model fit, aiming to achieve the most accurate adjustment possible by balancing goodness of fit with model simplicity. Next, we also considered two important metrics for prediction accuracy: RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error). The latest measures the square root of the average squared differences between the predicted and actual values. It gives more weight to larger errors due to the squaring of the differences, making it sensitive to outliers. So, a lower RMSE indicates that the model’s predictions are generally closer to the true values, with larger deviations penalized more severely. The former, on the other hand, measures the average of the absolute differences between the predicted and actual values, giving the same weight to small and large deviations. Lastly, a lower MAE indicates that, on average, the model’s predictions are closer to the true values, but it is less sensitive to large errors compared to RMSE.

Model	AIC	BIC	RMSE	MAE
ARMA(2,2)	694.4245	711.1494	4.1513	3.4209
MA(1)	693.9789	702.3413	4.2520	3.5387
MA(2)	693.7178	704.8678	4.2116	3.4965
ARMA(1,2)	692.7290	706.6664	4.1563	3.4254
ARMA(2,1)	692.5475	706.4850	4.1533	3.4213
AR(2)	692.3487	703.4986	4.1872	3.4737
AR(1)	691.9898	700.3523	4.2164	3.4954
ARMA(1,1)	691.8091	702.9591	4.1774	3.4642

Table 2: Model Performance Evaluation Metrics

When we consider together the information criteria and the statistical significance of the coefficients, the AR(1) model appears to be the best with a lower BIC. However, that alone is not enough — we can still look at other metrics to identify a good candidate, especially if our goal is forecasting. We can see in table 2 that the ARMA(1,1) model has the lowest AIC value of 691.8091, indicating it has the best balance between goodness of fit and model complexity among all models (recall - the lower the AIC, the best) the ARMA(1,1) model also has the lowest BIC value of 702.9591. Overall, we argue that, as the BIC penalizes model complexity more heavily than AIC, the ARMA(1,1) model's favorable BIC value, puts it as the most appropriate model based on this criterion as well.

Regards the MSE, we plot the residuals in figure 2 to a bettter understanding of how "well" each model is performing. If we look closely to the figure 2, and based on its MSE, the ARMA(2,1) model has the lowest RMSE at 4.1533, suggesting it provides the best predictive performance with the smallest prediction error when compared to the other models. Hence, we can say that is the most accurate in terms of minimizing the overall prediction error on the test data. On the other hand, the ARMA(2,2) model showed the lowest MAE value of 3.4209, suggesting that it is the most reliable in terms of average absolute errors, meaning its predictions are the closest to actual values on average.

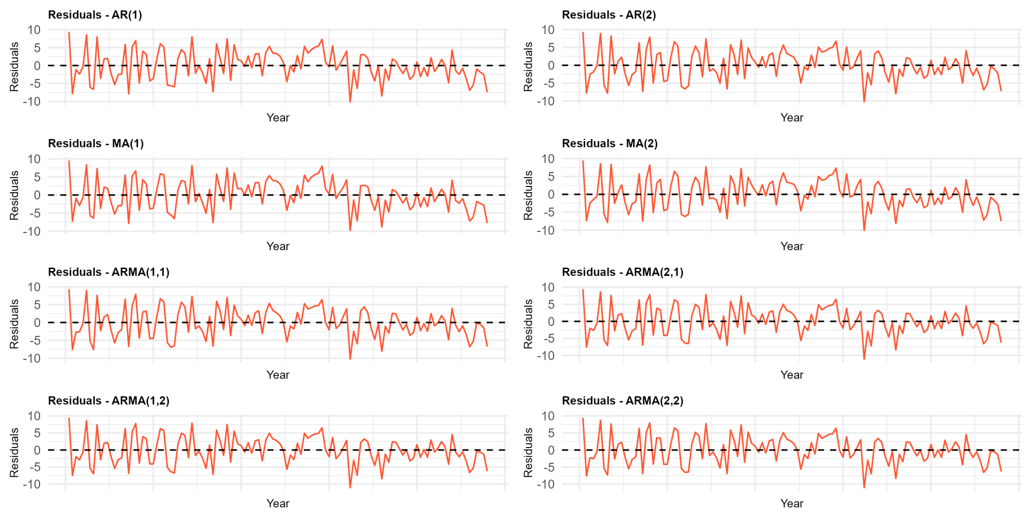


Figure 2: GDP forecast residuals for each model

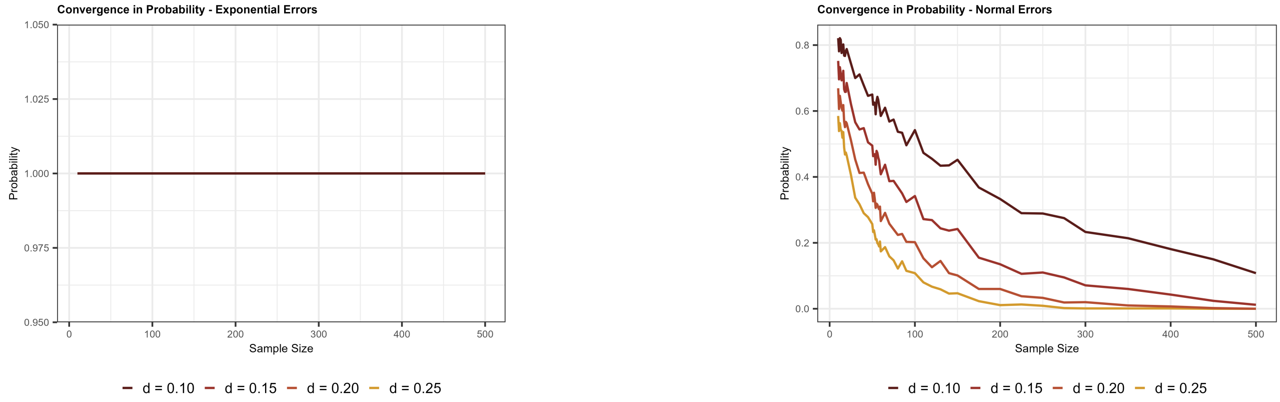
2 Question 2

Note: Before to start, just want to mention that my pc is not that good (i didn't learn how to run in parallell :/), so most of the simulations I did with 200 Monte Carlos simulations. The overall results indicate that choosing the exponential distribution to model the errors is quite problematic for convergence and test size. Most of the time, the exponential distribution slows down the convergence of the intercept and coefficients, mainly due to the bias introduced by its skewness and the fact that it has a strictly positive mean. In contrast, the normal errors behave as expected across all models, with convergence in probability and in distribution being properly achieved. Moreover, the test size is correctly adjusted in most models with normal errors, allowing for appropriate rejection of the null hypothesis when it is false

The intercept estimator of $Y_t = c + \theta\epsilon_{t-1} + \epsilon_t$, where c $= 0$, $\theta = 0.5$ and ϵ_t is $i.i.d$ $N(0,1)$ or $i.i.d$ $exp(1)$.

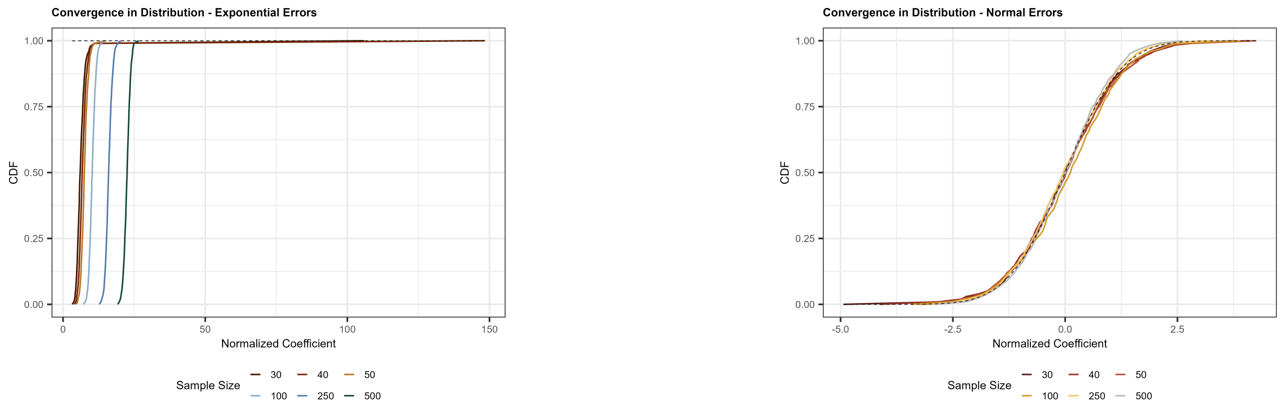
For this first simulation, we run 2000 Monte Carlo replications over a random sample of size 500. In figures 4b, 3b, and 5b, we present the results for the normal errors. We observe that the estimator performs well, and most metrics indicate asymptotic convergence of the error terms. Overall, the parameter estimate converges in probability and in distribution as the sample size increases. In the test size graph, we observe that the empirical rejection rate under the null hypothesis remains close to the 5% nominal level, indicating that the test is well-calibrated in terms of size under normality, and we are not "over rejecting" H_0.

On the other hand, when we examine the results under exponential errors in Figures 3a, 4a, and 5a, we observe a fundamental incompatibility between the zero-intercept assumption and the strictly positive support of exponential errors, which undermines convergence in probability. Moreover, there is substantial distortion in convergence in distribution, as the strict positivity and shape of the exponential distribution do not "respect" the regularity conditions required for asymptotic normality. Last, the test size becomes unreliable in this setting, resulting in a high incidence of Type I errors, as the null hypothesis H_0 is frequently rejected even when it is true.



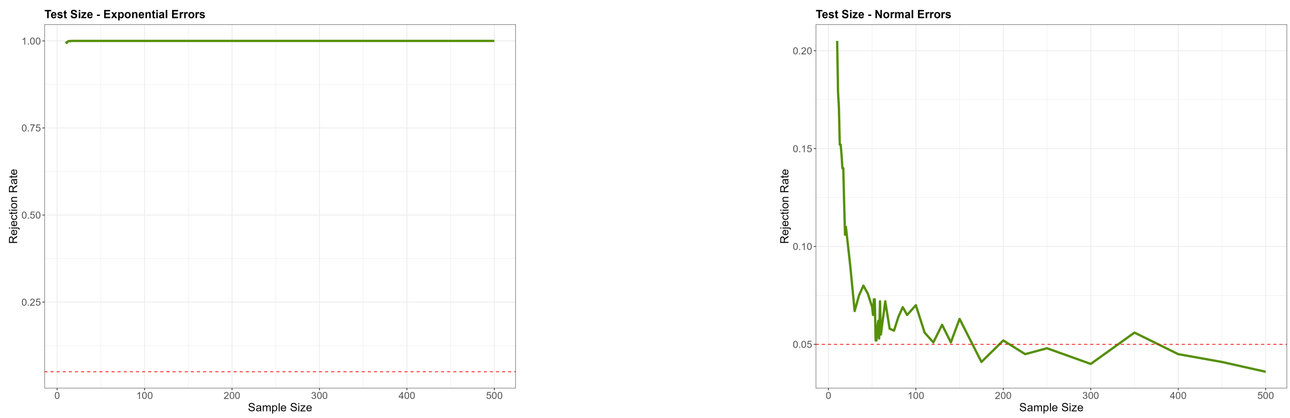
(a) Exponential errors **(b)** Normal errors

Figure 3: MA(1) Intercept probability convergence



(a) Exponential errors **(b)** Normal errors

Figure 4: MA(1) Intercept Convergence in Distribution

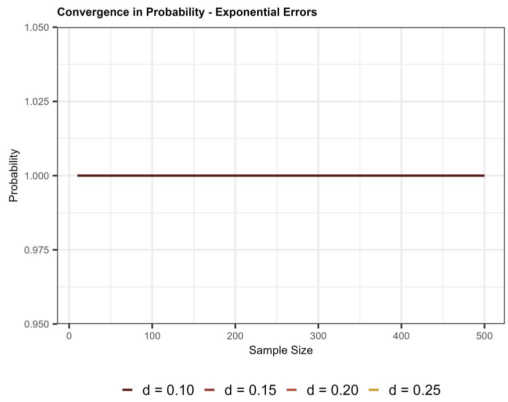


(a) Exponential errors **(b)** Normal errors

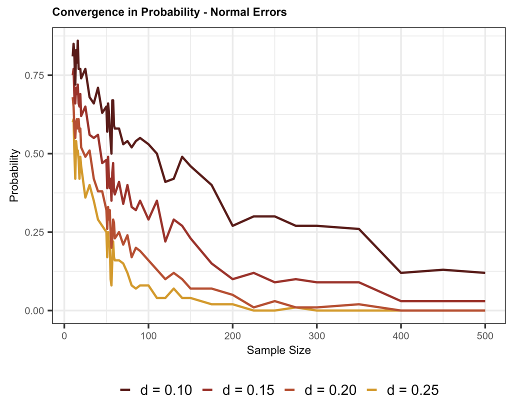
Figure 5: MA(1) Intercept Teste Size

The intercept estimator of $Y_t = c + \phi Yt - 1 + \epsilon_t$, where c = 0, $\theta = 0.5$ and ϵ_t is $i.i.d\ N(0,1)$ or $i.i.d\ exp(1)$.

From this simulation and on I ran 200 Monte Carlo simulations for these tests due to the limited performance of my computer. Nevertheless, as observed in the figures, the model performs satisfactorily under normal errors, although it takes a little bit longer to converge in both probability and distribution relative to the previous model. Moreover, the test size under normal errors is relatively accurate, although some noticeable deviations appear as the sample size increases. For the exponential errors, the same issues observed in the MA(1) model persist. The distribution is not appropriate for modeling the error terms, and all performance metrics deteriorate due to the positive skewness and shape of the distribution.

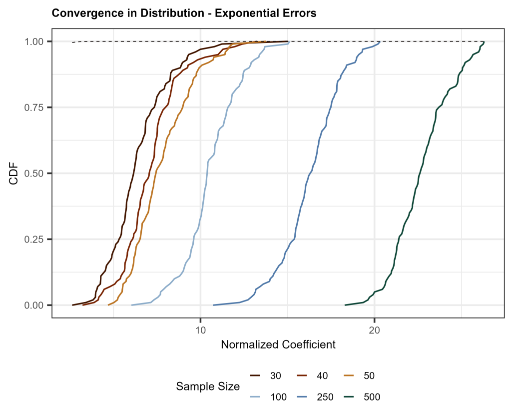


(a) Exponential errors

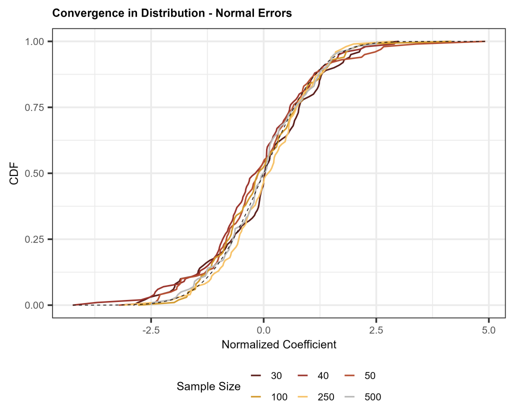


(b) Normal errors

Figure 6: AR(1) Intercept probability convergence

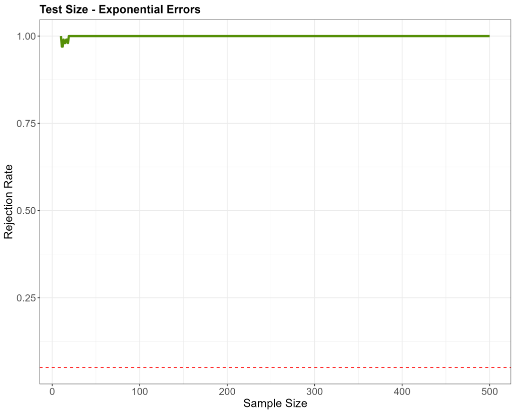


(a) Exponential errors

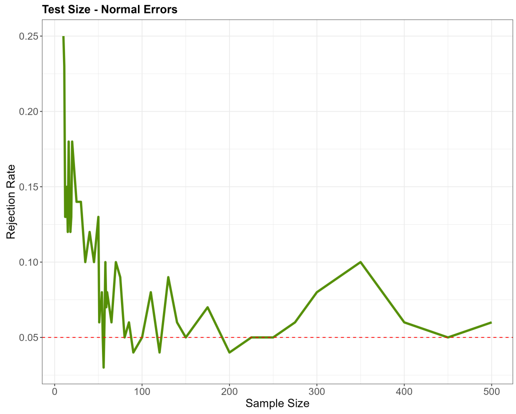


(b) Normal errors

Figure 7: AR(1) Intercept Convergence in Distribution



(a) Exponential errors

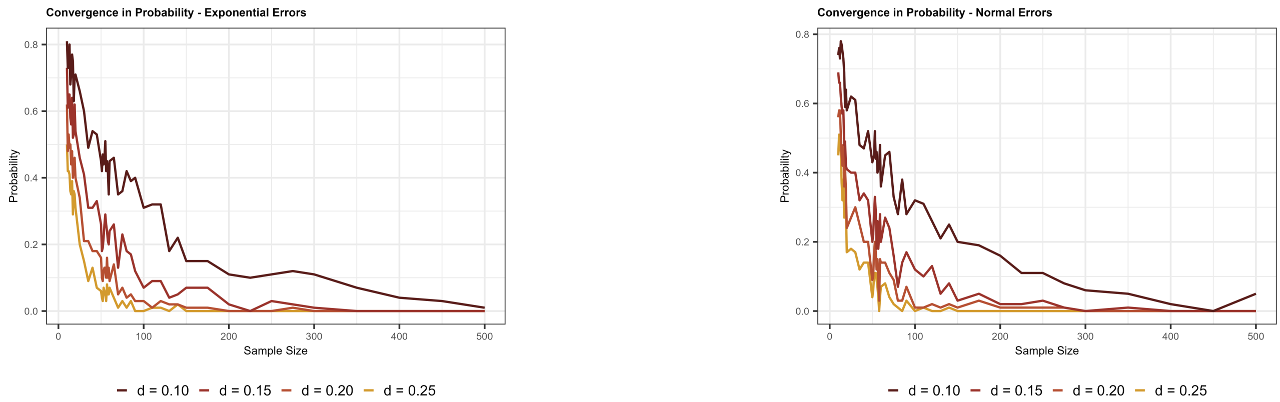


(b) Normal errors

Figure 8: AR(1) Intercept Teste Size

The estimator for the first autorregressive coefficient of $Y_t = c + \phi Yt - 1 + \epsilon_t$, where c = 0, θ = 0.5 and ϵ_t is $i.i.d$ $N(0,1)$ or $i.i.d$ $exp(1)$.

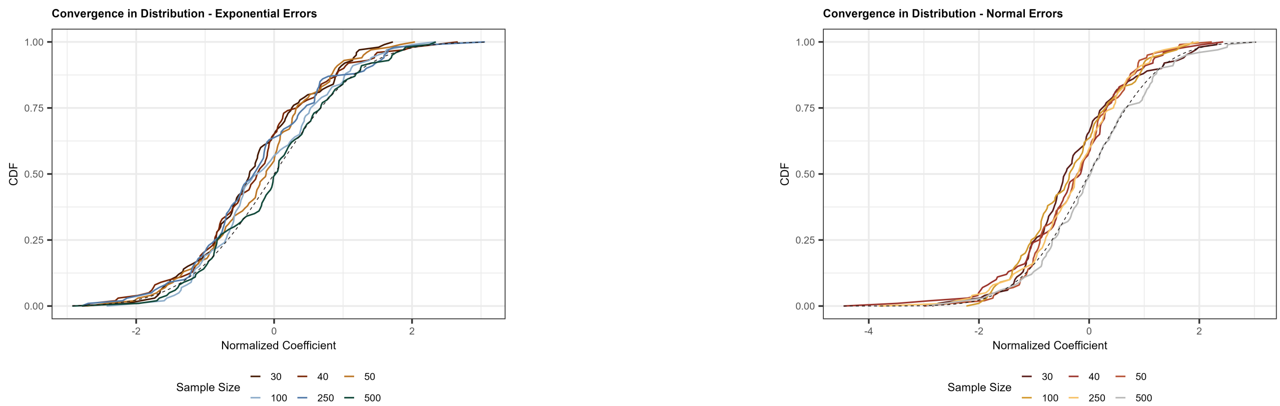
Under normal errors, as shown in the figures, the estimator for the coefficient converges slightly faster in probability than under exponential errors as the sample size increases. A similar pattern is observed in the convergence in distribution, where the estimator under normal errors appears more consistent than under exponential errors. Lastly, the test sizes seem to behave appropriately for both normal and exponential errors. I believe that if the number of simulations were increased, the green line would likely converge closely to the 5% level, indicating that we are correctly rejecting H_0 when we should.



(a) Exponential errors

(b) Normal errors

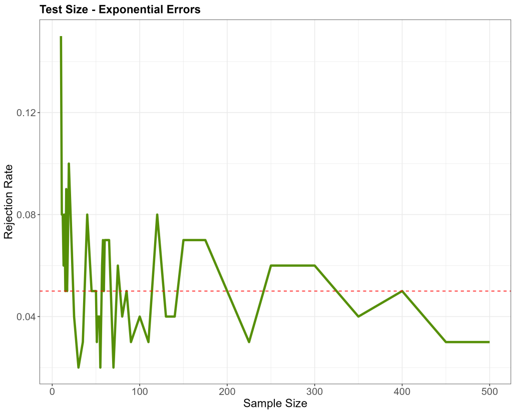
Figure 9: AR(1) Coefficient – Convergence in Probability



(a) Exponential errors

(b) Normal errors

Figure 10: AR(1) Coefficient – Convergence in Distribution



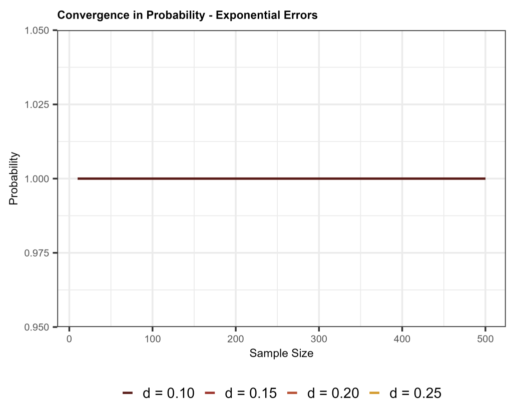
(a) Exponential errors

(b) Normal errors

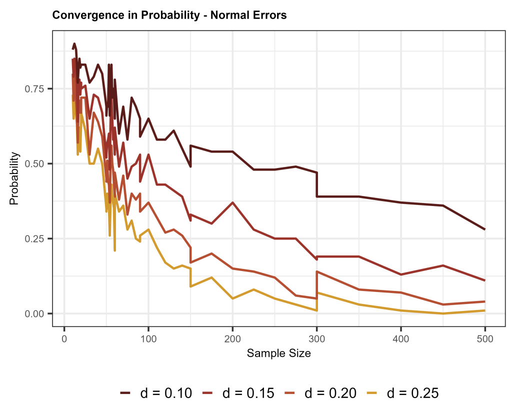
Figure 11: AR(1) Coefficient – Test Size

The intercept estimator of $Y_t = c + \phi Y_{t-1} + \theta\epsilon_{t-1} + \epsilon_t$, where c $= 0$, $\theta = 0.5$ and ϵ_t is $i.i.d$ $N(0,1)$ or $i.i.d$ $exp(1)$.

Again, we observe a more consistent pattern of convergence in probability for the coefficient under normal errors, while the results under exponential errors remain problematic due to the specific characteristics of the exponential distribution. Similarly, we find good performance in terms of convergence in distribution under normal errors, but poor convergence under exponential errors, once again due to the skewness and positivity of the error terms. Therefore, the test size under normal errors gives us confidence in correctly rejecting the null hypothesis. In contrast, the test size under exponential errors is highly unreliable, as the null hypothesis is almost always rejected.

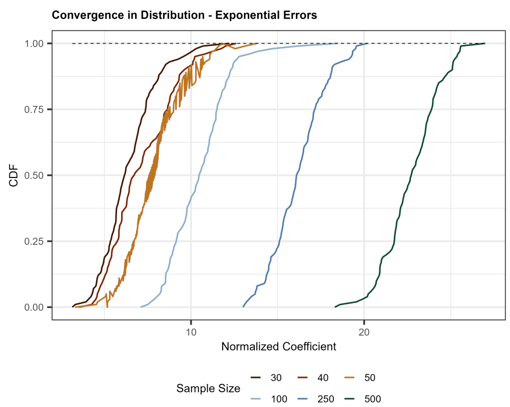


(a) Exponential errors

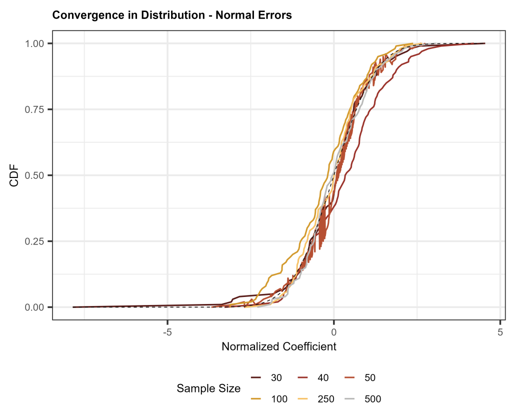


(b) Normal errors

Figure 12: ARMA(1,1) Intercept – Convergence in Probability

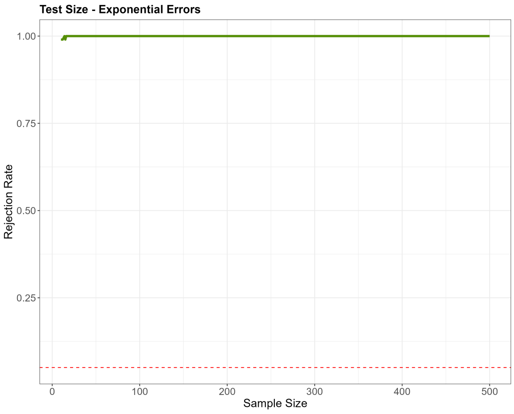


(a) Exponential errors

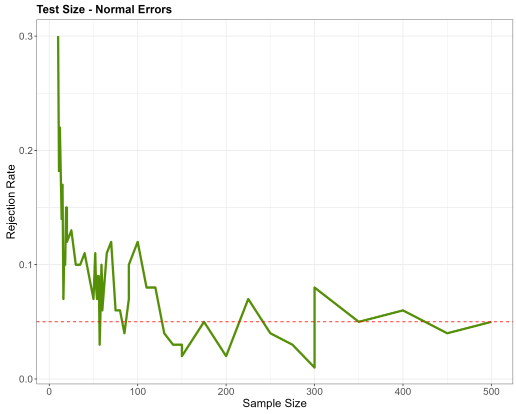


(b) Normal errors

Figure 13: ARMA(1,1) Intercept – Convergence in Distribution



(a) Exponential errors

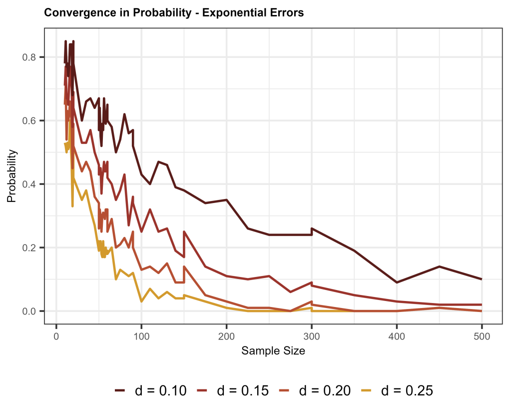


(b) Normal errors

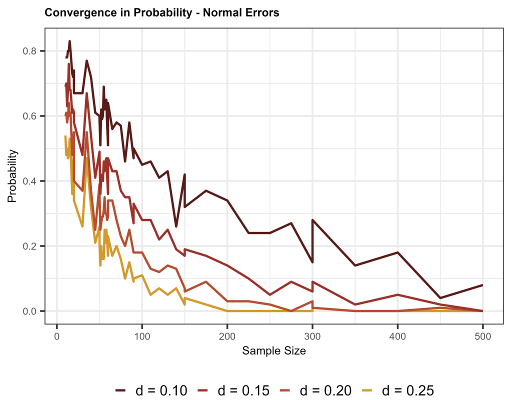
Figure 14: ARMA(1,1) Intercept – Test Size

The autoregressive coefficient of $Y_t = c + \phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t$, where c = 0, $\theta = 0.5$ and ϵ_t is $i.i.d$ $N(0, 1)$ or $i.i.d$ $exp(1)$.

In this setup, the autoregressive coefficient performs relatively similar under both normal and exponential errors in terms of convergence in probability. The distribuiton of the exponential erros seems to be more noisy and complex than the normal erros, exhibiting a slower pattern of convergence due to the limitations of the distribution itself. About the test sizes, we observe that as the sample size increases, both estimators tend to perform similar but still a little far from the optimal, but we can say that they still have a satisfactory rate of rejection of the null.

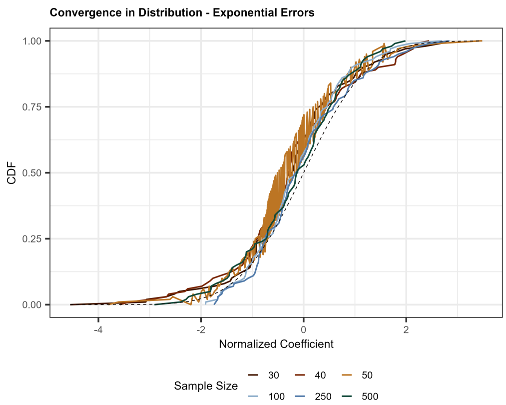


(a) Exponential errors

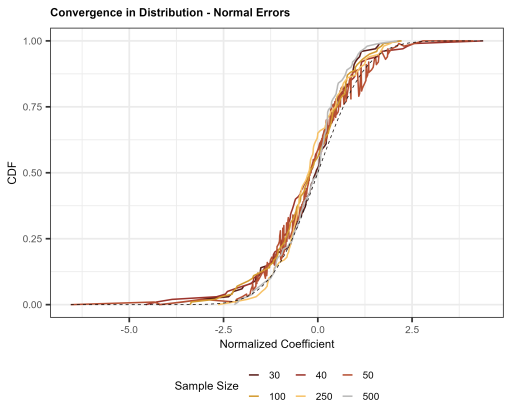


(b) Normal errors

Figure 15: ARMA(1,1) AR Coefficient – Convergence in Probability

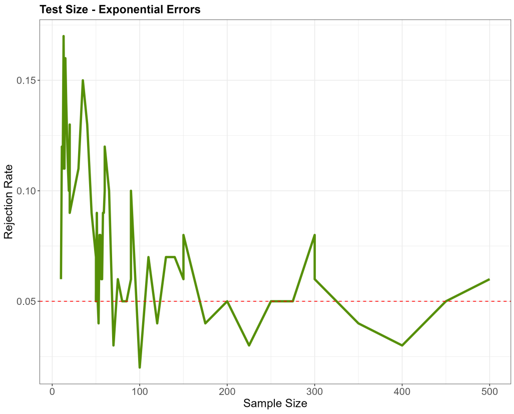


(a) Exponential errors

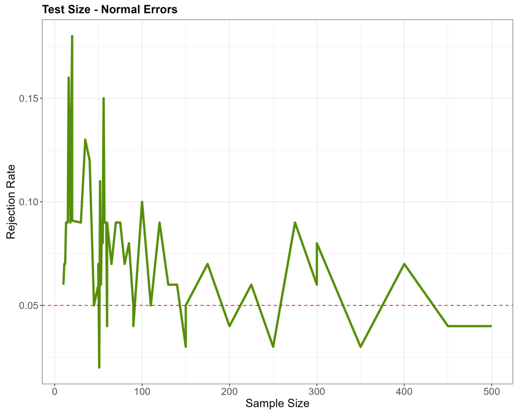


(b) Normal errors

Figure 16: ARMA(1,1) AR Coefficient – Convergence in Distribution



(a) Exponential errors

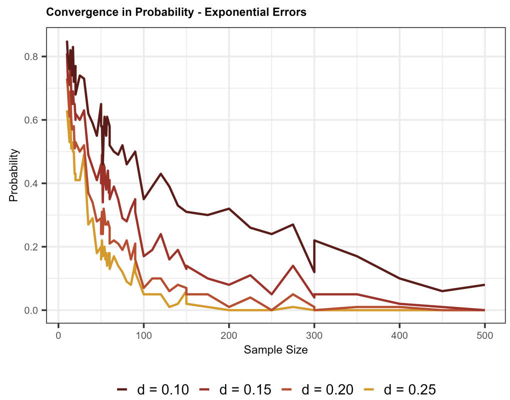


(b) Normal errors

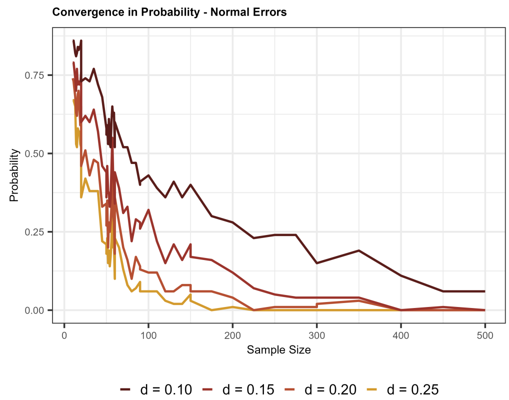
Figure 17: ARMA(1,1) AR Coefficient – Test Size

The moving average coefficient of $Y_t = c + \phi Y_{t-1} + \theta\epsilon_{t-1} + \epsilon_t$, where c = 0, $\theta = 0.5$ and ϵ_t is $i.i.d\ N(0,1)$ or $i.i.d\ exp(1)$.

It is easy to see that the exponential errors and the normal errors are quite similar in terms of convergence in probability, but in distribution the exponential erros take longer to converge, however we can see a more robust performance than the intercept estimator. We can also verify differences in the test size, where the exponential errors seem to perform poorly in terms of the power to reject the null hypothesis, and the normal erros show a more appropriate behavior to reject the null.

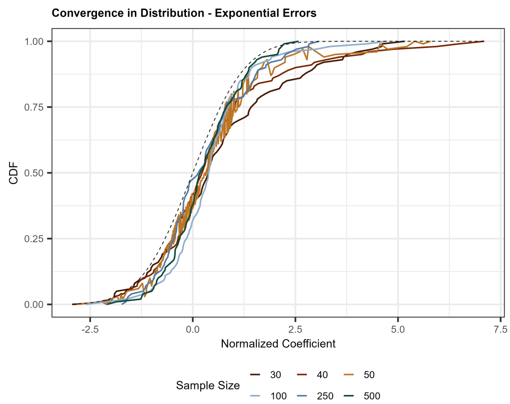


(a) Exponential errors

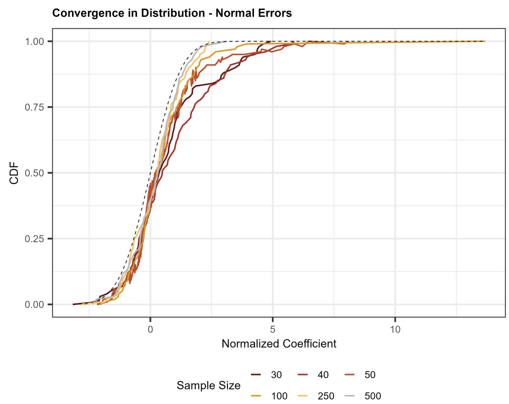


(b) Normal errors

Figure 18: ARMA(1,1) MA Coefficient – Convergence in Probability

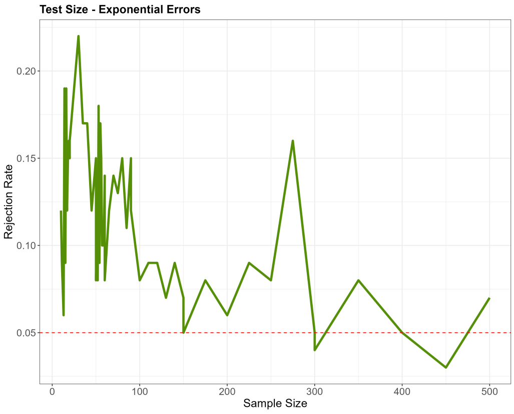


(a) Exponential errors

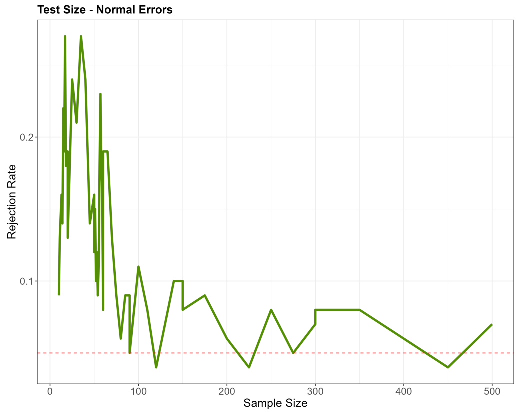


(b) Normal errors

Figure 19: ARMA(1,1) MA Coefficient – Convergence in Distribution



(a) Exponential errors



(b) Normal errors

Figure 20: ARMA(1,1) MA Coefficient – Test Size

Moreover, answer the following question (20 points): Based on the simulations above, do you feel comfortable with a sample size of 500 periods for any stochastic process? Explain your answer.

Honestly, I don't feel entirely confident in saying that 500 periods would be sufficient for modeling any time series. Even though it is a reasonably large sample, it might not be enough for processes with more complex or volatile dynamics. As we have observed in some of our models, we reach convergence in probability for certain values of δ, but not all. Moreover, processes with long memory or those that exhibit non-stationarity may require much larger sample sizes to achieve reliable statistical inference and to produce good-quality test sizes—or at least better than some of the estimates we observed here. That said, 500 observations is not a small sample, and some models do perform well. For example, the AR coefficient and intercept typically do not exhibit long memory, so we can rely on relatively short-period estimates in those cases with more confidence.

3 Question 3

Item a) Give one example of a weakly stationary stochastic process.

We can start with the definition of Hamilton (2020) to show that if neither the mean $\mu \in \mathbb{R}$, nor the autocovariances γ_j, depend on the date t, then the process for Y_t is said to be **covariance-stationary** or **weakly stationary** if:

$$\begin{gathered}\mathbb{E}(Y_t) = \mu \quad \text{for all } t,\\ \mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j \quad \text{for all } t \text{ and any } j.\\ Var(Y_t) = \gamma_0 < \infty\end{gathered}$$

For example, consider the Gaussian Process $Y_t = \mu + \epsilon_t$. Where μ is a constant, $\epsilon_t \sim WN(0, \sigma^2)$ is a white noise with $\mathbb{E}[\epsilon_t] = 0$, $Var[\epsilon_t] = \sigma^2$, and $Cov(\epsilon_t, \epsilon_s) = 0$ for $t \neq s$.

To show that the mean is constant and equal to μ for all t we make,

$$\begin{aligned}\mathbb{E}(Y_t) &= \mathbb{E}(\mu + \varepsilon_t)\\ &= \mathbb{E}(\mu) + \mathbb{E}(\varepsilon_t) \quad \text{(linearity of expectation)}\\ &= \mu + 0 \quad (\text{since } \mathbb{E}(\varepsilon_t) = 0)\\ &= \mu\end{aligned}$$

We compute the variance to show that it is constant and equal to σ^2 for all t

$$
\begin{aligned}
\text{Var}(Y_t) &= \text{Var}(\mu + \varepsilon_t) \\
&= \text{Var}(\varepsilon_t) \quad \text{(since } \mu \text{ is a constant)} \\
&= \sigma^2 \quad \text{(by definition of white noise)}
\end{aligned}
$$

Last, we show the covariance depends only on the lag j, not on time t. Let $j \in \mathbb{Z}$. We compute the covariance between Y_t and Y_{t-j}:

$$
\begin{aligned}
\text{Cov}(Y_t, Y_{t-j}) &= \text{Cov}(\mu + \varepsilon_t, \mu + \varepsilon_{t-j}) \\
&= \text{Cov}(\varepsilon_t, \varepsilon_{t-j}) \quad \text{(since } \mu \text{ is a constant)}
\end{aligned}
$$

Thus, by the definition of white noise:

$$
\text{Cov}(\varepsilon_t, \varepsilon_{t-j}) = \begin{cases} \sigma^2, & \text{if } j = 0 \\ 0, & \text{if } j \neq 0 \end{cases}
$$

Item b) Give one example of a stochastic process that is not stationary.

Accordingly to Hamilton (2020), one example of a non stationary process is a Random Walk. Just to stablish some definitions, consider the same stochastic process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ from previous item such that: (i) it has zero mean, i.e., $\mathbb{E}[\varepsilon_t] = 0$; (ii) it has constant variance, i.e., $\text{Var}(\varepsilon_t) = \sigma^2$ for all t, where $\sigma^2 < \infty$; and (iii) it has zero autocovariance at all nonzero lags, i.e., $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ for all $t \neq s$.
We can formaly define a Random Walk as,

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

It is easy to show that we can recursively iterate all the realizations to get,

$$
\begin{aligned}
Y_t &= Y_{t-1} + \varepsilon_t = Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t = Y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \\
&= \cdots = Y_0 + \sum_{s=1}^{t} \varepsilon_s
\end{aligned}
$$

Taking the expectation we can show that the mean is indeed constant an equal to μ:

$$
\mathbb{E}[Y_t] = \mathbb{E}\left[Y_0 + \sum_{s=1}^{t} \varepsilon_s\right] = \mathbb{E}[Y_0] + \sum_{s=1}^{t} \underbrace{\mathbb{E}[\varepsilon_s]}_{\text{zero mean}} = \mu
$$

We know the mean is constant for all realizations, so $\mathbb{E}[Y_0] = \mu$. And using the fact that:

$$
Y_t - \mu = (Y_0 - \mu) + \sum_{s=1}^{t} \varepsilon_s,
$$

we can multiply both sides by $Y_t - \mu$ taking the expectations to define

$$\mathbb{E}[(Y_t - \mu)^2] = \mathbb{E}[(Y_t - \mu)(Y_0 - \mu)] + \mathbb{E}\left[(Y_t - \mu)\sum_{s=1}^{t} \varepsilon_s\right].$$

Expanding the terms:

$$\mathbb{E}[(Y_t - \mu)^2] = \mathbb{E}\left[\underbrace{(Y_0 - \mu)^2}_{\text{Constant}=0} + \underbrace{(Y_0 - \mu)\sum_{s=1}^{t} \varepsilon_s}_{\sum_{s=1}^{t}\mathbb{E}[\varepsilon_s]=0}\right] + \mathbb{E}\left[\underbrace{(Y_0 - \mu)\sum_{s=1}^{t} \varepsilon_s}_{\sum_{s=1}^{t}\mathbb{E}[\varepsilon_s]=0} + \underbrace{\left(\sum_{s=1}^{t} \varepsilon_s\right)^2}_{\sum_{s=1}^{t}\mathbb{E}[\varepsilon_s^2]=t\sigma^2}\right]$$
$$= \mathbb{E}\left[\left(\sum_{s=1}^{t} \varepsilon_s\right)^2\right] = t\sigma^2$$

For the $Cov(\epsilon_t, \epsilon_s)$ we have

$$(Y_t - \mu)(Y_{t-j} - \mu) = \left[(Y_0 - \mu) + \sum_{s=1}^{t} \varepsilon_s\right]\left[(Y_0 - \mu) + \sum_{k=1}^{t-j} \varepsilon_k\right].$$

Multiplying the terms:

$$= \underbrace{(Y_0 - \mu)^2}_{\text{Constant } = 0} + \underbrace{(Y_0 - \mu)\sum_{k=1}^{t-j} \varepsilon_k}_{\sum_{s=1}^{t}\mathbb{E}[\varepsilon_k]=0} + \underbrace{(Y_0 - \mu)\sum_{s=1}^{t} \varepsilon_s}_{\mathbb{E}[\varepsilon_s]=0} + \sum_{s=1}^{t} \varepsilon_s \sum_{k=1}^{t-j} \varepsilon_k$$

Then we get:

$$\mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)] = \mathbb{E}\left[\sum_{s=1}^{t} \varepsilon_s \sum_{k=1}^{t-j} \varepsilon_k\right] = \sum_{s=1}^{t-j} \mathbb{E}[\varepsilon_s^2] = (t-j)\sigma^2.$$

Therefore, it is quite evident that we do not have a result independent of lags. Instead, we observe that the covariance depends on time or other lags, indicating that the process is indeed non-stationary.

4 Question 4

Let Y_t denote a weakly stationary process and γ_j denote its autocovariance. Show that $\gamma_j = \gamma_j$.

Let Y_t be a weakly stationary stochastic process, and let γ_j denote its autocovariance at lag j, defined as

$$\gamma_j = \mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)].$$

We can also define the covariance for all others $-j$ lag realizations such that:

$$\gamma_{-j} = \mathbb{E}[(Y_t - \mu)(Y_{t+j} - \mu)].$$

But observe that:

$$\gamma_{-j} = \mathbb{E}[(Y_t - \mu)(Y_{t+j} - \mu)] = \mathbb{E}[(Y_{t+j} - \mu)(Y_t - \mu)],$$

where we used the commutativity of multiplication and the linearity of expectation.

Since Y_t is weakly stationary for all t, the joint distribution of (Y_t, Y_{t+j}) is the same as that of (Y_{t-j}, Y_t), so:

$$\mathbb{E}[(Y_{t+j} - \mu)(Y_t - \mu)] = \mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j.$$

Thus,

$$\gamma_{-j} = \gamma_j.$$

Statement Regarding Artificial Intelligence Usage and Help of Classmates

This project made use of ChatGPT 4.0 for coding assistance and language correction. I was also inspired by Vinicius De Almeida's code for question 2.

References

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