MFE 230E Problem Set 3 - Solutions

Spring 2020

Note: This document only presents answers. Cf. the accompanying Jupyter Notebook for details and the corresponding code. Please email mfe230e@gmail.com if there are any corrections you would like to make to these solutions.

Question 1

(a) We estimate the process by OLS using an AR(1):

$$x_{t+1} = \phi x_t + e_{t+1}$$

Because the true process is an ARMA(1,1), the error e_{t+1} will not be White noise but instead will include MA terms ($\epsilon_{t+1} + 0.7\epsilon_t$). This leads to an endogeneity issue:

$$\mathbb{E}[x_t e_{t+1}] = \mathbb{E}[x_t(\epsilon_{t+1} + 0.7\epsilon_t)] = 0.7\mathbb{E}[x_t \epsilon_t] \neq 0$$

Therefore, the OLS estimator is inconsistent. This can be clearly seen on Figure 1 below, which shows the distribution of the OLS estimates over the 10,000 simulations. $\hat{\phi}_{OLS}$ is centered close to a mean of 0.894, and is far from the true value of $\phi = 0.8$. (Note that the inconsistency comes from the fact that there are MA terms in the error <u>and</u> the fact that we include the lagged dependent variable as regressor.)

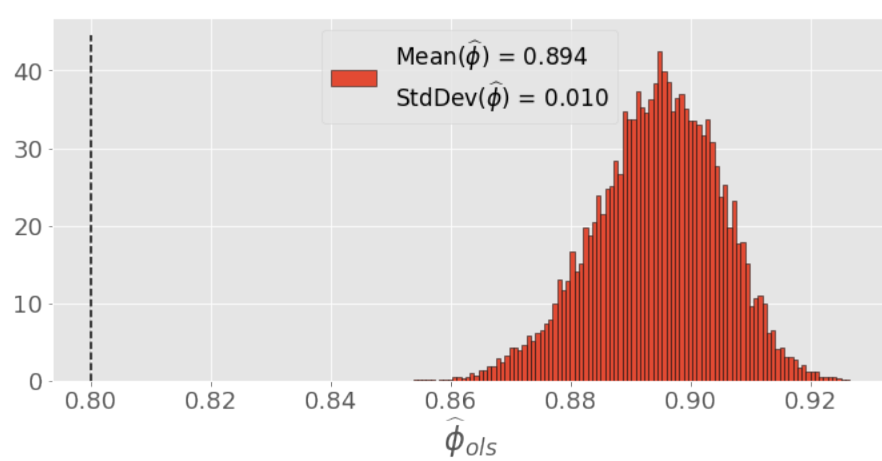

Figure 1: OLS estimation under endogeneity

(b) To confirm that there are some remaining MA terms, we can plot the autocorrelation of the errors. You can do this in a variety of ways, for instance, you can present the mean and standard deviation of the ACF at one or several lags across the 10,000 simulations. Figures 2 and 3 below report the distribution of the ACF and PACFs for lag 1 to 15 across the 10,000 simulations. Those clearly show that the residuals are very far from being White Noise and instead present correlation patterns for several lags.

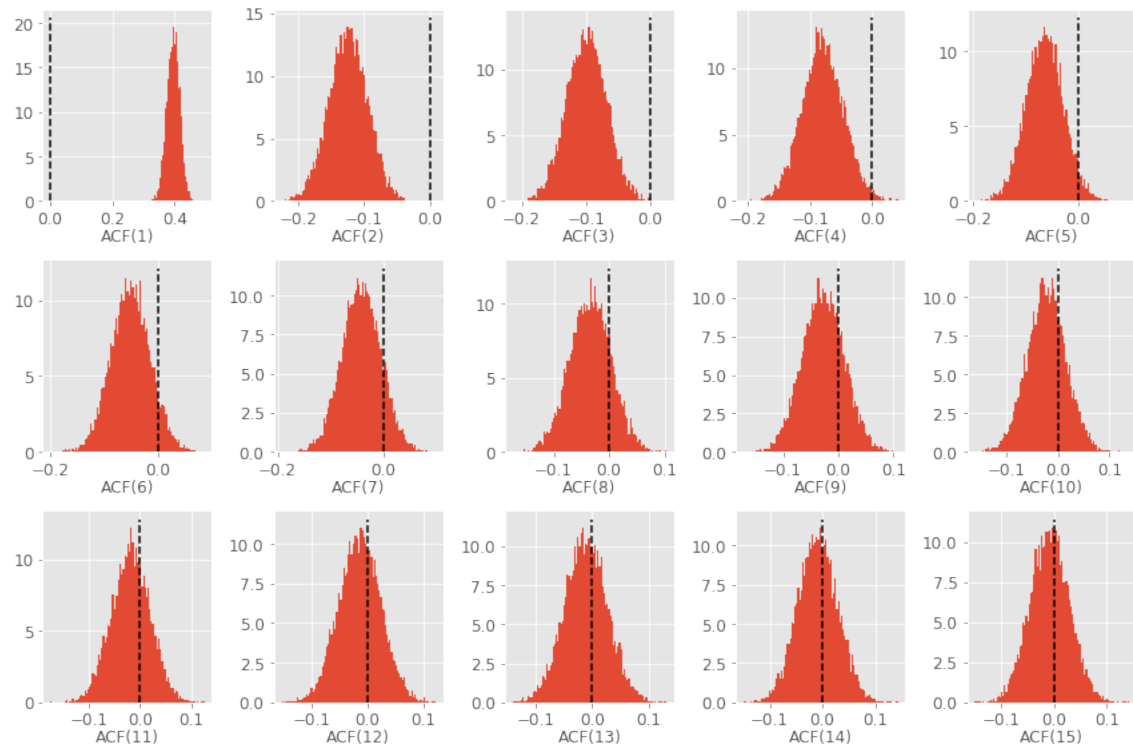

Figure 2: ACFs of the OLS errors

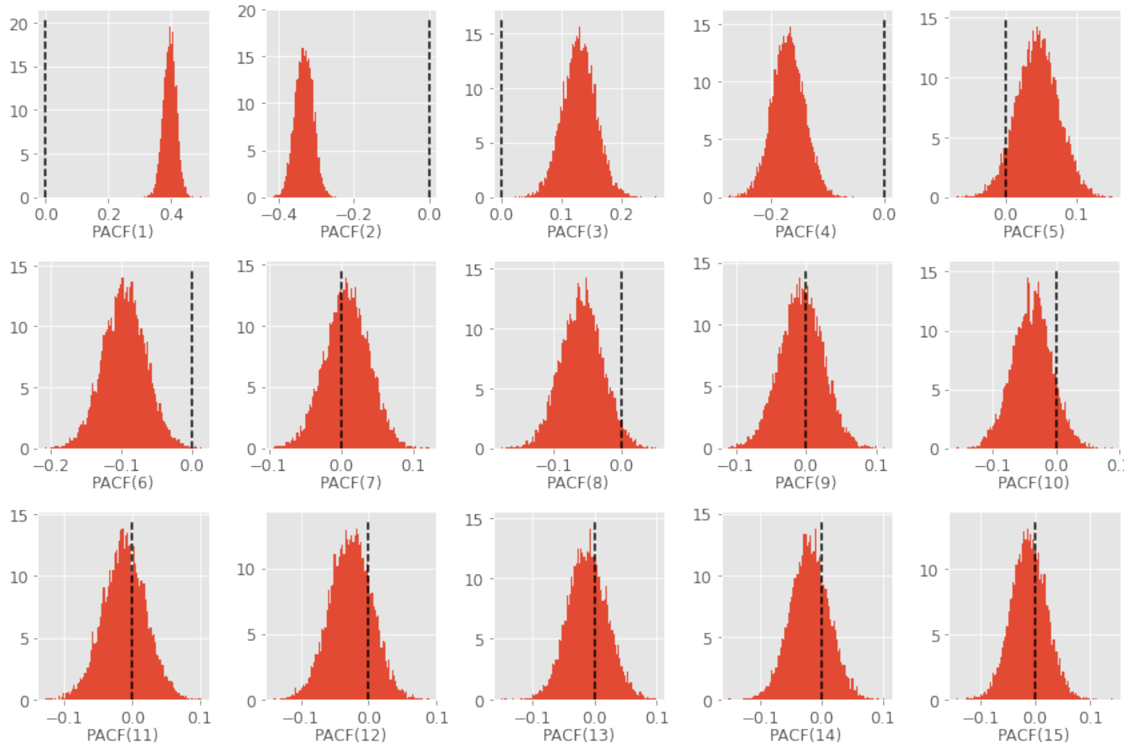

Figure 3: PACFs of the OLS errors

(c) To deal with the issues above, we use the lagged value x_{t-1} as an instrumental variable for x_t in the AR(1) regression above. x_{t-1} is orthogonal to the error $\mathbb{E}[x_{t-1}e_{t+1}] = 0$, so the IV is exogenous (or more formally satisfies the moment condition), but is correlated to x_{t-1} so that the IV is relevant. Thus, the IV estimator is consistent. This is clearly visible on Figure 4 below, in which the $\hat{\phi}_{IV}$ estimates center around the true value of $\phi = 0.8$ ($\mathbb{E}[\hat{\phi}_{IV}] \approx 0.796$).

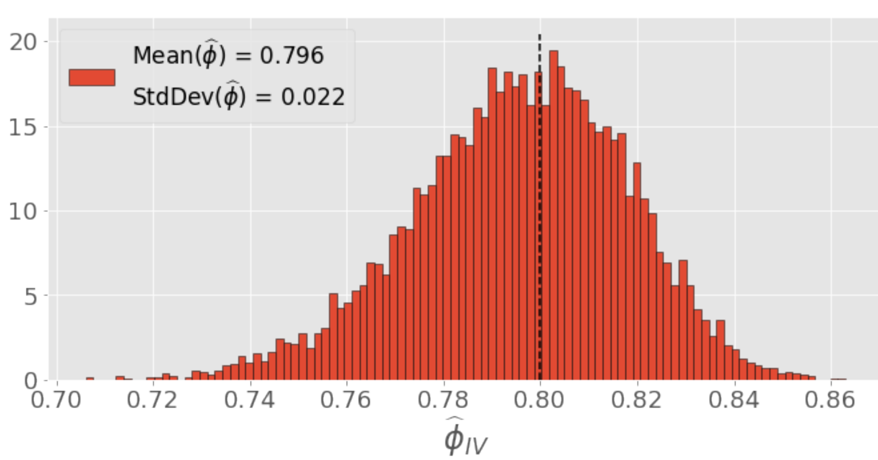

Figure 4: IV estimation

Question 2

(a) First, let us rewrite the process so that we get its AR coefficients. $\lambda_1 = 0.95, \lambda_2 = 0.9, \lambda_3 = 0.8$ imply:

$$(1 - \lambda_1 L)(1 - \lambda_2 L)(1 - \lambda_3 L)x_{t+1} = u_{t+1}$$

$$(1 - 2.65L + 2.335L^2 - 0.684L^3)x_{t+1} = u_{t+1}$$

$$\Rightarrow x_{t+1} = 2.65x_t - 2.335x_{t-1} + 0.684x_{t-2} + u_{t+1}$$

We can now simulate the AR(3) process. Below are 5 sample paths, to check the that the simulation worked correctly.

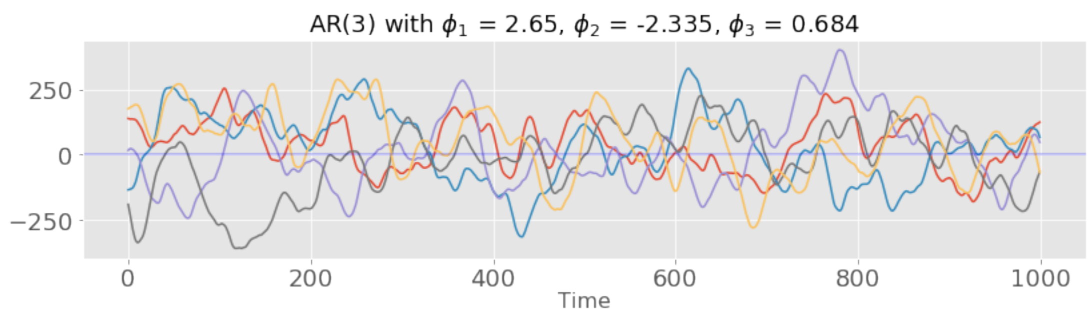

Figure 5: 5 sample paths among the 10,000 simulations

We estimate an AR(1) process:

$$x_{t+1} = \phi x_t + e_{t+1}$$

The histogram of the OLS estimates is shown in Figure 6.

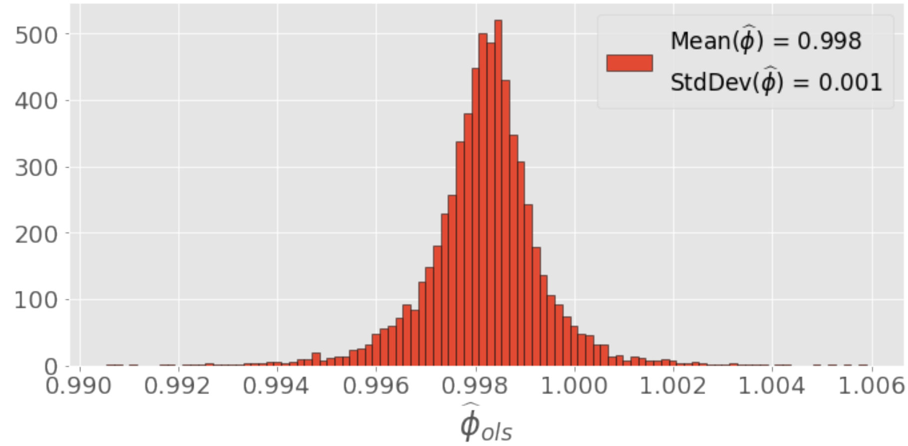

Figure 6: OLS estimation of an AR(1)

As expected, the OLS estimator is inconsistent, and is concentrated around $\mathbb{E}[\hat{\phi}_1] \approx 0.998$, which is quite different from the true value $\phi_1 = 2.65$. This is so because the error term e_{t+1}, which includes x_{t-1} and x_{t-2}, is correlated with the regressor x_t:

$$\mathbb{E}[x_t e_{t+1}] = \mathbb{E}[x_t(x_{t-1} + x_{t-2} + u_{t+1})] \neq 0$$

The distribution of ACFs and PACFs coefficients in Figures 7 and 8 confirm that the error term is not a White Noise. Because PACFs seem different from zero for lags $j = 1, 2$, and zero afterwards, this is suggestive of a remaining AR(2) in the error, as expected.

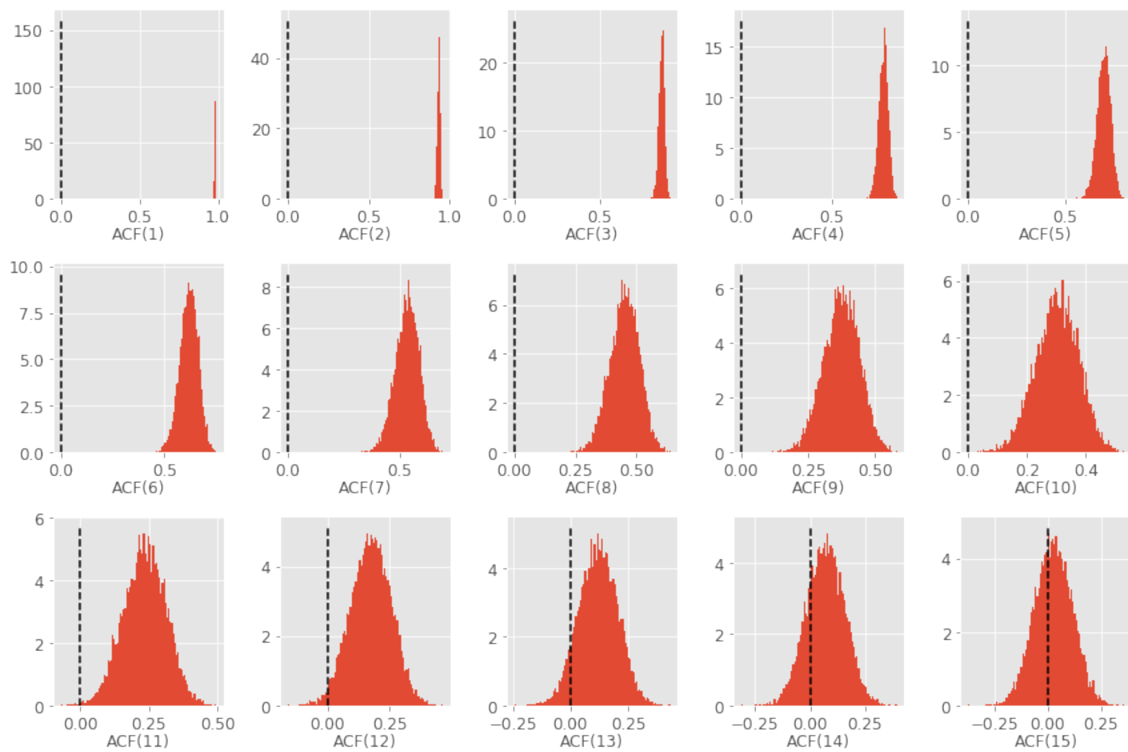

Figure 7: ACFs of the OLS errors for an AR(1)

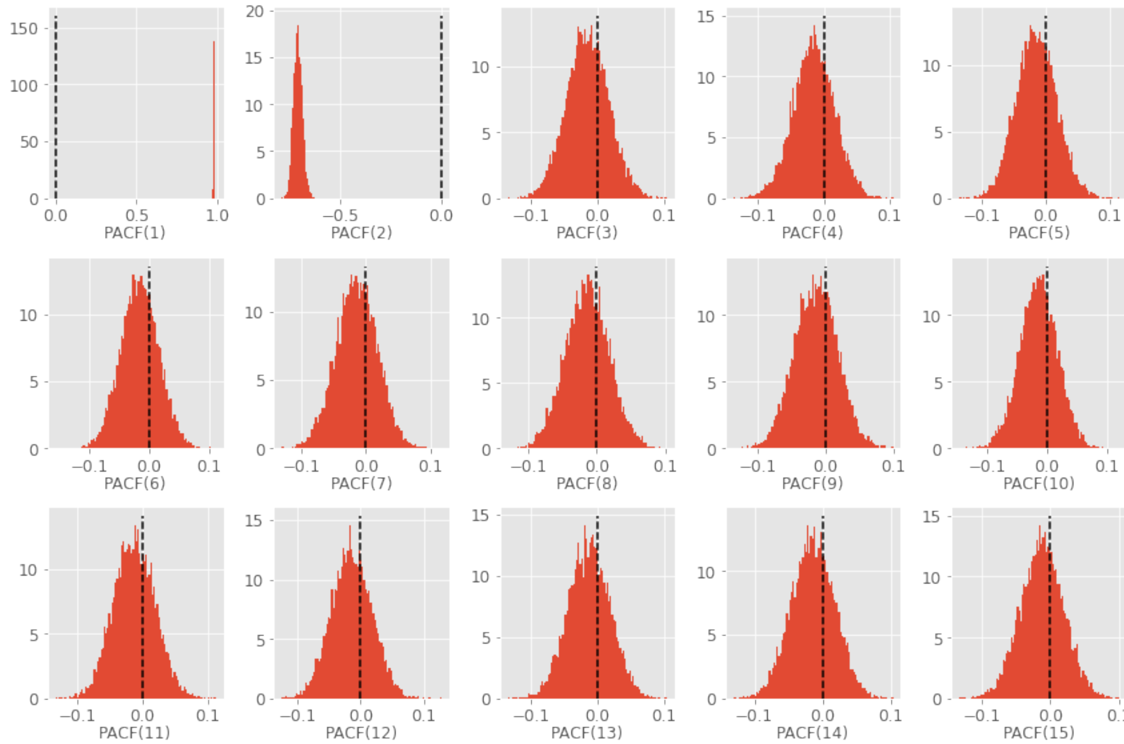

Figure 8: PACFs of the OLS errors for an AR(1)

(b) Figures 9, 10 and 11 show the histogram of the coefficients, the ACFs, and the PACFs when we estimate an AR(2) process instead. Even though OLS estimates get closer to the truth ($\mathbb{E}[\hat{\phi}_1] \approx 1.978$ vs. $\phi_1 = 2.65$), they are still quite far off because the OLS estimator is still not consistent due to the remaining correlation between e_{t+1} (which still contains x_{t-1}, x_{t-2}) and the regressors. Even though ACFs and PACFs remain somewhat different from zero at various lags, the PACF at lag $j = 1$ is the one that stands out as quite large. This is suggestive of a remaining AR(1) in the error, as expected.

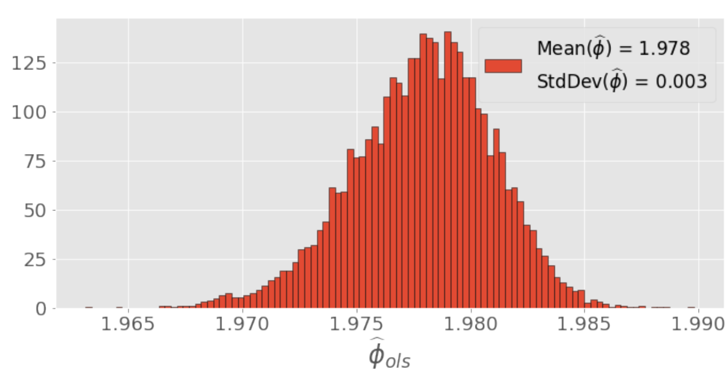

Figure 9: OLS estimation of an AR(2)

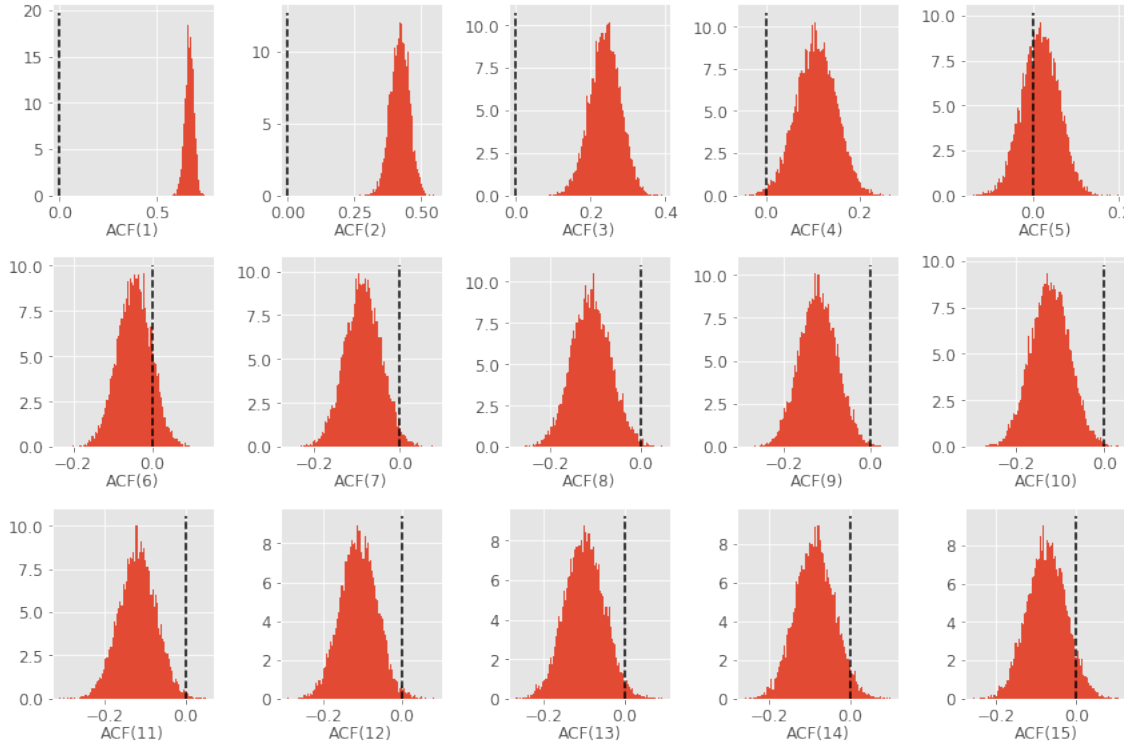

Figure 10: ACFs of the OLS errors for an AR(2)

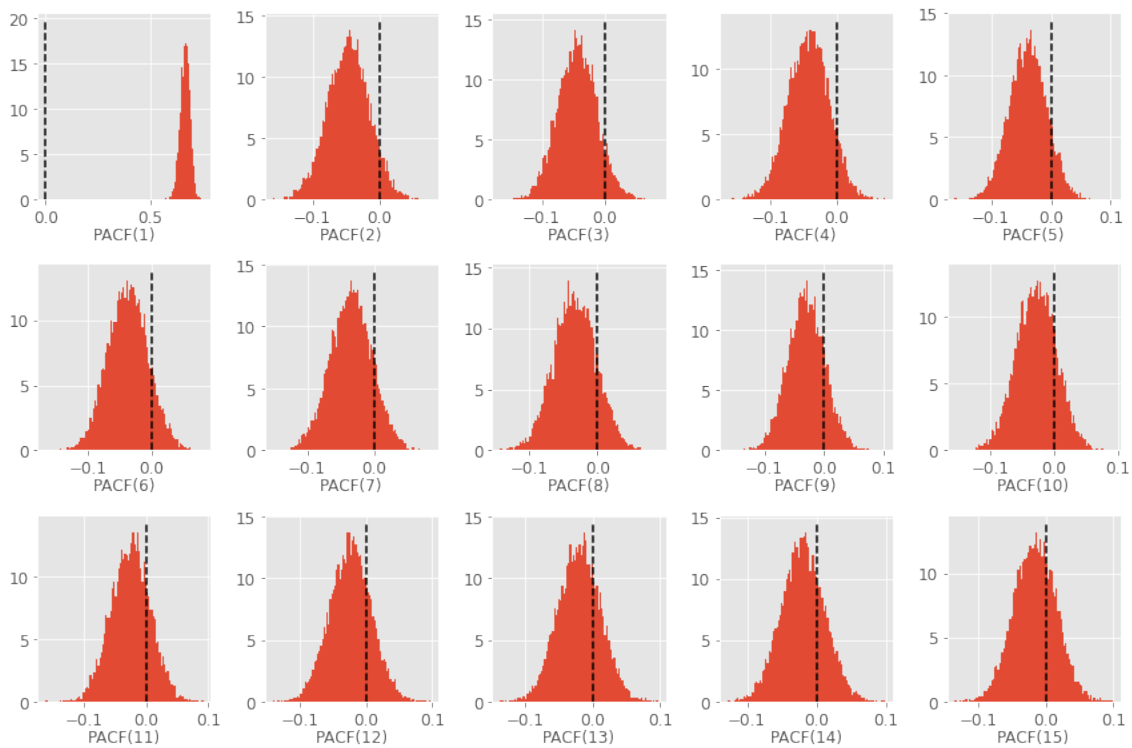

Figure 11: PACFs of the OLS errors for an AR(2)

Figures 12, 13 and 14 show the histogram of the coefficients, the ACFs, and the PACFs when we estimate an AR(3) process. Because we are estimate the appropriate process without omitted variables, the OLS estimator is consistent. This can be observed in the histogram, with the OLS estimates concentrating around $\mathbb{E}[\hat{\phi}_1] \approx 2.645$, which is close to the true value $\phi_1 = 2.65$. The ACFs and PACFs confirm that that there is no remaining pattern in the residuals, and that those are therefore White Noise, as desired.

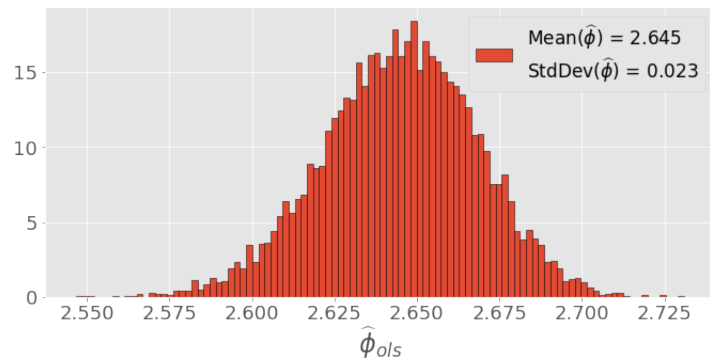

Figure 12: OLS estimation of an AR(3)

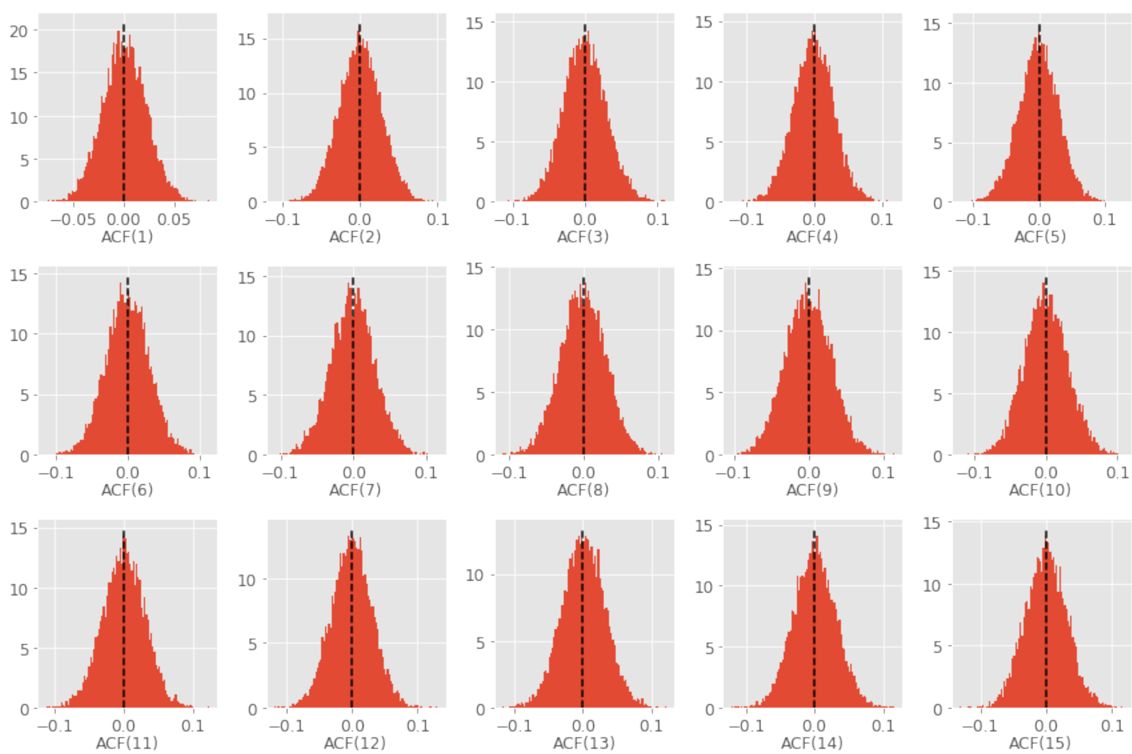

Figure 13: ACFs of the OLS errors for an AR(3)

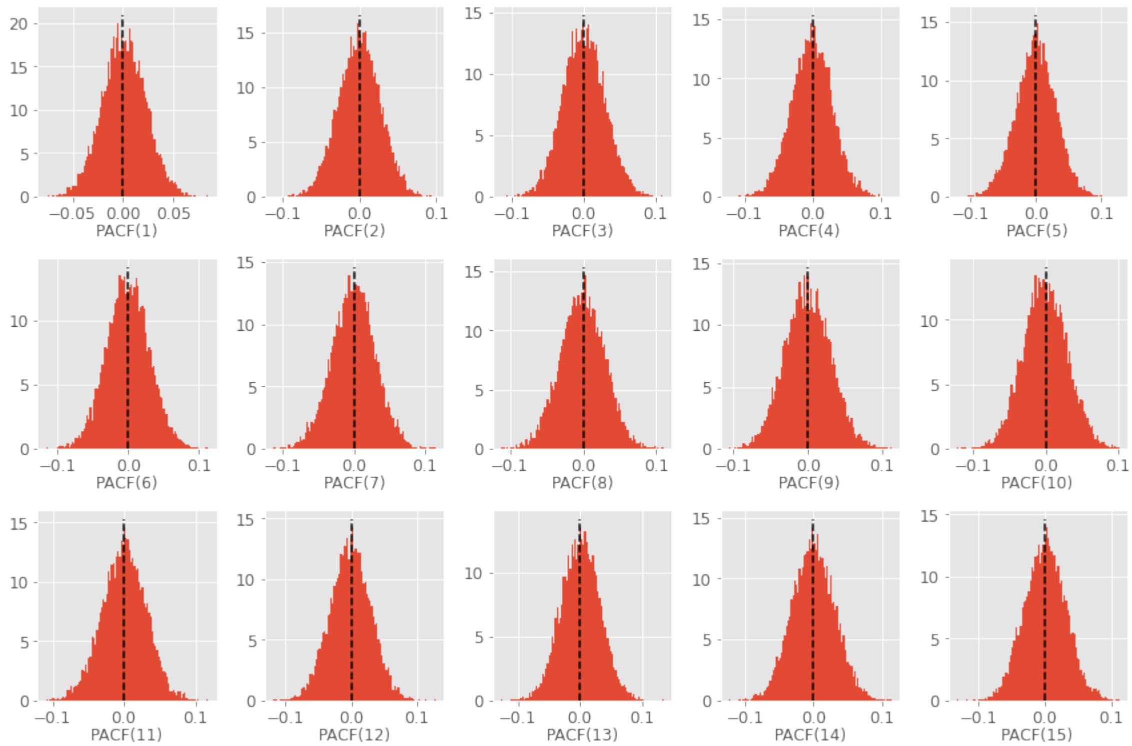

Figure 14: PACFs of the OLS errors for an AR(3)

Finally, Figures 12, 13 and 14 show the histogram of the coefficients, the ACFs, and the PACFs when we estimate an AR(4) process. There are no longer any endogeneity issue, so the OLS estimator is consistent. This is shown in the histogram with estimates concentrating around $\mathbb{E}[\hat{\phi}_1] \approx 2.646$, which is close to the true value $\phi_1 = 2.65$. The ACFs and PACFs confirm that there are no remaining patterns in the residuals, and that they are therefore White Noise, as desired. Compared to the AR(3) estimation, we include a variable, x_{t-3} that is not relevant. As expected, this leads to (slightly) larger standard errors: $std(\hat{\phi}_1 \approx 0.032$ vs. $std(\hat{\phi}_1 \approx 0.023$ when we estimate the AR(3).

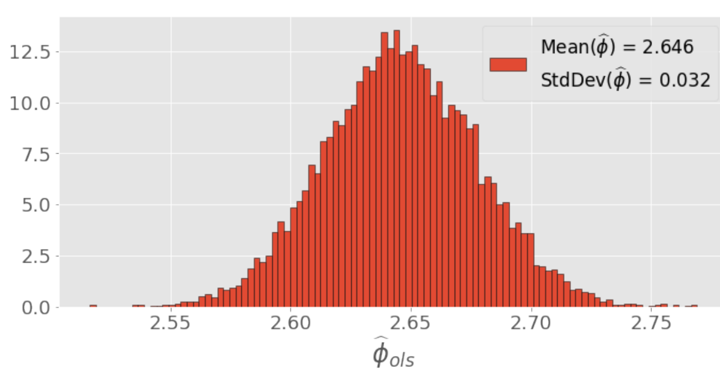

Figure 15: OLS estimation of an AR(4)

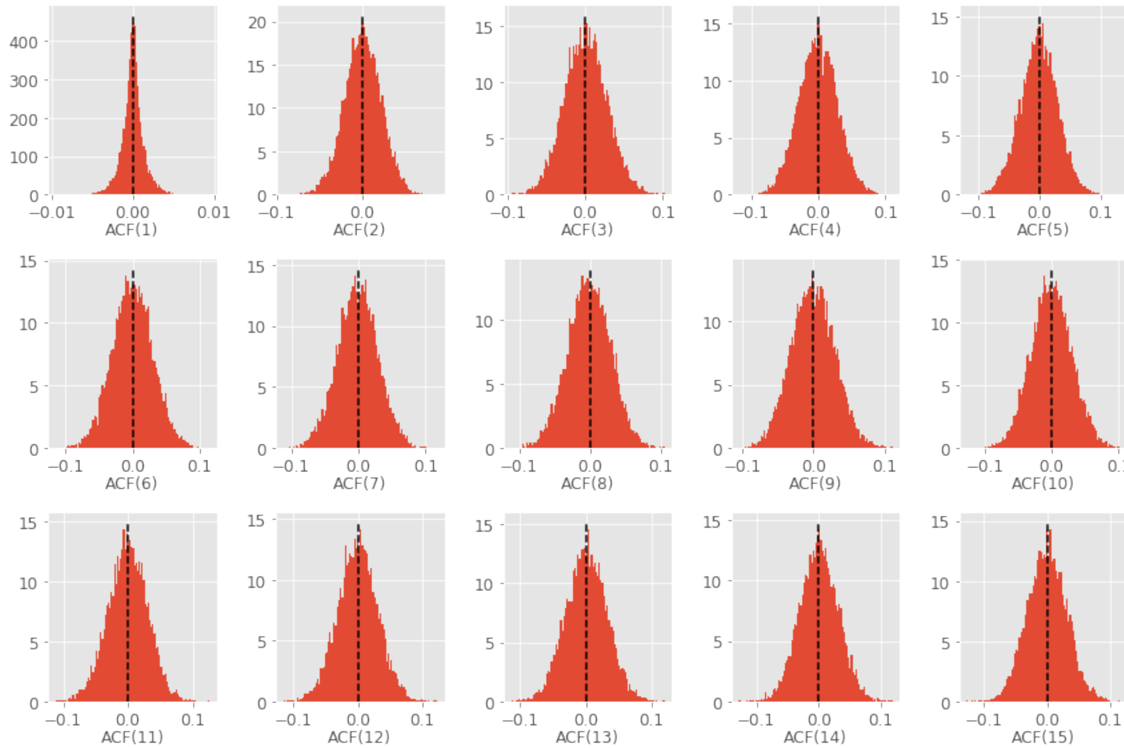

Figure 16: ACFs of the OLS errors for an AR(4)

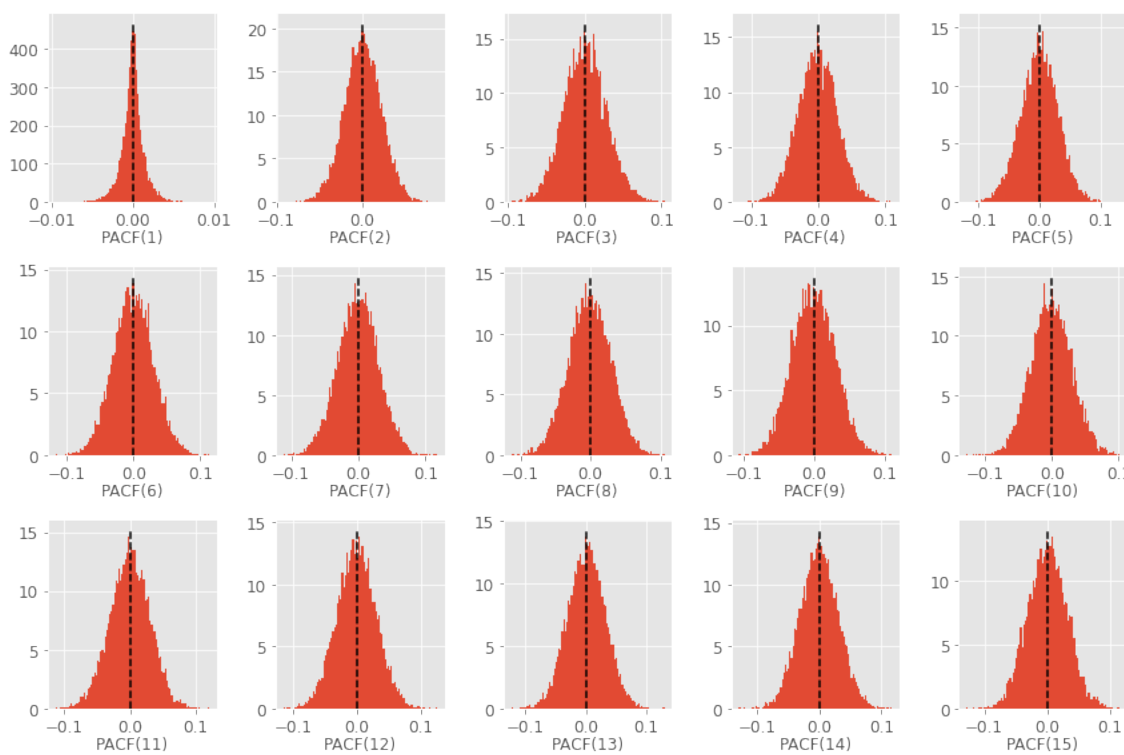

Figure 17: PACFs of the OLS errors for an AR(4)

(d) Taken together, those results highlight the importance of choosing the specification of ARMA processes, including the number of lags, very carefully. Issues are particularly serious when not enough lags are included, given that this creates an endogeneity issue due to the correlation between the residuals (which include the missing lagged variables) and the included regressors. In that case, OLS is inconsistent, i.e. it does not converge to the true parameter of interest. On the other hand, when too many lags are included, OLS remains consistent in this example, even though standard errors are unnecessarily increased. The latter option is clearly the lesser of the two issues in this case, and adding lags, or at least checking for the stability of the results when lags are added, could be a good idea. This does not necessarily have to be the case in general however, when there are MA terms or more complex dynamics.

Question 3

(a) If x_1 and x_2 are uncorrelated, the omission of x_2 in the regression does not cause a bias in the OLS estimation of the coefficient in front of x_1 ($\beta_1 = 1$). In the histogram in Figure 18, we see that the OLS estimates of β_1 are indeed centered around the truth 1 with $\mathbb{E}[\hat{\beta}] \approx 1.001$. This is expected: OLS is consistent since x_1 and x_2 satisfy the asymptotic OLS assumptions when they are orthogonal. In other words, there is no omitted variable "bias"[1] if the omitted variable is uncorrelated with regressors.

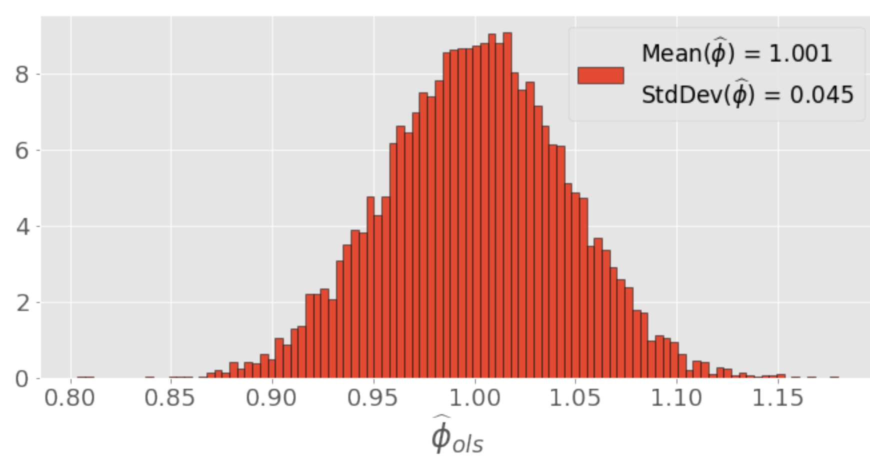

Figure 18: Omitted variable "bias" with uncorrelated regressors ($\Rightarrow$ no issue)

[1]Note: we put "bias" in quote because we are not technically focusing on small samples and biases, but rather on asymptotics and consistency.

(b),(c) If x_1 and x_2 are correlated, the omission of x_2 in the regression causes an omitted variable "bias" and OLS is inconsistent. In the histograms we see that the OLS estimates of β_1 are no longer centered around 1 but gather around $\mathbb{E}[\hat{\beta}] \approx 1.5$ for positive correlation 0.5 (overestimation) and are centered around $\mathbb{E}[\hat{\beta}] \approx 0.5$ for negative correlation -0.5 (underestimation).
This highlights that omitting variables when they are correlated to regressors (which is most of the time the case) is a very serious issue. This is particularly problematic because in reality, unlike in this simulation example, we have no way of knowing the variables that we omit!

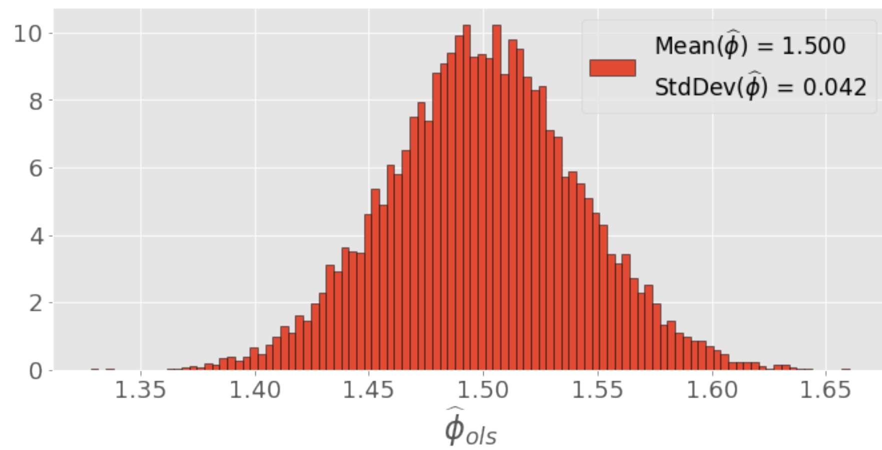

Figure 19: Omitted variable bias with positive correlation 0.5

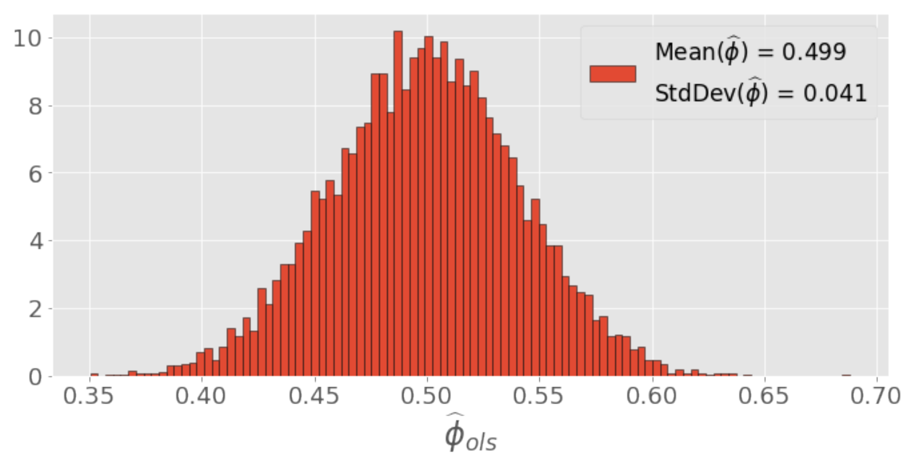

Figure 20: Omitted variable bias with negative correlation -0.5

Question 4

Y_t is stationary if the roots z of the equation:

$$\det\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} .5 & .1 \\ .4 & .5 \end{pmatrix} z - \begin{pmatrix} 0 & 0 \\ .25 & 0 \end{pmatrix} z^2\right] = 0$$

are outside the unitary circle ($|z_i| > 1 \; \forall i$). We can write this equation as:

$$\det\left[\begin{pmatrix} 1 - .5z & -.1z \\ -.4z - .25z^2 & 1 - .5z \end{pmatrix}\right] = 0$$

$$(1 - .5z)^2 = .1z^2(.4 + .25z)$$

$$0.025z^3 - 0.21z^2 + z - 1 = 0$$

Using Python, Matlab, or your favorite scientific package, we can easily find that the roots of this polynomial are:

$$\begin{cases} z_1 \approx 1.3 \;\Rightarrow\; |z_1| \approx 1.3 > 1 \\ z_2 \approx 3.55 + 4.26i \;\Rightarrow\; |z_2| \approx 5.5 > 1 \\ z_3 \approx 3.55 - 4.26i \;\Rightarrow\; |z_3| \approx 5.5 > 1 \end{cases}$$

We conclude that all the roots of the initial equation are outside the unit circle, and Y_t **is stationary**.

Question 5

(a) There are two possible ways to write the likelihood function of the MA(1) process z_t. In the first approach, we assume that we know for sure that $e_0 = 0$. Then:

$$e_1 = z_1 - \mu$$

$$e_2 = z_2 - \psi z_1 - (1 - \psi)\mu$$

$$\vdots$$

$$e_t = z_t - \psi z_{t-1} + \cdots + (-\psi)^{t-1} z_1 - (1 - \psi + \cdots + (-\psi)^{t-1})\mu$$

This implies that we can write the time t residual as a function of past realizations of the series z_t. Now, the full likelihood of a time series $z_1, z_2, ..., z_T$ conditional on $e_0 = 0$ is

$$f_{e_0=0}(z_1, z_2, \ldots, z_T) = f_{e_0=0}(z_1) f_{e_0=0}(z_2|z_1) f_{e_0=0}(z_3|z_2, z_1) \ldots f_{e_0=0}(z_T|z_{T-1}, \ldots, z_1)$$

Given that z_t is a MA(1) process $z_t = \mu + e_t + \psi e_{t-1}$, z_t depends on past values $(z_{t-1}, \ldots, z_1)$ only to the extent that they can predict the error term e_{t-1}. But e_{t-1} is actually fully determined by $(z_{t-1}, \ldots, z_1)$, and therefore:

$$f_{e_0=0}(z_t|z_{t-1}, \ldots, z_1) = f_{e_0=0}(z_t|e_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-e_t^2}{2\sigma^2}\right]$$

From this, it follows that the full log likelihood conditional on $e_0 = 0$ is:

$$l(\omega) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \sum_{t=1}^{T} \frac{e_t^2}{2\sigma^2}$$

where $\boldsymbol{\omega} = (\mu, \psi, \sigma)'$.

A second acceptable approach would be calculating the exact unconditional likelihood of the time series $\{z_1, z_2, ..., z_T\}$, without the assumption $e_0 = 0$. This could by done by remembering that $z_1, z_2, ..., z_T$ is a normal process, so

$$(z_1, z_2, ..., z_T)^\top \sim N(\mathbf{m}, \Sigma).$$

where $\mathbf{m}$ would be a vector with μ in all its entries, while Σ could be calculated using the covariance function of the MA(1) process.

Finally, e_0 can also simply be treated as an additional parameter to estimate.

(b) Figure 21 shows the path of length 100 that is the basis for the simulation. Note that in small samples, the actual sample mean and moments can be quite different from their true values.

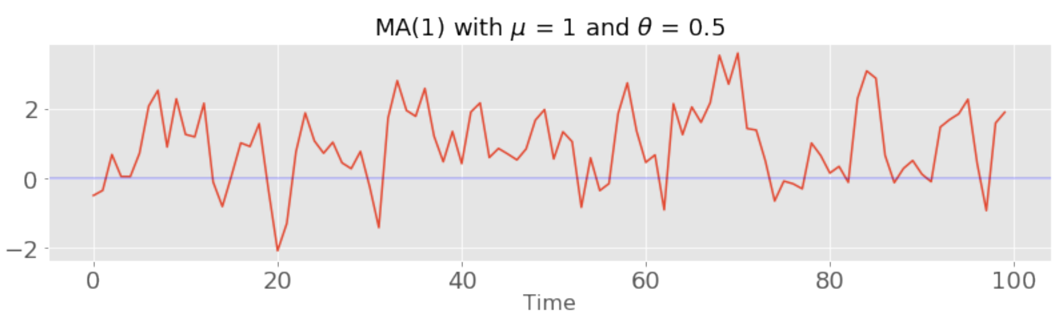

Figure 21: Sample for the MA(1) process used for the estimation

The log-likelihood function derived in **(a)** is maximized numerically using the *scipy.optimize* package. Specifically, we use the *minimize* on the opposite of the log-likelihood: $-l(\boldsymbol{\omega})$. We stick to the default optimization method, which is "BFGS" (Broyden, Fletcher, Goldfarb, and Shanno, 2006), and pick $(0.5, 0.5, 0.5)'$ as starting values. The log-likelihood behaved here so those choices make little difference here, even though they can matter tremendously in more complex cases.

We obtain:

$$\begin{cases} \hat{\mu} = 0.947 \\ \hat{\theta} = 0.535 \\ \hat{\sigma} = 0.963 \end{cases}$$

which is very close to the true values $\mu = 1, \theta = 0.5$, and $\sigma = 1$.

(c) The asymptotic variance of the MLE estimator is given by $\mathcal{I}^{-1}$, where $\mathcal{I}$ is the Fisher information matrix defined by:

$$\mathcal{I}_{jk} : \mathbb{E}\left[-\frac{\partial^2 \ln f_\theta(\mathbf{x}_i)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k}\right]$$

In practice, this is nothing but the Hessian returned by the optimizer. To obtain the standard error of each parameter, it suffices to take square root of each diagonal element. We obtain:

$$\begin{cases} \text{std}(\hat{\mu}) = 0.148 \\ \text{std}(\hat{\theta}) = 0.093 \\ \text{std}(\hat{\sigma}) = 0.068 \end{cases}$$

(d) The results from the estimation using *statsmodel* are in Figure 22 below. We use the standard method (Conditional Sum of Squares Likelihood) and hyper-parameters. Results are extremely close to our estimates, and their standard errors, which is good news! As mentioned previously, one could also treat e_0 as a parameter to estimate instead of assuming it to be 0. This is likely to be important in small samples (cf. Hamilton (1994) for much more details). In practice, it turns out to make little difference in this simple example.

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  100
Model:                     ARMA(0, 1)   Log Likelihood                -138.191
Method:                       css-mle   S.D. of innovations              0.962
Date:                Tue, 21 Apr 2020   AIC                            282.382
Time:                        22:07:32   BIC                            290.198
Sample:                             0   HQIC                           285.546

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.9432      0.149      6.331      0.000       0.651       1.235
ma.L1.y        0.5544      0.093      5.936      0.000       0.371       0.737
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
MA.1           -1.8037           +0.0000j            1.8037            0.5000
-----------------------------------------------------------------------------
```

Figure 22: MA(1) estimation by MLE using *statsmodel*