MGMT MFE 407: Empirical Methods in Finance
Homework 2: Solution

Prof. Lars A. Lochstoer
TAs: Chady Gemayel and Mahyar Kargar

January 24, 2019

Problem 1: AR(p) Processes

3. Consider an AR(2) process with $\phi_1 = 1.1$ and $\phi_2 = -0.25$

   (a) Plot the autocorrelation function for this process for lags 0 through 20.

   Suggested Solution: We know that $\rho_0 = 1$ we can further solve for ρ_1 using

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_1 \tag{1}$$

   and for $j \geq 2$

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} \tag{2}$$

   Figure 1 shows the bar plot for the autocorrelations for lags 0 through 20.

   (b) Is the process stationary? Explain why or why not.

   Suggested Solution: We solve for the characteristic roots for the AR(2) using

$$x_1, x_2 = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \tag{3}$$

   and that $\omega_1 = x_1^{-1}$, $\omega_2 = x_2^{-1}$. We have that $\omega_1 = 0.321$ and $\omega_2 = 0.779$. Both of the characteristic roots are less than zero, so this AR(2) is stationary.

   (c) Give the dynamic multiplier for a shock that occurred 6 periods ago. That is, calculate $\frac{\partial[r_{t+6}-\mu]}{\partial \epsilon_t}$

   Suggested Solution: We can rewrite the AR(2) process as

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + \epsilon_t \tag{4}$$

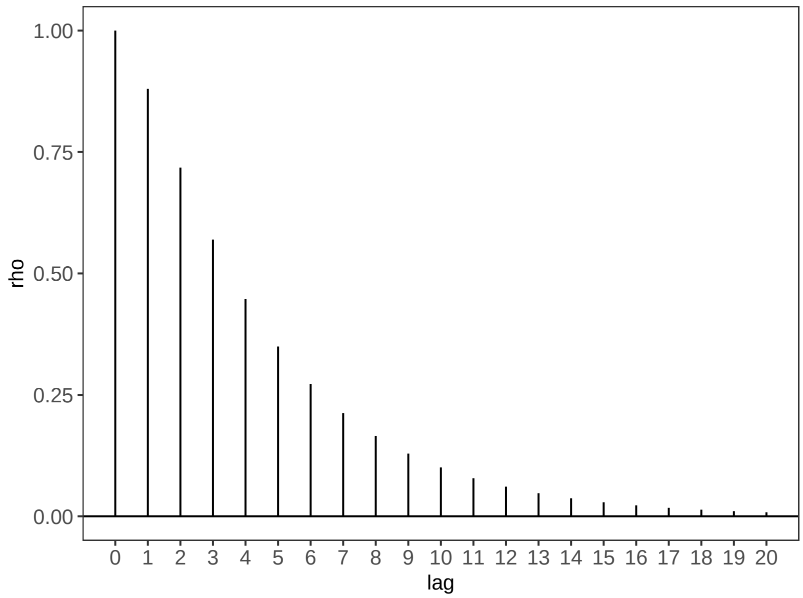

Figure 1: Autocorrelation for lags 0 through 20

We then iterate this formula

$$
\begin{aligned}
r_{t+1} - \mu &= \phi_1(r_t - \mu) + \phi_2(r_{t-1} - \mu) + \epsilon_{t+1} && (5)\\
&= (\phi_1^2 + \phi_2)(r_{t-1} - \mu) + \phi_1\phi_2(r_{t-2} - \mu) + \phi_1\epsilon_t + \epsilon_{t+1} && (6)\\
r_{t+2} - \mu &= \phi_1(r_{t+1} - \mu) + \phi_2(r_t - \mu) + \epsilon_{t+2} && (7)\\
&= ... + \phi_1^2\epsilon_t + \phi_2\epsilon_t && (8)\\
r_{t+3} - \mu &= \phi_1(r_{t+2} - \mu) + \phi_2(r_{t+1} - \mu) + \epsilon_{t+3} && (9)\\
&= ... + (\phi_1^3 + 2\phi_1\phi_2)\epsilon_t && (10)\\
r_{t+4} - \mu &= \phi_1(r_{t+3} - \mu) + \phi_2(r_{t+2} - \mu) + \epsilon_{t+4} && (11)\\
&= ... + (\phi_1^4 + 3\phi_1^2\phi_2 + \phi_2^2)\epsilon_t && (12)\\
r_{t+5} - \mu &= \phi_1(r_{t+4} - \mu) + \phi_2(r_{t+3} - \mu) + \epsilon_{t+5} && (13)\\
&= ... + (\phi_1^5 + 4\phi_1^3\phi_2 + 3\phi_1\phi_2^2)\epsilon_t && (14)\\
r_{t+6} - \mu &= \phi_1(r_{t+5} - \mu) + \phi_2(r_{t+4} - \mu) + \epsilon_{t+6} && (15)\\
&= ... + (\phi_1^6 + 5\phi_1^4\phi_2 + 6\phi_1^2\phi_2^2 + \phi_2^3)\epsilon_t && (16)
\end{aligned}
$$

Thus the dynamic multiplier for a shock that occurred 6 periods ago is 0.380.

(d) Now, instead assume $\phi_1 = 0.9$ and $\phi_2 = 0.8$. Give the dynamic multiplier for a shock that occurred 6 periods ago. Is the process stationary? Why/why not?

Suggested Solution: The dynamic multiplier for a shock that occurred 6 periods ago is 6.778. This process is not stationary as the influence of a shock does not seem to be going away as time goes on. In fact in this case, the characteristic roots are 1.451 and -0.551, one of which is greater than 1.

Problem 2: Applying the Box-Jenkins methodology[1]

In `PPIFGS.xls` you will find quarterly data for the Producer Price Index. Our goal is to develop a quarterly model for the PPI, so we can come up with forecasts. Our boss needs forecasts of inflation, because she wants to hedge inflation exposure. There is not a single 'correct' answer to this problem. Well-trained econometricians can end up choosing different specifications even though they are confronted with the same sample. However, there definitely are some wrong answers.

1. We look for a covariance-stationary version of this series. Using the entire sample, make a graph with four subplots:

   (a) Plot the PPI in levels.

   (b) Plot ΔPPI

   (c) Plot $\log PPI$

   (d) Plot $\Delta \log PPI$.

   Suggested Solution: Figure 2 presents the plots.

2. Which version of the series looks covariance-stationary to you and why? Let's call the covariance stationary version $y_t = f(PPI_t)$.

   Suggested Solution: It seems $\Delta \log PPI$ is covariance-stationary by looking at the time series graph. For ΔPPI, the unconditional variance seem to be changing. For $\Delta \log PPI$, the unconditional mean and variance seems quite stable. We'll need to do more tests to identify stationarity.

3. Plot the ACF of y_t for 12 quarters. What do you conclude? If the ACF converges very slowly, re-think whether y_t really is covariance stationary.

   Suggested Solution: Figure 3 shows the ACF for $\Delta \log PPI$. It converges to zero after 3 lags, though there is no clear cut off for lag 4-6. I will include 3 or 4 lags in the MA process.

[1] In Matlab, there is an **Econometrics Toolbox** and a series of functions : 'arima, estimate, forecast, infer, simulate, lbqtest' that can help you solve this problem. Alternatively, you can download Kevin Sheppard's **MFE toolbox**, which is freely available. You can just Google this and find it. In *R* there is a package called 'MTS' for *Multivariate Time Series*, by Ruey Tsay. This is a very useful package, that we will also use when estimating time-varying volatility models.

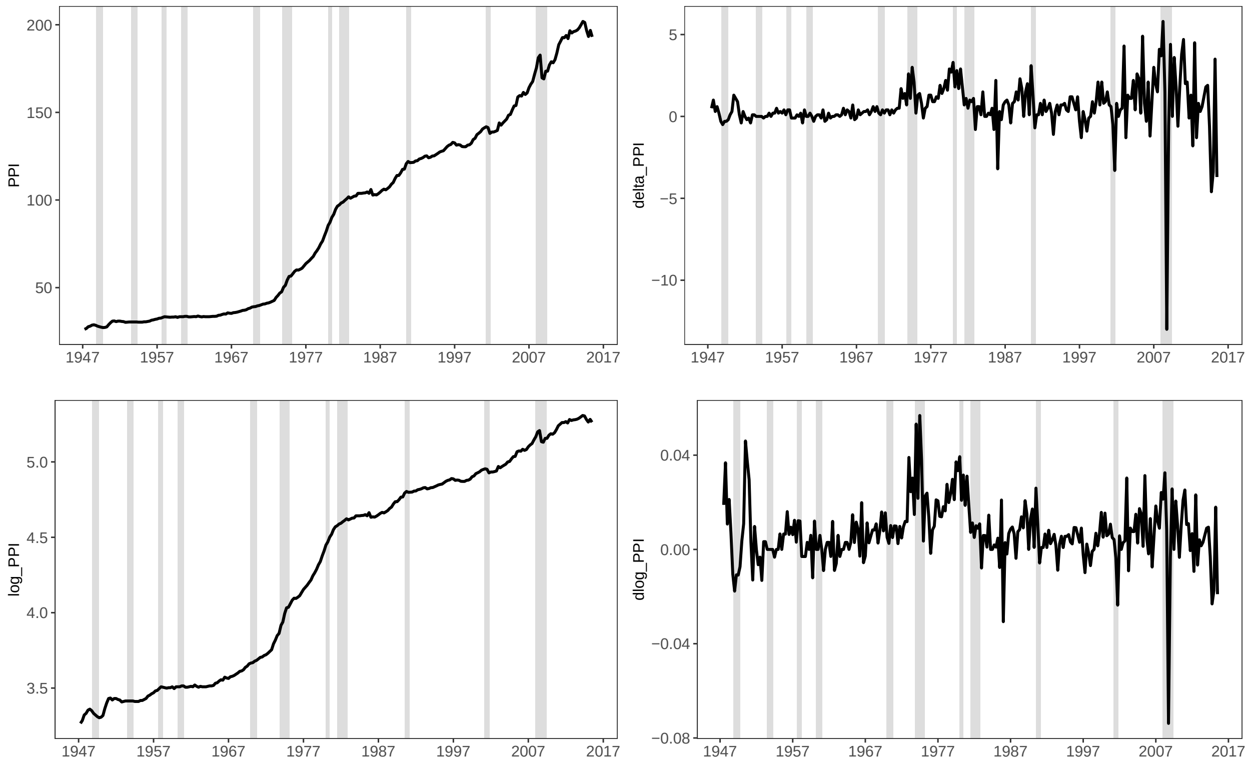

Figure 2: PPI Plots. The vertical shaded bars indicate NBER recessions.

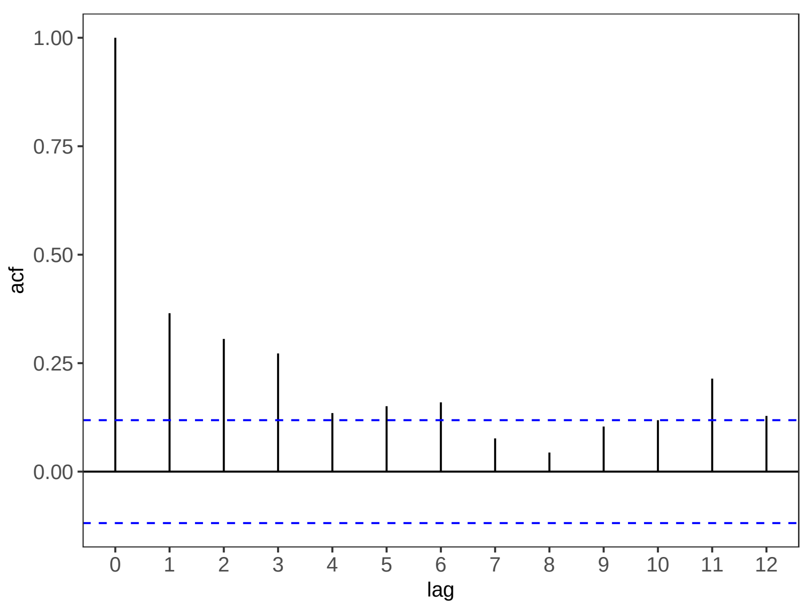

Figure 3: ACF of $\Delta \log PPI$

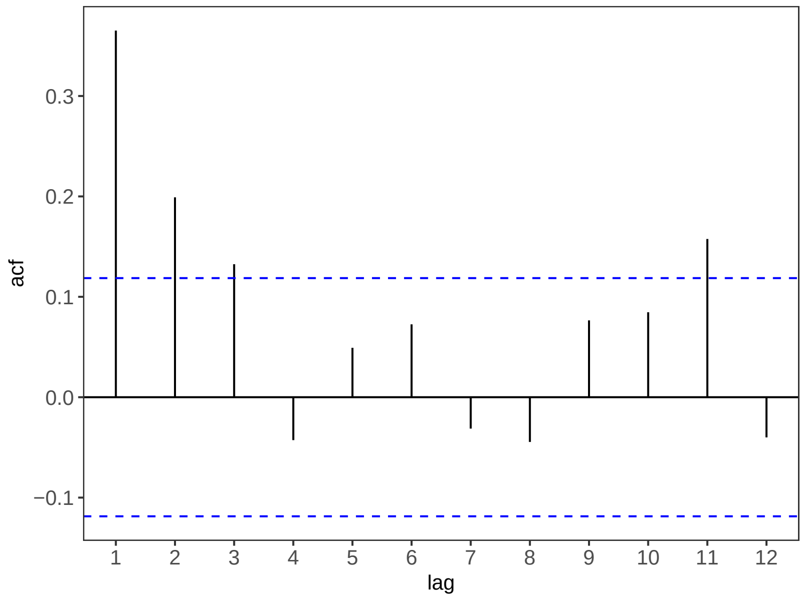

Figure 4: PACF of $\Delta \log PPI$

4. Plot the PACF of y_t for 12 quarters. What do you conclude?

   Suggested Solution: Figure 4 shows PACF for $\Delta \log PPI$. I will include 2-3 lags in the AR process.

5. On the basis of the ACF and PACF, select two different AR model specifications that seem the most reasonable to you. Explain why you chose these.

   (a) Using the entire sample, estimate each one of these. Report the coefficient estimates and standard errors. Check for stationarity of the parameter estimates.

   Suggested Solution: I will estimate the time series using AR(2) and AR (3). Table 1 gives the coefficients and standard errors for these models.

   We can check whether the AR part for each model is stationary by solving for the roots for polynomials, and then make sure 1/root<1.

$$AR(2): 1 - \phi_1 x - \phi_2 x^2 = 0 \tag{17}$$

$$AR(3): 1 - \phi_1 x - \phi_2 x^2 - \phi_3 x^3 = 0 \tag{18}$$

   The AR(2) model has 2 real roots (1.41 and -2.41) and the AR(3) has one real and two imaginary roots (1.32, $-1.16 + 1.84i$, and $-1.16 - 1.84i$), respectively, all are greater than 1, thus stationary:

Table 1: AR Estimates for $\Delta \log PPI$

	Dependent variable:	
	dlog(PPI)	
	AR(2)	AR(3)
	(1)	(2)
ar1	0.29***	0.27***
	(0.06)	(0.06)
ar2	0.20***	0.16**
	(0.06)	(0.06)
ar3		0.14**
		(0.06)
intercept	0.01***	0.01***
	(0.001)	(0.002)
Observations	273	273
Log Likelihood	822.29	824.87
σ^2	0.0001	0.0001
Akaike Inf. Crit.	−1,636.59	−1,639.75
Note:	*p<0.1; **p<0.05; ***p<0.01	

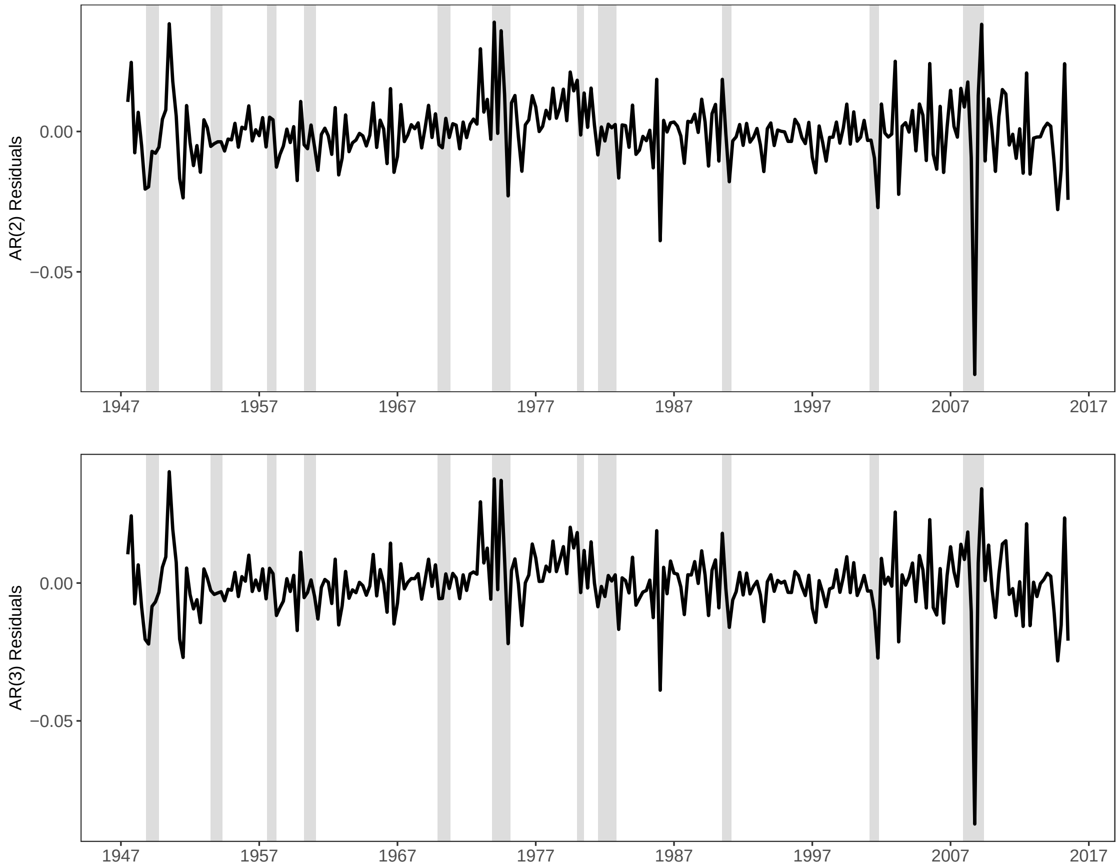

Figure 5: Residuals for AR(2) and AR(3) Models. The vertical shaded bars indicate NBER recessions.

```
model2 <- arima(y,order=c(2,0,0),method="ML")
model3 <- arima(y,order=c(3,0,0),method="ML")

polyroot(c(1,-model2$coef[1],-model2$coef[1]))

## [1]  1.410246-0i -2.410246+0i

Mod(polyroot(c(1,-model3$coef[1],-model3$coef[2],-model3$coef[2])))

## [1] 1.317121 2.173473 2.173473
```

(b) Plot the residuals. (Note: the residuals will have conditional heteroskedasticity or 'GARCH effects'. We will talk about this later. However, in well-specified models, the residuals should not be autocorrelated.)

Suggested Solutions: Figure 5 plots the residuals for AR(2) and AR(3) models. .

(c) Report the Q-statistic for the residuals for 8 and 12 quarters, as well as the AIC and BIC. Select a preferred model on the basis of these diagnostics. Explain your choice.

Suggested Solution: Table 2 gives the statistics for these four models. I will choose AR (3) model because the AIC is the smallest and the Q-stats are the lest significant.

```
## [1] -1636.589
## [1] -1639.746
## [1] -1622.151
## [1] -1621.699
##
##  Box-Ljung test
##
## data:  model2$residuals
## X-squared = 10.202, df = 6, p-value = 0.1164
##
##  Box-Ljung test
##
## data:  model3$residuals
## X-squared = 5.346, df = 5, p-value = 0.3751
##
##  Box-Ljung test
##
## data:  model2$residuals
## X-squared = 18.721, df = 10, p-value = 0.04396
##
##  Box-Ljung test
##
## data:  model3$residuals
## X-squared = 13.829, df = 9, p-value = 0.1285
```

6. Re-estimate the two models using only data up to the end of 2005 and compute the MSPE (mean squared prediction error) on the remainder of the sample for one-quarter ahead forecasts:

$$\frac{1}{H}\sum_{t=1}^{H} v_t^2$$

where H is the length of the hold-out sample, and v_i is the one-step ahead prediction error. Also report the MSPE assuming there is no predictability in y_t, i.e. assuming y_t follows a random walk. What do you conclude?

Suggested Solution: The MSPE for these models are shown in table 3. We fix the coefficient estimates using data prior to Dec 2005 and fit new data sequentially to get one period ahead

Table 2: AR Model Statistics

	AR(2)	AR(3)
Q-stats(8Q)	10.20	5.35
p-value	(0.12)	(0.38)
Q-stats(12Q)	18.72**	13.83
p-value	(0.04)	(0.13)
AIC	−1,636.59	−1,639.75
BIC	−1622.15	−1621.70
Note:	*p<0.1; **p<0.05; ***p<0.01	

prediction. For the random walk, we take the real value from last period as the prediction. We can see that the MSPE for random walk is the highest.

Table 3: AR MSPE. This table shows the MSPE for different models. The first line does the forecast without fitting in new data. The second line shows the forecast when we fix the model but keep fitting in new data for predictions.

	AR(2)	AR(3)	RW
without New Data	3.40e−4	3.38e−4	3.23e−4
with New Data	4.30e−4	4.17e−4	5.69e−4