Dr. Randall R. Rojas
Department of Economics
UCLA

Economics 144
Economic Forecasting
Spring, 2018

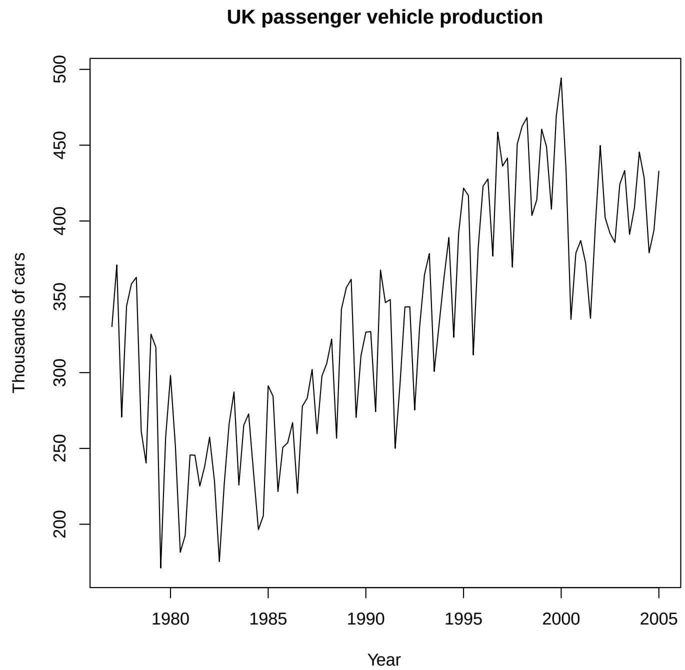
Midterm Exam
May 3, 2018

For full credit on a problem, you need to show all your work and the formula(s) used.

First Name	
Last Name	
UCLA ID #	

Please do not start the exam until instructed to do so.

1.(15%) We will look at quarterly UK passenger car production (thousands of cars) from 01/1977 to 01/2005. The figure below shows the data.



(a) Describe the model fit (in terms of the type of trend, seasonality and/or cycles) you would consider for this data set.

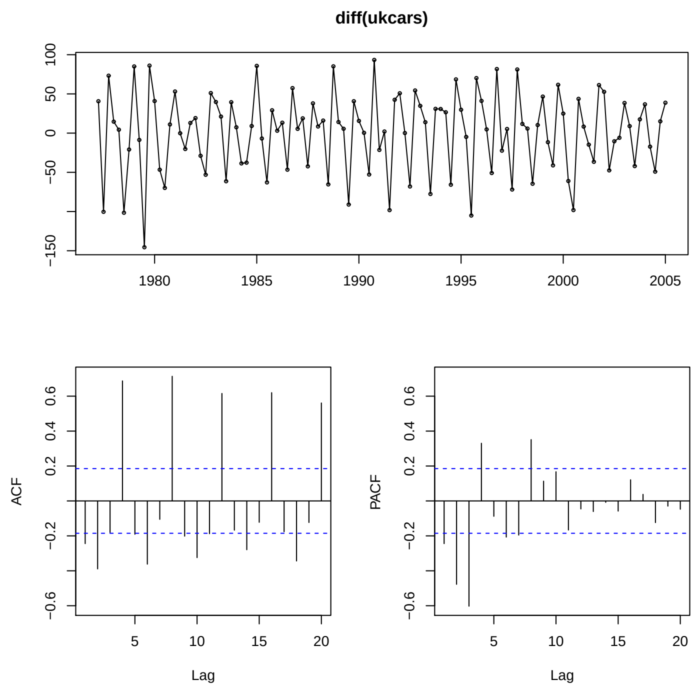
For the trend component I would propose a cubic polynomial (or some other high degree polynomial). For the seasonality, quarterly seasonal dummies should suffice. The does appear the be a strong cyclical component, therefore, maybe an AR model would work.

(b) Assume you fit a trend & seasonal model using `tslm`. Comment on the seasonal factors estimated in the table below.

	Estimate	Std. Error	t value	Pr(>|t|)
trend	1.8380	0.1230	14.94	0.0000
season1	254.7342	10.5799	24.08	0.0000
season2	250.4088	10.6049	23.61	0.0000
season3	183.1918	10.6853	17.14	0.0000
season4	225.5858	10.7664	20.95	0.0000

There only appears to be a sharp decrease in Q2 (season 2) and relatively constant for the other quarters (given their similar magnitiudes).

(c) Assume you decide to take the first difference of the data before fitting a model and obtained the results below. What model would you now recommend for fitting the new data?



The series appears to have zero mean with regular spikes in the ACF at $k = 4, 8, 12, \ldots$ and at $k = 4, 8, \ldots$ in the PACF. This suggests and S-AR(2). However since there are strong spikes in the PACF at lower lags such as at $k = 2, 3$, I would also add an AR(2). Therefore, the model would be ARIMA(2,0,0)(2,0,0)[4] with zero mean. Note: Please reference Leceture 9 slides.

2.(15%) Given the time-series below:

(a) Rewrite the following expression without using the lag operator and identify the process (e.g, AR(p), MA(q), or ARMA(p, q)):

$$y_t = y_{t-3} + \left(\frac{1}{L-2} + \frac{L}{L+2} - 1 \right) \varepsilon_{t-1} + \varepsilon_t$$

$$\begin{aligned}
\rightarrow y_t - y_{t-3} &= \left(\frac{2L - L^2 + 4}{(L-2)(L+2)} \right) \varepsilon_{t-1} + \varepsilon_t \\
\rightarrow (L^2 - 4)(y_t - y_{t-3}) &= (2L - L^2 + 4)\varepsilon_{t-1} + (L^2 - 4)\varepsilon_t \\
\rightarrow y_{t-2} - y_{t-5} - 4y_t + 4y_{t-3} &= 2\varepsilon_{t-2} - \varepsilon_{t-3} + 4\varepsilon_{t-1} + \varepsilon_{t-2} - 4\varepsilon_t \\
y_t &= \frac{-1}{4} y_{t-5} + y_{t-3} + \frac{1}{4} y_{t-2} + \frac{1}{4}\varepsilon_{t-3} - \frac{3}{4}\varepsilon_{t-2} - \varepsilon_{t-1} + \varepsilon_t \\
\rightarrow ARMA(5,3)
\end{aligned}$$

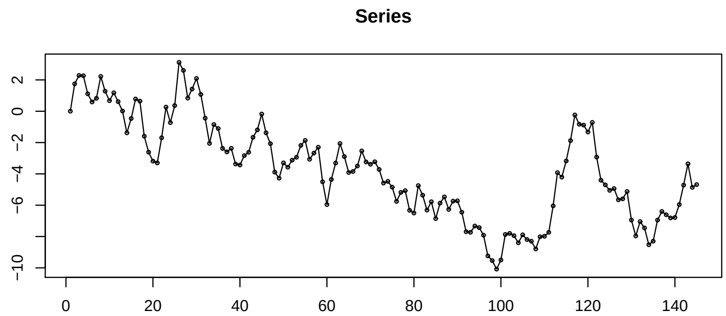
(b) Rewrite the following expression in lag operator form and identify the process (e.g, AR(p), MA(q), or ARMA(p, q)):

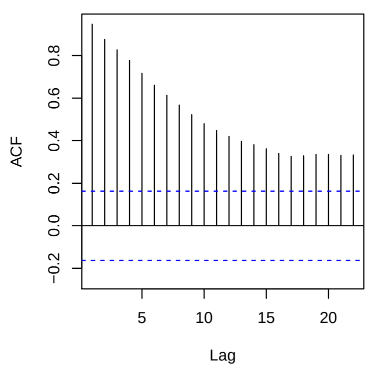
$$4y_{t-4} + y_t - 3y_{t-3} = 1 + \varepsilon_t + 5\varepsilon_{t-5} + 2\varepsilon_{t-2}$$

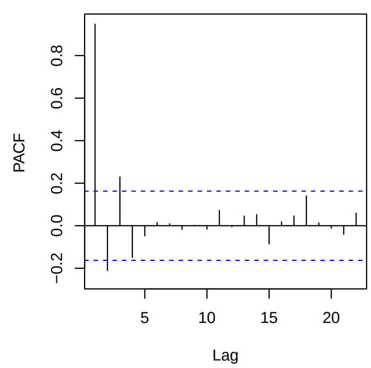
$$\rightarrow (4L^4 + 1 - 3L^3) y_t = 1 + (1 + 5L^5 + 2L^2)\varepsilon_t$$

$$\rightarrow ARMA(4,5)$$

3.(15%) The data below are monthly observations for an unknown series from 2000 to 2012. The ACF and PACF plots are based on the original observations given in the top plot.







(a) Based on the plot of the data alone, what trend model would you propose for this series? Does this process appear to be covariance stationary? Justify your answers.

I would suggest a quadratic (or other polynomial) model for the trend component. The series does not appear to be covariance stationary given the low mean reversion with high persistence.

(b) Based on the ACF and PACF plots, what model would you propose for this series in terms of an MA and/or AR process. Justify your answer.

I would recommend an AR(3) model given the decay to zero in the ACF and strong spikes at $k = 1, 2, 3$ in the PACF.

(c) Would you recommend taking the first difference of the data before fitting a model? Justify your answer.

Yes, in order to make it covariance stationary.

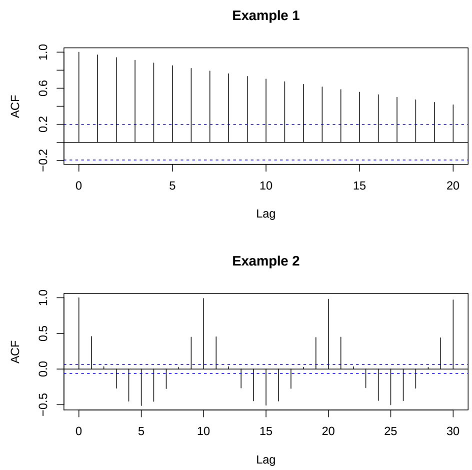
(d) Based on your answer to part (b), if you use that model to do a forecast, do you expect the 1-step, 2-step, and 3-setp-ahead forecast values to be different or equal to each other?

Since it looks like an AR model, then theses 3-step ahead forecasts should be different from each other.

4.(15%) (a) Describe an advantage of a rational distributed lag representation vs. a polynomial distributed lag representation? What is a common tradeoff between them?

With a rational distributed lag representation, you need fewer parameters than the polynomial representation. However, a common tradeoff is that for the rational polynomial, you need to include restrictions on the denominator to avoid any singularities.

(b) Sketch the ACF plot for a series that exhibits a very high degree of serial correlation.



(c) Explain in detail the output from: `stl(data,s.window=‘‘periodic’’)`, where `data` is your original time series.

S (Seasonal -typically just indicators variables for each season depending on the frequency of the data), T(Trend -what is left after removing the seasonal component), Remainder = what is left after removing the trend and seasonal dynamics from the data.

(d) Explain in simple words the ‘Rolling’ forecast algorithm.

The estimation sample in this scheme always has the same number of data points, once this number (call it n for now) is selected, we estimate starting from 1 to n, and then from 2 to $n+1$, and so on until we have exhausted all the observations. At each iteration, we compute the 1-step-ahed forecst.

5.(15%) Given the series $y_t = \frac{1}{2}\varepsilon_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim WN(0,4)$.

(a) Find the respective autoregressive representation.

This is an MA(1) process, where $\theta = 1/2$, $\mu = 0$, and $\sigma^2 = 4$ (see Lecture 7)

$$\begin{aligned} y_t - \mu &= \theta(y_{t-1} - \mu) - \theta^2(y_{t-2} - \mu) + \cdots + \varepsilon_t \\ \to y_t &= \frac{1}{2}y_{t-1} - \frac{1}{4}y_{t-2} + \cdots + \varepsilon_t \end{aligned}$$

(b) Find $E[y_t]$ and $var[y_t]$.

$$E[y_t] = 0, var[y_t] = (1+\theta^2)\sigma_\varepsilon^2 = (1+0.5^2) \times 4 = 5$$

(c) Find ACF($k=1$) and PACF($k=1$)

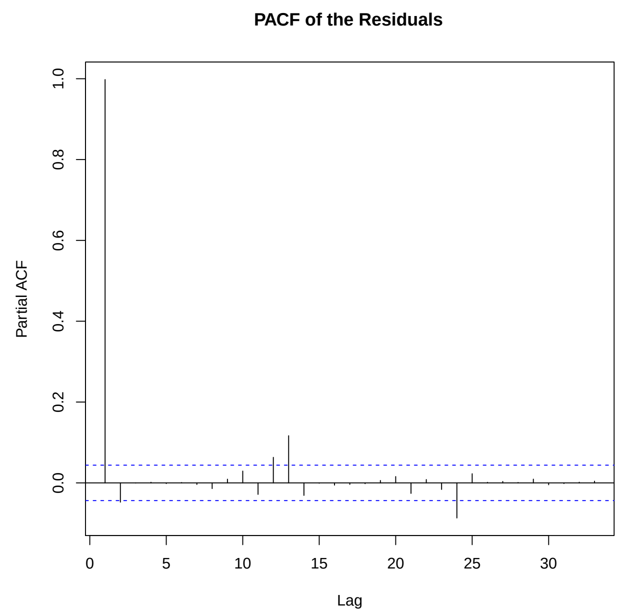
$$\begin{aligned} ACF(k=1) &= \rho_1 = \frac{\theta}{1+\theta^2} = \frac{0.5}{1+0.5^2} = 0.4 \\ PACF(k=1) &= ACF(k=1) = 0.4. \end{aligned}$$

(d) Does this process appear to be covariance stationary and/or invertible? Explain.

Since it is an MA process, then we do expect it to be covariance stationary, and since since $|\theta| = 0.5 < 1$, then it should also be invertible.

Questions 6-10 are multiple choice. Please select one answer per question only.

6.(5%) What model would you propose based only on the PACF (using monthly observations) below?



(a) AR(24)

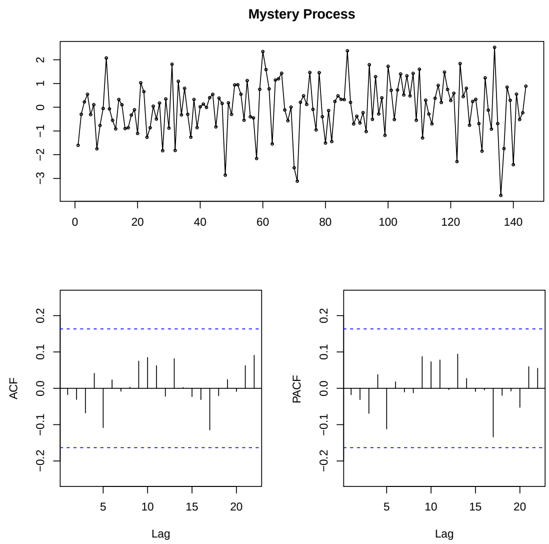
(b) S-AR(1)

(c) MA(24)

(d) S-MA(1)
Strong spikes at $k = 1, 12, 24$ in PACF with monthly observations is consistent with an S-MA(1) process.

(e) None of the above

7.(5%) Given the series and respective ACF and PACF plots, which ARMA simulation most likely produced it?



(a) $arima.sim(list(order = c(0, 0, 1), ma = 1), n = 144)$

(b) $arima.sim(list(order = c(0, 0, 1), ma = 0.0000005), n = 144)$
Based on the plots, the process looks like white noise. Therefore, Since this is an MA(1) process with $\theta = 0.0000005 << 1$, this best reflects a white noise process.

(c) $arima.sim(list(order = c(1, 0, 0), ar = 0.2), n = 144)$

(d) $arima.sim(list(order = c(1, 0, 0), ar = 1), n = 144)$

(e) None of the above

8.(5%) Which option best represents a covariance stationary process?

(a) AR(2)

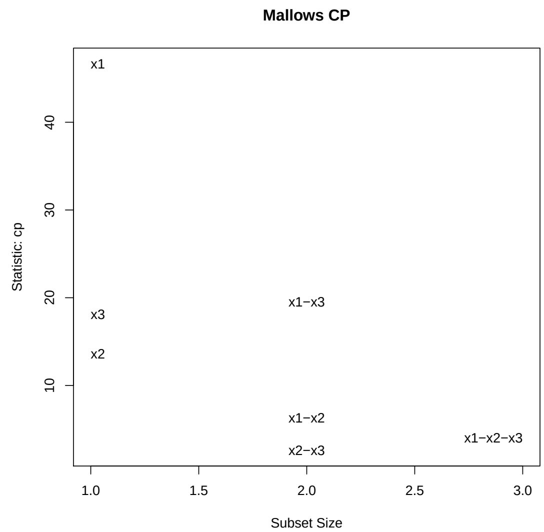
(b) $y_t = 4t^2 + \sin(2\pi t) + \varepsilon_t$

(c) MA(10)
By definition of an MA process, there is strong mean reversion, hence more covariance stationary.

(d) S&P 500 historical prices over the past 20 years

(e) None of the above

9.(5%) Based on the Mallows Cp plot below, which model is the preferred one?



(a) $x2 - x3$
This model has the lowest Mallows Cp.

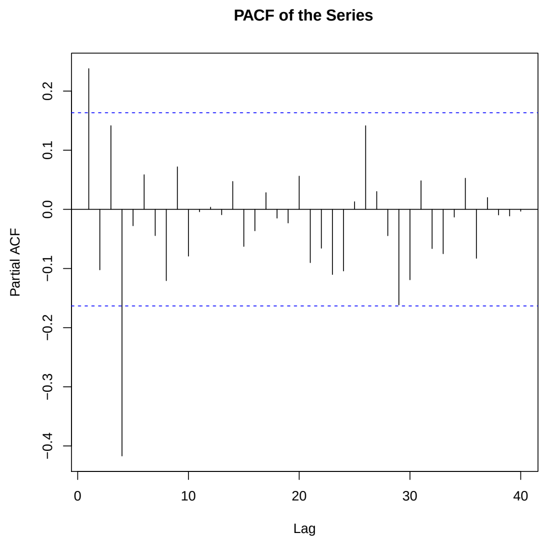
(b) $x1 - x3$

(c) $x1 - x2 - x3$

(d) $x1 - x2$

(e) $x1$

10.(5%) Which model corresponds to the PACF plot below?



(a) $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \varepsilon_t$

(b) $y_t = \phi_1 y_{t-1} + \phi_4 y_{t-4} + \varepsilon_t$
Only spikes at $k = 1$ and $k = 4$ are statisically significant.

(c) $y_t = \phi_4 y_{t-4} + \varepsilon_t$

(d) $y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} + \varepsilon_t$

(e) None of the above