Final Exam
Due: Sunday, July 20 by 11:59pm

Questions 1-4: 22.5 points
Question 5: 10 points
Data set is on Canvas.

1. You will find the data file **USMacro_Quarterly**, which contains quarterly data on several macroeconomic series for the United States; the data are described in the file USMacro_Description. The variable PCEP is the price index for personal consumption expenditures from the U.S. National Income and Product Accounts. In this exercise you will construct forecasting models for the rate of inflation, based on PCEP. For this analysis, use the sample period 1963:Q1–2012:Q4 (where data before 1963 may be used, as necessary, as initial values for lags in regressions).
   a. Compute the inflation rate, $Infl = 400 \times [\ln(PCEP_t) - \ln(PCEP_{t-1})]$.

   b. Plot the value of Infl from 1963:Q1 through 2012:Q4. Based on the plot, do you think that Infl has a stochastic trend? Explain.

   c. Compute the first four autocorrelations of Infl.

   d. Run an OLS regression of $\Delta Infl_t$ on $\Delta Infl_{t-1}$. Does knowing the change in inflation this quarter help predict the change in inflation next quarter?

   e. Estimate an AR(2) model for $\Delta Infl_t$. Is the AR(2) model better than an AR(1) model? Explain.

   f. Estimate an AR(p) model for p = 0,…, 8. What lag length is chosen by BIC? What lag length is chosen by AIC?

   g. Use the AR(2) model to predict the change in inflation from 2012:Q4 to 2013:Q1—that is, predict the value of $\Delta Infl_{2013:Q1}$?

2. The index of industrial production (IP_t) is a monthly time series that measures the quantity of industrial commodities produced in a given month. This problem uses data on this index for the United States. All regressions are estimated over the sample period 1986:M1 to 2013:M12 (that is, January 1986 through December 2013). Let

$$Y_t = 1200 \times \ln\left(\frac{IP_t}{IP_{t-1}}\right)$$

   a. The forecaster states that Y_t shows the monthly percentage change in IP, measured in percentage points per annum. Is this correct? Why?

   b. Suppose that a forecaster estimates the following AR(4) model for Y_t:

$$\hat{Y}_t = 0.787 + 0.052Y_{t-1} + 0.185Y_{t-2} + 0.234Y_{t-3} + 0.164Y_{t-4}.$$

   Use this AR(4) to forecast the value of Y_t in January 2014, using the following values of IP for July 2013 through December 2013:

Date	2013:M7	2013:M8	2013:M9	2013:M10	2013:M11	2013:M12
IP	99.016	99.561	101.196	100.374	101.034	101.359

   c. Worried that she might have included too few or too many lags in the model, the forecaster estimates AR(p) models for p = 0, 1, …, 6 over the same sample period. The sum of squared residuals from each of these estimated models is shown in the table below. Use the BIC to estimate the number of lags that should be included in the autoregression. Do the results differ if you use the AIC?

Final Exam

Due: Sunday, July 20 by 11:59pm

Questions 1-4: 22.5 points
Question 5: 10 points

Data set is on Canvas.

AR Order	0	1	2	3	4	5	6
SSR	19,533	18,643	17,377	16,285	15,842	15,824	15,824

3.

Consider the following data generating process

$$Y_t = \phi_3 Y_{t-3} + \varepsilon_t$$

where ε_t is a white noise process. Assuming $0 < \phi_3 < 1$, answer the following:

(a) Derive the expected value of this process.
(b) Derive the variance of this process.
(c) Derive the covariance of this process for $j = 1, 2, 3$.
(d) Derive the autocorrelation function of this process for $j = 1, 2, 3$.

4.

Consider the pieces of R output listed below and suppose the last two values for y_t are $y_T = 0.50$ and $y_{T-1} = 1.50$, respectively, where $t = 1, 2, \ldots, T$. With this information, answer the following:

```
model.fit = arima(data,order=c(1,0,2),method='ML')
model.fit
Coefficients:
          ar1      ma1      ma2     intercept
          0.0488   0.4705   0.2496  -0.0190
s.e.      0.7467   0.7376   0.3428   1.7343
SSR 405.1: log likelihood = -140.76, aic = 291.53, bic = 300.33

resid = model.fit$residuals
data = resid

resid.fit = arima(data,order=c(3,0,0),method='ML')
resid.fit
Coefficients:
          ar1      ar2      ar3     intercept
          0.0345   0.0590   0.0459  -0.0190
s.e.      0.7467   0.7376   0.3428   1.7343
SSR 405.1: log likelihood = -140.76, aic = 291.53, bic = 300.33
```

(a) Write the equation for the first model?

(b) Derive the one-step-ahead forecast from the first model.

(c) What is the h-step ahead forecast for the first model when h tends to infinity?

(d) Write the equation for the second model?

(e) What can you conclude from the second model?

5. In this exercise you will conduct a Monte Carlo experiment to study the phenomenon of spurious regression. In a Monte Carlo study, artificial data are generated using a computer, and then those artificial data are used to calculate the statistics being studied. This makes it possible to compute the distribution of statistics for known models when mathematical expressions for those distributions are complicated (as they are here) or even unknown. In this exercise, you will generate data so that two series, Y_t and X_t, are independently distributed random walks. The specific steps are as follows:

Final Exam
Due: Sunday, July 20 by 11:59pm

Questions 1-4: 22.5 points
Question 5: 10 points

Data set is on Canvas.

i. Use your computer to generate a sequence of T = 100 i.i.d. standard normal random variables. Call these variables $e_1, e_2, \cdots, e_{100}$. Set $Y_1 = e_1$ and $Y_t = Y_{t-1} + e_t$ for $t = 2,3,\cdots,100$.
ii. Use your computer to generate a new sequence of T = 100 i.i.d. standard normal random variables. Call these variables $a_1, a_2, \cdots, a_{100}$. Set $X_1 = a_1$ and $X_t = X_{t-1} + a_t$ for $t = 2,3,\cdots,100$.
iii. Regress Y_t onto a constant and X_t . Compute the OLS estimator, the regression R^2, and the (homoskedastic-only) t-statistic testing the null hypothesis that β_1 (the coefficient on X_t) is zero.

Use this algorithm to answer the following questions:

a. Run the algorithm (i) through (iii) once. Use the t-statistic from (iii) to test the null hypothesis that $\beta_1 = 0$, using the usual 5% critical value of 1.96. What is the R^2 of your regression?
b. Repeat (a) 1000 times, saving each value of R^2 and the t-statistic. Construct a histogram of the R^2 and t-statistic. What are the 5%, 50%, and 95% percentiles of the distributions of the R^2 and the t-statistic? In what fraction of your 1000 simulated data sets does the t-statistic exceed 1.96 in absolute value?