Financial Econometrics
Shangai Jiao Tong University Summer School 2025-2026
Problem Set - Lecture 7

Fatih Kansoy

Reference: Stock, J.H. and Watson, M.W., Introduction to Econometrics (4th Edition, Global Edition)

Chapter 15: Introduction to Time Series Regression and Forecasting

Question 1. 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:M1–2017:M12. Let $Y_t = 1200 \times \ln(IP_t/IP_{t-1})$.

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

(b) Suppose she estimates the following AR(4) model for Y_t:

$$\hat{Y}_t = 0.749 + 0.071Y_{t-1} + 0.170Y_{t-2} + 0.216Y_{t-3} + 0.167Y_{t-4}$$

with standard errors (0.488), (0.088), (0.053), (0.078), (0.064) respectively.

Use this AR(4) to forecast the value of Y_t in January 2018, using the following values of IP for July 2017 through December 2017:

Date	2017:M7	2017:M8	2017:M9	2017:M10	2017:M11	2017:M12
IP	105.01	104.56	104.82	106.58	106.86	107.30

(c) Worried about potential seasonal fluctuations, she adds Y_{t-12} to the autoregression. The estimated coefficient on Y_{t-12} is -0.061, with a standard error of 0.043. Is this coefficient statistically significant?

(d) Worried about a potential break, she computes a QLR test (with 15% trimming). The resulting QLR statistic is 1.80. Is there evidence of a break?

(e) The sum of squared residuals from AR(p) models for $p = 0, 1, \ldots, 6$ are given below. Use the BIC to estimate the number of lags. Do the results differ using AIC?

AR Order	0	1	2	3	4	5	6
SSR	21,045	20,043	18,870	17,838	17,344	17,337	17,306

Question 2. 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.

Generate data so that two series, Y_t and X_t, are **independently distributed random walks**:

(i) Generate a sequence of $T = 100$ i.i.d. standard normal random variables $\varepsilon_1, \ldots, \varepsilon_{100}$. Set $Y_1 = \varepsilon_1$ and $Y_t = Y_{t-1} + \varepsilon_t$ for $t = 2, \ldots, 100$.

(ii) Generate a new sequence $\alpha_1, \ldots, \alpha_{100}$ of i.i.d. standard normal variables. Set $X_1 = \alpha_1$ and $X_t = X_{t-1} + \alpha_t$.

(iii) Regress Y_t onto a constant and X_t. Compute the OLS estimator, the R^2, and the t-statistic testing $H_0 : \beta_1 = 0$.

(a) Run the algorithm once. Use the t-statistic to test $H_0 : \beta_1 = 0$ at 5%. What is R^2?

(b) Repeat 1000 times. Construct histograms of R^2 and the t-statistic. What are the 5%, 50%, and 95% percentiles? In what fraction do you reject H_0?

(c) Repeat (b) for $T = 50$ and $T = 200$. Does the rejection rate approach 5% as T increases?