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

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

Solution:

(a) **Interpretation of Y_t:**

Yes, the statement is correct.

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

- $\ln(IP_t/IP_{t-1}) \approx (IP_t - IP_{t-1})/IP_{t-1}$ = monthly growth rate (for small changes)
- Multiplying by 100 converts to percentage points
- Multiplying by 12 annualizes (12 months per year)
- Combined: $1200 \times \ln(\cdot)$ gives **annualized percentage growth**

(b) **Forecasting January 2018:**

First, compute Y values from the IP data:

$$\begin{aligned}
Y_{2017:M8} &= 1200 \times \ln(104.56/105.01) = 1200 \times (-0.00429) = -5.15 \\
Y_{2017:M9} &= 1200 \times \ln(104.82/104.56) = 1200 \times 0.00249 = 2.99 \\
Y_{2017:M10} &= 1200 \times \ln(106.58/104.82) = 1200 \times 0.01666 = 20.00 \\
Y_{2017:M11} &= 1200 \times \ln(106.86/106.58) = 1200 \times 0.00262 = 3.15 \\
Y_{2017:M12} &= 1200 \times \ln(107.30/106.86) = 1200 \times 0.00411 = 4.93
\end{aligned}$$

Forecast for January 2018:

$$\begin{aligned}
\hat{Y}_{2018:M1} &= 0.749 + 0.071(4.93) + 0.170(3.15) + 0.216(20.00) + 0.167(-5.15) \\
&= 0.749 + 0.350 + 0.536 + 4.320 - 0.860 \\
&= \boxed{5.10}
\end{aligned}$$

Interpretation: IP is forecast to grow at about 5.1% per annum in January 2018.

(c) **Significance of seasonal lag:**

$t = \frac{-0.061}{0.043} = -1.42$

Since $|-1.42| < 1.96$, the coefficient is **not statistically significant** at the 5% level.

(d) **QLR test for structural break:**

QLR statistic = 1.80

From Table 15.5 (15% trimming, $q = 5$ coefficients), the 10% critical value is 3.26.

Since 1.80 < 3.26, we **fail to reject** the null of parameter stability.

Conclusion: No evidence of a structural break.

(e) **Model selection using BIC and AIC:**

With $T = 32 \times 12 = 384$ observations:

$$BIC(p) = \ln(SSR/T) + (p+1)\frac{\ln T}{T}$$

$$AIC(p) = \ln(SSR/T) + (p+1)\frac{2}{T}$$

p	$\ln(SSR/T)$	BIC	AIC
0	3.998	4.014	4.003
1	3.949	3.980	3.959
2	3.889	3.935	3.905
3	3.832	3.894	3.853
4	3.804	3.881	3.830
5	3.804	3.896	3.835
6	3.802	3.910	3.838

BIC selects: $p = 4$ (minimum BIC = 3.881)

AIC selects: $p = 4$ (minimum AIC = 3.830)

Both criteria select **AR(4)**.

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 $\varepsilon_1, \dots, \varepsilon_{100}$. Set $Y_1 = \varepsilon_1$ and $Y_t = Y_{t-1} + \varepsilon_t$ for $t = 2, \dots, 100$.

(ii) Generate a new sequence $\alpha_1, \dots, \alpha_{100}$ 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?

Solution:

This exercise demonstrates **spurious regression**—finding significant relationships between independent random walks.

(a) **Single simulation (answers will vary):**

In one typical run with $T = 100$:

- $R^2 \approx 0.25$ (could range from near 0 to 0.8+)
- t-statistic ≈ 4.5 (could be anywhere from -15 to $+15$)
- If $|t| > 1.96$, we (incorrectly) reject H_0

(b) **1000 replications with $T = 100$:**

Typical results from simulation:

	5th percentile	50th percentile	95th percentile
R^2	0.00	0.19	0.73
t-statistic	-13.0	0.02	13.0

Rejection rate: Approximately **76%** of simulations reject H_0 at 5% level.

Key insight: Even though Y and X are **completely independent**, we reject the null ~76% of the time—far more than the nominal 5%!

(c) **Different sample sizes:**

T	R^2 (5%, 50%, 95%)	t (5%, 50%, 95%)	Rejection rate
50	(0.00, 0.16, 0.68)	(-8.3, 0.20, 7.8)	66%
100	(0.00, 0.19, 0.73)	(-13, 0.02, 13)	76%
200	(0.00, 0.17, 0.68)	(-17, 0.76, 17)	83%

Key findings:

- The rejection rate does **NOT** converge to 5%—it **increases** with T!
- The R^2 distribution is essentially unchanged with sample size.
- The t-statistic distribution becomes **more dispersed** as T grows.

Why spurious regression occurs:

Random walks are **non-stationary**. Standard regression asymptotics assume stationarity. With I(1) series:

- $\hat{\beta}_1$ does not converge to a normal distribution.
- The t-statistic has a non-standard distribution (Dickey-Fuller type).
- Using normal critical values (1.96) leads to massive over-rejection.

Solution: Test for unit roots first. If both series are I(1), either: (1) test for cointegration, or (2) difference the data before regression.