Diego Escobari
The University of Texas Rio Grande Valley
http://faculty.utrgv.edu/diego.escobari

Business and Economics Forecasting

Class Notes

ECON 3342

November 19, 2019

Contents

Chapter 1
Introduction to Forecasting

1.1 Introduction

What would happen if we could know more about the future? Forecasting is very important for:

- Business. Forecasting sales, prices, inventories, new entries.
- Finance. Forecasting financial risk, volatility forecasts. Stock prices?
- Economics. Unemployment, GDP, growth, consumption, investment.
- Governments. Tax revenues, population, infrastructure.

Use of data to forecast and types of data:

- Cross-section.
- Time series.
- Panel data.

Time-series data is a structure where observations of a variable or several variables are ordered in time (e.g., stock prices, money supply, consumer price index). Unlike cross-section data, observations are related. For example, knowing something about the GDP in the past can tell you something about the GDP in the future.

Data Frequency: Daily / weekly / monthly / quarterly / annually

Seasonal Patterns: Sales during Christmas / agricultural data.

Forecasting Methods: Before forecasting we need to build a statistical model.

Statistical Model. Describes the relationship between variables. It's parameters are estimated using historical data.
Forecasting Model. Characterization of what we expect on the present, conditional on the past. It can be used to infer about the future.

Table 1.1 Data for Texas

Observation	Year	Unemployment Rate	GDP	Population
1	1951	6.7%	543	8.11
2	1952	7.2%	549	8.21
3	1953	7.5%	551	8.27
4	1954	6.8%	556	8.31
⋮	⋮	⋮	⋮	⋮
65	2016	4.4%	1,498	26.91
66	2017	4.7%	1,524	27.22
67	2018	4.0%	1,547	28.35
68	2019	3.4%	1,581	28.74

GDP in Billions of US$. Population in millions.

Components of a time series model:

Trend. Long-term movement.
Seasonal. Movement that repeats every season.
Cycle. Irregular dynamic behavior.

Chapter 2
Main Statistical Concepts

2.1 Random Variables

Goals:

- Working with data.
- Become familiar with the data in hand.

Random Experiment: Process leading to two or more possible outcomes, with uncertainty as to which outcome will occur.

· Flip a coin. $\rightarrow$ 2 outcomes. Head (H) or Tail (T).
· Flip two coins. $\rightarrow$ 4 outcomes. (HH, HT, TH, TT).

Random Variable: Variable that takes numerical values determined by the outcome or a random experiment.

Random variable Y: Number of tails observed when flipping two coins.

Y: Random variable.
y: Realizations of the random variable.
$y = 0, 1, 2$.

Event: Subset of outcomes.

Sample Space: Sample space S is the set of all outcomes of the random experiment.

Probability: Given a random experiment, we want to determine the probability that a particular event will occur.

Probability is measured from 0 to 1.
$0 \rightarrow$ the event will not occur.

$1 \rightarrow$ the event is certain.

When all events are equally likely, the probability of event A is:

$$P(A) = \frac{1}{N} \tag{2.1}$$

where N is the number of outcomes in the sample space S.

Example 1) Flip a coin:

Define event A: "Head", then:

$$P(A) = \frac{1}{2} \tag{2.2}$$

where $N = 2$ is the number of outcomes "Head" or "Tail".
Example 2) Winning the lottery:

Define event B: Winning the lottery.

You buy 2 tickets from a total of 1,000 existing tickets. Then:

$$P(B) = \frac{2}{1,000} = 0.002 \tag{2.3}$$

There is a 1/500 chance that you win the lottery.

If A is an event in the sample space S,

$$0 \leq P(A) \leq 1 \tag{2.4}$$

Probability distribution function: $f(\cdot)$. The probability distribution function (p.d.f.) assigns a probability to each of the realizations of a random variable.

Example 3) Flip two coins: (HH, HT, TH, TT).

Define the random variable Y as the number of Tails. Hence:

$y = 0, 1, 2.$

$f(Y = 0) = 0.25$
$f(Y = 1) = 0.5$
$f(Y = 2) = 0.25$

Example 4) Toss a die.

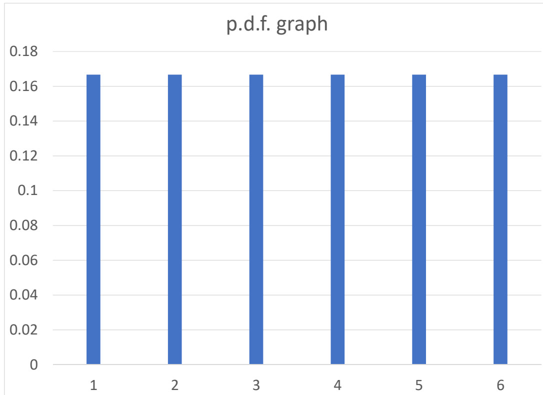


Fig. 2.1 Probability Density Function.

Define the random variable X as the number resulting from tossing a die. Hence:

$x = 1,2,3,4,5,6.$

$f(Y=1) = 1/6$
$f(Y=2) = 1/6$
$\vdots$
$f(Y=6) = 1/6$

Properties of the p.d.f.:

1) $0 \leq P(x_i) \leq 1$ for any x
2) $\sum_i P(x_i) = 1$

p.d.f. graph, $P(X=x)$, see Figure 2.1.

Mean of a random variable:

$$E(y) = \sum_i p_i y_i = \sum_i P(y_i) y_i \tag{2.5}$$

where $p_i = P(Y = y_i)$.

Example) Toss a die.

$$E(X) = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 = 3.5$$

$\mu = E(X)$ is a measure of central tendency.

Variance of a random variable:

$$\sigma^2 = Var(Y) = E(y-\mu)^2 \tag{2.6}$$

$\sigma^2 = Var(Y)$ is a measure of dispersion.

Example) Toss a die.

$$\begin{aligned} Var(X) &= \sum_i (x_i-\mu)^2 p(x_i) \\ &= (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + \cdots + (6-3.5)^2 \cdot \frac{1}{6} \\ &= 2.916 \end{aligned}$$

Standard deviation of a random variable: It is simply the square root of the variance.

$$\sigma = \sqrt{Var(Y)} = \sqrt{E(y-\mu)^2} \tag{2.7}$$

2.2 Multivariate Random Variables

What if instead of observing a single random variable X, we now jointly observe two random variables X and Y.

$f(X,Y) \rightarrow$ denotes the joint distribution of X and Y. It gives you the probability associated with each possible pair x and y.

Covariance: How are these two variables associated?

$$Cov(X,Y) = E[(y_t-\mu_y)(x_t-\mu_x)] \tag{2.8}$$

$$\begin{aligned} &Cov(X,Y) > 0 \quad \text{move together.} \\ &Cov(X,Y) < 0 \quad \text{move in opposite directions.} \end{aligned}$$

Correlation: Units-free measure of the association between variables.

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y} \tag{2.9}$$

where σ_x and σ_y are the standard deviations of X and Y respectively.

$$-1 \le Corr(X,Y) \le 1$$

Conditional distribution: What is the distribution of Y conditional on observing X?

$$f(Y|X) = \frac{f(X,Y)}{f(X)} \tag{2.10}$$

2.3 Statistics

Note that we do not know the true $f(X)$, $f(X,Y)$, $f(Y|X)$.

We have the sample $\{y_t\}_{t=1}^{T} \sim f(Y)$, where T is the sample size.

From these data we can obtain the following.

Sample mean:

$$\hat{\mu}_y = \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \tag{2.11}$$

Sample variance:

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2 \tag{2.12}$$

$$s^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})^2 \tag{2.13}$$

2.4 Regression Analysis

2.5 Simple Regression Model

Two variables: X and Y. See Figure 2.2.

X: Education.

Y: Wage.

The regression equation holds for every observation t:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{2.14}$$

β_0 and β_1 are unknown parameters.

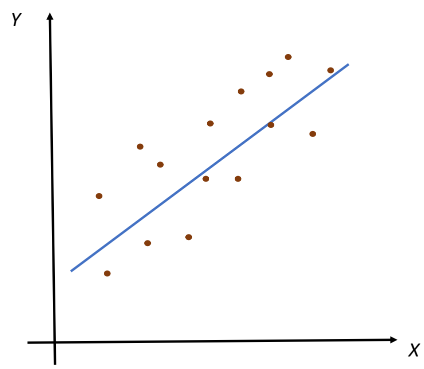


Fig. 2.2 Fitted regression line.

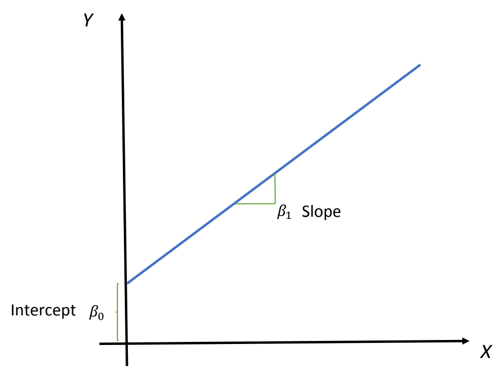


Fig. 2.3 Intercept and slope.

We need to estimate β_0 and β_1 from the data. See Figure 2.3.

The regression fitted values are given by:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t \tag{2.15}$$

Figure 2.4 illustrates the actual and the fitted values.

$$e_t = y_t - \hat{y}_t \tag{2.16}$$

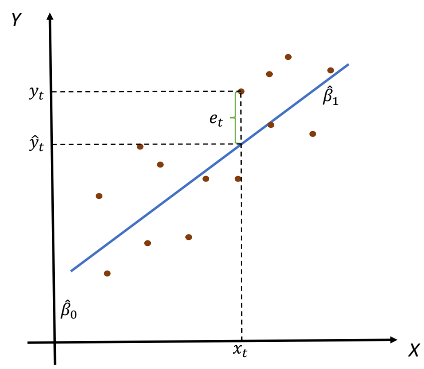


Fig. 2.4 Estimating β_0 and β_1.

where:
e_t: residuals or in-sample forecast errors.
y_t: actual values / true values.
$\hat{y}_t$: fitted values or in-sample forecast.

Ordinary Least Squares: obtains $\hat{\beta}_0$ and $\hat{\beta}_1$ by minimizing:

$$\min_{\beta_0,\beta_1} \sum_{t=1}^{T} \left(y_t - \beta_0 - \beta_1 x_t\right)^2 \tag{2.17}$$

In this simple case where there is a single right-hand side variable, the slope coefficient is obtained using:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{T} \left(x_t - \bar{x}\right)\left(y_t - \bar{y}\right)}{\sum_{t=1}^{T} \left(x_t - \bar{x}\right)^2} \tag{2.18}$$

and the constant is obtained from:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \tag{2.19}$$

Keep in mind that:

$$\left.\begin{array}{lll} \beta_0 & \text{and} & \beta_1 \end{array}\right\} \quad \text{are the true unknown parameters.}$$
$$\left.\begin{array}{lll} \hat{\beta}_0 & \text{and} & \hat{\beta}_1 \end{array}\right\} \quad \text{are the estimators of } \beta_0 \text{ and } \beta_1.$$

Specific values of $\hat{\beta}_0$ and $\hat{\beta}_1$ are called estimates (these are the ones obtained using econometrics software).
$\hat{\beta}_0$ and $\hat{\beta}_1$ are random variables and depend on the sample.

Hence, $\hat{\beta}_0$ and $\hat{\beta}_1$ have standard errors.

2.6 Multiple Regression Model

In the multiple regression model we have more than one right-hand side variables. In a model with two regressors x and z we have:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + \varepsilon_t. \tag{2.20}$$

Then the fitter values are:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{\beta}_2 z_t. \tag{2.21}$$

The error terms are assumed to be independent and identically distributed with mean zero and variance σ_ε^2:

$$\varepsilon_t \overset{\text{iid}}{\sim} (0, \sigma_\varepsilon^2) \tag{2.22}$$

The $\hat{\beta}_j$ in a multiple regression model can easily be obtained with econometrics software.

t-statistics: Provides a test that the true, but unknown, parameter β is equal to zero. That is: $H_0 : \beta = 0$.

$$\text{t-statistic} = \frac{\text{Coefficient}}{\text{Standard Error}} = \frac{\hat{\beta}}{\text{Std.Error}(\hat{\beta})} \tag{2.23}$$

Then you would need to compare it with the t-distribution.

Probability value: The p-value comes from comparing the t-statistics with the table t-distribution. It is the minimum confidence level at which the null $H_0 : \beta = 0$ is rejected.

Interpretation of β: Consider the following example. Here, wage_i is the hourly wage in US\$, while educ_i is the number of years of formal education.

$$\text{wage}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{educ}_i + \varepsilon_i$$

$\hat{\beta}_0$: This is the hourly wage of an individual with no formal education. That is, when $\text{educ}_i = 0$.
$\hat{\beta}_1$: This is the marginal effect of educ_i on wage_i. For every additional year of education, the hourly wage increases by $\hat{\beta}_1$.

Sum of Squared Residuals: (SSR) the amount of variance in the dependent variable (y) that is not explained by a regression model:

$$\text{SSR} = \sum_{t=1}^{T} e_t^2$$

where

$$e_t = y_t - \hat{y}_t.$$

We can add and subtract $\bar{y}$ from the right-hand side to get:

$$e_t = y_t - \bar{y} - (\hat{y}_t - \bar{y}).$$

We then square and sum across all observations in the sample to obtain:

$$\sum_{t-1}^{T} e_t^2 = \sum_{t-1}^{T} (y_t - \bar{y})^2 - \sum_{t-1}^{T} (\hat{y}_t - \bar{y})^2 + 0$$

Rearranging terms:

$$\sum_{t-1}^{T} (y_t - \bar{y})^2 = \sum_{t-1}^{T} e_t^2 + \sum_{t-1}^{T} (\hat{y}_t - \bar{y})^2 \tag{2.24}$$

we have that:

$\sum_{t-1}^{T} (y_t - \bar{y})^2$: is the Total Sum of Squares (TSS).

$\sum_{t-1}^{T} e_t^2$: is the Sum of Square Residuals (SSR).

$\sum_{t-1}^{T} (\hat{y}_t - \bar{y})^2$: is the Model Sum of Squares (MSS).

From Equation 2.24 we can observe that the total variation (TSS) on the left-hand side variable can be broken down into variation *not* explained by the more (SSR) and the variation that is explained by the model (MSS). This is also illustrated in Figure 2.5.

R-squared: Captures the proportion of the variation in y that is explained by the model:

$$R^2 = \frac{\sum_{t-1}^{T} (\hat{y}_t - \bar{y})^2}{\sum_{t-1}^{T} (y_t - \bar{y})^2} = 1 - \frac{\sum_{t-1}^{T} e_t^2}{\sum_{t-1}^{T} (y_t - \bar{y})^2}$$

Of course, $0 \leq R^2 \leq 1$.

Adjusted R-squared: Adjusted the R^2 to account for the degrees of freedom used in fitting the model:

$$\bar{R}^2 = 1 - \frac{\frac{1}{T-k} \sum_{t-1}^{T} e_t^2}{\frac{1}{T-1} \sum_{t-1}^{T} (y_t - \bar{y})^2}$$

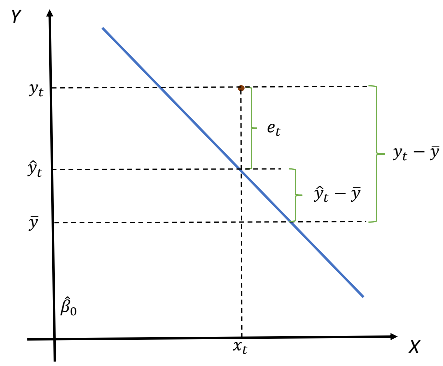


Fig. 2.5 Variation in the dependent variable y.

As more variables are included in the model, the R^2 will always increase. However, the $\bar{R}^2$ can either increase or decrease. Both, the R^2 and $\bar{R}^2$, are used as measures of the model fit.

Akaike Information Criterion: (AIC) it is effectively an estimate of the out-of-sample forecast variance. It has a high penalty for degrees of freedom:

$$\text{AIC} = e^{\frac{2k}{T}} \frac{\sum_{t-1}^{T} e_t^2}{T}.$$

Schwarz Information Criterion: (SIC) it is an alternative to the AIC, but has an even harsher degrees-of-freedom penalty:

$$\text{SIC} = T^{\frac{k}{T}} \frac{\sum_{t-1}^{T} e_t^2}{T}.$$

F-statistic: The most popular F-statistic is to test if all the slope coefficients are jointly equal to zero. That is, $H_0 : \beta_1 = \beta_2 = \cdots = \beta_j = 0$.

$$F = \frac{\left(\text{SSR}_{\text{restricted}} - SSR\right)/(k-1)}{SSR/(T-k)}$$

where T is the total number of observations, k is the number of slope coefficients, and SSR is the Sum of Squared Residuals. This F-statistic has also an associated p-value. Its interpretation is similar to the p-value of the t-statistic.

Dependent Variable: Y
Method: Least Squares
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	-5.515772	1.147782	-4.805594	0.0000
Z	11.18922	0.416949	26.83592	0.0000

R-squared	0.926941	Mean dependent var	885.0800
Adjusted R-squared	0.925419	S.D. dependent var	407.7874
S.E. of regression	111.3649	Akaike info criterion	12.30268
Sum squared resid	595303.0	Schwarz criterion	12.37916
Log likelihood	-305.5670	Hannan-Quinn criter.	12.33180
Durbin-Watson stat	0.176587		

Fig. 2.6 EViews regression output.

Consider the example presented in Figure 2.6. This computer output shows how the econometrics software will help us to quickly obtain all the statistics needed for the analysis.

Chapter 3
EViews: Basics

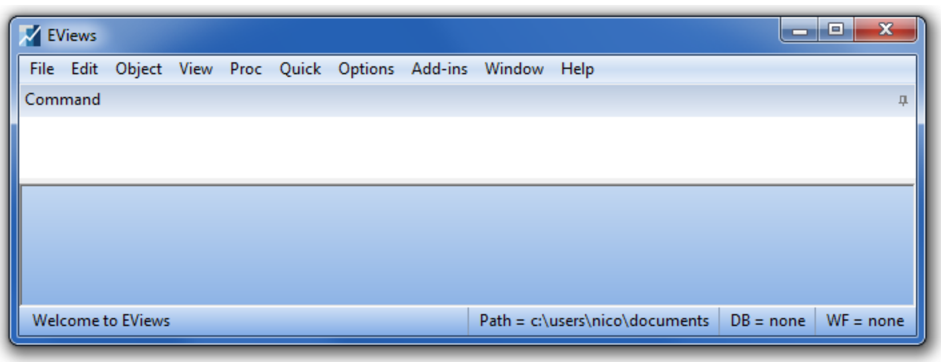
This chapter will cover the following points:

1. To get you familiar with EViews basics.
2. Learn how to import data to EViews.
3. Learn some basic commands to obtain summary statistics, line graphs, histograms.

3.1 Simple and multiple regression

EViews is a general purpose statistical software package. It is relatively easy for beginners who are starting with econometrics/time-series, but has some many more advance built-in procedures you may want to consider studying in the future.[1]

Once you open EViews, you will get the following screen:

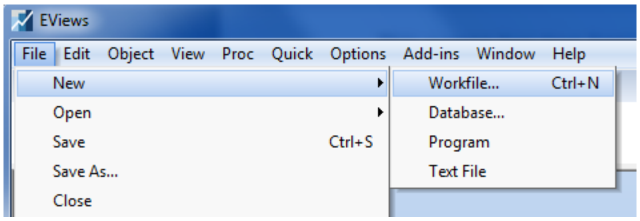


[1] These include time series analysis, panel data models, survival analysis, nonparametric methods, limited dependent variables and many more.

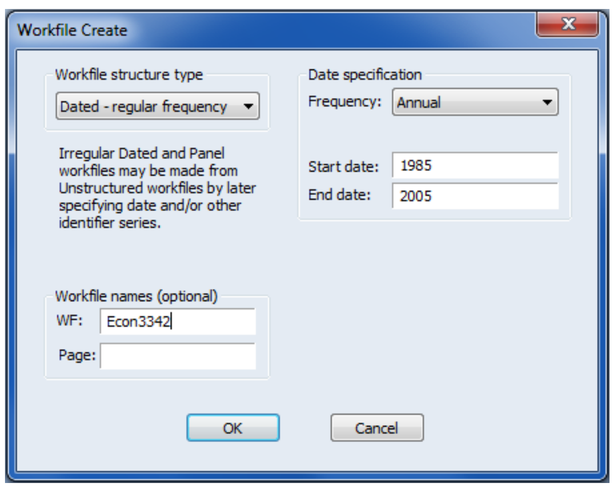
This screen is basically divided into two windows. The upper white portion is to type the commands and the lower portion of the screen is for the output and where you will see the data.

How to create a Workfile.

Before you are able to perform any operation, you need to create an EViews "Workfile."

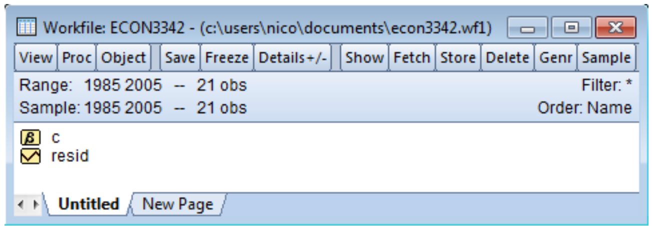


Recall the types of data econometricians work with? (1) Cross-section, (2) Time-series, and (3) Panel data. This class is all about time-series data, so you have to select "Dated - regular frequency."[2] For this example, we will be working with 21 yearly observations from 1985 to 2005.

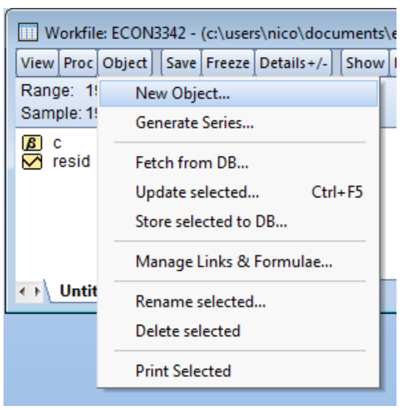


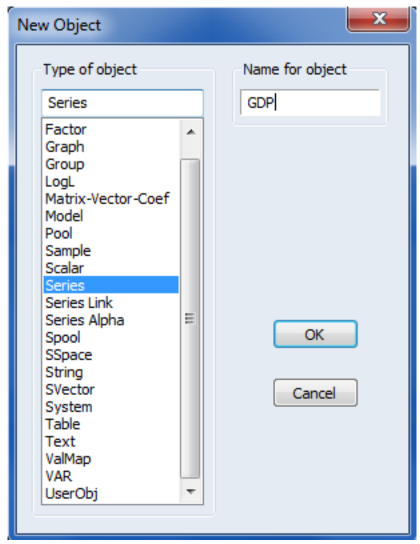
You should then have the following screen:

[2] Different versions of EViews may have a different outlay, but they should all perform these operations.



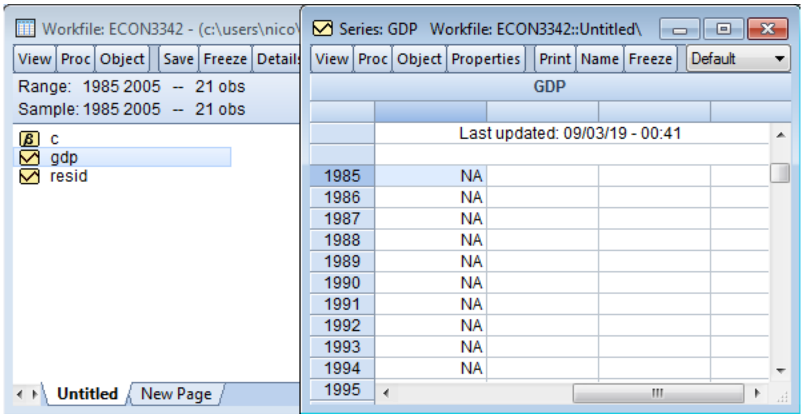
In order to create a new series, let's say GDP, you need to go to "Object" and select "New Object."





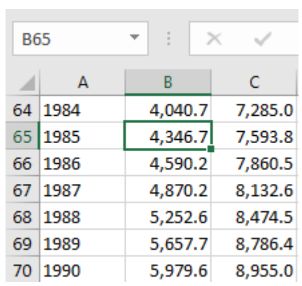
On a second screen you have to select "Series" as the type of object and select a name. In this case we decide the new name will be GDP.

If you click twice in the newly created series you will be able to see its content. Editing the series is simple and can be done by simply clicking the icon "edit." Then, typical features like "copy" and "paste" will be allowed, making it very easy to import data from any web page or, for example, MS Excel.



Let's get some real data! The Bureau of Economic Analysis website has a MS Excel file with real GDP data since the Great Depression. You can get the file directly from the following link:

http://www.bea.gov/national/xls/gdplev.xls.

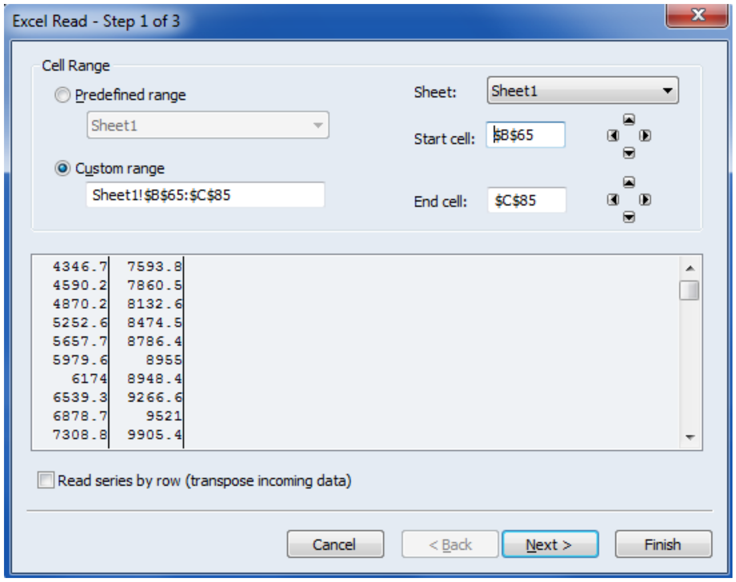


B65

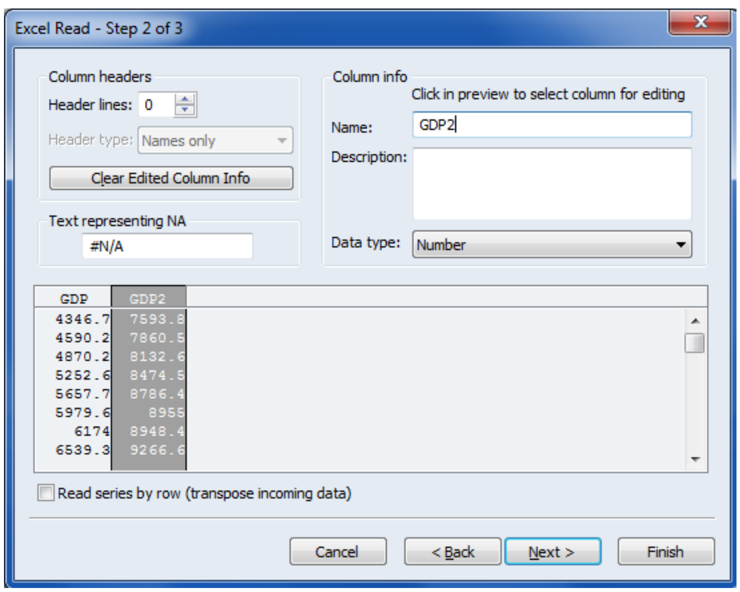
	A	B	C
64	1984	4,040.7	7,285.0
65	1985	4,346.7	7,593.8
66	1986	4,590.2	7,860.5
67	1987	4,870.2	8,132.6
68	1988	5,252.6	8,474.5
69	1989	5,657.7	8,786.4
70	1990	5,979.6	8,955.0

Save the Excel file on your computer to be able to import it with EViews. To get the GDP series into EViews go to "File", then to "Import" and select "Import from file..."

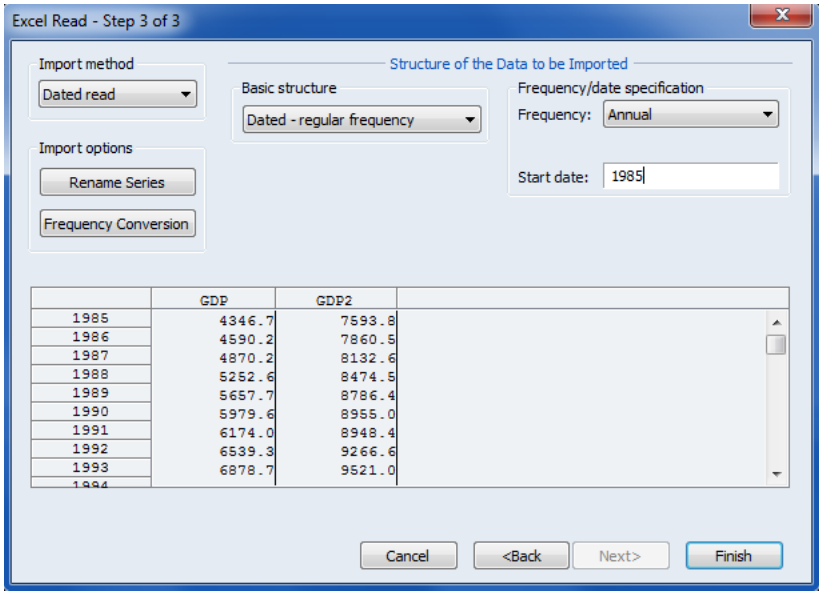
After selecting the Excel file from your computer you will be able to select the cells where the data starts and finishes.



Then select the names of the series.

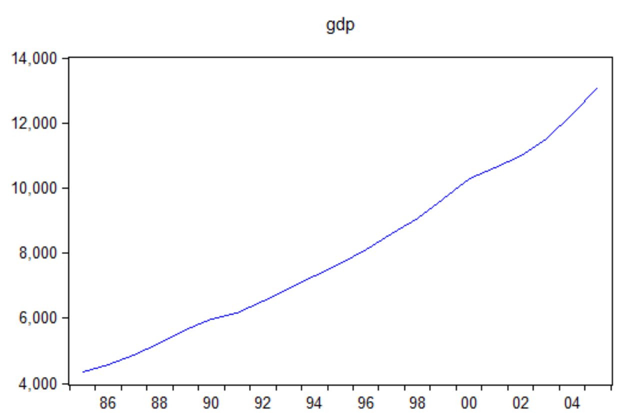


To finally tell EViews where the data starts. In this example, we selected it to start in 1985. Make sure you always correctly match the starting cell in Excel with the correct starting date.

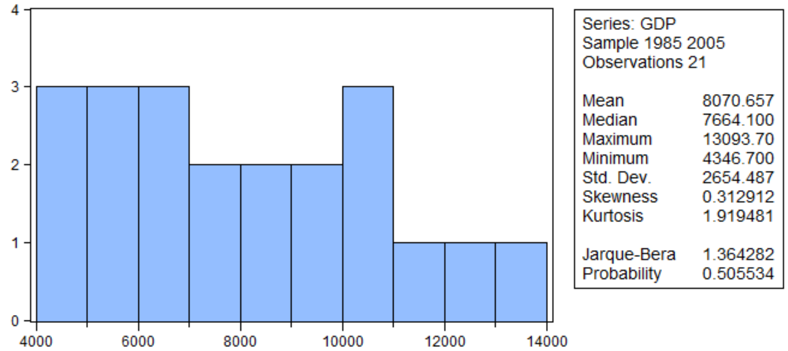


Note that there are various ways to successfully import data from an external source. We just described one way to do it. I encourage you to try other options to make sure you understand the steps.

Once your data is in EViews, playing with the options is very intuitive. For example, if you want a time-series graph of the GDP series, you just need to open the series and then select "View", then "Graphs...", and click OK on the default settings. You should be getting the following graph:



One easy way to obtain the sample descriptive statistics is to go to "View", then "Descriptive Statistics & Tests", and select "Histogram and Stats". The resulting is the following:



From this output you can see the sample (1985-2005), number of observations, and some simple statistics such as the mean, median, standard deviation, minimum and maximum.

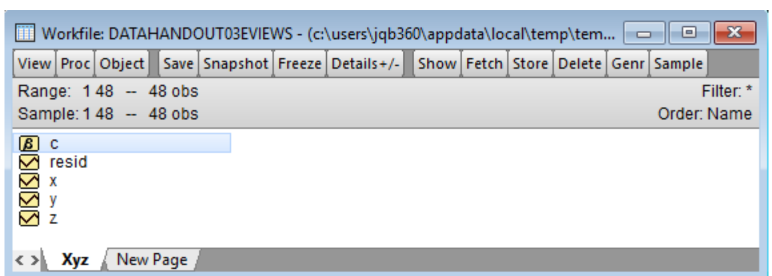
Chapter 4
EViews: Estimating a Regression Equation

This chapter will cover the following points:

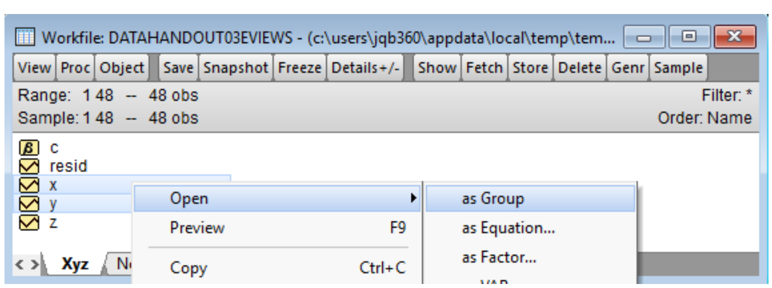
1. Scatter plots.
2. Linear regressions.

4.1 Scatter plots

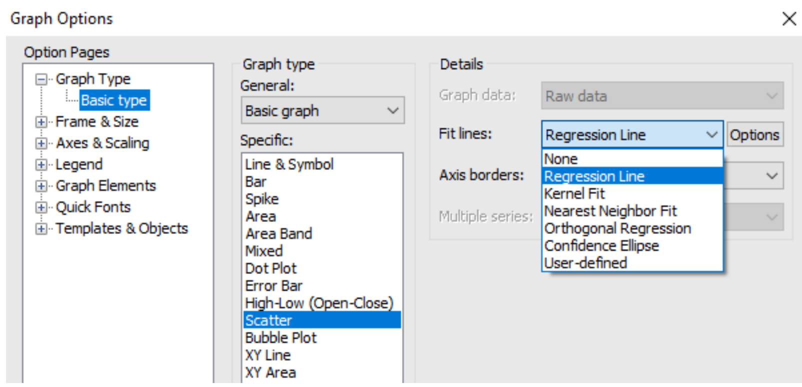
We will be using the data set under Handout 3 from the class website. The data set is already formatted for EViews (or gretl) and contains three variables: x, y and z:



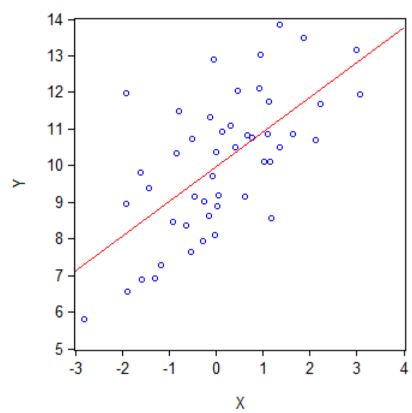
Open variables x and y as a group:



Then select "View," "Graph...," "Scatter," and then select the "Scatter" with "Regression Line" options.



You will then obtain the following figure. This one shows the data points in the sample along with the linear regression of y as a function of x.



4.2 Regression output

How is the linear regression line obtained? This is done easily by typing the following command:

```
LS Y C X Z
```

This is basically telling EViews to run a linear regression using Least Squares (LS) with y as the dependent variable and on a constant and on variables x and z. The regression output is as follows:

Dependent Variable: Y
Method: Least Squares
Sample: 1 48
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.884732	0.190297	51.94359	0.0000
X	1.073140	0.150341	7.138031	0.0000
Z	−0.638011	0.172499	−3.698642	0.0006
R-squared	0.552928	Mean dependent var		10.08241
Adjusted R-squared	0.533059	S.D. dependent var		1.908842
S.E. of regression	1.304371	Akaike info criterion		3.429780
Sum squared resid	76.56223	Schwarz criterion		3.546730
Log likelihood	−79.31472	Hannan-Quinn criter.		3.473976
F-statistic	27.82752	Durbin-Watson stat		1.506278
Prob(F-statistic)	0.000000			

Chapter 5
Considerations to Successful Forecasting

5.1 Decision Environment and Loss Function

· Forecasts are made to guide decisions.
· Getting the wrong answer is costly.

Example: Forecast airline demand.

· The seller needs to select between two aircrafts (big vs. small).
· There are two states of the demand (high vs. low).

	High Demand	Low Demand
100-seat aircraft	\$0	\$10,000
80-seat aircraft	\$10,000	\$0

Need to forecast the demand to decide whether to schedule the 100-seat aircraft or the 80-seat aircraft.

In this example there are only two demand states. What if we have a continuous range of values? Then, we need to consider:

$$e_t = y_t - \hat{y}_t \tag{5.1}$$

where:

· e_t: forecast error.
· y_t: actual value.
· $\hat{y}_t$: forecast.

Loss function: $L(e)$, a function of the forecast errors (e) that gives us the loss associated to forecasting.

We want three conditions for $L(e)$:

1. $L(0) = 0$: Perfect forecast gives us zero loss.
2. $L(e)$ is a continuous function.

3. $L(e)$ should punish $(+)$ as well as $(-)$ deviations.

   Quadratic loss: $L(e) = e^2$. Large errors are penalized more.

   Absolute loss: $L(e) = |e|$. All errors are penalized equally.

   In general $L(y, \hat{y})$. For example, in financial assets returns:

$$L(y,\hat{y}) = \begin{cases} 0 & \text{if} \quad sign(\Delta y) = sign(\Delta \hat{y}) \\ 1 & \text{if} \quad sign(\Delta y) \neq sign(\Delta \hat{y}) \end{cases}$$

   No loss if the sign is forecasted correctly (Note that $\Delta y = y_t - y_{t-1}$).

5.2 Forecast Object

a) Event outcome forecast. An event is certain but the outcome is uncertain.

Example: Event − Sunday weather. Outcome − rain / shine.

b) Event timing forecast. En event is certain and the outcome is known, but the timing is uncertain.

Example: It is not raining today and we know it will rain in the future, but we do not know when. Forecast when it will rain.

c) Time-series forecast. Project future values of a series.

Example: Forecast the amount of rain each month for the next 12 months given that we have historical data.

5.3 Forecast Statement

a) Point forecast. Forecast a single number.

Example: The inflation rate next month is forecasted at 0.3%

b) Interval forecast. A range in which we expect the realized value to fall.

Example: The 95% confidence interval forecast for the GDP growth rate is $[-2.6\%, 4.7\%]$.

c) Density forecast. Forecast the probability distribution.

Example:

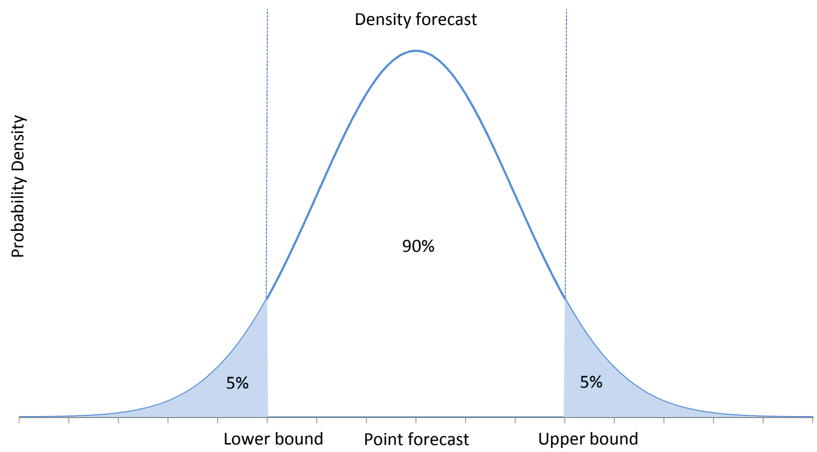


Fig. 5.1 Interval forecast and forecasting the probability distribution.

d) Probability forecast. Forecasts a probability (number between 0 and 1) of an event.

Example: Forecast the probability that it will rain on Sunday.

5.4 Forecast Horizon

The data set goes from $t = 1, 2, \ldots, T$.

The forecast could be for one period: $T + 1$ (1 step), or for two periods: $T + 2$ (2 steps).

h-step-ahead forecast is the forecast at period $T + h$ (only period $T + h$).

h-step-ahead extrapolation forecast is for h periods up until $T+h$ (all steps from 1 to h).

5.5 Information Set

Forecasts are conditional of the information set.

To forecast y_{T+1} we can use:

a) Univariate information set:

$$\Omega^{\text{Univariate}} = \{y_T, y_{T-1}, \ldots, y_2, y_1\} \tag{5.2}$$

a) Multivariate information set:

$$\Omega^{\text{Multivariate}} = \{x_T, x_{T-1}, \ldots, x_2, x_1, y_T, y_{T-1}, \ldots, y_2, y_1\} \tag{5.3}$$

5.6 Methods and Complexity

Key: Use the correct tool for the task in hand.

Parsimony principle: Simpler models are preferred. They are easier to estimate and interpret.

Shrinkage principle: Imposing restrictions on the forecast usually improves performance.

Chapter 6
EViews: In-sample Forecast

This chapter will cover the following points:

1. Simple and multiple regression.
2. In-sample forecast.
3. In-sample forecast errors.

6.1 Simple and multiple regression

We will be using the data set under Handout 4 from the class website. The data set is already formatted for EViews and contains for key components of U.S. real GDP: Manufacturing, retail, services, and agriculture. The series correspond to annual data from 1960 to 2001 measured in millions of dollars.

We want to estimate the following model to see how the agricultural GDP has been changing over the years:

$$agriculture_t = \beta_0 + \beta_1 year_t + \varepsilon_t \tag{6.1}$$

The variable $year_t$ takes the value of the corresponding year: 1960, 1961, ..., 2001.

To generate the variable year you have to type the following command:

```
genr year = @year
```

Now, to estimate the model in Equation 6.1, you have to type the command:

```
LS agriculture c year
```

to obtain the following regression output: Notice that the interpretation of the slope coefficient β_1 is the same as before: If year increases by one unit, then the agricultural GDP (aGDP) will increase by 3.12 million dollars. This means that in a given year the aGDP is about 3.12 million dollars greater than the aGDP the year before. The p-value indicates that the variable $year_t$ is statistically significant and the R^2 shows that time ($year_t$) explains 97% of the variation in aGDP.

Dependent Variable: AGRICULTURE
Method: Least Squares
Sample: 1960 2001
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−6119.273	165.4182	−36.99275	0.0000
YEAR	3.126007	0.083522	37.42740	0.0000

R-squared	0.972238	Mean dependent var	71.78352
Adjusted R-squared	0.971544	S.D. dependent var	38.89304
S.E. of regression	6.560855	Akaike info criterion	6.646567
Sum squared resid	1721.793	Schwarz criterion	6.729313
Log likelihood	−137.5779	Hannan-Quinn criter.	6.676897
F-statistic	1400.810	Durbin-Watson stat	1.298698
Prob(F-statistic)	0.000000		

What happened in the year zero? The aGDP is estimated to be negative 6,119 million dollars. Does that make sense? No! That's why you have to be very careful in using these type of models to predict out-of-sample values.

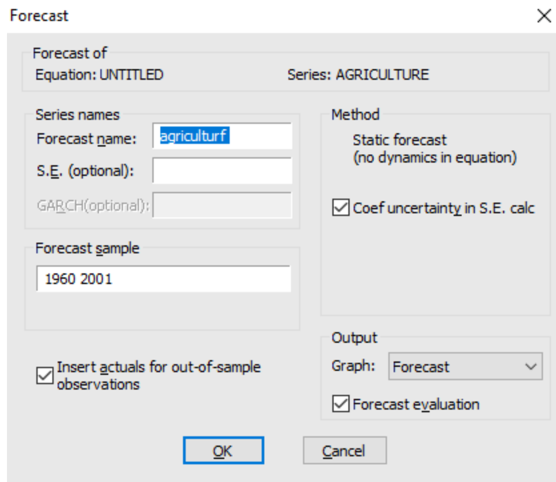
6.2 In-sample Forecast

Let's obtain the in-sample forecasted values for aGDP (agriculture):

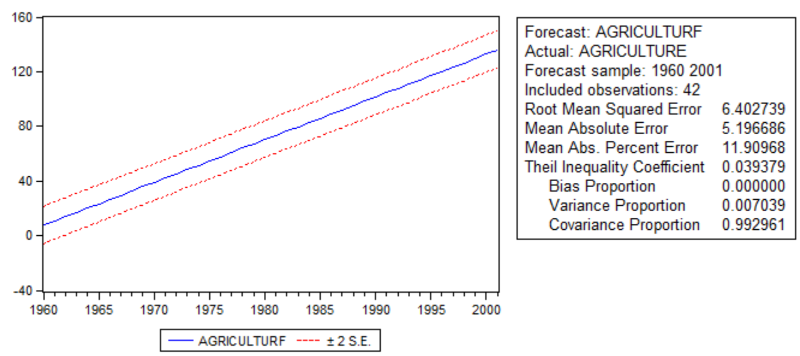
$$\widehat{agriculture}_t = \hat{\beta}_0 + \hat{\beta}_1 year_t \tag{6.2}$$

$$\widehat{agriculture}_t = 6,119.273 + 3.126 year_t$$

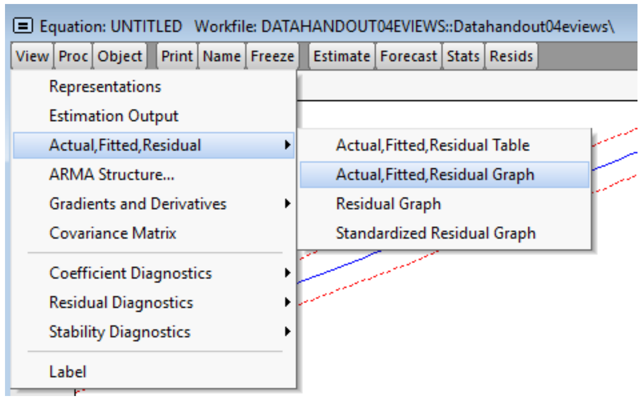
This can be done by simply selecting the "Forecast" icon while keeping the default options:



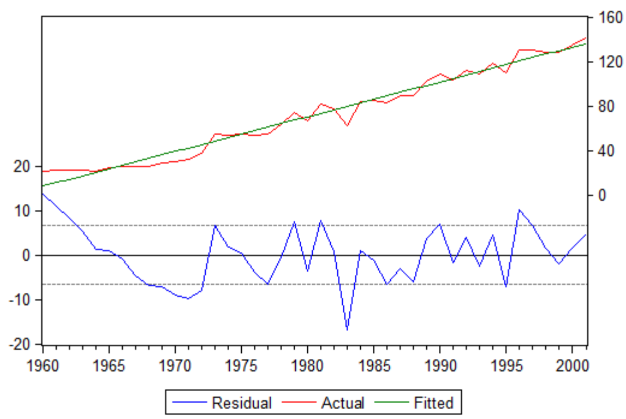
EViews will obtain:



and more importantly, EViews generated the variable “agricurturf” that contains the in-sample forecasted values. The difference between ”agriculture” and ”agriculturf” corresponds to the forecasting errors and this variable is automatically stored under “resid.” You can obtain a graph of all these three components (actual value = agriculture, fitted value = agriculturf, forecasting error = resid) by selecting the following option:


Equation: UNTITLED Workfile: DATAHANDOUT04EVIEWS::Datahandout04eviews\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
Representations
Estimation Output
Actual,Fitted,Residual
ARMA Structure...
Gradients and Derivatives
Covariance Matrix
Coefficient Diagnostics
Residual Diagnostics
Stability Diagnostics
Label
Actual,Fitted,Residual Table
Actual,Fitted,Residual Graph
Residual Graph
Standardized Residual Graph

To obtain:


160
120
80
40
0
20
10
0
-10
-20
1960 1965 1970 1975 1980 1985 1990 1995 2000
Residual Actual Fitted

Chapter 7
EViews: Importance of Graphics for Forecasting

This chapter will show the importance of using graphical tool before engaging into sophisticated statistical forecasting.

Consider the following variables, available under Handout 5 on the class website:

	X1	X2	X3	X4	Y1	Y2	Y3	Y4
1	10.00000	10.00000	10.00000	8.000000	8.040000	9.140000	7.460000	6.580000
2	8.000000	8.000000	8.000000	8.000000	6.950000	8.140000	6.770000	5.760000
3	13.00000	13.00000	13.00000	8.000000	7.580000	8.740000	12.74000	7.710000
4	9.000000	9.000000	9.000000	8.000000	8.810000	8.770000	7.110000	8.840000
5	11.00000	11.00000	11.00000	8.000000	8.330000	9.260000	7.810000	8.470000
6	14.00000	14.00000	14.00000	8.000000	9.960000	8.100000	8.840000	7.040000
7	6.000000	6.000000	6.000000	8.000000	7.240000	6.130000	6.080000	5.250000
8	4.000000	4.000000	4.000000	19.00000	4.260000	3.100000	5.390000	12.50000
9	12.00000	12.00000	12.00000	8.000000	10.84000	9.130000	8.150000	5.560000
10	7.000000	7.000000	7.000000	8.000000	4.820000	7.260000	6.420000	7.910000
11	5.000000	5.000000	5.000000	8.000000	5.680000	4.740000	5.730000	6.890000

In these data you have four pairs of y and x variables. Let's estimate the following model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{7.1}$$

Using any of the different y and x pairs, you will obtain the following output:

Dependent Variable: Y1
Method: Least Squares
Sample: 1 11
Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.000091	1.124747	2.667348	0.0257
X1	0.500091	0.117906	4.241455	0.0022

R-squared	0.666542	Mean dependent var	7.500909
Adjusted R-squared	0.629492	S.D. dependent var	2.031568
S.E. of regression	1.236603	Akaike info criterion	3.425579
Sum squared resid	13.76269	Schwarz criterion	3.497924
Log likelihood	−16.84069	Hannan-Quinn criter.	3.379976
F-statistic	17.98994	Durbin-Watson stat	3.212290
Prob(F-statistic)	0.002170		

That corresponds to the following estimated equation:

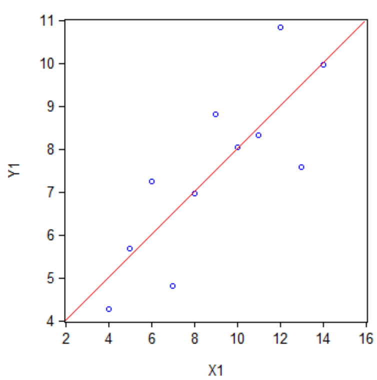
$$\hat{y1}_i = 3 + 0.5x1_i$$

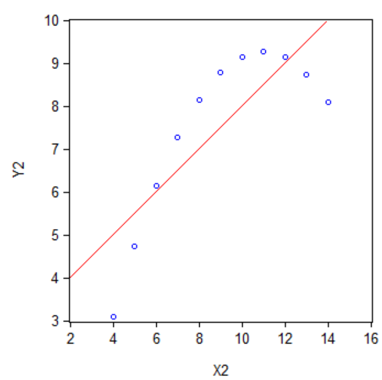
which holds for any pair. That is:

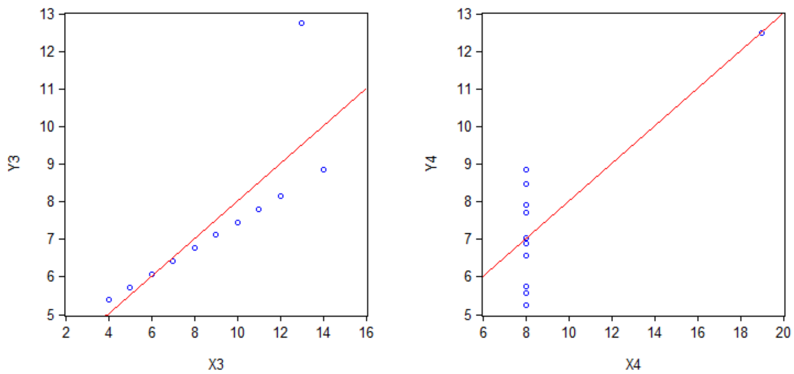
$$\hat{y2}_i = 3 + 0.5x2_i \qquad \hat{y3}_i = 3 + 0.5x3_i \qquad \hat{y4}_i = 3 + 0.5x4_i$$

Moreover, you will also get the same R^2 as well as the same standard errors, t-statistics and p-values.

What's the problem with this? There doesn't seem to be any problem, you may think, as different pairs of x and y can give exactly the same regression equation. The problem becomes clear when you graph the data:






Y3
X3
Y4
X4

Chapter 8
Modeling and Forecasting Trend

8.1 Modeling Trend

Trend: Long-run evolution in a variable.

The dynamics of a series can be broadly separated into a trend, a seasonal component, and the cyclical component.

Deterministic Trend: It is a predicable trend.

Linear Trend.

$$T_t = \beta_0 + \beta_1 TIME_t, \tag{8.1}$$

where β_0 is the intercept and β_1 is the slope (so we can have an increasing or decreasing series).

Quadratic Trend.

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^2. \tag{8.2}$$

It is a local approximation of a “U-shaped” trend.

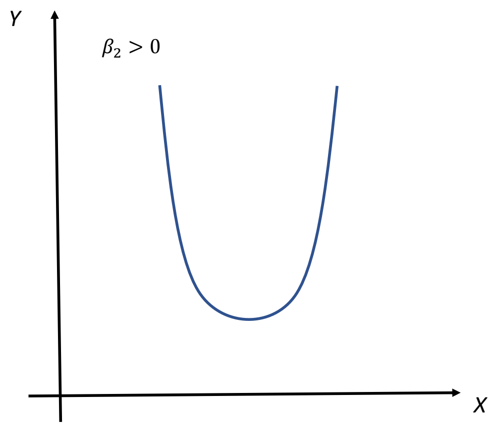


Fig. 8.1 Quadratic trend with $\beta_2 > 0$.

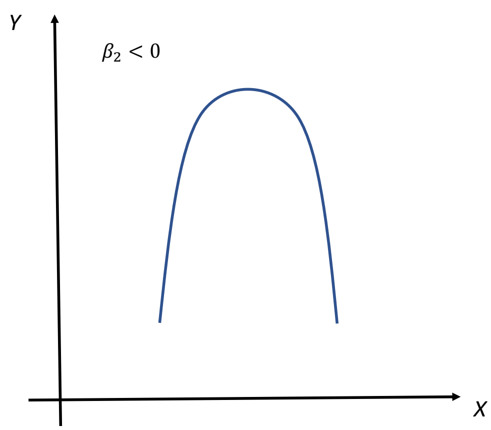


Fig. 8.2 Quadratic trend with $\beta_2 > 0$

Cubic Trend.

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^2 + \beta_3 TIME_t^3. \tag{8.3}$$

Exponential of Log-linear Trend. Economic variables sometimes grow at a constant rate β_1.

$$T_t = \beta_0 e^{\beta_1 TIME_t}, \tag{8.4}$$

where the trend is a exponential function of time. Taking natural logarithms of both sides we have:

$$\log(T_t) = \log(\beta_0) + \beta_1 \log(e^{TIME_t}) \tag{8.5}$$

$$\log(T_t) = \log(\beta_0) + \beta_1 TIME_t \tag{8.6}$$

as $\log(e) = 1$.

8.2 Estimating Trend Models

We can easily fit various trend models using ordinary least squares. Any computer software should be able to estimate:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_t - T_t(\theta)\right)^2, \tag{8.7}$$

where θ is just the set of parameters to be estimated. For example, in the quadratic trend of Equation 8.2, $\theta = (\beta_0, \beta_1, \beta_2)$. In this case the computer will find:

$$\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right) = \underset{\beta_0,\beta_1,\beta_2}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_t - \beta_0 - \beta_1 TIME_t - \beta_2 TIME_t^2\right)^2. \tag{8.8}$$

8.3 Forecasting Trend

Consider the following linear trend model:

$$y_t = \beta_0 + \beta_1 TIME_t + \varepsilon_t, \tag{8.9}$$

which holds for any time t. Hence, for time $T+h$ in the future we have:

$$y_{T+h} = \beta_0 + \beta_1 TIME_{T+h} + \varepsilon_{T+h}. \tag{8.10}$$

After obtaining estimates of β_0 and β_1 via least squares, on the right-hand side of this equation we have:

$TIME_{T+h} \rightarrow$ known at time T.
$\varepsilon_{T+h} \rightarrow$ unknown at time T.

We replace ε_{T+h} with 0 in Equation 8.10 as it has expected value zero.

Point Forecast: We can use the following point forecast:

$$\hat{y}_{T+h,T} = \hat{\beta}_0 + \hat{\beta}_1 TIME_{T+h}. \tag{8.11}$$

where the subscript "$T+h,T$" on $\hat{y}_{T+h,T}$ just emphasizes that the forecast of period $T+h$ is done at period T.

Interval Forecast: If we assume that the trend regression disturbance is normally distributed, in which case a 95% interval forecast is:

$$y_{T+h,T} \pm 1.96\sigma, \tag{8.12}$$

where σ is the standard deviation of the disturbance term. To make this operational we use:

$$\hat{y}_{T+h,T} \pm 1.96\hat{\sigma}, \tag{8.13}$$

with $\hat{\sigma}$ being an estimate of σ.

Density Forecast: Under the assumption that the trend regression is normally distributed, the density forecast is given by:

$$N\left(\hat{y}_{T+h,T}, \hat{\sigma}^2\right) \tag{8.14}$$

8.4 Model Selection Criteria

How do we select between competing models? Minimizing the Mean Squared Error (MSE):

$$MSE = \frac{\sum_{t=1}^{T} e_t^2}{T}, \tag{8.15}$$

is the same as maximizing the R^2:

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} \left(y_t - \bar{y}\right)^2} \tag{8.16}$$

Moreover, improving the "fit" of historical data usually does not help in improving the out-of-sample forecasting. Hence, alternative involve the adjusted R^2 ($\bar{R}^2$), which adjusts for the degrees of freedom.

We can use the Akaike Information Criterion (AIC):

$$\text{AIC} = e^{\frac{2k}{T}} \frac{\sum_{t-1}^{T} e_t^2}{T},$$

and the Schwarz Information Criterion (SIC):

$$\text{SIC} = T^{\frac{k}{T}} \frac{\sum_{t-1}^{T} e_t^2}{T}.$$

where k is the number of parameters to be estimated and $(2k/T)$ and (k/T) work as penalty factors. The idea is to select the model that gives the smallest AIC or SIC.

Chapter 9
EViews: Modeling and Forecasting Trend

This chapter will compare models with different trend structures and illustrate the use of the AIC and the SIC as two forms of selection criteria.

9.1 Comparing Trend Models

The variable of interest is the volume on the New York Stock Exchange.

Linear trend: Type and run the command:

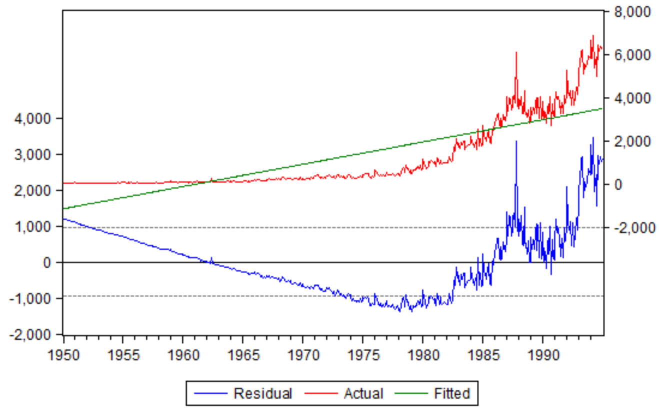
```
ls nysevol c @trend
```

To obtain:

Dependent Variable: NYSEVOL
Method: Least Squares
Sample: 1950M01 1994M12
Included observations: 540

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−6311.367	227.6358	−27.72572	0.0000
@TREND	8.592274	0.257692	33.34316	0.0000

R-squared	0.673893	Mean dependent var	1159.615
Adjusted R-squared	0.673287	S.D. dependent var	1633.118
S.E. of regression	933.4706	Akaike info criterion	16.51939
Sum squared resid	$4.69E+08$	Schwarz criterion	16.53529
Log likelihood	−4458.236	Hannan-Quinn criter.	16.52561
F-statistic	1111.766	Durbin-Watson stat	0.113092
Prob(F-statistic)	0.000000		



Quadratic trend: Type and run the command:

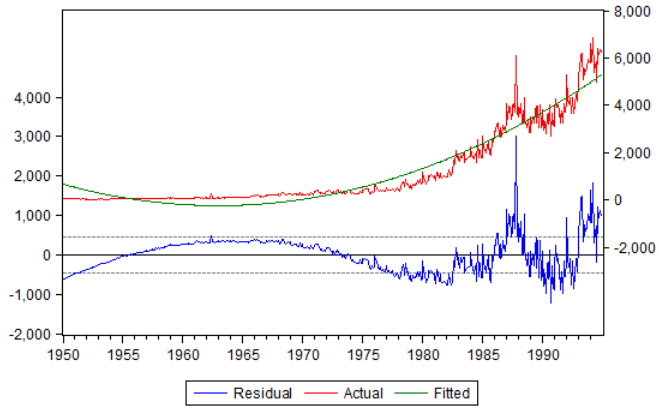
```
ls nysevol c @trend @trend^2
```

To obtain:

Dependent Variable: NYSEVOL
Method: Least Squares
Sample: 1950M01 1994M12
Included observations: 540

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	21239.88	656.3047	32.36284	0.0000
@TREND	−56.88488	1.543046	−36.86532	0.0000
@TREND^2	0.037652	0.000884	42.56987	0.0000

R-squared	0.925456	Mean dependent var	1159.615
Adjusted R-squared	0.925178	S.D. dependent var	1633.118
S.E. of regression	446.7168	Akaike info criterion	15.04727
Sum squared resid	$1.07E+08$	Schwarz criterion	15.07111
Log likelihood	−4059.762	Hannan-Quinn criter.	15.05659
F-statistic	3333.379	Durbin-Watson stat	0.493887
Prob(F-statistic)	0.000000		



Cubic trend: Type and run the command:

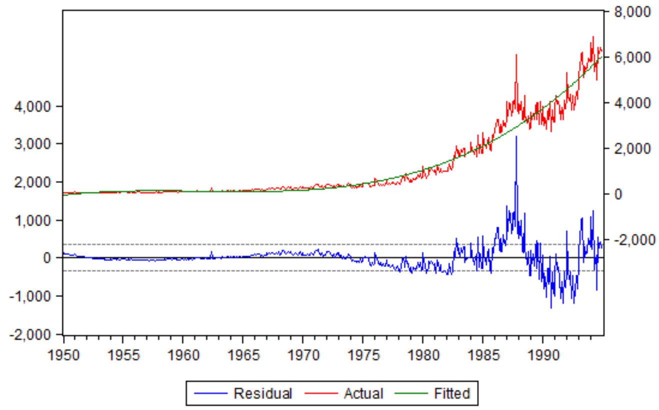
```
ls nysevol c @trend @trend^2 @trend^3
```

To obtain:

Dependent Variable: NYSEVOL
Method: Least Squares
Sample: 1950M01 1994M12
Included observations: 540

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−37461.26	3141.303	−11.92539	0.0000
@TREND	153.9406	11.19722	13.74810	0.0000
@TREND^2	−0.209583	0.013074	−16.03063	0.0000
@TREND^3	$9.48E-05$	$5.01E-06$	18.93661	0.0000

R-squared	0.955336	Mean dependent var	1159.615
Adjusted R-squared	0.955086	S.D. dependent var	1633.118
S.E. of regression	346.1037	Akaike info criterion	14.53873
Sum squared resid	64206230	Schwarz criterion	14.57052
Log likelihood	−3921.458	Hannan-Quinn criter.	14.55117
F-statistic	3821.611	Durbin-Watson stat	0.823825
Prob(F-statistic)	0.000000		



Fourth power trend: Type and run the command:

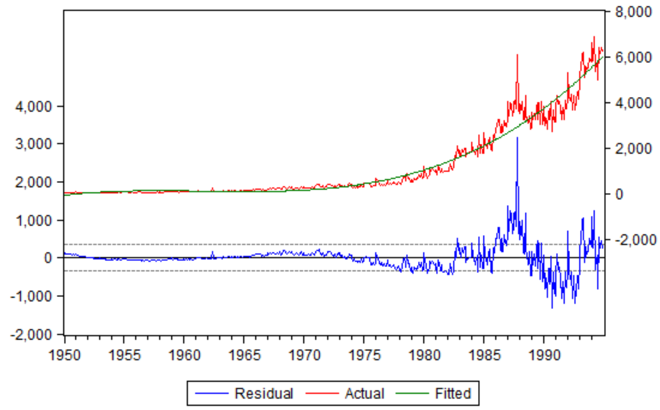
```
ls nysevol c @trend @trend^2 @trend^3 @trend^4
```

To obtain:

Dependent Variable: NYSEVOL
Method: Least Squares
Sample: 1950M01 1994M12
Included observations: 540

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-40938.43	19576.47	-2.091206	0.0370
@TREND	170.6429	93.48719	1.825307	0.0685
@TREND^2	-0.239225	0.165235	-1.447789	0.1483
@TREND^3	0.000118	0.000128	0.919407	0.3583
@TREND^4	$-6.63E-09$	$3.68E-08$	-0.179956	0.8573

R-squared	0.955339	Mean dependent var	1159.615
Adjusted R-squared	0.955005	S.D. dependent var	1633.118
S.E. of regression	346.4165	Akaike info criterion	14.54238
Sum squared resid	64202344	Schwarz criterion	14.58211
Log likelihood	-3921.442	Hannan-Quinn criter.	14.55792
F-statistic	2861.042	Durbin-Watson stat	0.823879
Prob(F-statistic)	0.000000		



Comparing the fit of different models for the trend we have:

	Linear	Quadratic	Cubic	Four	Five
R-squared	0.6739	0.9255	0.9553	0.9553	0.9561
Adjusted R-squared	0.6733	0.9252	0.9551	0.9550	0.9557
S.E. of regression	933.4706	446.7168	346.1037	346.4165	343.6152
Akaike info criterion (AIC)	16.5194	15.0473	14.5387	14.5424	14.5280
Schwarz critetion (SIC)	16.5353	15.0711	14.5705	14.5821	14.5757

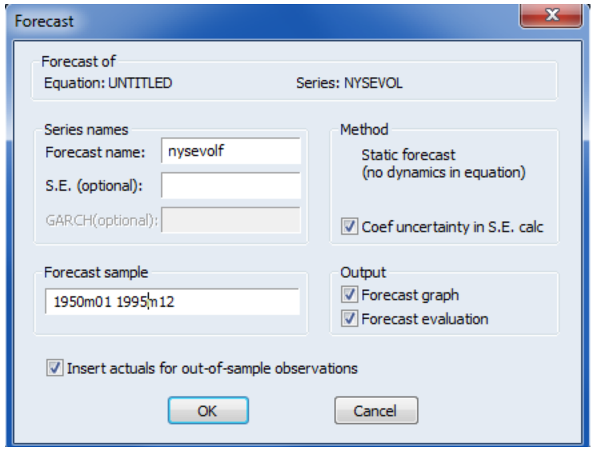
The R-squared will always increase as we include more variables into the model, hence does not work as a model selection criterion.

The Adjusted R-squared and the Standard Error of the regression do penalize for the inclusion of more variables into the model (which decreases the degrees of freedom), but the penalty is not severe enough. They can increase or decrease as more variables are included.

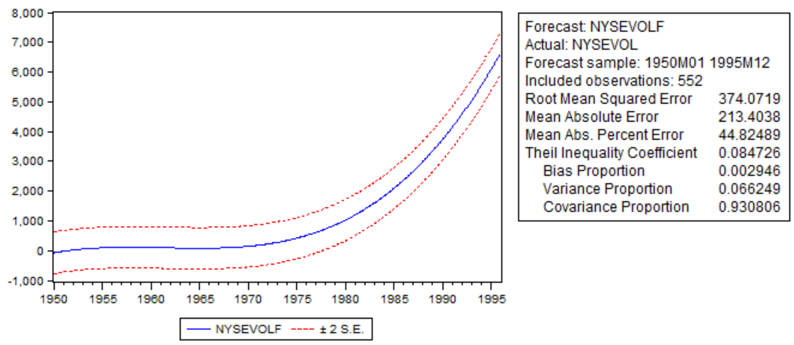
The AIC and the SIC can increase or decrease as more variables are included. The selected model should be the one that has the smallest AIC and SIC. When they do not select the same model, the parsimonious model should be selected. That is, the one with the least number of estimated parameters and this will be given by the SIC. In the models above, AIC selects the fifth specification, but SIC selects the cubic specification. We pick the parsimonious model: the cubic trend model.

9.2 Forecasting

With the cubic trend as our selected model we now aim at getting the out-of-sample point forecast values. After estimating the equation, just click on “Forecast” and make sure the “Forecasting sample” contains some values into the future:



To obtain:



The dotted red lines are the one standard deviation confidence intervals. Notice that the forecast spans for an additional year (the twelve months of 1995). Moreover, remember that the variable NYSEVOLF contains the values of the point forecasts.

Chapter 10
Modeling and Forecasting Seasonality

10.1 Nature and Sources of Seasonality

Seasonality: A seasonal pattern is one that repeats itself every year (or season, week, month).

Deterministic Seasonality: The annual repetition can be exact. This is different from *stochastic seasonality* in which the repetition is approximate. This chapter focuses on deterministic seasonality.

Examples:

· Retail sales are usually higher during the Christmas season.
· More travelers fly during weekends.
· Tax collection peaks in April.
· Weather $\rightarrow$ Summer / winter.

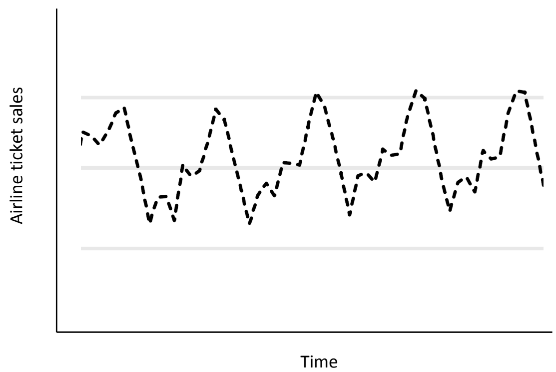


Fig. 10.1 Seasonality in air ticket sales.

10.2 Modeling Seasonality

Regression using seasonal dummies.

Dummy variable = 1 during some periods (e.g., weekends).
Dummy variable = 0 the rest of the time.

Consider the following example.

Table 10.1 Quarterly Sales Data

Observation	Sales	Quarter	Year	D_1	D_2	D_3	D_4
1	58	1	2017	1	0	0	0
2	63	2	2017	0	1	0	0
3	72	3	2017	0	0	1	0
4	53	4	2017	0	0	0	1
5	57	1	2018	1	0	0	0
6	62	2	2018	0	1	0	0
7	75	3	2018	0	0	1	0
8	58	4	2018	0	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Sales in thousands of $.

The dummies will capture the deterministic seasonal effect.

$$\text{Sales}_t = \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + \gamma_4 D_{4t} + \varepsilon_t \tag{10.1}$$

The pure seasonal dummy model is:

$$y_t = \sum_{t=1}^{s} \gamma_i D_{it} + \varepsilon_t \tag{10.2}$$

Hence, for $s = 4$, Equation 10.2 reduces to Equation 10.1 as $\sum_{i=1}^{s} \gamma_i D_{it} = \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + \gamma_4 D_{4t}$. Note that we can modify Equation 10.2 to additionally include a trend:

$$y_t = \beta_1 TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t \tag{10.3}$$

Holiday variation: Dummies for specific holidays. For example,

$HD = 1$: if Thanksgiving.
$HD = 0$: otherwise.

10.3 Forecasting Seasonal Series

Consider the model:

$$y_t = \beta_1 TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \sum_{i=1}^{v} \delta_i HD_{it} + \varepsilon_t \tag{10.4}$$

where $TIME_t$ is the linear time trend, $\sum_{i=1}^{s} \gamma_i D_{it}$ captures the seasonal variation, and $\sum_{i=1}^{v} \delta_i HD_{it}$ captures the holiday variation. ε_t is the remaider stochastic term.

Equation 10.4 holds for every time t, so at time $T + h$ we have:

$$y_{T+h} = \beta_1 TIME_{T+h} + \sum_{i=1}^{s} \gamma_i D_{i,T+h} + \sum_{i=1}^{v} \delta_i HD_{i,T+h} + \varepsilon_{T+h} \tag{10.5}$$

At time T (i.e., the moment we forecast), we have:

$TIME_{T+h} \rightarrow$ known at time T.
$D_{i,T+h} \rightarrow$ known at time T.
$HD_{i,T+h} \rightarrow$ known at time T.
$\varepsilon_{T+h} \rightarrow$ unknown at time T.

We replace ε_{T+h} with 0 in Equation 10.4 as it has expected value zero.

The forecast of y_{T+h} made at time T is:

$$\hat{y}_{T+h,T} = \hat{\beta}_1 TIME_{T+h} + \sum_{i=1}^{s} \hat{\gamma}_i D_{i,T+h} + \sum_{i=1}^{v} \hat{\delta}_i HD_{i,T+h} \tag{10.6}$$

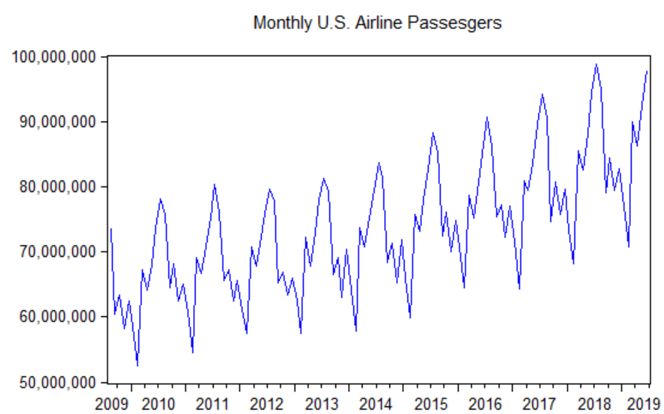
where $\hat{\beta}_1$, $\hat{\gamma}_i$, and $\hat{\delta}_i$ denote the estimates obtained via ordinary least squares using historical data.

Chapter 11
EViews: Modeling and Forecasting Seasonality

This chapter will show the use of dummy variables to model and forecast seasonality.

Let the variable of interest be the monthly total number of air passengers transported in U.S. domestic and international flights. We have data from August, 2009 up until June, 2019. This is a typical variable that has seasonal fluctuations in addition to a potential trend.

The following time series graph of illustrates the importance of seasonal component in this variable:



This variable is contained in the EViews file “passengers.wf1” along with some dummy variables. Part of the data showing the dummy variables is as follows:

	AIRPASS	D01	D02	D03	D04	D05
2009M08	73607921	0	0	0	0	0
2009M09	60512481	0	0	0	0	0
2009M10	63325757	0	0	0	0	0
2009M11	58170882	0	0	0	0	0
2009M12	62377082	0	0	0	0	0
2010M01	58655574	1	0	0	0	0
2010M02	52438942	0	1	0	0	0
2010M03	67304853	0	0	1	0	0
2010M04	64062751	0	0	0	1	0
2010M05	67970934	0	0	0	0	1

Note the 0/1 nature of the dummy variables. For example, $D2$ is equal to one when the month is February, zero otherwise.

11.1 Failing to Model Seasonality

If we estimate a naive econometric model that just accounts for a linear trend we would type:

```
ls airpass c @trend
```

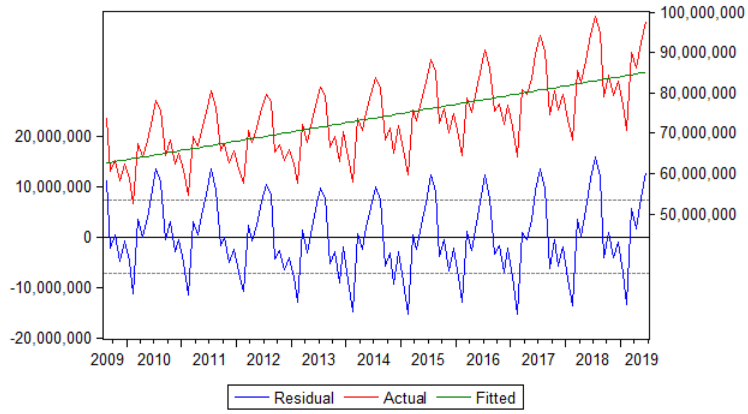
To obtain:

Dependent Variable: AIRPASS
Method: Least Squares
Sample: 2009M08 2019M06
Included observations: 119

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	62547562	1341965.	46.60894	0.0000
@TREND	190748.7	19656.29	9.704206	0.0000

R-squared	0.445948	Mean dependent var	73801737
Adjusted R-squared	0.441213	S.D. dependent var	9853554.
S.E. of regression	7365736.	Akaike info criterion	34.47924
Sum squared resid	$6.35E+15$	Schwarz criterion	34.52595
Log likelihood	-2049.515	Hannan-Quinn criter.	34.49821
F-statistic	94.17161	Durbin-Watson stat	0.982941
Prob(F-statistic)	0.000000		

This regression model yields the following actual, fitted and residuals graph:



This model allows controlling for the trend, but it still misses to account for the systematic fluctuations that appear every year.

11.2 Modeling Seasonality with Dummies

The econometric model that accounts for the seasonal variation is:

```
ls airpass d01 d02 d03 d04 d05 d06 d07 d08 d09 d10 d11 d12
```

With the regression output given by:

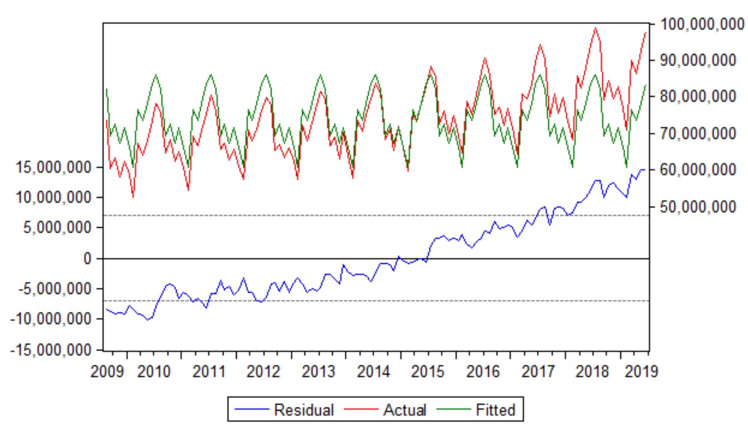
Dependent Variable: AIRPASS
Method: Least Squares
Sample: 2009M08 2019M06
Included observations: 119

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D01	66392856	2222332.	29.87530	0.0000
D02	60738127	2222332.	27.33080	0.0000
D03	76392041	2222332.	34.37472	0.0000
D04	73361346	2222332.	33.01097	0.0000
D05	78099360	2222332.	35.14297	0.0000
D06	83312966	2222332.	37.48898	0.0000
D07	86113225	2342544.	36.76056	0.0000
D08	82046928	2222332.	36.91929	0.0000
D09	69226001	2222332.	31.15016	0.0000
D10	72449763	2222332.	32.60078	0.0000
D11	67211336	2222332.	30.24360	0.0000
D12	71508042	2222332.	32.17702	0.0000

R-squared	0.538753	Mean dependent var	73801737
Adjusted R-squared	0.491335	S.D. dependent var	9853554.
S.E. of regression	7027632.	Akaike info criterion	34.46398
Sum squared resid	$5.28E+15$	Schwarz criterion	34.74423
Log likelihood	-2038.607	Hannan-Quinn criter.	34.57778
Durbin-Watson stat	0.035139		

Table 11.1 Regression model of air passengers transported as a function of seasonal dummies.

In this Table 11.1 we can see how the coefficients of the dummy variables explain about 53.9% of the total variation in air passengers. Note that we not include a constant to avoid having a problem of multicollinearity.

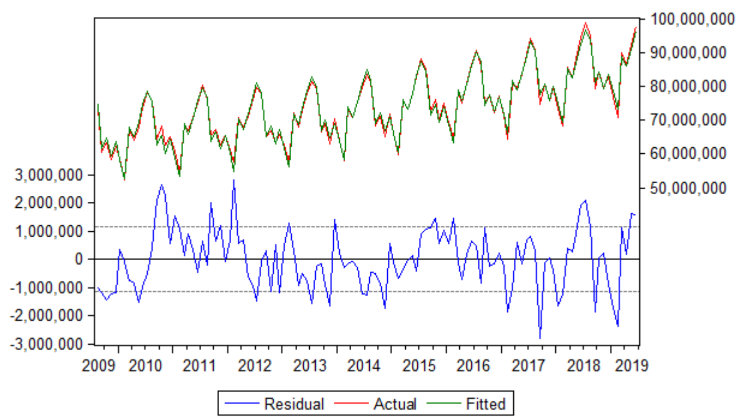


The actual, fitted, and residual graph below shows that while the model accounts for the seasonal variation, there is still a trend that needs to be modeled. The code on EViews to jointly estimate a model with seasonal dummies and a quadratic trend is:[1]

```
ls airpass @trend @trend^2 d01 d02 d03 d04 d05 d06 d07 d08 d09 d10 d11 d12
```

We omit the regression output as it is similar to the one reported in Table 11.1. Both coefficients on `@trend` and `@trend^2` are statistically significant, and the R^2 of the model is 98.8%.

The actual, fitted, and residual graph is:

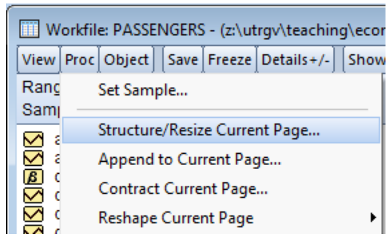


In this graph we can observe how the fitted values (green line), follow very closely the actual values (red line). Moreover the residuals measured on the left-hand side do not appears to have any remaining seasonal pattern or trend.

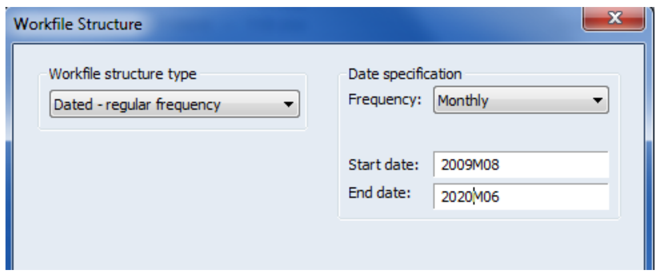
11.3 Forecasting Seasonality

To be able to obtain the out-of-sample forecast, we need to first increase the workfile size to be able to include observations beyond period T. To do this, go to “Proc” and then “Structure/Resize Current Page”.

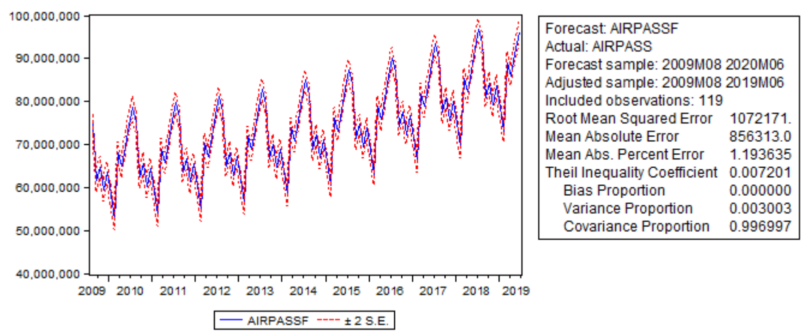
[1] You still need to select between competing models to assess whether the linear, quadratic or a different trend model better explains the data.



To then select some date in the future (i.e., beyond "T"). In this case se use June, 2020 given that out data stops at June, 2019.



For the forecasting graph and the forecasting series, we follow the steps in the previous handouts to obtain:



11.4 How to Create Dummy Variables

If the dummy variables d1, d2, ..., d12 are not readily available in the data set, they can easily be created using the following command:

```
genr dum1 = @seas(1)
```

This generates the dummy for the first month. You have to repeat this for all 12 months in the sample: genr dum2 = @seas(2)... until genr dum12 = @seas(12).

Chapter 12
Characterizing Cycles

Cycles: Any sort of dynamics not captured by the trend or seasonality.

- Only need some persistence.
- Are more sophisticated than the trend and seasonal components.

12.1 Covariance Stationary Time Series

Consider the following realizations of a time series:

$$\{\dots y_{-3}, y_{-2}, y_{-1}, y_0, y_1, y_2, y_3, \dots\}$$

which are ordered in time.

We only observed a sample path:

$$\{y_1, y_2, y_3, \dots, y_T\}$$

To forecast we need:

- That the probabilistic structure of the series be the same in the future.
- At the minimum we want the covariance structure to be stable over time (we call this *covariance stationary*).

Covariance Stationary: We want the mean and the covariance structure of the series to be stable.

For the mean to be stable we need:

$$E(y_t) = \mu \tag{12.1}$$

where μ does not have a t subscript as it is constant over time.

To assess if the covariance structure is stable we will use the autocovariance function, the autocorrelation function, and the partial autocorrelation function.

Autocovariance Function: It is defined as the covariance between y_t and $y_{t-\tau}$ at different values of the displacement τ. Formally, the autocovariance function is given by

$$\gamma(t,\tau) = cov(y_t, y_{t-\tau}) \tag{12.2}$$

where τ is the displacement, and $cov(y_t, y_{t-\tau})$ is just the covariance between y_t and $y_{t-\tau}$.

If the autocovariance is stable it should depend only on τ, not on t. That is,

$$\gamma(t,\tau) = \gamma(\tau) \quad \text{for all } t. \tag{12.3}$$

The autocovariance is symmetric:

$$\gamma(\tau) = \gamma(-\tau) \tag{12.4}$$

Moreover, the autocovariance at displacement zero, $\tau = 0$, is equal to the variance,

$$\gamma(0) = var(y_t) \tag{12.5}$$

where $var(y_t)$ denotes the variance of y_t.

Autocorrelation Function: For practical purposes is it better to focus on the autocorrelation function, which is units free and defined as:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} \quad \text{for } \tau = 0,1,2,\ldots \tag{12.6}$$

where $\gamma(0)$ is the variance of y_t, and $\gamma(\tau)$ is the autocovariance at displacement τ. We can view $\rho(\tau)$ as the correlation coefficient between y_t and $y_{t-\tau}$.
Note that at displacement zero, $\rho(0) = \frac{\gamma(0)}{\gamma(0)} = 1$.

Partial Autocorrelation Function: It is denoted by $p(\tau)$ and measures the association between y_t and $y_{t-\tau}$ after controlling for $y_{t-1}, y_{t-2}, \ldots, y_{t-\tau+1}$.

It is obtained by regressing y_t on $y_{t-1}, y_{t-2}, \ldots, y_{t-\tau}$. Then $p(\tau)$ is the slope coefficient on $y_{t-\tau}$.

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_\tau y_{t-\tau} + \varepsilon_t \tag{12.7}$$

where in this model $p(\tau) = \beta_\tau$.

The partial autocorrelation function contrasts with the autocorrelation function, which does not control for other lags.

The covariance stationary processes that we will study have autocorrelation and partial autocorrelation functions that approach to 0.

12.2 White Noise

Suppose that:

$$y_t = \varepsilon_t \tag{12.8}$$

where $\varepsilon_t \sim (0, \sigma^2)$. There are no dynamics in the process.

We say that ε_t is serially uncorrelated. That is, we cannot predict ε_t based on its past observations, $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$

We say that y_t is a white noise process when:

$$y_t \overset{\text{iid}}{\sim} (0, \sigma^2) \qquad \text{or} \qquad y_t \sim WN(0, \sigma^2).$$

where iid means independent and identically distributed. Figure 12.1 illustrates the dynamics of a white-noise process.

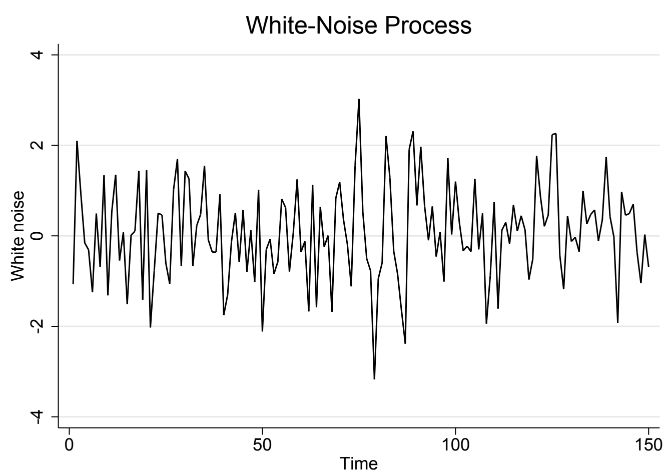


Fig. 12.1 White-noise process $y_t \overset{\text{iid}}{\sim} (0, \sigma^2)$ or $y_t \sim WN(0, \sigma^2)$.

A Gaussian white noise process is:

$$y_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

where $N(0, \sigma^2)$ just denotes the normal (or Gaussian) distribution so that the mean and variance are given by:

$$E(y_t) = 0$$
$$var(y_t) = \sigma^2$$

The autocovariance function for a white noise process is:

$$\gamma(\tau) = \sigma^2 \quad \text{for} \quad \tau = 0$$
$$= 0 \quad \text{for} \quad \tau \geq 1$$

The autocorrelation function for a white noise process is:

$$\rho(\tau) = 1 \quad \text{for} \quad \tau = 0$$
$$= 0 \quad \text{for} \quad \tau \geq 1$$

Because $y_{t-1}, y_{t-2}, y_{t-3}, \ldots$ have no information to predict y_t, the partial autocorrelation function of a white noise process is:

$$\rho(\tau) = 1 \quad \text{for} \quad \tau = 0$$
$$= 0 \quad \text{for} \quad \tau \geq 1$$

The conditional mean and variances are:

$$E(y_t|\Omega_{t-1}) \; = 0$$
$$var(y_t|\Omega_{t-1}) = \sigma^2$$

12.3 Lag Operator

$$L^m y_t = y_{t-m} \tag{12.9}$$
$$L^1 y_t = y_{t-1} \tag{12.10}$$
$$L^2 y_t = y_{t-2}$$
$$B(L) = b_0 + b_1 L + b_2 L^2 + b_3 L^3 + \cdots + b_n L^n = \sum_{i=0}^{\infty} b_i L^i.$$

12.4 Wold's Theorem

What model should we use after controlling for the trend and seasonal components?

Let $\{y_t\}$ be a zero-mean and covariance-stationary process. Then $\{y_t\}$ can be written in its Wold representation form:

$$y_t = \sum_{i=0}^{\infty} b_i L^i; \quad b_0 = 1; \quad \sum_{i=0}^{\infty} b_i^2 < 0; \quad \text{and} \quad \varepsilon_t = WN(0, \sigma^2) \tag{12.11}$$

or

$$y_t = b_0 + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + b_3 \varepsilon_{t-3} + \cdots \tag{12.12}$$

In summary, the Wold's Theorem indicates that any stationary process has this seemingly special representation of Equation 12.12. As we will see later on, this is called the moving average representation of a covariance-stationary process.

12.5 Estimation of μ, $\rho(\tau)$, and $p(\tau)$

For μ, we use the sample mean:

$$\hat{\mu} = \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t. \tag{12.13}$$

For the autocorrelation $\rho(\tau)$, we use sample autocorrelation function:

$$\hat{\rho}(\tau) = \frac{\sum_{t=\tau+1}^{T}(y_t - \bar{y})(y_{t-\tau} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}. \tag{12.14}$$

What if we are interested in knowing if the series is a good approximation of a white noise process? This is an important question to assess the quality of a forecasting model. Once we select a forecasting model, the regression residuals need to be a good approximation of a white noise process. A simple test would be to assess if the all the autocorrelations are zero. For example, we can plot the sample autocorrelations along their two-standard-error bands and assess if 95% of the sample autocorrelations fall within this band. If so, the series can be said to be white noise.

Box-Pierce Q-statistic: It is a formal test that y_t is white noise.

$$Q_{BP} = T\sum_{\tau=1}^{m}\hat{\rho}^2(\tau) \sim \chi^2_m, \tag{12.15}$$

where m is the number of autocorrelations, which is also equal to the number of degrees of freedom. Moreover, "$\sim$" means that the Q_{BP} approximates a chi-squared distribution with m degrees of freedom (χ^2_m). We reject the null hypothesis of white noise if the p-value is less than α (e.g., $\alpha = 0.05$).

The Box-Pierce Q-statistic is essentially a test that all autocorrelations are zero. If we fail to reject the null hypothesis of y_t being white noise, then we can conclude that the series y_t is unpredictable.

Ljung-Pierce Q-statistic: In small samples, we use the Ljung-Pierce Q-statistic instead of the Box-Pierce Q-statistic. This is because the Ljung-Pierce Q-statistic presents a small sample correction.

$$Q_{LP} = T(T+2)\sum_{\tau=1}^{m}\Big(\frac{1}{T-\tau}\hat{\rho}^2(\tau)\Big)\chi^2_m. \tag{12.16}$$

For the partial autocorrelation function $p(\tau)$, we use:

$$\hat{y}_t = \hat{c} + \hat{\beta}_1 y_{t-1} + \cdots + \hat{\beta}_\tau y_{t-\tau}, \tag{12.17}$$

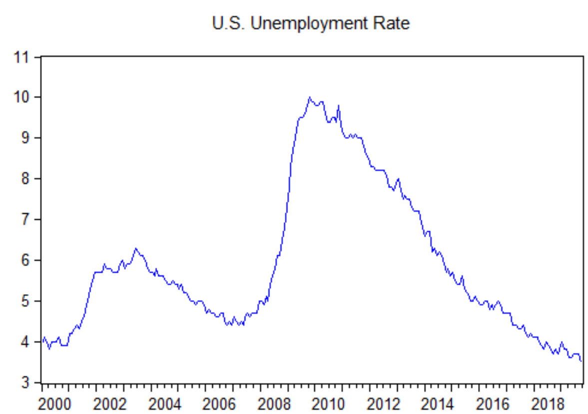
where $\hat{p}(\tau) \equiv \hat{\beta}_\tau$.

Chapter 13
EViews: Characterizing Cycles

This chapter will show how to obtain the correlogram.

13.1 Unemployment Rate

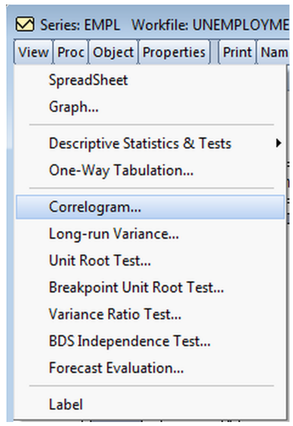
Figure 13.1 presents the U.S. monthly unemployment rate between January 2000 and September 2019. The data comes from the Bureau of Labor Statistics.[1] This variable is contained in the EViews file “unemploymentrate.wf1”



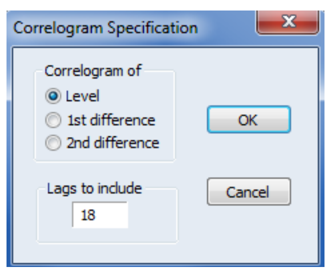
[1] It considers individuals who are 16 years old and over.

13.2 Correlogram of a Series

The correlogram of this unemployment rate is obtained on EViews by opening the series and then selecting "View" and "Correlogram..."



Then the we need to have "Level" (the default option) and select the lags to include. The default is 12, but for this example we use 18.[2]



The resulting correlogram is:

[2] Given the persistence of the series, a selection of lags 36 (3 years) would have been more appropriate. We selected 18 for space purposes.

Sample: 2000M01 2019M09
Included observations: 237

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.990	0.990	235.18	0.000
		2	0.979	-0.046	466.16	0.000
		3	0.965	-0.132	691.77	0.000
		4	0.949	-0.130	910.85	0.000
		5	0.932	-0.068	1122.7	0.000
		6	0.911	-0.125	1326.2	0.000
		7	0.889	-0.060	1520.7	0.000
		8	0.865	-0.043	1705.9	0.000
		9	0.840	-0.045	1881.1	0.000
		10	0.813	-0.052	2046.2	0.000
		11	0.784	-0.098	2200.4	0.000
		12	0.754	-0.045	2343.6	0.000
		13	0.724	0.050	2476.3	0.000
		14	0.694	0.003	2598.6	0.000
		15	0.663	-0.017	2710.8	0.000
		16	0.633	0.035	2813.5	0.000
		17	0.601	-0.087	2906.4	0.000
		18	0.569	-0.003	2990.2	0.000

There are various important elements in this computer output:

- The numbers that go from 1 to 18 on the unlabeled column are the different displacements τ that we introduced on Chapter 12.
- The bars under "Autocorrelation" along the autocorrelation point estimates under "AC" are obtained using Equation 12.14 at different displacements τ.
- The bars under "Partial Correlation" along with the partial autocorrelations point estimates under "PAC" follow Equation 12.17 for different displacements τ.
- The column "Q-Stat" is reporting the Ljung-Pierce Q-statistic from Equation 12.16 at different displacements τ. This statistic serves to test the null hypothesis that the underlying series follows a white-noise process. The p-values reported under "Prob" provide strong empirical evidence that the U.S. unemployment rate is not a white-noise process. That is, we reject the null at different τs.

Chapter 14
Modeling Cycles: MA, AR and ARMA Models

There are three approximation of the Wold representation of a covariance-stationary series y_t:

MA: Moving average.
AR: Autoregressive.
ARMA: Autoregressive moving average.

We will use $\rho(\tau)$, $p(\tau)$, AIC, and BIC to select the model.

14.1 Moving Average (MA) Models

14.1.1 The MA(1) Process

MA(1) process:

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}, \tag{14.1}$$

where,

$$\varepsilon_t \sim WN(0, \sigma^2).$$

That is, the shocks ε_t follow a white noise process with mean zero and variance σ^2.

Equation 14.1 shows how a shock affects the series y_t contemporaneously, and then again after one period.

The idea in MA models is that y_t is modeled as a function of current and lagged values of the unobserved shocks.

Expected Value:

$$E(y_t) = 0 \tag{14.2}$$

Variance:

$$Var(y_t) = \sigma^2(1 + \theta^2) \tag{14.3}$$

Autocorrelation:

$$\rho(\tau) = \frac{\theta}{1+\theta^2} \quad \text{if} \quad \tau = 1$$
$$\rho(\tau) = 0 \quad \text{if} \quad \tau > 1$$

Note that the MA(1) process:

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1},$$

holds for every t, hence we can write:

$$\begin{aligned} y_{t-1} &= \varepsilon_{t-1} + \theta\varepsilon_{t-2} \\ y_{t-2} &= \varepsilon_{t-2} + \theta\varepsilon_{t-3} \\ y_{t-3} &= \varepsilon_{t-3} + \theta\varepsilon_{t-4} \end{aligned}$$

and so forth. We can then substitute backwards in the MA(1) process to obtain:

$$y_t = \varepsilon_t + \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} - \cdots, \tag{14.4}$$

which is essentially y_t as a function of its own lags and the contemporaneous shock ε_t.

To illustrate the role of θ in the dynamics of an MA(1) process, consider the following two MA(1) processes:

$$\begin{aligned} y_t &= \varepsilon_t + 0.08\varepsilon_{t-1}, \\ x_t &= \varepsilon_t + 0.98\varepsilon_{t-1}. \end{aligned}$$

Both of these processes are illustrated in Figure 14.1. Note that consistent with Equation 14.3, the variance of x_t is higher than the variance of y_t as the x_t process has a higher θ.

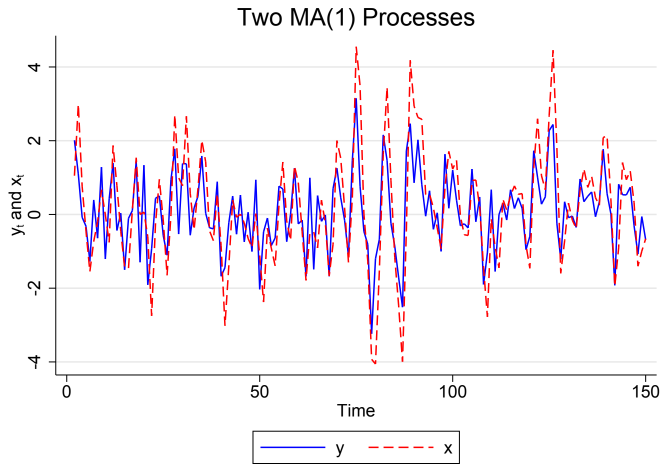


Fig. 14.1 Two MA(1) processes: $y_t = \varepsilon_t + 0.08\varepsilon_{t-1}$ and $x_t = \varepsilon_t + 0.98\varepsilon_{t-1}$.

14.1.2 The MA(q) Process

The MA(q) process is a finite order moving average process of order q. It can be written as:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \tag{14.5}$$

where,

$$\varepsilon_t \sim WN(0, \sigma^2).$$

We can see from Equation 14.5 that a shock affects the series for q periods. MA(1) is a special case where $q = 1$.

14.2 Autoregressive (AR) Models

14.2.1 The AR(1) Process

AR(1) process:

$$y_t = \phi y_{t-1} + \varepsilon_t, \tag{14.6}$$

where,

$$\varepsilon_t \sim WN(0, \sigma^2).$$

In Equation 14.6 shows how in an AR(1) process, the current value of a series linearly depends on the past values plus a random shock.

Expected Value:

$$E(y_t) = 0 \tag{14.7}$$

Variance:

$$Var(y_t) = \frac{\sigma^2}{1-\phi^2} \tag{14.8}$$

Autocorrelation:

$$\rho(\tau) = \phi^\tau \quad \text{for} \quad \tau = 0, 1, 2, \dots$$

Partial autocorrelation:

$$\begin{aligned} p(\tau) &= \phi \quad \text{if} \quad \tau = 1 \\ p(\tau) &= 0 \quad \text{if} \quad \tau > 1 \end{aligned}$$

Note that the AR(1) process:

$$y_t = \phi y_{t-1} + \varepsilon_t,$$

holds for every t, hence we can write:

$$\begin{aligned} y_{t-1} &= \phi y_{t-2} + \varepsilon_{t-1} \\ y_{t-2} &= \phi y_{t-3} + \varepsilon_{t-2} \\ y_{t-3} &= \phi y_{t-4} + \varepsilon_{t-3} \end{aligned}$$

and so forth. We can then substitute backwards for lagged y's on the right-hand side of the AR(1) process to obtain:

$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \cdots, \tag{14.9}$$

which is essentially an AR(1) represented as an MA(∞). This representation is convenient if and only if $|\phi| < 1$. This is the condition for covariance stationarity in an AR(1) process.

To illustrate the role of ϕ in the dynamics of an AR(1) process, consider the following two AR(1) processes:

$$\begin{aligned} y_t &= 0.2y_{t-1} + \varepsilon_t, \\ x_t &= 0.9x_{t-1} + \varepsilon_t. \end{aligned}$$

Both of these processes are illustrated in Figure 14.2. Note that consistent with Equation 14.8, the variance of x_t is higher than the variance of y_t as the x_t process has a higher ϕ.

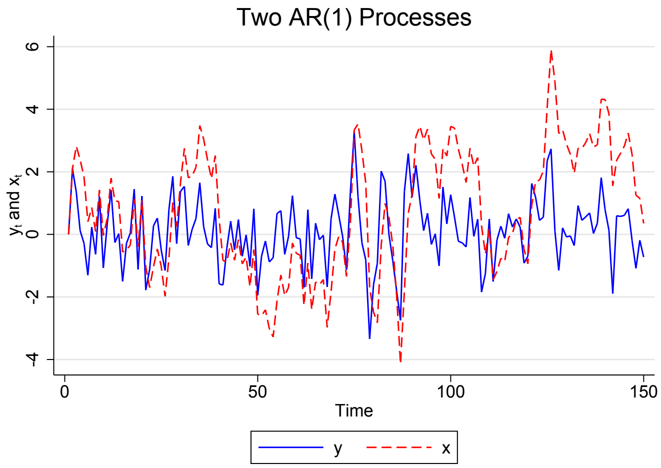


Fig. 14.2 Two AR(1) processes: $y_t = 0.2y_{t-1} + \varepsilon_t$ and $x_t = 0.9x_{t-1} + \varepsilon_t$.

14.2.2 The AR(p) Process

The AR(q) process is a finite order autoregressive process of order p. It can be written as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{14.10}$$

where,

$$\varepsilon_t \sim WN(0, \sigma^2).$$

We can see from Equation 14.10 that y_t is a function of its own lagged values for p periods. AR(1) is a special case where $p = 1$.

14.3 Autoregressive Moving Average (ARMA) Models

14.3.1 The ARMA(1,1) Process

ARMA(1,1) process:

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \tag{14.11}$$

where,

$$\varepsilon_t \sim WN(0, \sigma^2).$$

In Equation 14.11 we can see that an ARMA(1,1) process is just the combination on an AR(1) and an MA(1) process.

14.3.2 The ARMA(p,q) Process

ARMA(p,q) process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \quad (14.12)$$

where,

$$\varepsilon_t \sim WN(0, \sigma^2).$$

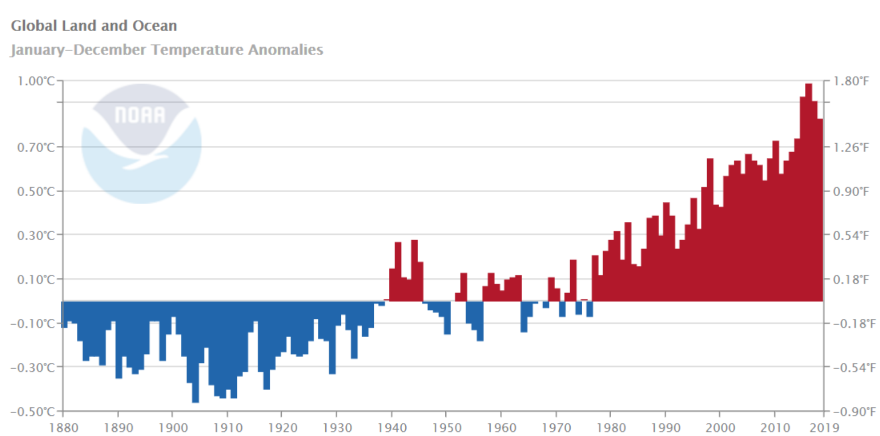
In Equation 14.12 we can see that an ARMA(p,q) process is just the combination on an AR(p) and an MA(q) process. The ARMA(1,1) is just a special case of an ARMA(p,q) where $p = q = 1$.

Chapter 15
EViews: MA, AR and ARMA Models

This chapter will show how to obtain the correlogram.

15.1 Climate Change

Consider the following time series information on Global Land and Ocean January-December Temperature anomalies. These are global and hemispheric anomalies with respect to the 20th century average. They are measured in ^{o}C. The data is in the EViews file "Temperatures."



The correlogram of the series for a twenty year period shows:

Sample: 1880 2018
Included observations: 139

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.927	0.927	122.04	0.000
		2	0.868	0.059	229.73	0.000
		3	0.832	0.143	329.40	0.000
		4	0.806	0.086	423.68	0.000
		5	0.769	-0.051	510.11	0.000
		6	0.748	0.120	592.65	0.000
		7	0.731	0.024	671.91	0.000
		8	0.714	0.033	748.09	0.000
		9	0.682	-0.076	818.24	0.000
		10	0.667	0.085	885.74	0.000
		11	0.640	-0.083	948.46	0.000
		12	0.616	0.010	1006.9	0.000
		13	0.591	-0.003	1061.3	0.000
		14	0.575	0.012	1113.2	0.000
		15	0.551	-0.034	1161.2	0.000
		16	0.525	-0.034	1205.1	0.000
		17	0.511	0.075	1247.1	0.000
		18	0.499	-0.017	1287.3	0.000
		19	0.469	-0.071	1323.3	0.000
		20	0.450	0.035	1356.7	0.000

From the Ljung-Pierce Q-statistics we can say that this series it not White Noise. Moreover, the autocorrelations at various displacements τ show important dynamics.

Consider estimating the following quadratic trend model:

$$TEMP_t = \beta_0 + \beta_1 TREND_t + \beta_2 TREND_t^2 + \varepsilon_t \tag{15.1}$$

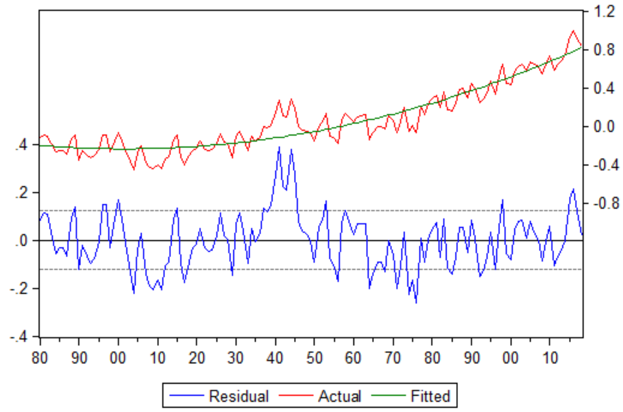
where the regression output is:

Dependent Variable: TEMP
Method: Least Squares
Sample: 1880 2018
Included observations: 139

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.201013	0.030103	-6.677582	0.0000
@TREND	-0.003403	0.001008	-3.376312	0.0010
@TREND^2	7.79E-05	7.07E-06	11.01747	0.0000

R-squared	0.875870	Mean dependent var	0.060360
Adjusted R-squared	0.874045	S.D. dependent var	0.338144
S.E. of regression	0.120008	Akaike info criterion	-1.381174
Sum squared resid	1.958655	Schwarz criterion	-1.317840
Log likelihood	98.99156	Hannan-Quinn criter.	-1.355436
F-statistic	479.8136	Durbin-Watson stat	0.850351
Prob(F-statistic)	0.000000		

with the corresponding actual, fitted and residuals:



Moreover, note that as soon as you run a regression, EViews will generate the series `resid` that corresponds to the estimated regression residuals $\hat{\varepsilon}_t$ from Equation 15.1. The correlogram of those residuals for a window of up to ten displacements is given by:

Sample: 1880 2018
Included observations: 139

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.573	0.573	46.641	0.000
		2	0.291	-0.056	58.732	0.000
		3	0.237	0.140	66.842	0.000
		4	0.261	0.117	76.755	0.000
		5	0.139	-0.114	79.590	0.000
		6	0.133	0.120	82.177	0.000
		7	0.137	-0.001	84.955	0.000
		8	0.133	0.028	87.612	0.000
		9	0.057	-0.051	88.096	0.000
		10	0.079	0.062	89.033	0.000

Based on these results you reject the null hypothesis of White Noise error terms. This is because the autocorrelation and the partial autocorrelation for various values of the displacement fall outside the two-standard deviation bands. Moreover, the Q-statistic (Ljung-Box Q-statistic) which is the weighted sum of squared autocorrelations has large values when compared to the χ^2 distribution with the corresponding degrees of freedom (the p-values are below $\alpha = 0.05$). Hence, the model of a quadratic trend still leaves some elements in the residuals $\hat{\varepsilon}_t$ of Equation 15.1 that can be forecasted.

Consider estimating the following quadratic trend models with an MA(1), an AR(1), and ARMA(1,1) components:

$$TEMP_t = \beta_0 + \beta_1 TREND_t + \beta_2 TREND_t^2 + \theta \varepsilon_{t-1} + \varepsilon_t \quad (15.2)$$
$$TEMP_t = \beta_0 + \beta_1 TREND_t + \beta_2 TREND_t^2 + \phi TEMP_{t-1} + \varepsilon_t \quad (15.3)$$
$$TEMP_t = \beta_0 + \beta_1 TREND_t + \beta_2 TREND_t^2 + \phi TEMP_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t \quad (15.4)$$

These equations are estimated in EViews with the following commands:

```
ls temp c @trend @trend^2 ma(1)
```

```
ls temp c @trend @trend^2 ar(1)
```

```
ls temp c @trend @trend^2 ar(1) ma(1)
```

The regression output for the ARMA(1,1) model is:

Dependent Variable: TEMP
Method: ARMA Maximum Likelihood (BFGS)
Sample: 1880 2018
Included observations: 139
Convergence achieved after 10 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.196062	0.061133	-3.207166	0.0017
@TREND	-0.003551	0.001949	-1.821994	0.0707
@TREND^2	7.88E-05	1.37E-05	5.766724	0.0000
AR(1)	0.474254	0.141648	3.348114	0.0011
MA(1)	0.142941	0.170470	0.838512	0.4032
SIGMASQ	0.009401	0.001201	7.826003	0.0000

R-squared	0.917182	Mean dependent var	0.060360
Adjusted R-squared	0.914068	S.D. dependent var	0.338144
S.E. of regression	0.099124	Akaike info criterion	-1.739765
Sum squared resid	1.306793	Schwarz criterion	-1.613097
Log likelihood	126.9137	Hannan-Quinn criter.	-1.688290
F-statistic	294.5860	Durbin-Watson stat	2.001175
Prob(F-statistic)	0.000000		

Inverted AR Roots	.47
Inverted MA Roots	-.14

and the correlogram of the residuals is:

Sample: 1880 2018
Included observations: 139

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.003	-0.003	0.0010	0.975
		2	-0.044	-0.044	0.2780	0.870
		3	0.019	0.019	0.3311	0.954
		4	0.215	0.214	7.0726	0.132
		5	-0.060	-0.060	7.6055	0.179
		6	0.050	0.070	7.9678	0.240
		7	0.040	0.028	8.2056	0.315
		8	0.109	0.073	9.9747	0.267
		9	-0.077	-0.055	10.870	0.285
		10	0.112	0.098	12.773	0.237

which shows that the regression residuals of Equation 15.4 are White Noise. Hence in this model there is nothing left in the error term that can be forecasted. The orders of p and q in Equation 14.12 need to be selected based on the AIC and BIC. After the selection of the model, the regression residual needs to be White Noise.

15.2 MA(1) Simulated Processes

To simulate a process the first step is to create a workfile. Please review Section 6.1 on how to to this. On your workfile you are free to select any time frequency, just make sure you have about 100 observations. In this example we selected to have 121 yearly observations from 1900 to 2020. Now, it we want to generate a White Noise process ε with mean zero and variance one, the command is:

```
genr epsilon=nrnd
```

If we graph this ε sequence, we obtain:

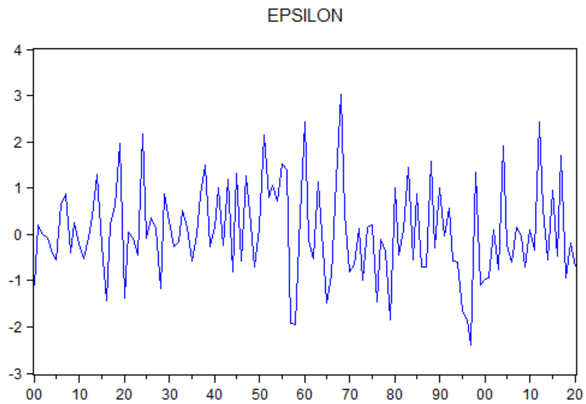


Fig. 15.1 White-noise process $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$ or $\varepsilon_t \sim WN(0,1)$.

which is equivalent to the one presented in Figure 12.1.

Based on this ε_t sequence, we can type the following commands in EViews to generate three different MA(1) processes:

```
genr Y1=epsilon+0.08*epsilon(-1)
genr Y2=epsilon+0.98*epsilon(-1)
genr Y3=epsilon-0.98*epsilon(-1)
```

A graph of $Y1$ and $Y2$ shows that $Y2$ is more volatile than $Y1$, consistent with the variance formula for an MA(1) process. This was shown in Figure 14.1.

To further study the dynamics of these three series, we obtain their autocorrelation and the partial autocorrelation functions as presented in Figures 15.2, 15.2, 15.2 below.

For the MA(1) process $Y1_t = \varepsilon_t + 0.08\varepsilon_{t-1}$, we can see that because $\theta = 0.08$ is very small, we cannot statistically distinguish it from a White Noise process. The Ljung-Pierce Q-statistic show p-values greater than 0.05. Moreover, most of the correlation and partial correlation estimates are within the 95% confidence bands and there is really no distinguishable pattern on these estimates.

Sample: 1900 2020
Included observations: 120

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.086	0.086	0.9077	0.341
		2	-0.071	-0.079	1.5373	0.464
		3	-0.071	-0.058	2.1613	0.540
		4	0.024	0.030	2.2324	0.693
		5	0.137	0.125	4.6313	0.463
		6	-0.043	-0.068	4.8640	0.561
		7	-0.006	0.025	4.8691	0.676
		8	0.191	0.207	9.6541	0.290
		9	-0.034	-0.089	9.8047	0.367
		10	-0.096	-0.085	11.036	0.355
		11	-0.152	-0.104	14.121	0.226
		12	0.036	0.036	14.294	0.282

Fig. 15.2 MA(1) process: $Y1_t = \varepsilon_t + 0.08\varepsilon_{t-1}$ with $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$.

For the MA(1) process $Y2_t = \varepsilon_t + 0.98\varepsilon_{t-1}$, we can see that with a positive and relatively large $\theta = 0.98$, the The Ljung-Pierce Q-statistics clearly reject the null hypothesis of White Noise. The first autocorrelation is positive, while the partial autocorrelations flip from positive to negative. This is always the case when $\theta > 0$.

Note that for this $Y2$, from the theoretical formula we have that $\rho(\tau = 1) = \frac{\theta}{1+\theta^2} = \frac{0.98}{1+0.98^2} = 0.499$. The simulated series gives as a $\hat{\rho}(\tau = 1) = 0.469$, which is very close to the theoretical value. The theory also predicts that $\rho(\tau) = 0$ for $\tau > 1$, which also appears to hold in these estimates.

Sample: 1900 2020
Included observations: 120

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.469	0.469	27.009	0.000
		2	-0.096	-0.404	28.146	0.000
		3	-0.086	0.255	29.066	0.000
		4	0.062	-0.077	29.546	0.000
		5	0.123	0.145	31.463	0.000
		6	0.004	-0.176	31.465	0.000
		7	0.052	0.326	31.810	0.000
		8	0.173	-0.122	35.723	0.000
		9	0.020	-0.035	35.774	0.000
		10	-0.172	-0.143	39.734	0.000
		11	-0.165	0.036	43.382	0.000
		12	-0.000	-0.022	43.382	0.000

Fig. 15.3 MA(1) process: $Y2_t = \varepsilon_t + 0.98\varepsilon_{t-1}$ with $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$.

For the MA(1) process $Y3_t = \varepsilon_t - 0.98\varepsilon_{t-1}$, we can see that with a negative and relatively large $\theta = -0.98$, the The Ljung-Pierce Q-statistics also clearly reject the null hypothesis of White Noise. The first autocorrelation is negative, and the partial autocorrelations are also all negative and decrease in magnitude as we increase the displacement τ.

Ones again, note that from the theoretical formula we have that $\rho(\tau=1) = \frac{\theta}{1+\theta^2} = \frac{-0.98}{1+(-0.98)^2} = -0.499$. In this case the simulated series gives as a $\hat{\rho}(\tau=1) = -0.457$, which is again very close to the theoretical value. Moreover, as predicted by the theory, the rest of the correlations are not distinguishable from zero.

Sample: 1900 2020
Included observations: 120

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.457	-0.457	25.703	0.000
		2	-0.043	-0.318	25.931	0.000
		3	-0.035	-0.290	26.085	0.000
		4	-0.028	-0.317	26.187	0.000
		5	0.170	-0.072	29.879	0.000
		6	-0.118	-0.117	31.678	0.000
		7	-0.093	-0.278	32.798	0.000
		8	0.238	0.041	40.210	0.000
		9	-0.114	0.035	41.921	0.000
		10	0.015	0.037	41.952	0.000
		11	-0.131	-0.115	44.254	0.000
		12	0.087	-0.052	45.277	0.000

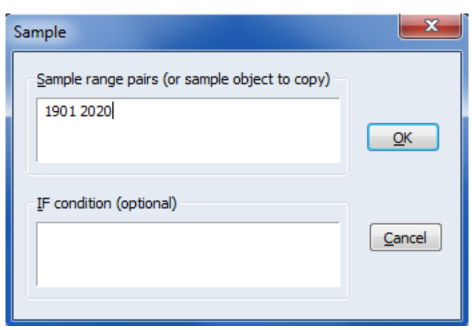
Fig. 15.4 MA(1) process: $Y3_t = \varepsilon_t - 0.98\varepsilon_{t-1}$ with $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$.

15.3 AR(1) Simulated Processes

Let's now generate some artificial AR(1) processes. As before, we first need to generate the random variable ε. Then, we create the following series:

```
genr Z1 = 0
genr Z2 = 0
genr Z3 = 0
genr Z4 = 0
```

Next, we need to modify the sample to get rid of the first observation.



Now, proceed to general the series:

```
genr Z1 = +0.90*Z1(-1) + epsilon
genr Z2 = +0.20*Z2(-1) + epsilon
genr Z3 = -0.90*Z3(-1) + epsilon
genr Z4 = -0.20*Z4(-1) + epsilon
```

To see how a simple difference in the sign and the magnitude (size) of the autoregressive coefficient ϕ can have important differences in the series, let's graph $Z1$ and $Z2$:

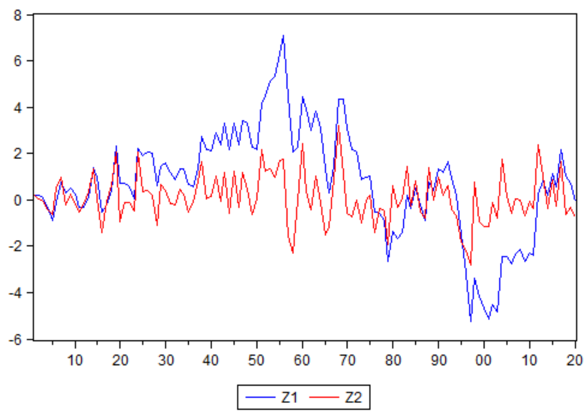


Fig. 15.5 AR(1) processes: $Z1_t = 0.9 \cdot Z1_{t-1} + \varepsilon_t$ and $Z2_t = 0.2 \cdot Z2_{t-1} + \varepsilon_t$ with $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$.

And we can also graph $Z3$ and $Z4$:

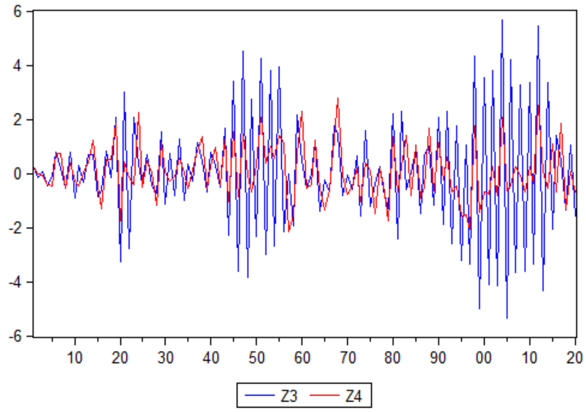
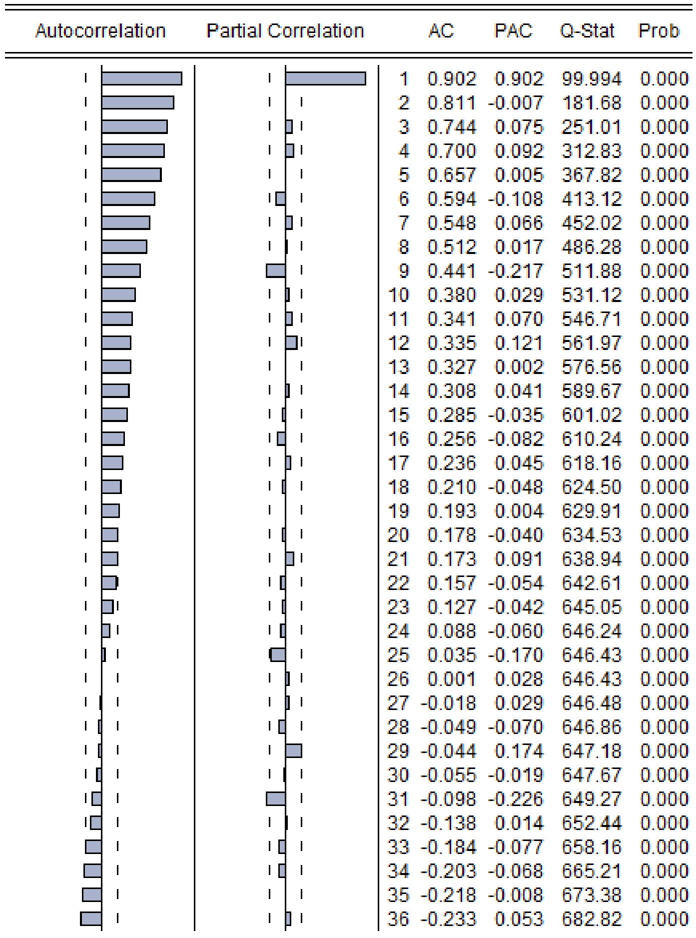


Fig. 15.6 AR(1) processes: $Z3_t = -0.9 \cdot Z3_{t-1} + \varepsilon_t$ and $Z4_t = -0.2 \cdot Z4_{t-1} + \varepsilon_t$ with $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$.

We can easily get three important insights from these Figures 15.3 and 15.3. First, series $Z1$ and $Z2$, which have a positive autoregressive coefficient ($\phi > 0$) are more likely to have longer periods of consecutive negative and positive values. Second, the series $Z3$ and $Z4$, which have a negative autoregressive coefficient ($\phi < 0$) are constantly switching from negative to positive and vice versa. Third, the larger the magnitude of the autoregressive coefficient, $|\phi|$, the more volatile the series (higher variance, $var(Z)$).

From the autocorrelations and the partial autocorrelations presented in Figure 15.3 for the simulated process $Z1_t = +0.9 \cdot Z1_{t-1} + \varepsilon_t$, we can that AR(1) models with a high ϕ have a long memory. In this particular example it takes up to 26 periods for a shock to dissipate. Of course, the Ljung-Pierce Q-statistics clearly reject the null of White Noise.



Sample: 1901 2020
Included observations: 120

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.902	0.902	99.994	0.000
		2	0.811	-0.007	181.68	0.000
		3	0.744	0.075	251.01	0.000
		4	0.700	0.092	312.83	0.000
		5	0.657	0.005	367.82	0.000
		6	0.594	-0.108	413.12	0.000
		7	0.548	0.066	452.02	0.000
		8	0.512	0.017	486.28	0.000
		9	0.441	-0.217	511.88	0.000
		10	0.380	0.029	531.12	0.000
		11	0.341	0.070	546.71	0.000
		12	0.335	0.121	561.97	0.000
		13	0.327	0.002	576.56	0.000
		14	0.308	0.041	589.67	0.000
		15	0.285	-0.035	601.02	0.000
		16	0.256	-0.082	610.24	0.000
		17	0.236	0.045	618.16	0.000
		18	0.210	-0.048	624.50	0.000
		19	0.193	0.004	629.91	0.000
		20	0.178	-0.040	634.53	0.000
		21	0.173	0.091	638.94	0.000
		22	0.157	-0.054	642.61	0.000
		23	0.127	-0.042	645.05	0.000
		24	0.088	-0.060	646.24	0.000
		25	0.035	-0.170	646.43	0.000
		26	0.001	0.028	646.43	0.000
		27	-0.018	0.029	646.48	0.000
		28	-0.049	-0.070	646.86	0.000
		29	-0.044	0.174	647.18	0.000
		30	-0.055	-0.019	647.67	0.000
		31	-0.098	-0.226	649.27	0.000
		32	-0.138	0.014	652.44	0.000
		33	-0.184	-0.077	658.16	0.000
		34	-0.203	-0.068	665.21	0.000
		35	-0.218	-0.008	673.38	0.000
		36	-0.233	0.053	682.82	0.000

Fig. 15.7 AR(1) process: $Z1_t = +0.9 \cdot Z1_{t-1} + \varepsilon_t$ with $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$.

For the $Z2_t = +0.2 \cdot Z2_{t-1} + \varepsilon_t$ process presented in Figure 15.3, we observe that a low ϕ means the series is close to a White Noise. Only for $\tau < 3$ we fail to reject the null of White Noise at a 10% significance level (see the p-values on last column).

Sample: 1901 2020
Included observations: 120

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.196	0.196	4.7027	0.030
		2	-0.037	-0.078	4.8726	0.087
		3	-0.064	-0.043	5.3879	0.146
		4	0.030	0.052	5.5051	0.239
		5	0.132	0.114	7.7131	0.173
		6	-0.019	-0.073	7.7608	0.256
		7	0.014	0.051	7.7854	0.352
		8	0.180	0.190	12.027	0.150
		9	-0.024	-0.124	12.106	0.207
		10	-0.110	-0.088	13.713	0.187
		11	-0.154	-0.086	16.901	0.111
		12	0.024	0.050	16.981	0.150

Fig. 15.8 AR(1) process: $Z2_t = +0.2 \cdot Z2_{t-1} + \varepsilon_t$ with $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$.

Figure 15.3 presents the simulated process $Z3_t = -0.9 \cdot Z3_{t-1} + \varepsilon_t$. A negative ϕ (here $\phi = -0.9$) shows that autocorrelations flip between positive and negative while they slowly decrease in magnitude. This is consistent with Figure 15.3.

Sample: 1901 2020
Included observations: 120

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.893	-0.893	98.125	0.000
		2	0.801	0.016	177.70	0.000
		3	-0.732	-0.067	244.69	0.000
		4	0.659	-0.045	299.44	0.000
		5	-0.594	-0.006	344.30	0.000
		6	0.568	0.157	385.73	0.000
		7	-0.562	-0.117	426.71	0.000
		8	0.561	0.070	467.83	0.000
		9	-0.524	0.156	503.99	0.000
		10	0.476	-0.080	534.13	0.000
		11	-0.450	-0.092	561.36	0.000
		12	0.397	-0.102	582.70	0.000
		13	-0.336	0.051	598.16	0.000
		14	0.295	-0.024	610.19	0.000
		15	-0.250	0.097	618.88	0.000
		16	0.218	0.051	625.58	0.000
		17	-0.195	-0.015	630.96	0.000
		18	0.175	0.013	635.38	0.000
		19	-0.159	0.026	639.05	0.000
		20	0.143	0.007	642.05	0.000
		21	-0.134	-0.091	644.71	0.000
		22	0.145	0.109	647.83	0.000
		23	-0.136	0.037	650.63	0.000
		24	0.138	0.064	653.54	0.000

Fig. 15.9 AR(1) process: $Z3_t = -0.9 \cdot Z3_{t-1} + \varepsilon_t$ with $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$.

Lastly, Figure 15.3 presents the process $Z4_t = -0.2 \cdot Z4_{t-1} + \varepsilon_t$. Due to a relatively small (and negative) ϕ, it is hard to distinguish this series from a White Noise. The Ljung-Pierce Q-statistic fails to reject the null of White Noise for all displacements except the first.

Sample: 1901 2020
Included observations: 120

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.177	-0.177	3.8406	0.050
		2	-0.017	-0.050	3.8773	0.144
		3	-0.057	-0.071	4.2829	0.232
		4	-0.004	-0.029	4.2847	0.369
		5	0.146	0.141	7.0057	0.220
		6	-0.070	-0.023	7.6290	0.267
		7	-0.042	-0.052	7.8568	0.345
		8	0.205	0.216	13.370	0.100
		9	-0.077	-0.014	14.160	0.117
		10	-0.033	-0.071	14.302	0.160
		11	-0.144	-0.132	17.076	0.106
		12	0.058	0.013	17.538	0.130

Fig. 15.10 AR(1) process: $Z4_t = -0.2 \cdot Z4_{t-1} + \varepsilon_t$ with $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$.

Chapter 16
Forecasting Cycles

Information set at time T:

$$\Omega_T = \{y_T, y_{T-1}, y_{T-2}, \dots\} \tag{16.1}$$

Can be expressed in terms of current and past shocks:

$$\Omega_T = \{\varepsilon_T, \varepsilon_{T-1}, \varepsilon_{T-2}, \dots\} \tag{16.2}$$

Hence, Ω_T can be written as:

$$\Omega_T = \{y_T, y_{T-1}, y_{T-2}, \dots, \varepsilon_T, \varepsilon_{T-1}, \varepsilon_{T-2}, \dots\} \tag{16.3}$$

Based on Ω_T we want the optimal forecast of y at time $T+h$. This is the same as saying that we want the smallest loss. The forecast can be expressed as:

$$E(y_{T+h}|\Omega_T) \tag{16.4}$$

We will use the linear projection:

$$P(y_{T+h}|\Omega_T) \tag{16.5}$$

If errors are normally distributed:

$$E(y_{T+h}|\Omega_T) = P(y_{T+h}|\Omega_T) \tag{16.6}$$

16.1 Forecasting an MA Process

16.1.1 Optimal Point Forecasts

Consider the following finite order MA process:

$$\text{MA(2):} \quad y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \tag{16.7}$$

$$\varepsilon_t = WN(0, \sigma^2)$$

At time $T+1$ (one step ahead):

$$y_{T+1} = \varepsilon_{T+1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} \tag{16.8}$$

ε_{T+1}: Unknown.
$\theta_1 \varepsilon_T$: Known.
$\theta_2 \varepsilon_{T-1}$: Known.

So we can write the forecast of y_{T+1} forecasted at time T as,

$$y_{T+1,T} = +\theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} \tag{16.9}$$

At time $T+2$ (two steps ahead):

$$y_{T+2} = \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1} + \theta_2 \varepsilon_T \tag{16.10}$$

ε_{T+2}: Unknown.
$\theta_1 \varepsilon_{T+1}$: Unknown.
$\theta_2 \varepsilon_T$: Known.

So we can write the forecast of y_{T+2} forecasted at time T as,

$$y_{T+2,T} = \theta_2 \varepsilon_T \tag{16.11}$$

At time $T+3$ (three steps ahead):

$$y_{T+3} = \varepsilon_{T+3} + \theta_1 \varepsilon_{T+2} + \theta_2 \varepsilon_{T+1} \tag{16.12}$$

ε_{T+3}: Unknown.
$\theta_1 \varepsilon_{T+2}$: Unknown.
$\theta_2 \varepsilon_{T+1}$: Unknown.

So we can write the forecast of y_{T+3} forecasted at time T as,

$$y_{T+3,T} = 0 \tag{16.13}$$

For all $h > 0$ we have $y_{T+h} = 0$.
An MA(q) process is not forecastable more than q steps ahead.

- The forecast errors increase with h.

- The forecast error variance also increases with h.

Infinite order MA process, $q = \infty$. The Wold representation of y_t is:

$$y_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i} \tag{16.14}$$

$$\varepsilon_t = WN(0, \sigma^2)$$

$$b_0 = 1 \quad \text{and} \quad \sigma^2 \sum_{i=0}^{\infty} b_i^2 < \infty$$

We can first write out the process at the future times of interest, $T+h$:

$$y_{T+h} = \varepsilon_{T+h} + b_1 \varepsilon_{T+h-1} + b_2 \varepsilon_{T+h-2} + \cdots + b_h \varepsilon_T + b_{h+1} \varepsilon_{T-1} + \cdots \tag{16.15}$$

The first terms on the left-hand side of the equation are unknown, but the last terms are known.

Hence, we can see that the process can be forecasted:

$$y_{T+h,T} = b_h \varepsilon_T + b_{h+1} \varepsilon_{T-1} + \cdots \tag{16.16}$$

16.1.2 Interval and Density Forecasts

For an MA(2):

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \tag{16.17}$$

$$\varepsilon_t = WN(0, \sigma^2)$$

The 95% interval forecast:

$$y_{T+1,T} = (\theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}) \pm 1.96\sigma \tag{16.18}$$

which assumes normality of the forecast.

The density forecast is:

$$N(\theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}, \sigma^2) \tag{16.19}$$

which also assumes normality.

16.2 Forecasting an AR Process

Consider the following AR(1) process:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \tag{16.20}$$

$$\varepsilon_t = WN(0, \sigma^2)$$

16.2.1 Optimal Point Forecasts

To construct the 1-step-ahead forecast, we can write out the process for time $T+1$:

$$y_{T+1} = \phi_1 y_T + \varepsilon_{T+1} \tag{16.21}$$

Then, projecting the right-hand side on the time-T information set:

$$y_{T+1,T} = \phi_1 y_T + \varepsilon_{T+1}, \tag{16.22}$$

where ε_{T+1} has an expected value of zero. We can write the process for time $T+2$:

$$y_{T+2} = \phi_1 y_{T+1} + \varepsilon_{T+2}$$

to then project directly on the time-T information set:

$$y_{T+2,T} = \phi_1 y_{T+1,T}.$$

As before, future shocks are replaced by 0. The process for time $T+3$:

$$y_{T+3} = \phi_1 y_{T+2} + \varepsilon_{T+3}$$

than when projected on the time-T information set, we obtain:

$$y_{T+3,T} = \phi_1 y_{T+2,T}.$$

with the required 2-step-ahead forecast already constructed.

If we keep dong this we can forecast any of the future periods. This is called the "chair rule of forecasting." For an AR(1) process, only the most recent lag of y_t is used to obtain the optimal forecast. For a general AR(p) process, we need the p most recent values.

Consider obtaining the 2-step-ahead point forecast of the following ARMA(1,1) process.

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}. \tag{16.23}$$

$$\varepsilon_t = WN(0, \sigma^2)$$

The one-step ahead forecast is given by:

$$y_{T+1} = \phi y_T + \varepsilon_{T+1} + \theta \varepsilon_T. \tag{16.24}$$

ϕy_T: Known.
ε_{T+1}: Unknown.
$\theta \varepsilon_T$: Known.

and the 2-step ahead forecast is:

$$y_{T+2} = \phi y_{T+1} + \varepsilon_{T+2} + \theta \varepsilon_{T+1}. \tag{16.25}$$

ϕy_{T+1}: Known.
ε_{T+2}: Unknown.
$\theta\varepsilon_{T+1}$: Unknown.

Replacing y_{T+1} from Equation 16.24 on Equation 16.25:

$$\begin{aligned} y_{T+2,T} &= \phi(\phi y_T + \theta\varepsilon_T) \\ &= \phi^2 y_T + \phi\theta\varepsilon_T \end{aligned} \tag{16.26}$$

16.2.2 Interval and Density Forecasts

The chair-rule helps to simplify the point forecasts. However, for density forecasts we require the *h*-step-ahead forecast of the error variance as well. We can obtain it from the moving average representation on an AR process. It is written as:

$$\sigma_h^2 = \sigma^2 \sum_{i=0}^{h-1} b_i^2. \tag{16.27}$$

Because we do not know the values for the parameters σ^2 and b_i, we need to use the following instead:

$$\hat{\sigma}_h^2 = \hat{\sigma}^2 \sum_{i=0}^{h-1} \hat{b}_i^2. \tag{16.28}$$

While there are many b_is that we would need to estimate via the MA representation of the process, the good news is that we only estimate an AR and then solve backward to solve for as many *b*s as needed.

Consider again the following example of an ARMA(1,1) process:

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}. \tag{16.29}$$

We want to construct its 2-step-ahead 95% interval forecast. The 2-step-ahead point forecast was already presented in Equation 16.26, but we additionally need to construct the 2-step-ahead forecast error variance. From Equation 16.29, we substitute backward to get:

$$y_t = \phi(\phi y_{t-2} + \varepsilon_{t-1} + \theta\varepsilon_{t-2}) + \varepsilon_t + \theta\varepsilon_{t-1}. \tag{16.30}$$

$$= \varepsilon_t + (\phi + \theta)\varepsilon_{t-1} + \cdots. \tag{16.31}$$

We do not need to move back any further, because the 2-step-ahead forecast error variance is $\sigma_2^2 = \sigma^2(1 + b_1^2)$, where b_1 is the coefficient on ε_{t-1} in the moving average representation of the ARMA(1,1) process. In this case this is just $(\phi + \theta)$. Hence, the 2-step-ahead interval forecast is:

$$y_{T+2,T} \pm 1.96\sigma_2, \tag{16.32}$$

or

$$(\phi^2 y_T + \phi\theta\varepsilon_T) \pm 1.96\sigma\sqrt{1+(\phi+\theta)^2}. \tag{16.33}$$

Assuming normality, the density forecast is:

$$N\big(\phi^2 y_T + \phi\theta\varepsilon_T, \sigma^2(1+(\phi+\theta)^2)\big). \tag{16.34}$$

Chapter 17
EViews: Forecasting Cycles

This chapter will cover an empirical application on how to forecast cycles.

The variable we want to forecast is the Canadian employment.

17.1 Moving Average Models

Before we estimate the models, lets make sure we all have the same sample:

```
smpl 1962q1 1993q4
```

The preferred MA model is an MA(4), so the E-Views command is:

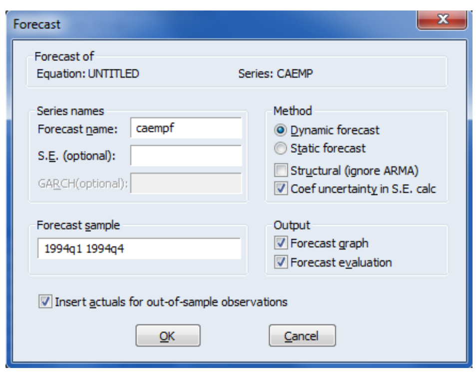
```
ls caemp c ma(1) ma(2) ma(3) ma(4)
```

Dependent Variable: CAEMP
Method: ARMA Maximum Likelihood (BFGS)
Sample: 1962Q1 1993Q4
Included observations: 128
Convergence achieved after 20 iterations
Coefficient covariance computed using outer product of gradients

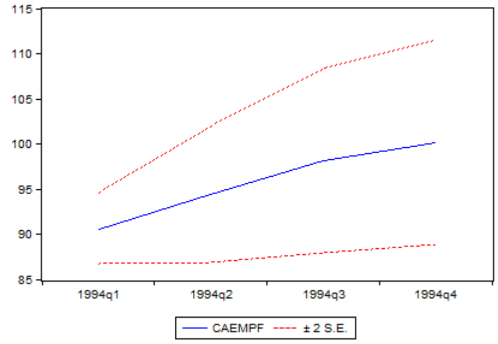
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	100.6692	1.043603	96.46314	0.0000
MA(1)	1.640307	0.067239	24.39506	0.0000
MA(2)	1.734850	0.110991	15.63054	0.0000
MA(3)	1.245124	0.117703	10.57857	0.0000
MA(4)	0.523848	0.078383	6.683164	0.0000
SIGMASQ	3.599362	0.479745	7.502664	0.0000

R-squared	0.935493	Mean dependent var	101.0176
Adjusted R-squared	0.932849	S.D. dependent var	7.499163
S.E. of regression	1.943291	Akaike info criterion	4.241168
Sum squared resid	460.7184	Schwarz criterion	4.374857
Log likelihood	−265.4347	Hannan-Quinn criter.	4.295486
F-statistic	353.8540	Durbin-Watson stat	1.674683
Prob(F-statistic)	0.000000		

The simplest way to forecast the values between the first quarter of 1994 and the fourth quarter of 1994 is to go to click on the icon "forecast" and then choose the correct forecast sample:



This will yield the forecast presented in the following figure:



Forecast: CAEMPF	
Actual: CAEMP	
Forecast sample: 1994Q1 1994Q4	
Included observations: 4	
Root Mean Squared Error	5.474956
Mean Absolute Error	4.843200
Mean Abs. Percent Error	5.294005
Theil Inequality Coefficient	0.029311
Bias Proportion	0.782535
Variance Proportion	0.205870
Covariance Proportion	0.011595

A second more interesting way to obtain the same forecast is to follow these steps:

1. Select the sample to estimate the model:

```
smpl 1962q1 1993q4
```

2. Estimate the model:

```
equation ma4.ls caemp c ma(1) ma(2) ma(3) ma(4)
```

3. Generate a variable with the historical values:

```
genr history = caemp
```

4. Modify the your sample to include the period you want to forecast:

```
smpl 1994:1 1996:4
```

5. Forecast your values (stored in yhat) and the standard errors (stored in se):

```
ma4.forecast yhat se
```

6. Generate the variable that will store the forecasted values:

```
genr fcst=yhat
```

7. Generate the 95% confidence intervals:

```
genr yhatplus=yhat+1.96*se
genr yhatminus=yhat-1.96*se
```

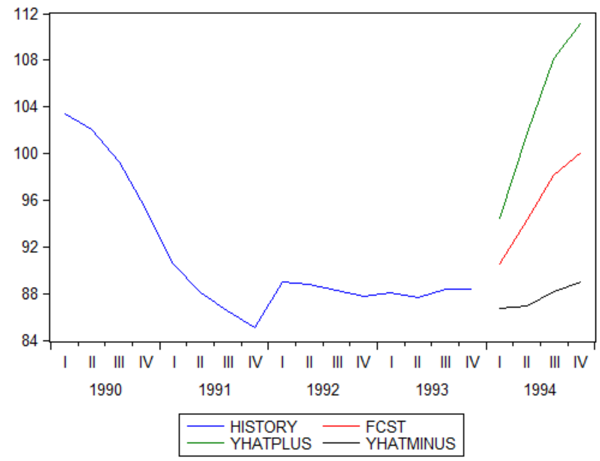
8. Modify the sample to include what you want to see in the graph:

```
smpl 1990:1 1994:4
```

9. Open the history, the forecast and the lower and upper limits all in one group:

```
group group01 history fcst yhatplus yhatminus
```

10. Just open the group and graph them all together:

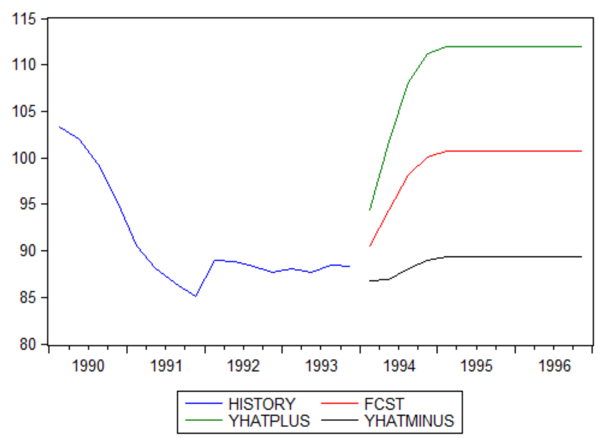


What if we want to forecast all the values until the fourth quarter of 1996?

11. Just select the sample for your graph:

```
smpl 1990:1 1996:4
```

12. Open the group you produced in step (9) and graph it.



Note that the forecast becomes flat after when forecasting beyond the fourth period. This is because MA models have a short memory and in this MA(4) case, beyond the fourth period the estimated equation just does not have any information to forecast. We covered this in detail in the previous chapter for an MA(2). What if you want to compare the actual values with the forecast? Remember that we do have the data for the following years.

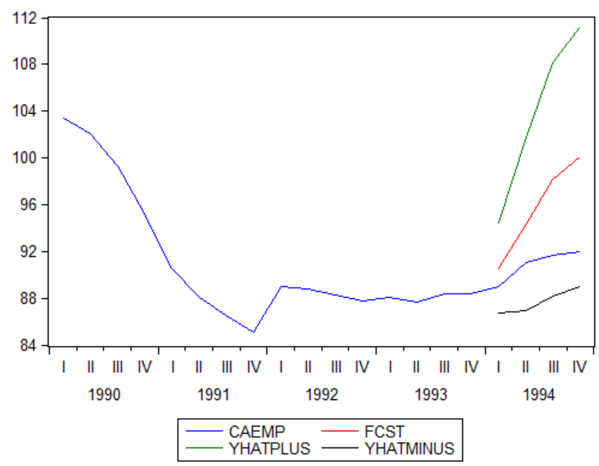
13. Modify the sample again to cover the periods of the forecast and where we have actual data:

```
smpl 1990:1 1994:4
```

14. Create another group. This time with the actual data (caemp) instead of the history.

```
group group02 caemp fcst yhatplus yhatminus
```

15. Then open the group and graph:



17.2 Autoregressive Models

Before we start, let's make sure we have the correct sample we will use to estimate the model:

```
smpl 1962:1 1993:4
```

Based on different model selection criteria, our preferred AR model was an AR(2) model:

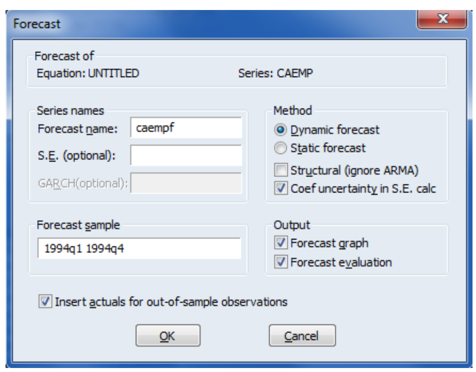
```
ls caemp c ar(1) ar(2)
```

Dependent Variable: CAEMP
Method: ARMA Maximum Likelihood (BFGS)
Sample: 1962Q1 1993Q4
Included observations: 128
Convergence achieved after 7 iterations
Coefficient covariance computed using outer product of gradients

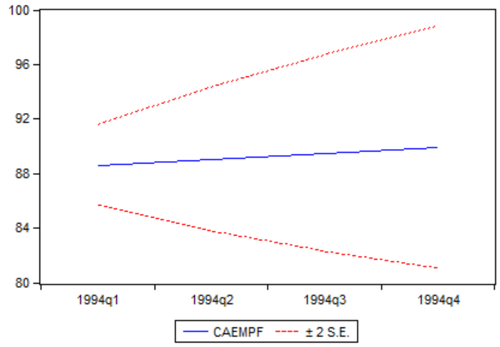
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	98.03049	3.812684	25.71167	0.0000
AR(1)	1.448340	0.064717	22.37973	0.0000
AR(2)	−0.476697	0.064966	−7.337611	0.0000
SIGMASQ	2.088639	0.166810	12.52109	0.0000

R-squared	0.962568	Mean dependent var	101.0176
Adjusted R-squared	0.961662	S.D. dependent var	7.499163
S.E. of regression	1.468337	Akaike info criterion	3.666458
Sum squared resid	267.3458	Schwarz criterion	3.755584
Log likelihood	−230.6533	Hannan-Quinn criter.	3.702671
F-statistic	1062.889	Durbin-Watson stat	2.054328
Prob(F-statistic)	0.000000		

The simplest way to forecast the values between the first quarter of 1994 and the fourth quarter of 1994 is to go to click on the icon “forecast” and then choose the correct forecast sample:



This will yield the following forecast:



Forecast: CAEMPF	
Actual: CAEMP	
Forecast sample: 1994Q1 1994Q4	
Included observations: 4	
Root Mean Squared Error	1.847093
Mean Absolute Error	1.691644
Mean Abs. Percent Error	1.850444
Theil Inequality Coefficient	0.010251
Bias Proportion	0.838765
Variance Proportion	0.133755
Covariance Proportion	0.027480

A second more interesting way to obtain the same forecast is to follow these steps:

1. Select the sample you want to use for your model:

```
smpl 1962:1 1993:4
```

2. Estimate the AR(2) model and store your estimation under the name ar2:

```
equation ar2.ls caemp c ar(1) ar(2)
```

3. Select the sample you want to include in your forecast:

```
smpl 1994:1 2010:4
```

4. Generate the forecast and the standard error of the forecast:

```
ar2.forecast yhat se
```

5. Generate the variable that will store the forecasted values:

```
genr fcst2=yhat
```

6. Generate the upper and lower levels for your 95% confidence intervals:

```
genr yhatplus2=yhat+1.96*se
genr yhatminus2=yhat-1.96*se
```

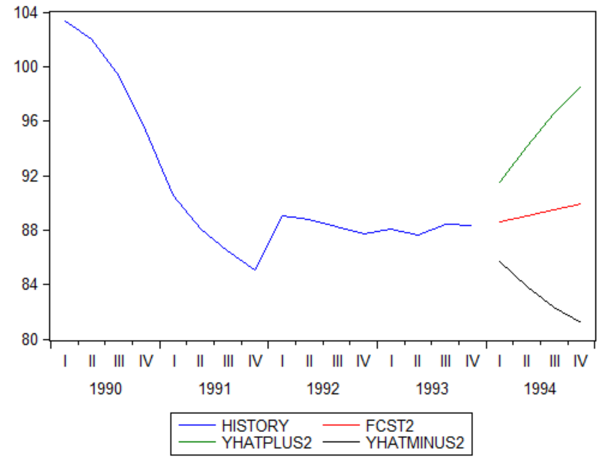
7. Select the sample you want to see in your forecast graph:

```
smpl 1990:1 1994:4
```

8. Create a group of all the variables you want to include in your graph:

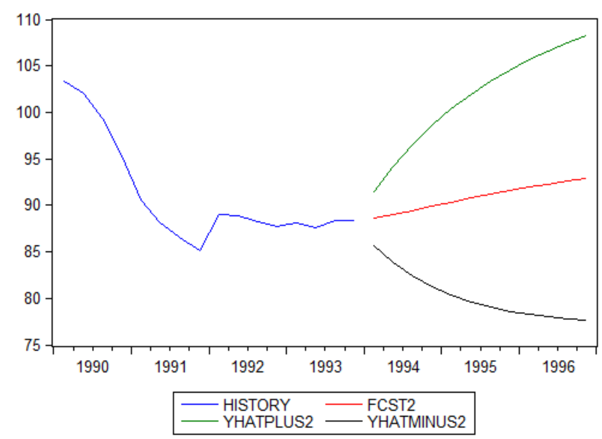
```
group group03 history fcst2 yhatplus2 yhatminus2
```

9. Open the group and graph all variables together:



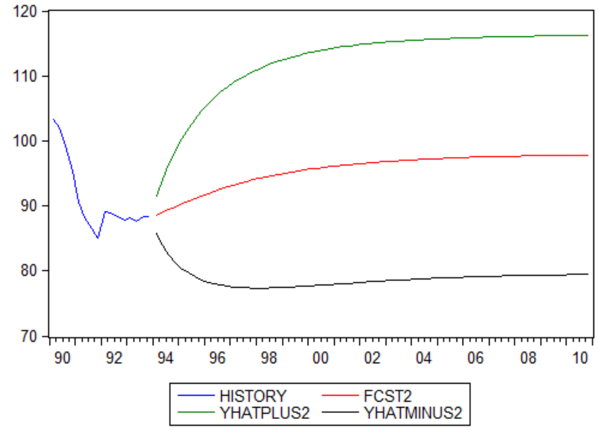
If you want to see the forecast all the way until the end of 1996, just modify the sample size:

```
smpl 1990:1 1996:4
```



10. For the forecast that includes the values until 2010.

```
smpl 1990:1 2010:4
```



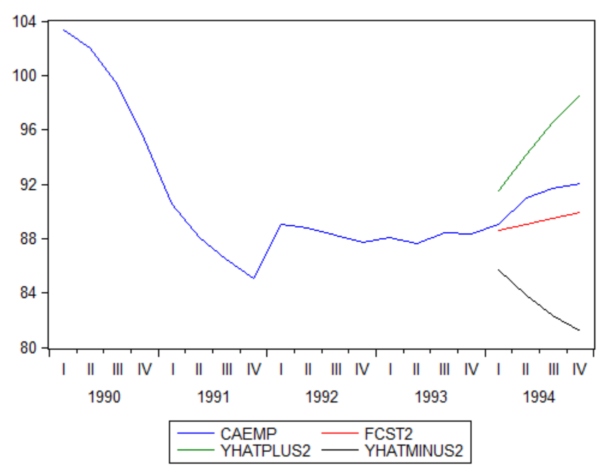
11. If you want to include the actual values in the forecast, select the sample that contains actual values first:

```
smpl 1990:1 1994:4
```

12. Then create a group with the actual values (caempl), the forecast and the 95% upper and lower confidence intervals:

```
group group04 caemp fcst2 yhatplus2 yhatminus2
```

13. Finally, open the group and generate the line graph will all the variables:



Notice that the forecast lies very close to the actual values. This AR(2) model appears to be a better forecasting model than the MA(4) model presented earlier.

Chapter 18
Forecasting with Trend, Seasonal, and Cyclical Components

18.1 Structure

Consider the following model:

$$y_t = T_t(\beta) + \sum_{i=1}^{s} \gamma_i D_{it} + \sum_{i=1}^{v} \delta_i HD_{it} + \nu_t \tag{18.1}$$

where $T_t(\beta)$ is the trend, $\sum_{i=1}^{s} \gamma_i D_{it} + \sum_{i=1}^{v} \delta_i HD_{it}$ is the seasonal component, and ν_t is the cycle and it is given by:

$$\begin{aligned} \nu_t \;&= \phi_1 \nu_{t-1} + \phi_2 \nu_{t-2} + \cdots + \phi_p \nu_{t-p} + \varepsilon_t \\ &\quad + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \end{aligned}$$

$$\varepsilon_t = WN(0, \sigma^2)$$

The time trend can be modeled, for example, as:

$$\begin{aligned} T_t(\beta) &= \beta_1 TIME_t, \text{ or} \\ T_t(\beta) &= \beta_1 TIME_t + \beta_2 TIME_t^2 \end{aligned}$$

The h-step ahead forecast (at time $T+h$) is:

$$y_{T+h} = T_{T+h}(\beta) + \sum_{i=1}^{s} \gamma_i D_{i,T+h} + \sum_{i=1}^{v} \delta_i HD_{i,T+h} + \nu_{T+h} \tag{18.2}$$

$T_{T+h}(\beta)$:	Known.
$\sum_{i=1}^{s} \gamma_i D_{i,T+h}$:	Known.
$\sum_{i=1}^{v} \delta_i HD_{i,T+h}$:	Known.
ν_{T+h}:	Known/unknown.

Consider the following AR(1) example to understand the difference between dynamic and static forecasts:

$$y_t = \phi y_{t-1} + \varepsilon_t \quad (18.3)$$

The 2-step-ahead forecast, made at time T is:

$$y_{T+2,T} = \phi y_{t+1} \quad (18.4)$$

Dynamic forecast: Uses the forecasted values for y_{t+1}, obtained from the one-step-ahead forecast.

Static forecast: Uses the actual values for y_{t+1}.

18.2 Recursive Estimation Procedures for Diagnosing and Selecting Forecasting Models

Recursive estimation:

· Beginning with a small sample → estimate the model.
· Add one observation → estimate the model again.
· Repeat until the whole sample is used.

Why is this useful?

· Stability assessment.
· Model selection.

To assess the stability of a model, we use the recursive residuals.
We assumed that the parameters β, γ, δ, ϕ, and θ in Equation 18.1 are stable (i.e., they do not change over time).
What if they are not stable? The model will provide a poor forecast.

Recursive Parameter Estimation: Consider the following model:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (18.5)$$

$$\varepsilon_t = WN(0, \sigma^2)$$

for $t = 1, 2, \ldots, T$

· Start with two observations → estimate β_0 and β_1.
· Use three observations → estimate β_0 and β_1 again.
· Obtain the sequence of $T-1$ estimates of β_0 and β_1.

Recursive Residuals: Each time we estimate β_0 and β_1, we compute the 1-step-ahead forecast $\hat{y}_{t+1,t} = \hat{\beta}_0 + \hat{\beta}_1 x_{t+1}$, to then compute the residuals:

$$\hat{\varepsilon}_{t+1,t} = y_{t+1} - \hat{y}_{t+1,t} \quad (18.6)$$

The $\hat{\varepsilon}_{t+1,t}$ in Equation 18.6 are the recursive residuals. After obtaining the residuals:

- Plot the residuals with two standard error bands.
- If many of the residuals fall outside the bands $\rightarrow$ one or more parameters are unstable.

CUSUM: Cumulative sum of standardized recursive residuals.

$$CUSUM_t \equiv \sum_{\tau=2}^{t} w_{\tau+1,\tau} \quad \text{for} \quad t = 2, 3, \ldots, T-1 \tag{18.7}$$

$$w_{t+1,t} \overset{\text{iid}}{\sim} N(0,1)$$

- Plot the $w_{t+1,t}$ with the 95% probability bounds.
- If violated the bounds $\rightarrow$ evidence of parameter inestability.

Chapter 19
EViews: Forecasting with Trend, Seasonal, and Cyclical Components

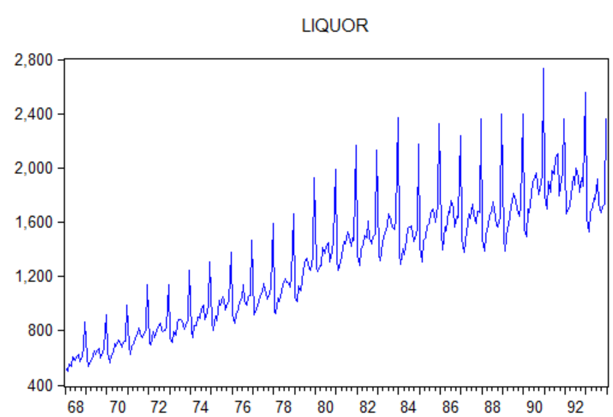
This chapter will cover an example on how to forecast a model with trend, seasonal component, and cyclical component.

19.1 Forecasting Sales

The variable we will use is monthly U.S. liquor sales from January 1968 until December 1993. We use the sample

```
smpl 1967m1 1993m12
```

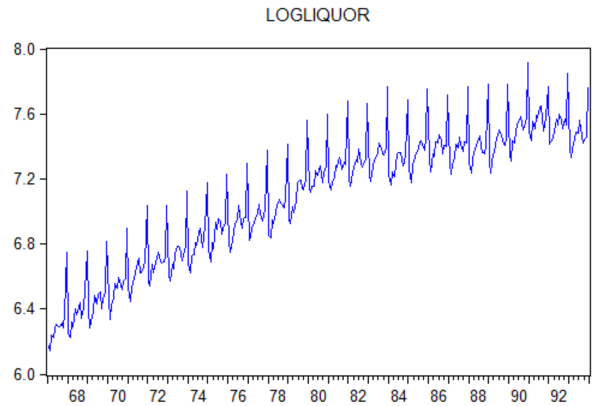
The time series graph of the series is:



Because the variance seems to be larger for larger values of sales, we will work with the variable in logarithms:

```
smpl 1967m1 1998m12
genr logliquor = log(liquor)
```

The time series graph of the logarithm of liquor is:



The series shows a strong seasonal component with sales being higher in December. As a first step, let's estimate the model with a quadratic trend:

```
smpl 1968m1 1993m12
ls logliquor c @trend @trend^2
```

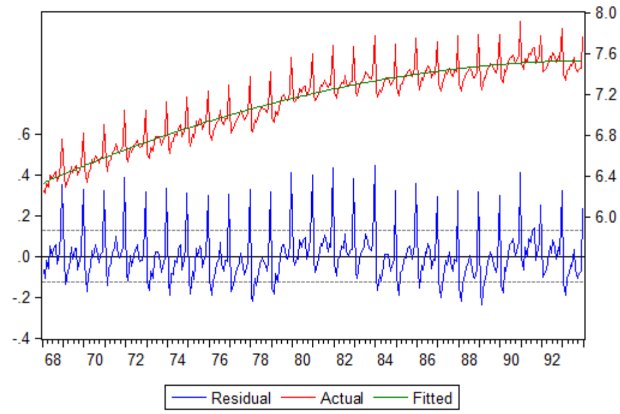
That yields the following regression output:

Dependent Variable: LOGLIQUOR
Method: Least Squares
Sample: 1968M01 1993M12
Included observations: 312

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.237356	0.024496	254.6267	0.0000
@TREND	0.007690	0.000336	22.91552	0.0000
@TREND^2	$-1.14E-05$	$9.74E-07$	-11.72695	0.0000

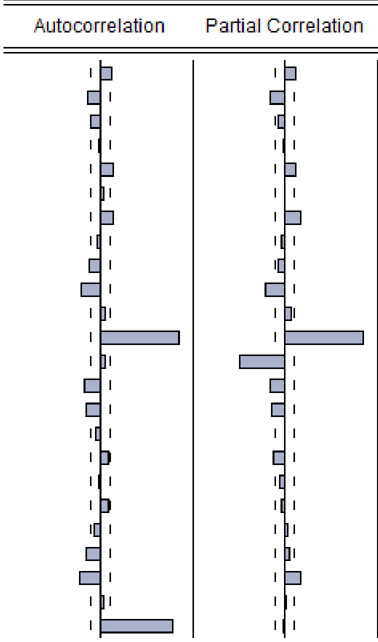
R-squared	0.892394	Mean dependent var	7.112383
Adjusted R-squared	0.891698	S.D. dependent var	0.379308
S.E. of regression	0.124828	Akaike info criterion	-1.314196
Sum squared resid	4.814823	Schwarz criterion	-1.278206
Log likelihood	208.0146	Hannan-Quinn criter.	-1.299812
F-statistic	1281.296	Durbin-Watson stat	1.752858
Prob(F-statistic)	0.000000		

with the following in-sample forecast, forecast errors:



The seasonal component (and any potential cyclical component) is still in the error term. Let's look at the autocorrelation and the partial autocorrelation function for various values of the displacement:

Sample: 1968M01 1993M12
Included observations: 312



	AC	PAC	Q-Stat	Prob
1	0.117	0.117	4.3158	0.038
2	-0.149	-0.165	11.365	0.003
3	-0.106	-0.069	14.943	0.002
4	-0.014	-0.017	15.007	0.005
5	0.142	0.125	21.449	0.001
6	0.041	-0.004	21.979	0.001
7	0.134	0.175	27.708	0.000
8	-0.029	-0.046	27.975	0.000
9	-0.136	-0.080	33.944	0.000
10	-0.205	-0.206	47.611	0.000
11	0.056	0.080	48.632	0.000
12	0.888	0.879	306.26	0.000
13	0.055	-0.507	307.25	0.000
14	-0.187	-0.159	318.79	0.000
15	-0.159	-0.144	327.17	0.000
16	-0.059	-0.002	328.32	0.000
17	0.091	-0.118	331.05	0.000
18	-0.010	-0.055	331.08	0.000
19	0.086	-0.032	333.57	0.000
20	-0.066	0.028	335.03	0.000
21	-0.170	0.044	344.71	0.000
22	-0.231	0.180	362.74	0.000
23	0.028	0.016	363.00	0.000
24	0.811	-0.014	586.50	0.000

Notice the seasonal displacements at 12 and 24 and some evidence of cyclical dynamics. If we estimate the model with the monthly dummies we have:

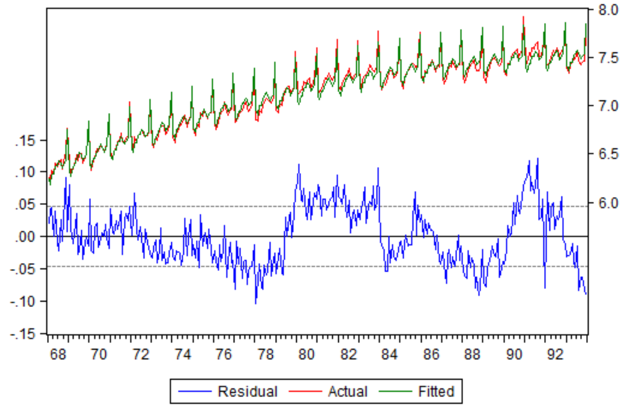
```
ls logliquor @trend @trend^2 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12
```

Dependent Variable: LOGLIQUOR
Method: Least Squares
Sample: 1968M01 1993M12
Included observations: 312

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@TREND	0.007656	0.000123	62.35882	0.0000
@TREND^2	$-1.14E-05$	$3.56E-07$	-32.06823	0.0000
D1	6.147456	0.012340	498.1699	0.0000
D2	6.088653	0.012353	492.8890	0.0000
D3	6.174127	0.012366	499.3008	0.0000
D4	6.175220	0.012378	498.8970	0.0000
D5	6.246086	0.012390	504.1398	0.0000
D6	6.250387	0.012401	504.0194	0.0000
D7	6.295979	0.012412	507.2402	0.0000
D8	6.268043	0.012423	504.5509	0.0000
D9	6.203832	0.012433	498.9630	0.0000
D10	6.229197	0.012444	500.5968	0.0000
D11	6.259770	0.012453	502.6602	0.0000
D12	6.580068	0.012463	527.9819	0.0000

R-squared	0.986111	Mean dependent var	7.112383
Adjusted R-squared	0.985505	S.D. dependent var	0.379308
S.E. of regression	0.045666	Akaike info criterion	-3.291086
Sum squared resid	0.621448	Schwarz criterion	-3.123131
Log likelihood	527.4094	Hannan-Quinn criter.	-3.223959
Durbin-Watson stat	0.586187		

The graph with the in-sample forecasting errors is:



and the correlogram of the residuals:

Sample: 1968M01 1993M12
Included observations: 312

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.700	0.700	154.34	0.000
		2	0.686	0.383	302.86	0.000
		3	0.725	0.369	469.36	0.000
		4	0.569	-0.141	572.36	0.000
		5	0.569	0.017	675.58	0.000
		6	0.577	0.093	782.19	0.000
		7	0.460	-0.078	850.06	0.000
		8	0.480	0.043	924.38	0.000
		9	0.466	0.030	994.46	0.000
		10	0.327	-0.188	1029.1	0.000
		11	0.364	0.019	1072.1	0.000
		12	0.355	0.089	1113.3	0.000
		13	0.225	-0.119	1129.9	0.000
		14	0.291	0.065	1157.8	0.000
		15	0.211	-0.119	1172.4	0.000
		16	0.138	-0.031	1178.7	0.000
		17	0.195	0.053	1191.4	0.000
		18	0.114	-0.027	1195.7	0.000
		19	0.055	-0.063	1196.7	0.000
		20	0.134	0.089	1202.7	0.000
		21	0.062	0.018	1204.0	0.000
		22	-0.006	-0.115	1204.0	0.000
		23	0.084	0.086	1206.4	0.000
		24	-0.039	-0.124	1206.9	0.000

The seasonality disappeared, but there is still a strong cyclical component. With an AR(3) model for the cycle we have:

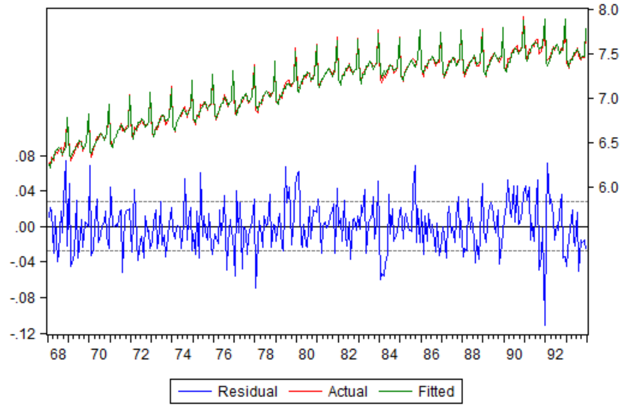
```
ls logliquor @trend @trend^2 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 AR(1) AR(2) AR(3)
```

Dependent Variable: LOGLIQUOR
Method: ARMA Maximum Likelihood (BFGS)
Sample: 1968M01 1993M12
Included observations: 312
Convergence achieved after 6 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@TREND	0.007780	0.000706	11.01819	0.0000
@TREND^2	$-1.21E-05$	$1.91E-06$	-6.344009	0.0000
D1	6.148880	0.056367	109.0868	0.0000
D2	6.090196	0.056318	108.1392	0.0000
D3	6.175964	0.056660	109.0008	0.0000
D4	6.177579	0.056817	108.7283	0.0000
D5	6.248742	0.056660	110.2840	0.0000
D6	6.253401	0.056659	110.3694	0.0000
D7	6.299354	0.057039	110.4397	0.0000
D8	6.271793	0.056571	110.8657	0.0000
D9	6.207966	0.057248	108.4392	0.0000
D10	6.233549	0.057165	109.0456	0.0000
D11	6.264644	0.055956	111.9571	0.0000
D12	6.585369	0.056719	116.1046	0.0000
AR(1)	0.272320	0.051777	5.259525	0.0000
AR(2)	0.236852	0.048599	4.873586	0.0000
AR(3)	0.391816	0.052596	7.449559	0.0000
SIGMASQ	0.000712	$5.39E-05$	13.22494	0.0000

R-squared	0.995032	Mean dependent var	7.112383
Adjusted R-squared	0.994745	S.D. dependent var	0.379308
S.E. of regression	0.027496	Akaike info criterion	-4.288442
Sum squared resid	0.222271	Schwarz criterion	-4.072500
Log likelihood	686.9970	Hannan-Quinn criter.	-4.202137
Durbin-Watson stat	1.887695		

The in-sample forecasting errors are:



Where the residuals appear to be White Noise. The corresponding correlogram of the residuals is:

Sample: 1968M01 1993M12
Included observations: 312

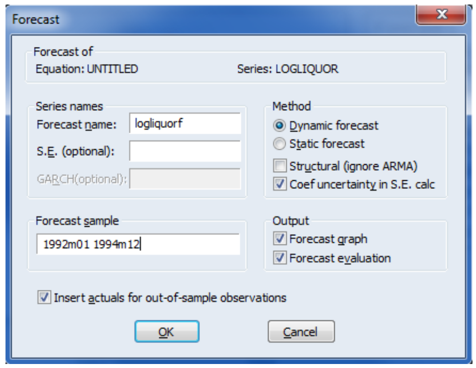
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.054	0.054	0.9318	0.334
		2	0.041	0.038	1.4591	0.482
		3	0.026	0.022	1.6810	0.641
		4	-0.086	-0.091	4.0397	0.401
		5	-0.006	0.001	4.0531	0.542
		6	0.068	0.075	5.5120	0.480
		7	-0.037	-0.040	5.9428	0.546
		8	0.080	0.072	8.0129	0.432
		9	0.089	0.081	10.552	0.308
		10	-0.153	-0.159	18.125	0.053
		11	-0.006	-0.005	18.135	0.079
		12	0.144	0.175	24.888	0.015
		13	-0.081	-0.083	27.063	0.012
		14	0.149	0.112	34.364	0.002
		15	-0.040	-0.061	34.882	0.003
		16	-0.091	-0.063	37.632	0.002
		17	0.058	0.045	38.749	0.002
		18	-0.064	-0.049	40.102	0.002
		19	-0.111	-0.076	44.199	0.001
		20	0.100	0.056	47.576	0.000
		21	0.039	0.044	48.084	0.001
		22	-0.115	-0.111	52.565	0.000
		23	0.151	0.131	60.268	0.000
		24	-0.072	-0.039	62.031	0.000

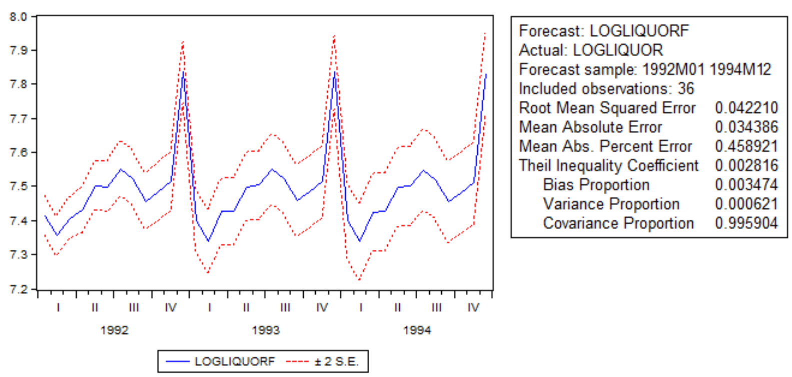
The Ljung-Pierce Q-statistic fails to reject the null hypothesis of White Noise for displacements below 10.

To obtain the forecast we need first to modify the sample to be able to include the forecasted values:

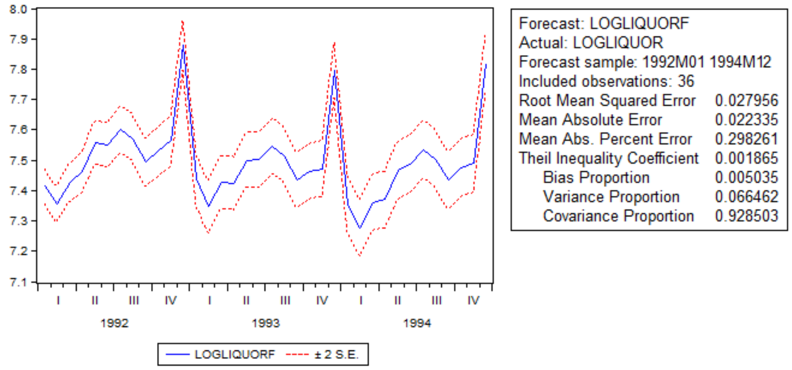
```
smpl 1968m1 1994m12
```

1) The dynamic forecast:





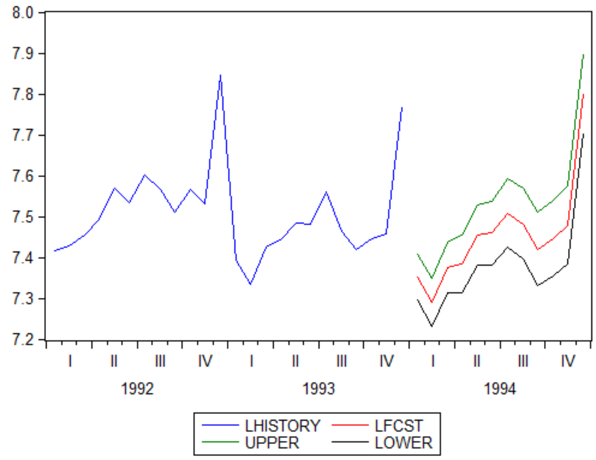
2) The static forecast:



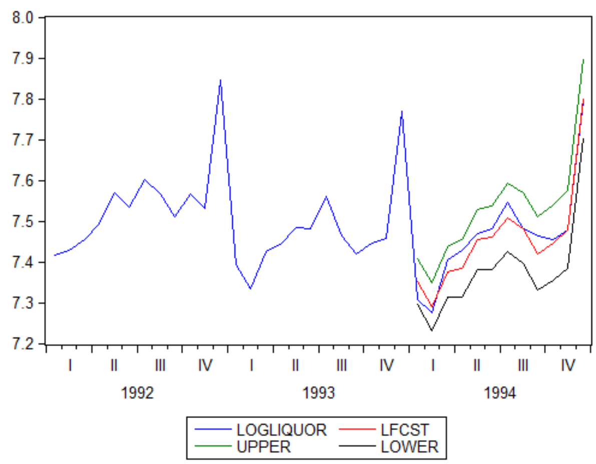
As explained before, the difference between the dynamic and the static forecast is that the dynamic forecast uses previously forecasted values of the lagged dependent variables in forming forecasts of the current value. The static forecast calculates the sequence of one-step ahead forecasts, using actual, rather than forecasted values for lagged dependent variables, if available.

A step-by-step approach to obtain the static forecast is:

```
smpl 1966:1 1993:12
genr lhistory=logliquor
smpl 1994:1 1998:12
forecast yhat se
genr lfcst=yhat
genr fcst=@exp(yhat)
genr upper = yhat + 1.96*se
genr lower = yhat - 1.96*se
smpl 1992:1 1994:12
group group01 lhistory lfcst upper lower
```




```
group group02 logliquor lfcst upper lower
```



For the details behind these steps, please refer to Chapter 17.

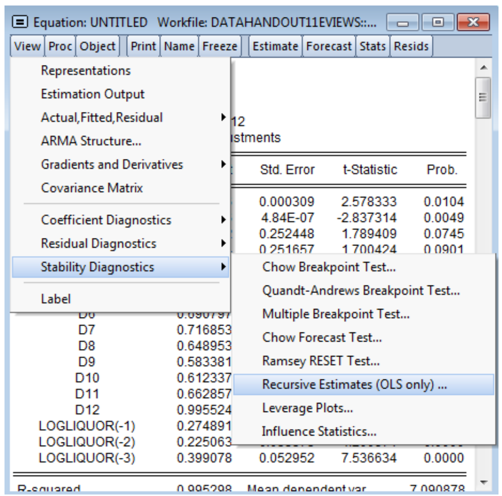
19.2 Recursive Estimation Procedures

Recursive estimation procedures can only be estimated when the model was estimated by ordinary least squares. Usual estimation of ARMA models use nonlinear least squares or maximum likelihood procedures that are not compatible with recursive estimation procedures.

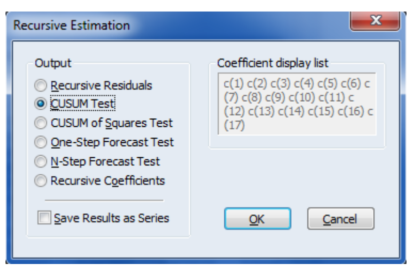
To be able to work with our model, estimate it again, but use the lag operators rather than the AR(1) notation:

```
smpl 1966:1 1993:12
ls logliquor @trend @trend^2 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12
 logliquor(-1) logliquor(-2) logliquor(-3)
```

Go to "View," then "Stability Diagnostics," and finally "Recursive Estimates."

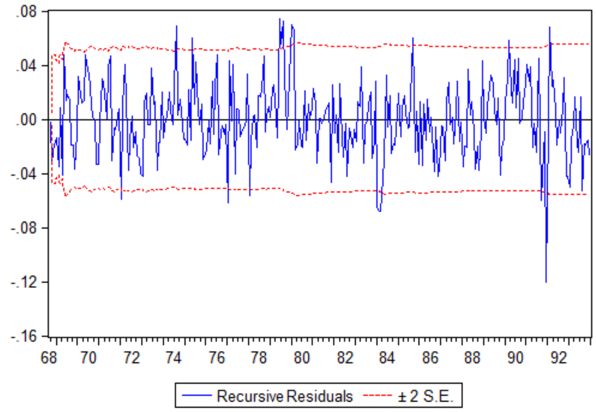


You will obtain the following menu:

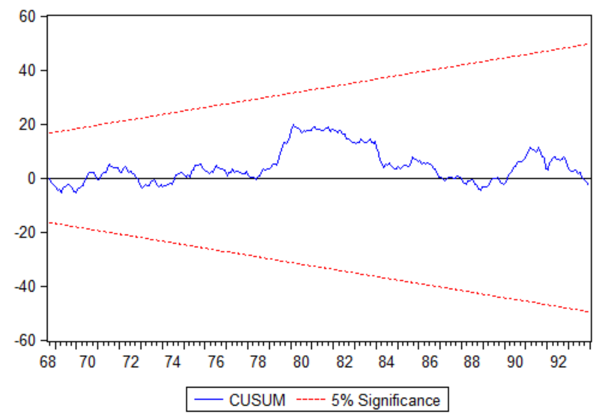


The options we discussed in Chapter 18 are (1) Recursive residuals, (2) CUSUM test, and (3) Recursive coefficients.

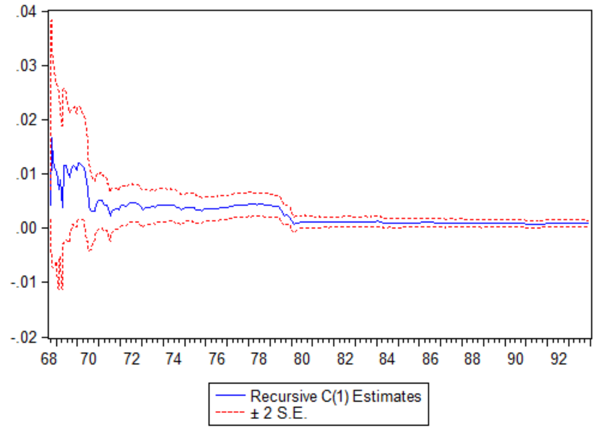
(1) Recursive residuals.



(2) CUSUM test.



(3) Recursive coefficients, where we selected to have the results only for C(1), the first coefficient in the regression table results.


.04
.03
.02
.01
.00
-.01
-.02
68 70 72 74 76 78 80 82 84 86 88 90 92
Recursive C(1) Estimates
± 2 S.E.

Chapter 20
Forecasting with Regression Models

20.1 Conditional Forecasting Models

So far we have been focusing on univariate models. However, other variables (e.g., x_t) can help predict future values of y_t. For example,

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{20.1}$$

$$\varepsilon_t \sim WN(0, \sigma^2) \tag{20.2}$$

The idea in conditional forecasting models is to generate the forecast of y conditional on an assumed future value of x (scenario analysis).

Let x^*_{T+h} be the h-step ahead forecast of x. Then, the h-step-ahead forecast of y given x^*_{T+h} is:

$$y_{T+h,T} | x^*_{T+h} = \beta_0 + \beta_1 x^*_{T+h} \tag{20.3}$$

Parameter Uncertainty:

- *Specification Uncertainty*: Our model is a simplification of the real world.
- *Innovation Uncertainty*: Future shocks ε_t are unknown when the forecast is made.
- *Parameter Uncertainty*: The parameters in our models θ, ϕ, β are unknown. Hence, the coefficients we use are just estimates $\hat{\theta}$, $\hat{\phi}$, $\hat{\beta}$, which are subject to sample variability.

20.2 Unconditional Forecasting Models

When using the following model to forecast y_{T+h},

$$y_{T+h,T} = \beta_0 + \beta_1 x_{T+h,T} \tag{20.4}$$

we face the problem that we do not have the value for $x_{T+h,T}$. How about forecast it with an ARMA model? Perhaps it is easier to just to forecast y with an ARMA model.

A feasible model is:

$$y_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t \tag{20.5}$$

where we use the lagged known value of x_t. This is good for a 1-step=ahead forecast.

x may be perfectly deterministic, e.g., trend or seasonal components.

20.3 Vector Autoregressions

AR(p): Univariate autoregression of order p.
VAR(p): Vector (mutivariate) autoregression of order p.

- N variables.
- N equations.
- p lags on every other variable.
- Allows for cross-variable dynamics.

Example: 2 variable VAR(p), $y_{1,t}$ and $y_{2,t}$, with $p = 1$:

$$y_{1,t} = \phi_{11} y_{1,t-1} + \phi_{12} y_{2,t-1} + \varepsilon_{1,t} \tag{20.6}$$

$$y_{2,t} = \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + \varepsilon_{2,t} \tag{20.7}$$

$$\varepsilon_{1,t} \sim WN(0, \sigma_1^2)$$
$$\varepsilon_{2,t} \sim WN(0, \sigma_2^2)$$

$$cov(\varepsilon_{1,t}, \varepsilon_{2,t}) = \sigma_{12}$$

Model selection: How do we select p? With multivariate versions of the AIC and SIC.

Forecasting: Same as the AR $\rightarrow$ use the chair rule of forecasting.

Predictive Causality: It has two principles:

1. Cause should occur before effect.
2. A causal series should contain information not available in other series.

Unrestricted VAR: Everything causes everything.

20.4 Impulse-Response Functions

The Impulse-Response Function (IRF) helps us learn about the dynamics of a variable.

How does a unit innovation to a series affects it now and in the future?

Unit shock = one standard deviation of ε_t.

Consider the VAR(1) model of Equations 20.6 and 20.7. Remember from previous chapters that the AR had an MA representation. Same works for a VAR. The moving average representation of the VAR(1)of Equations 20.6 and 20.7 is:

$$y_{1,t} = \varepsilon_{1,t} + \phi_{11}\varepsilon_{1,t-1} + \phi_{12}\varepsilon_{2,t-1} + \cdots \quad (20.8)$$

$$y_{2,t} = \varepsilon_{2,t} + \phi_{21}\varepsilon_{1,t-1} + \phi_{22}\varepsilon_{2,t-1} + \cdots \quad (20.9)$$

Cholesky Decomposition: Need to decide the order of the variables "cause and effect."

If y_1 is ordered first. That is, y_1 occurs first (y_1 causes y_2):

$$y_{1,t} = b_{11}^0\varepsilon'_{1,t} \qquad + b_{11}^1\varepsilon'_{1,t-1} + b_{12}^1\varepsilon'_{2,t-1} + \cdots \quad (20.10)$$

$$y_{2,t} = b_{21}^0\varepsilon'_{1,t} + b_{22}^0\varepsilon'_{2,t} + b_{21}^1\varepsilon'_{1,t-1} + b_{22}^1\varepsilon'_{2,t-1} + \cdots \quad (20.11)$$

where b are normalized coefficients, and

$$\varepsilon'_{1,t} \sim WN(0,1)$$
$$\varepsilon'_{2,t} \sim WN(0,1)$$

$$cov(\varepsilon'_{1,t}, \varepsilon'_{2,t}) = 0$$

What Equations 20.10 and 20.11 basically say is that shocks (unexpected changes) to y_1 or y_2 affect the path of both y_1 and y_2. The only restriction is that at time t, y_2 does not affect y_1.

Four different Impulse-Response Functions:

- IRF of y_1 to a a shock in y_1, ε'_1: $\{b_{11}^0, b_{11}^1, b_{11}^2, \ldots\}$
- IRF of y_1 to a a shock in y_2, ε'_2: $\{b_{12}^1, b_{12}^2, b_{13}^3, \ldots\}$
- IRF of y_2 to a a shock in y_1, ε'_1: $\{b_{21}^0, b_{21}^1, b_{21}^2, \ldots\}$
- IRF of y_2 to a a shock in y_2, ε'_2: $\{b_{22}^0, b_{22}^1, b_{22}^2, \ldots\}$

We will obtain a graphical representation of these IRFs in the following chapter.

Chapter 21
EViews: Vector Autoregressions

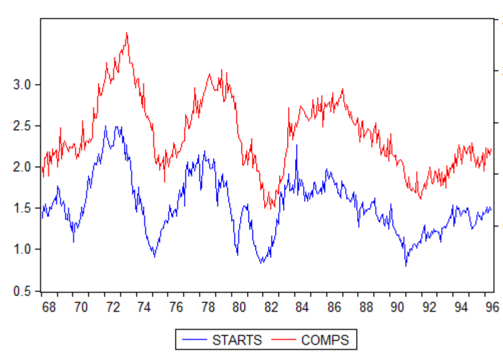
This chapter covers the computer commands for the estimation of Vector Autoregressions (VAR), forecasting with regression models, and Impulse-response functions (IRF).

21.1 Estimation of Vector Autoregressions

The data we will use includes two variables: (1) the seasonally adjusted housing starts and (2) housing completions. These are monthly observations from January 1968 through June 1996. A graph of both variables can be obtained with:

```
group both starts comps
both.line(d)
```

To obtain:



We will use the data from January 1968 through December 1991 for model estimation and the forecast will be done for the period from January 1992 through June 1996.

The correlograms for both variables are:

Sample: 1968M01 1996M06
Included observations: 342

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.947	0.947	309.40	0.000
		2	0.939	0.406	614.29	0.000
		3	0.919	0.073	907.27	0.000
		4	0.904	0.027	1191.5	0.000
		5	0.870	-0.196	1455.6	0.000
		6	0.848	-0.057	1707.4	0.000
		7	0.813	-0.121	1939.8	0.000
		8	0.784	-0.061	2156.3	0.000
		9	0.748	-0.052	2354.1	0.000
		10	0.714	-0.056	2534.6	0.000
		11	0.672	-0.090	2694.9	0.000
		12	0.630	-0.092	2836.6	0.000
		13	0.591	-0.015	2961.3	0.000
		14	0.549	-0.022	3069.2	0.000
		15	0.511	0.050	3163.1	0.000
		16	0.465	-0.056	3241.1	0.000
		17	0.427	0.010	3307.1	0.000
		18	0.377	-0.112	3358.6	0.000
		19	0.335	-0.055	3399.4	0.000
		20	0.287	-0.050	3429.4	0.000
		21	0.237	-0.109	3450.1	0.000
		22	0.194	0.037	3463.9	0.000
		23	0.146	-0.060	3471.7	0.000
		24	0.097	-0.061	3475.2	0.000

Sample: 1968M01 1996M06
Included observations: 342

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.940	0.940	304.66	0.000
		2	0.914	0.263	593.59	0.000
		3	0.888	0.070	866.99	0.000
		4	0.853	-0.062	1120.5	0.000
		5	0.816	-0.086	1352.9	0.000
		6	0.778	-0.061	1564.7	0.000
		7	0.737	-0.062	1755.3	0.000
		8	0.689	-0.097	1922.6	0.000
		9	0.648	-0.004	2071.0	0.000
		10	0.594	-0.120	2196.2	0.000
		11	0.553	0.038	2304.9	0.000
		12	0.510	0.008	2397.6	0.000
		13	0.475	0.076	2478.3	0.000
		14	0.422	-0.146	2542.0	0.000
		15	0.375	-0.064	2592.5	0.000
		16	0.321	-0.116	2629.8	0.000
		17	0.268	-0.063	2655.9	0.000
		18	0.221	-0.013	2673.5	0.000
		19	0.174	0.017	2684.5	0.000
		20	0.127	-0.022	2690.4	0.000
		21	0.079	-0.017	2692.7	0.000
		22	0.041	0.038	2693.3	0.000
		23	-0.005	-0.025	2693.3	0.000
		24	-0.047	-0.060	2694.1	0.000

Both show a strong cyclical component.

The cross-correlation function shows the correlation between a variable and the lags of another variable. To obtain it open both variables as a group, then go to "view" and then "cross-correlation" to obtain:

Sample: 1968M01 1996M06
Included observations: 342
Correlations are asymptotically consistent approximations

STARTS,COMPS(-i)	STARTS,COMPS(+i)	i	lag	lead
		0	0.7789	0.7789
		1	0.7311	0.8095
		2	0.6827	0.8417
		3	0.6389	0.8676
		4	0.5938	0.8918
		5	0.5445	0.9025
		6	0.5034	0.9105
		7	0.4529	0.9039
		8	0.4067	0.8967
		9	0.3726	0.8828
		10	0.3217	0.8679
		11	0.2722	0.8455
		12	0.2257	0.8195
		13	0.1904	0.7946
		14	0.1370	0.7585
		15	0.1018	0.7293
		16	0.0547	0.6977
		17	0.0138	0.6665
		18	-0.0272	0.6225
		19	-0.0645	0.5886
		20	-0.1096	0.5427
		21	-0.1399	0.5026
		22	-0.1702	0.4569
		23	-0.2015	0.4126
		24	-0.2269	0.3698

This cross-correlation shows there is a strong correlation between the lags and leads of these variables. This is evidence of dynamic interaction between the two that can be modeled with a vector autoregression model.

The VAR(4) as presented in Equations 20.6 and 20.7 can be estimated using EViews in two different ways: (1) Equation by equation and (2) jointly. Equation by equation can we simply type the following command:

```
smpl 1968m01 1991m12
ls starts c starts(-1) starts(-2) starts(-3) starts(-4) comps(-1) comps(-2) comps(-3) comps(-4)
ls comps c starts(-1) starts(-2) starts(-3) starts(-4) comps(-1) comps(-2) comps(-3) comps(-4)
```

That give us the following estimation output:

Dependent Variable: STARTS
Method: Least Squares
Sample (adjusted): 1968M05 1991M12
Included observations: 284 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.146871	0.044235	3.320264	0.0010
STARTS(-1)	0.659939	0.061242	10.77587	0.0000
STARTS(-2)	0.229632	0.072724	3.157587	0.0018
STARTS(-3)	0.142859	0.072655	1.966281	0.0503
STARTS(-4)	0.007806	0.066032	0.118217	0.9060
COMPS(-1)	0.031611	0.102712	0.307759	0.7585
COMPS(-2)	−0.120781	0.103847	−1.163069	0.2458
COMPS(-3)	−0.020601	0.100946	−0.204078	0.8384
COMPS(-4)	−0.027404	0.094569	−0.289779	0.7722

R-squared	0.895566	Mean dependent var	1.574771
Adjusted R-squared	0.892528	S.D. dependent var	0.382362
S.E. of regression	0.125350	Akaike info criterion	−1.284241
Sum squared resid	4.320952	Schwarz criterion	−1.168605
Log likelihood	191.3622	Hannan-Quinn criter.	−1.237880
F-statistic	294.7796	Durbin-Watson stat	1.991908
Prob(F-statistic)	0.000000		

Dependent Variable: COMPS
Method: Least Squares
Sample (adjusted): 1968M05 1991M12
Included observations: 284 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.045347	0.025794	1.758045	0.0799
STARTS(-1)	0.074724	0.035711	2.092461	0.0373
STARTS(-2)	0.040047	0.042406	0.944377	0.3458
STARTS(-3)	0.047145	0.042366	1.112805	0.2668
STARTS(-4)	0.082331	0.038504	2.138238	0.0334
COMPS(-1)	0.236774	0.059893	3.953313	0.0001
COMPS(-2)	0.206172	0.060554	3.404742	0.0008
COMPS(-3)	0.120998	0.058863	2.055593	0.0408
COMPS(-4)	0.156729	0.055144	2.842160	0.0048

R-squared	0.936835	Mean dependent var	1.547958
Adjusted R-squared	0.934998	S.D. dependent var	0.286689
S.E. of regression	0.073093	Akaike info criterion	−2.362995
Sum squared resid	1.469205	Schwarz criterion	−2.247359
Log likelihood	344.5453	Hannan-Quinn criter.	−2.316634
F-statistic	509.8375	Durbin-Watson stat	2.013370
Prob(F-statistic)	0.000000		

For the correlogram of the residuals we have:

Sample: 1968M01 1991M12
Included observations: 284

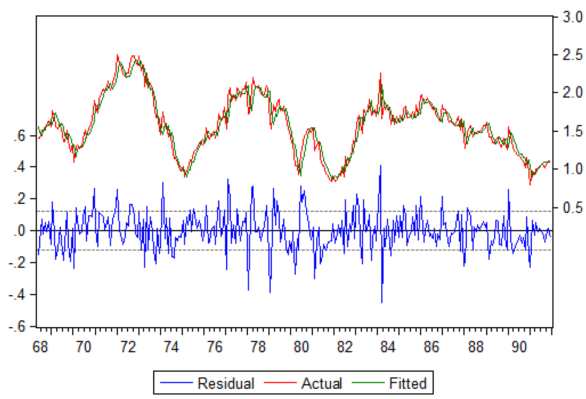
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.001	0.001	0.0004	0.985
		2	0.003	0.003	0.0029	0.999
		3	0.006	0.006	0.0119	1.000
		4	0.023	0.023	0.1650	0.997
		5	-0.013	-0.013	0.2108	0.999
		6	0.022	0.021	0.3463	0.999
		7	0.038	0.038	0.7646	0.998
		8	-0.048	-0.048	1.4362	0.994
		9	0.056	0.056	2.3528	0.985
		10	-0.114	-0.116	6.1868	0.799
		11	-0.038	-0.038	6.6096	0.830
		12	-0.030	-0.028	6.8763	0.866
		13	0.192	0.193	17.947	0.160
		14	0.014	0.021	18.010	0.206
		15	0.063	0.067	19.199	0.205
		16	-0.006	-0.015	19.208	0.258
		17	-0.039	-0.035	19.664	0.292
		18	-0.029	-0.043	19.927	0.337
		19	-0.010	-0.009	19.959	0.397
		20	0.010	-0.014	19.993	0.458
		21	-0.057	-0.047	21.003	0.459
		22	0.045	0.018	21.644	0.481
		23	-0.038	0.011	22.088	0.515
		24	-0.149	-0.141	29.064	0.218

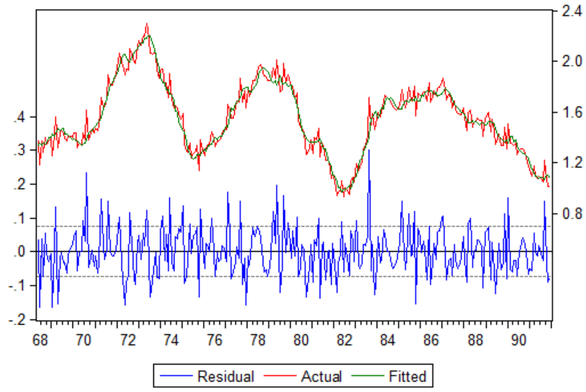
Sample: 1968M01 1991M12
Included observations: 284

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.009	-0.009	0.0238	0.877
		2	-0.035	-0.035	0.3744	0.829
		3	-0.037	-0.037	0.7640	0.858
		4	-0.088	-0.090	3.0059	0.557
		5	-0.105	-0.111	6.1873	0.288
		6	0.012	-0.000	6.2291	0.398
		7	-0.024	-0.041	6.4047	0.493
		8	0.041	0.024	6.9026	0.547
		9	0.048	0.029	7.5927	0.576
		10	0.045	0.037	8.1918	0.610
		11	-0.009	-0.005	8.2160	0.694
		12	-0.050	-0.046	8.9767	0.705
		13	-0.038	-0.024	9.4057	0.742
		14	-0.055	-0.049	10.318	0.739
		15	0.027	0.028	10.545	0.784
		16	-0.005	-0.020	10.553	0.836
		17	0.096	0.082	13.369	0.711
		18	0.011	-0.002	13.405	0.767
		19	0.041	0.040	13.929	0.788
		20	0.046	0.061	14.569	0.801
		21	-0.096	-0.079	17.402	0.686
		22	0.039	0.077	17.875	0.713
		23	-0.113	-0.114	21.824	0.531
		24	-0.136	-0.125	27.622	0.276

From the different reported Q-statistics, we see both series are White Noise. This is evidence to validate our VAR(4) model.

The actual, fitted, and residuals graphs are:





The graphs from the residuals are consistent White Noise processes.

21.2 Impulse Response Functions

To estimate both equations of the VAR(4) at the same time we need to type the following command:

```
var bookfigure.ls 1 4 starts comps
```

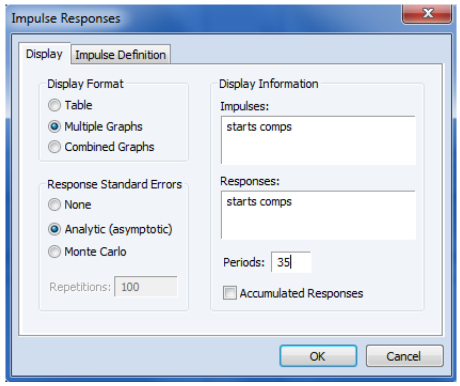
This estimates both equations and stores the VAR(4) in the workfile under the name "bookfigure." The output is the following:

Vector Autoregression Estimates
Sample (adjusted): 1968M05 1991M12
Included observations: 284 after adjustments
Standard errors in ( ) & t-statistics in [ ]

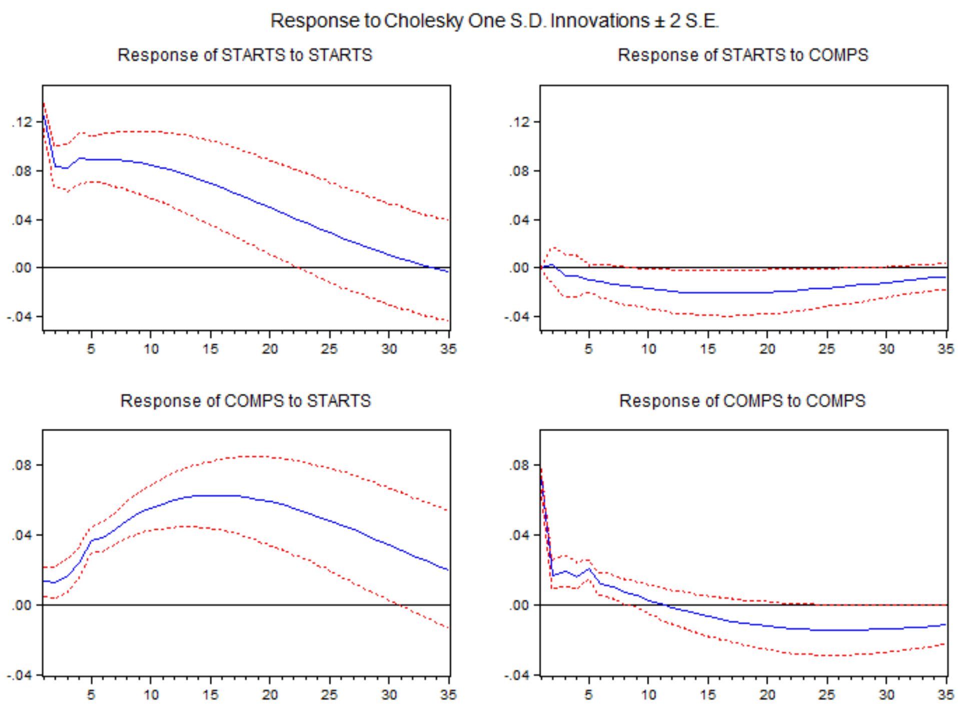
	STARTS	COMPS
STARTS(-1)	0.659939	0.074724
	(0.06124)	(0.03571)
STARTS(-2)	0.229632	0.040047
	(0.07272)	(0.04241)
STARTS(-3)	0.142859	0.047145
	(0.07265)	(0.04237)
STARTS(-4)	0.007806	0.082331
	(0.06603)	(0.03850)
COMPS(-1)	0.031611	0.236774
	(0.10271)	(0.05989)
COMPS(-2)	−0.120781	0.206172
	(0.10385)	(0.06055)
COMPS(-3)	−0.020601	0.120998
	(0.10095)	(0.05886)
COMPS(-4)	−0.027404	0.156729
	(0.09457)	(0.05514)
C	0.146871	0.045347
	(0.04423)	(0.02579)
R-squared	0.895566	0.936835
Adj. R-squared	0.892528	0.934998
Sum sq. resids	4.320952	1.469205
S.E. equation	0.125350	0.073093
F-statistic	294.7796	509.8375
Log likelihood	191.3622	344.5453
Akaike AIC	−1.284241	−2.362995
Schwarz SC	−1.168605	−2.247359
Mean dependent	1.574771	1.547958
S.D. dependent	0.382362	0.286689

Determinant resid covariance (dof adj.)	$8.11E-05$
Determinant resid covariance	$7.61E-05$
Log likelihood	540.7183
Akaike information criterion	−3.681115
Schwarz criterion	−3.449842

Notice that this is exactly the same result we obtained before. The benefit from this second approach is that the impulse-response functions can then be easily estimated by going to "View"' and then "Impulse Response":



To obtain:



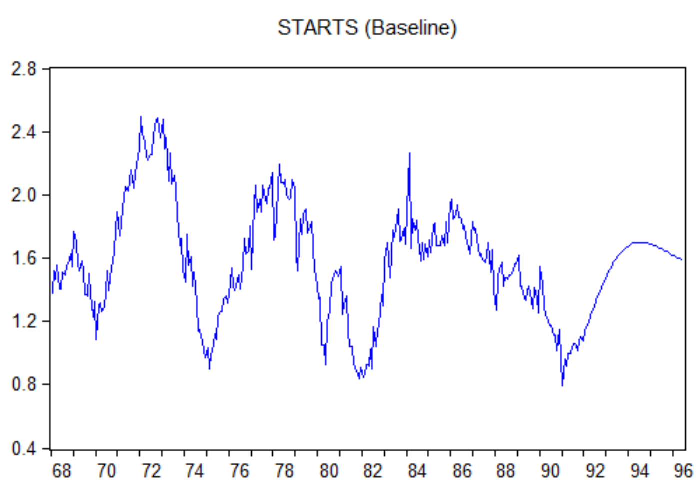
For example, from upper left quadrant we see that starts responds positively to starts. A one standard deviation shock in starts has a positive effect on starts that lasts about 23 months. The marginal effect is given by the blue line, while the red bands are approximately the 95% confidence intervals. Once the intervals include zero, the effect is no longer statistically significant.

On the lower left quadrant we see that completions responds positively to starts. A one standard deviation shock in starts has a positive (and increasing, at the beginning) effect on completions. The effects last for about 30 months.

21.3 Forecasting with Regression Models

For forecasting using the estimated VAR(4) we need to do the following:

```
bookfigure.makemodel(varmod) @prefix s_
smpl 1992m01 1996m06
varmod.solveopt(s=d, d=d)
solve varmod
smpl 1968m01 1996m06
varmod.makegraph(g=v) finalfigure starts
```



Alternatively, one can use the VAR(4) and obtain forecasts equation by equation using the same tools described in previous handouts. Just go to “Forecast”’ right after the estimation of each of the VAR equations.