



**Уральский
федеральный
университет**

имени первого Президента
России Б.Н.Ельцина

**Высшая школа
экономики
и менеджмента**

**О. С. МАРИЕВ
А. Л. АНЦЫГИНА**

ПРИКЛАДНАЯ ЭКОНОМЕТРИКА ДЛЯ МАКРОЭКОНОМИКИ

APPLIED ECONOMETRICS FOR MACROECONOMICS

Учебное пособие

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИМЕНИ ПЕРВОГО ПРЕЗИДЕНТА РОССИИ Б. Н. ЕЛЬЦИНА

О. С. Мариев, А. Л. Анцыгина

ПРИКЛАДНАЯ ЭКОНОМЕТРИКА ДЛЯ МАКРОЭКОНОМИКИ

APPLIED ECONOMETRICS FOR MACROECONOMICS

Рекомендовано методическим советом УрФУ
в качестве учебного пособия для студентов,
обучающихся по программе магистратуры
по направлению подготовки 080100 «Экономика»

Екатеринбург
Издательство Уральского университета
2014

ББК Ув631
М261

Рецензенты:

сектор инфраструктурного развития и экономико-математического
моделирования Института экономики УрО РАН (заведующий сектором
кандидат экономических наук, доцент С. Н. К о т л я р о в а);

Л. М. К а п у с т и н а, доктор экономических наук, профессор,
заведующий кафедрой маркетинга и международного менеджмента
(Уральский государственный экономический университет)

Мариев, О. С.

М261 Прикладная эконометрика для макроэкономики = Applied
econometrics for macroeconomics : [учеб. пособие] / О. С. Мариев,
А. Л. Анцыгина ; М-во образования и науки Рос. Федерации, Урал.
федер. ун-т. — Екатеринбург : Изд-во Урал. ун-та, 2013. — 152 с. —
Загл. парал. рус., англ. — Текст англ.

ISBN 978-5-7996-1303-7

Учебное пособие включает теоретико-методологический блок, вопросы для самоконтроля, глоссарий, список рекомендуемой литературы по всему курсу. Пособие сочетает в себе традиционные учебные формы и элементы методической новизны. Дается авторская интерпретация и сравнительный анализ статистических тестов, приведены конкретные команды для реализации ряда техник в эконометрическом пакете Econometric Views.

Для студентов, изучающих эконометрику, а также для специалистов по прикладной макроэкономике и финансам.

ББК Ув631

CONTENTS

Introduction	5
Chapter 1. INTRODUCTION TO TIME SERIES ANALYSIS	6
1.1. Main Characteristics of Time Series.....	6
1.2. Adjustment of Time Series	10
1.3. Stationary Time Series.....	19
Chapter 2. UNIT ROOT TESTS AND TESTS FOR STRUCURAL BREAKS	25
2.1. The Idea and the Dickey — Fuller Test.....	25
2.2. The Phillips — Perron Test.....	28
2.3. The Kwiatkowski — Phillips — Schmidt — Shin (KPSS) Test.....	30
2.4. Seasonal Unit Root and Unit Roots in Panel Data	32
2.5. Tests for Structural Breaks.....	36
Chapter 3. MODELS OF STATIONARY TIME SERIES	48
3.1. ARMA Model	48
3.2. ARCH and GARCH models	53
Chapter 4. INTERVENTION ANALYSIS	61
4.1. General Facts about Intervention Analysis.....	61
4.2. Vector Autoregression (VAR) model	65
4.3. Granger Causality	76
Chapter 5. COINTEGRATION AND ERROR CORRECTION MODELS	84
5.1. Cointegrated Time Series.....	84
5.2. The Error Correction Model	89
5.3. Methods for Identification of Cointegration	92
Chapter 6. FORECASTING	99
6.1. Main Principles of Forecasting.....	99
6.2. Forecasting With the Help of Different Models	104
6.3. Combined Forecasts	108

Chapter 7. APPLIED ECONOMETRICS FOR MACROECONOMICS	
IN STATISTICAL PACKAGES	113
7.1. Adjustment Data in Statistical Packages	113
7.2. Unit Root Tests in Applied Statistical Packages.....	117
7.3. Tests for Structural Breaks in Applied Statistical Programs.....	120
7.4. Models for Stationary Time Series	123
7.5. Intervention Analysis in Applied Statistical Packages	127
7.6. Cointegration	130
7.7. Forecasting.....	131
Conclusion	135
Appendix	137
References	142
Glossary	144

INTRODUCTION

This book is devoted to applied econometrics for macroeconomics. It presents all basic method for time series analysis. Such methods should be quite useful in modern economics conditions. For instance, we should make different forecasts in constantly changing macroeconomic conditions. For this purpose it is necessary to have special skills in working with time series. Moreover, different data types require particular tools for their interpretation and analysis. This book should provide current master students with all instruments for achieving the purposes indicated above.

As the course is devoted to applied aspects, we will not present extra difficult formulas and estimation methods in this book. It is more important to understand key features of time series and main approaches to analyzing and forecasting. This goal can be achieved without the deepest digressions into mathematics.

The book includes seven chapters. This first chapter is devoted to main characteristics of time series. Also it covers topics connected with seasonality, data smoothing and transformations, the concept of stationarity. The second chapter presents different modifications of unit root tests. The third chapter covers models of stationary time series. In particular, ARMA, ARCH/GARCH models are presented here. The fourth chapter includes intervention analysis. It covers topics in event representation, vector of autoregression models and Granger causality relationships. The fifth chapter is devoted to the concept of cointegration. There are different methodologies for identification of cointegration and error correction model. The sixth chapter includes topics in forecasting. For example, there is an estimation of forecast's quality etc. The seventh chapter is included to cover topics connected with time series analysis in applied statistical packages. It contains examples of estimating different time series models in Stata and EViews. The book is also provided with the list of references and appendixes which include tables with critical values and may be useful for practical works.

Chapter 1

INTRODUCTION TO TIME SERIES ANALYSIS

1.1. Main Characteristics of Time Series

It is known that Econometrics works with data sets, and such sets can describe economic objects in different states. Traditionally researchers indicate three data types:

- Cross-sectional data;
- Time series;
- Panel data*.

The first type represents a set of different objects at a particular time period. Time series, in opposite, describe one object during several moments of time. Panel data are a combination of two previous types: they characterize a set of objects during several periods. Naturally, applied econometrics for macroeconomics works with time series. So, this course is devoted to data analysis in dynamics and forecasting. Further we will consider only time series, their main features and special method for interpretation of such data.

Time series have several main characteristics:

- Frequency;
- Mean;
- Variance;
- Covariance.

Frequency is one of the most important features of any time series. It reflects how often the data were collected within one time period. For example, we can consider a year as a unit of time. If we collect the data every day, we will obtain series with daily frequency. If we do it only once a month, we will have monthly data etc. In our case daily data are

* Often pooled cross-sectional data are named as the forth type. We will not mark this one because it is a combination of all presented data types.

characterized with high frequency, and monthly series will have lower frequency.

There are several problems connected with data frequency. The first problem is seasonality. Often data with high frequency are characterized with seasonal waves. It means that the result of any observation depends on particular moment of time when such observation has been made. In most cases seasonality is objective. So, if we consider some groups of goods (shoes, clothes etc.), we will mark that there are peaks and falls. Such fluctuations can be explained by features of demand functions.

Seasonality takes place in data with high frequency. So, annual data should not be corrected for seasonality because it does not appear in such case. But seasonality can be observed within a year when we collect data with higher frequency (quarterly, monthly, daily etc.). We should expect seasonal trends here and correct data to eliminate.

Let's consider an example which reflects seasonal fluctuations. Fig. 1.1 represents the dynamics of consumer price index (CPI) for Russian Federation in 2011*.

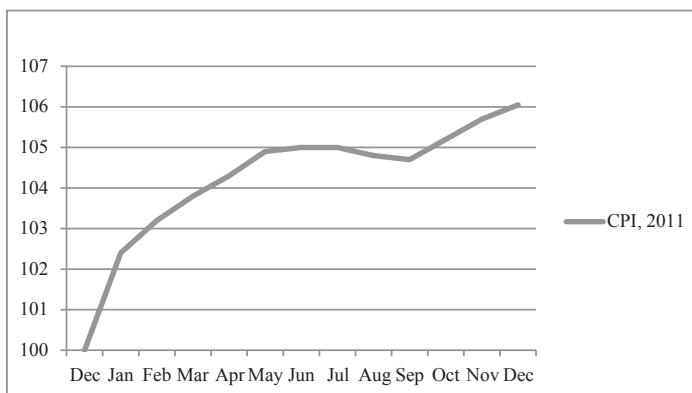


Fig. 1.1. Monthly CPI Dynamics for Russia in 2011

* All values are obtained on the end of the period as a fraction to December of a previous year.

There are particular seasonal tendencies for Russian CPI. We can see increasing dynamics, but some corrections take place. For example, growth in CPI stops and then it demonstrates decreasing tendency in the summer and in the beginning of the autumn (June, July, August, and September). This fact reflects seasonal fluctuations. Sometimes seasonality is brighter in the data, and it makes problems for our analysis. The reason is seasonality increases variance in the data. As a result, estimators become less efficient, and we can lose accuracy. It leads to mistakes in forecasting. So, we have to eliminate seasonality from time series to improve estimators. The ways how this process can be provided, will be discussed below.

The second problem which appears with frequency also is explained by higher variance in the data. Often it concerns data with high frequency (daily or more frequent observations). For such data we should use special models and tools to estimate parameters and create forecasts.

Another feature of time series is a mean value. There can be two possibilities:

- Constant mean;
- Non-Constant mean.

The first case supposes that the mean does not change during the time. Formally we can present:

$$\mu_t = \mu.$$

Such situation is desirable for the analysis, but it does not take place in most cases.

Often we should consider the second possibility when the mean depends on a particular moment of time:

$$\mu_t = f(t).$$

Such function characterizes both linear and non-linear dynamics. The main example when the mean is determined by the time is a trend. There can be different types of trends. Linear, exponential, logarithmic functions can reflect trends in time series. Such trends describe common dynamics and tendencies which appear in particular time series. Also trends can be used for forecasting purposes.

There are several ways how to eliminate trend form the data. These methods will be described and discussed below.

The third characteristic of time series is variance. In general, it measures a dispersion of observations around the mean. For statistical reasons variance should be constant:

$$\sigma_t = \sigma.$$

In this case the variance does not depend on a particular moment of time. Such condition can guarantee that our statistical estimators will be efficient (will have the least variance in a particular estimators' class).

But sometimes we can face with the opposite case when the variance is characterized by a special skedastic function:

$$\sigma_t = g(t).$$

Such situation is called heteroskedasticity. This function also can belong to different classes. In the indicated case we should use special correction instruments to make our estimation properly. Sometimes it is also necessary to use special models for the data with such variance. In any case if we do not correct heteroskedasticity, we will have false standard errors of estimators and make incorrect inferences about their significance.

The forth feature of time series is covariance between observations made at different moments of time. In general, it exists for any time series because the data are characterized with memory. In other words, a particular value today depends on previous actions and observations which took place yesterday, the day before yesterday etc. The existence of such memory makes some problems for the analysis. We face with autocorrelation. In general sense autocorrelation presents a dependence of current values from previous ones. There are negative sequences of autocorrelation. It leads to inconsistent and biased estimators which cannot be used for forecasting. It means that autocorrelation should be corrected in the model.

Applied Econometrics for Macroeconomics provides a range of autocorrelation models which can be used for the analysis and further forecasting. But not all data can be estimated initially. For statistical

purposes we should provide some transformations to clean time series from negative features. The first problem as it was indicated above is seasonality. Then we will consider different ways how to solve this one.

1.2. Adjustment of Time Series

As it was marked above, data with high frequency (quarterly, monthly, daily etc.) are characterized with seasonality. It increases the variance in time series and leads to loss in estimators' quality. Also we should analyze data with a constant mean. As a result, it is necessary to correct time series for trends and eliminate seasonality.

In general, any time series can be decomposed on determined and random parts. The first component includes:

- Trend;
- Seasonality;
- Cycle.

Trends and seasonality were briefly discussed above. Cycles are connected with fluctuations which is longer than seasonal corrections. A random part cannot be accounted. All indicated parts can interact in two ways:

- Additive;
- Multiplicative.

To show these types formally, we will introduce following labels. Let's y_t will be an initial range, f_t will represent a trend, s_t — a seasonal part, c_t — a cyclical component, e_t — a random part. Then additive interaction will look like:

$$y_t = f_t + s_t + c_t + e_t. \quad (1.1)$$

Multiplicative form can be presented as:

$$y_t = f_t \cdot s_t \cdot c_t \cdot e_t. \quad (1.2)$$

These two forms can be represented graphically. The difference concerns a seasonal component. So, if such fluctuations are constant (equal peaks and falls around the mean), we can use additive models.

In opposite case multiplicative model takes place. Graphically additive and multiplicative models are presented on the Fig. 1.2.

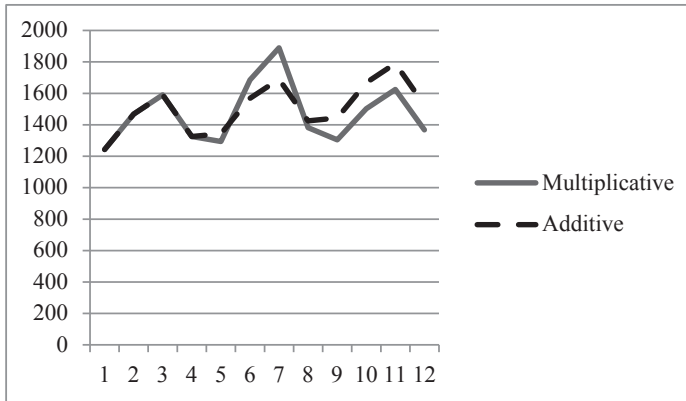


Fig. 1.2. Additive and Multiplicative Models of Time Series

Main problems for time series analysis can be formulated in the following way:

- To determine which part problem is present in particular data (trend, seasonality, cycles);

- To evaluate unknown parameters for deterministic components;

- To find the best model for random component approximation and estimate its parameters.

The solution of all indicated problems helps to create a proper forecast with minimum errors.

Now we will consider ways how to find and eliminate deterministic component from time series.

Initially it is always useful to present the data graphically and provide a visual analysis. Let's look at a previous example with Russian CPI (Fig. 1.1). As it was marked earlier, there is an increasing trend with seasonal fluctuations. We can say nothing about a cyclical component because this series is not quite large. So, from such observation we have established two determined parts in the dynamic data.

Now we should remove these components from our series. Let's start from the trend. What can we say about this one? We definitely know that there is increasing dynamics. Also it would be useful to have a function which could represent the trend. For example, it may be a linear function*:

$$y_t = \alpha + \beta t + \varepsilon_t, \quad (1.3)$$

where α is an intercept, βt presents a trend component, and ε_t includes an error term. In our example we have a linear trend.

Also a trend function can be non-linear. For instance, there can be a quadratic specification of the model:

$$y_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t. \quad (1.4)$$

This form makes sense if there is a maximum or minimum value in the data. For economic sets the order higher than two (quadratic function) is not used.

Non-linear trends can be presented as a half-logarithmic or logarithmic function:

$$y_t = \alpha + \beta \ln t + \varepsilon_t; \quad (1.5)$$

$$\ln y_t = \alpha + \beta \ln t + \varepsilon_t. \quad (1.6)$$

In any case a non-linear trend can be transformed into linear. For this purpose it is sufficient to take a logarithm of initial series, and a new range will be characterized with a linear trend.

One more type of trend is an autoregression trend. It shows the dependence of current values from previous ones. It is harder to find such kind of trends graphically. As a result, it makes sense to check its existence analytically.

There are three main ways how to exclude the trend:

- Differencing;
- Detrending;
- Centered moving average.

* The usage of the intercept is not necessary. Significance of such trend equation can be checked with standard tests.

To apply differencing, we should transform a non-linear trend into a linear one. The matter is that we create a new range from the first differences, and this new series should be without a trend. Every element of this new data set can be obtained according to the next formula:

$$\Delta y_t = y_t - y_{t-1}. \quad (1.7)$$

If these data keep a trend, it is possible to repeat the procedure and take the second difference:

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}. \quad (1.8)$$

In general, it is possible to take a finite number of differences and work with the last range as with initial data. But there are two problems. The first one is on every step we miss one observation. So, if we take a lot of differences, we will loss the same number of observations. There is no problem if a data set includes 10 000 observations, and we take 50 differences. But if we have only 100 observations, in the same situation we will loss a half of them.

The second problem is that we should interpret results of the analysis. Differences with high orders are quite difficult and sometimes impossible for interpretation. For economic data the order higher than 2 is not used because of problems with interpretation. Moreover, in most cases it is enough for removing a trend.

Detrending means evaluation of a trend equation and separation of a random component. If we suppose a general view of a trend function, it is possible to estimate its unknown parameters using standard tools of regression analysis. For example, if we use a linear model, we should compute estimators for α and β^* :

$$y_t = \hat{\alpha} + \hat{\beta}t + \varepsilon_t. \quad (1.9)$$

By obtaining these estimators, we find a deterministic part of the series. But in the same equation, indicated above, we can see a random component of the series:

$$y_t - \hat{\alpha} - \hat{\beta}t = \hat{\varepsilon}_t. \quad (1.10)$$

* Significance of estimators can be checked with standard statistical toolkit.

This component is deprived from a trend because a deterministic part was removed from the series. Now we can work with a new range like with the initial data*.

The third way, a calculation of centered moving average, also allows us to decrease an additional variance in time series. The idea is that we use previous, current and future values to estimate a mean for every moment of time. Depending on data frequency, it is necessary to choose a number of periods, τ . If we deal with quarterly data, τ will be equal to 4. For monthly frequency this number will be 12 etc. For calculation we should use the formula for a mean chronological value. Let's consider quarterly data. Then a centered moving average for every period will look like:

$$\tilde{y}_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}. \quad (1.11)$$

Weights before the first element and the fifth one are connected with an even number of period within a year. In opposite case these parts would be included with weights equal to 1.

Such average is calculated for every period of time except the first and the last two. It means that we also loose some data. But a resulting curve will be smoother than initial one because we decrease a variance in the data using centered moving average.

Now we should consider how it is possible to remove seasonality form time series. It allows us to exclude an additional variance from the data and work only with non-systematic components in the range.

There are three main methods for seasonality separation:

- Fourier spectral analysis;
- Dummy variables;
- Seasonal indexes.

The first algorithm is based on a standard spectral analysis. The idea is to present seasonality as a set of harmonic (sinusoidal) functions with a particular wave. Formally it looks like:

* It is possible to interpret such transformation and its dynamics as an initial data set.

$$y_t = \sum_{k=1}^{\infty} (M_k \sin \omega_k t + N_k \cos \omega_k t). \quad (1.12)$$

Where:

- $E(M_k) = E(N_k) = 0$;
- $Var(M_k) = Var(N_k) = \sigma$;
- $Corr(M_k, N_k) = 0$;
- and ω_k represents a length of the wave (frequency).

This method allows us to reflect seasonal dynamics properly, but it is quite hard for manual calculations. Fourier's decomposition is realized in many statistical packages and can be provided automatically.

The second method uses with dummy variables which correspond to every season. Let's consider quarterly data. Then we should present three dummy variables:

$$d_1 = \begin{cases} 1, & \text{if an observation took place at the first quarter} \\ 0, & \text{otherwise,} \end{cases}$$

$$d_2 = \begin{cases} 1, & \text{if an observation took place at the second quarter} \\ 0, & \text{otherwise,} \end{cases}$$

$$d_3 = \begin{cases} 1, & \text{if an observation took place at the third quarter} \\ 0, & \text{otherwise.} \end{cases}$$

We do not include a variable for the fourth quarter to prevent a perfect multicollinearity problem. Then this quarter will be a base level in the model.

Using presented labels, we can write a specification:

$$y_t = \alpha_0 + \alpha_1 d_1 + \alpha_2 d_2 + \alpha_3 d_3 + \varepsilon_t. \quad (1.13)$$

It is a multiple linear regression model which can be estimated with the ordinary least squares method (OLS). Significance of coefficients and the regression can be checked with standard statistical tests. If the

regression is not significant, we can suppose that seasonality does not take place in the model.

The third method is connected with constructing of seasonal indexes. We will consider a case with quarterly data again. To calculate seasonal indexes, it is necessary to smooth the range with a centered moving average method. As a result, we will have two data sets: y_t will present an initial range, and \hat{y}_t will describe smoothed one.

Now we will present an example with artificial data to present this algorithm visually. We have quarterly data for 3 years reflected on the graph (Fig. 1.3).

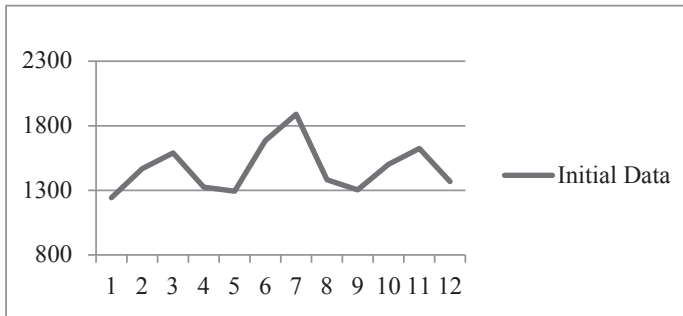


Fig. 1.3. Graphical Representation of initial Data

This graph demonstrates seasonality in the data. It means that we should remove it for making quality forecasts.

Before constructing indexes, we should determine in what kind different parts of time series interact with each other. In other words, what kind of models we should use: additive or multiplicative. To solve this problem it is enough to study a graphical presentation of the data and mark their characteristic features. As it was indicated above, if fluctuations are equal, it is reasonable to use an additive model. In opposite case a multiplicative model should be applied. In our case data do not demonstrate equal fluctuations. It means that different components of time series interact in multiplicative form, and we should use this model for further analysis.

Now we need to calculate a difference between real and smoothed values, z_t . Formally it can be presented in the next way:

$$z_t = y_t - \tilde{y}_t \text{ for additive model,}$$

$$z_t = \frac{y_t}{\tilde{y}_t} \text{ for multiplicative model.}$$

On this step we have got deviations of real values from centered average means. Now we can calculate a mean deviation for any particular quarter. For example, if we have 3 years, we observe each quarter three times during this period. But smoothed values for each quarter will be available only two times because we do not have enough observations to calculate these numbers. To present it visually, look at the Tab. 1.1 with the data and all previous calculations.

Table 1.1

Initial Data and All Calculations

Year	Quarter	Order	y_t	\tilde{y}_t	z_t
Year 1	Q_1	1	1243	—	—
	Q_2	2	1468	—	—
	Q_3	3	1590	1412,9	1,125
	Q_4	4	1325	1446,4	0,916
Year 2	Q_1	5	1294	1511	0,856
	Q_2	6	1685	1555,6	1,083
	Q_3	7	1890	1564	1,203
	Q_4	8	1382	1542,4	0,896
Year 3	Q_1	9	1304	1486,4	0,877
	Q_2	10	1502	1451,5	1,0348
	Q_3	11	1625	—	—
	Q_4	12	1368	—	—

Now we can calculate the first set of indexes which will present a mean deviation for each quarter. For this purpose we use the formula of an arithmetic mean:

$$i_1 = \frac{z(Q_1, \text{Year 2}) + z(Q_1, \text{Year 3})}{2},$$

$$i_2 = \frac{z(Q_2, \text{Year 2}) + z(Q_2, \text{Year 3})}{2},$$

$$i_3 = \frac{z(Q_3, \text{Year 1}) + z(Q_3, \text{Year 2})}{2},$$

$$i_4 = \frac{z(Q_4, \text{Year 1}) + z(Q_4, \text{Year 2})}{2}.$$

Now we should correct these indexes to reflect the lack of seasonality within a year. So, the sum of all indexes should be equal to 0 for additive model, and their product should be equal to 1 for multiplicative model. To solve this problem it is necessary to find a correcting number:

$$\pi = \frac{i_1 + i_2 + i_3 + i_4}{4} \text{ for additive model;}$$

$$\pi = \sqrt{i_1 \cdot i_2 \cdot i_3 \cdot i_4} \text{ for multiplicative model.}$$

And now it is possible to calculate final seasonal indexes with the help of the correcting number:

$$s_k = i_k - \pi, \quad k = 1, 2, 3, 4 \text{ for additive model;}$$

$$s_k = \frac{i_k}{\pi}, \quad k = 1, 2, 3, 4 \text{ for multiplicative model.}$$

The graph constructed from seasonal indexes is called a seasonal wave. And now it is possible to remove seasonality from the data set.

For this every observation should be divided on a corresponding index for multiplicative case, and the index is differenced from the initial value in additive model. For our example initial and final graphs are presented together on the Fig. 1.4.

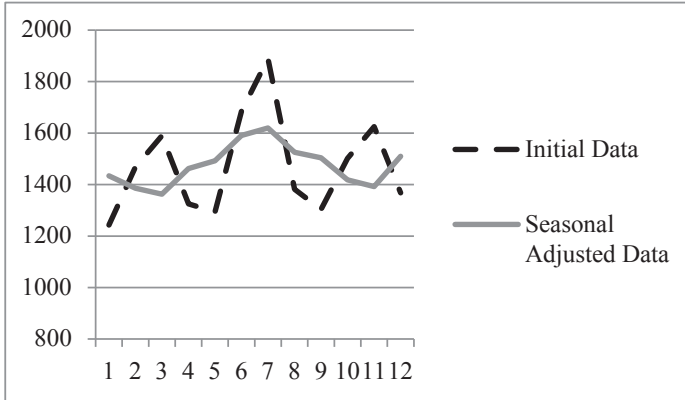


Fig. 1.4. Graphs for Initial and Seasonal Adjusted Data

A new curve is smoother than initial one. It means that seasonality was successfully removed from the data, and now this set can be used for further estimating and forecasting.

Thus, we have discussed different ways how to remove deterministic components and save irregular part from time series. And now it is necessary to speak about stationarity and its importance for time series analysis.

1.3. Stationary Time Series

Time series analysis is connected with the concept of stationarity closely. In general sense a range is stationary if three conditions are satisfied:

- $E(y_t) = \mu = \text{const} < \infty$;
- $\text{Var}(y_t) = \sigma = \text{const} < \infty$;
- $\text{Cov}(y_t, y_{t-s}) = \text{Cov}(y_{t-j}, y_{t-j-s}) = \gamma = \text{const} < \infty$.

It means that all parameters of this distribution do not depend on time.

In economic sense stationary processes are quite important. For example, such data always converge to the constant mean. This fact allows us to analyze a steady system, observe different shocks and the process of their neutralization. Also only stationary data can be estimated with standard econometric tools, and it will be possible to apply asymptotic test to obtained estimators. Moreover, it is impossible to make a good forecast using non-stationary data because false estimators and a lack of convergence.

Let's consider several examples of stationary processes. The first one is a white noise:

$$y_t = \varepsilon_t, \varepsilon_t \sim iid(0, \sigma). \quad (1.14)$$

This distribution of the error term is desirable because it signalizes about proper estimation and good quality of estimators. But often time series are not distributed this way, and we should use other models to present the data correctly.

Another stationary process is an autoregression process with order 1:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t. \quad (1.15)$$

But this process is stationary with one assumption:

$$|\alpha_1| < 1. \quad (1.16)$$

Only in this case the process will converge to the mean. If this restriction is satisfied as equality, we will have a unit root case, and the process will not be stationary. If the sign of the restriction will be changed, we will have the lack of convergence in the process. It will be non-stationary too.

It is possible to show that the given process with the presented assumption has constant mean, variance and covariance, and these values do not depend on time. In opposite case when the assumption is not satisfied parameters of this distribution will not be constant.

Now we can consider non-stationary processes too. The first case is a random walk:

$$y_t = \alpha_0 + y_{t-1} + \varepsilon_t. \quad (1.17)$$

This process was briefly discussed above when we were speaking about AR(1) process. This case is equivalent the next situation:

$$|\alpha_1| = 1. \quad (1.18)$$

In other words, the process has a unit root and is not stationary. The name “random walk” is connected with unpredictable dynamics which is generated with such process. All changes are included into the error term, ε_t . It is the reason why the work with such processes is complicated.

Another example of non-stationary process is a trend:

$$y_t = \beta t + \varepsilon_t, \varepsilon_t \sim iid(0, \sigma). \quad (1.19)$$

In this case a mean of such process will not be constant:

$$E y_t = E(\beta t + \varepsilon_t) = \beta t = f(t). \quad (1.20)$$

The mean depends on time. The same will not be true for variance and covariance of this process (it also can be shown analytically).

This process will also be non-stationary:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta t + \varepsilon_t, |\alpha_1| < 1. \quad (1.21)$$

Despite stationarity of this autoregression process, the trend component breaks the property. Also it is possible to combine different trends with the random walk and other processes to generate non-stationary one.

Initially we do not know a formal representation of the process which exists in the data. But we can make some assumptions about stationarity with the help of visual analysis. Graphically stationary process looks like Fig. 1.5*.

* Growth rate for American GDP per capita was used here to construct the graph (yearly data, 1947–2010). Source: www.bea.gov.

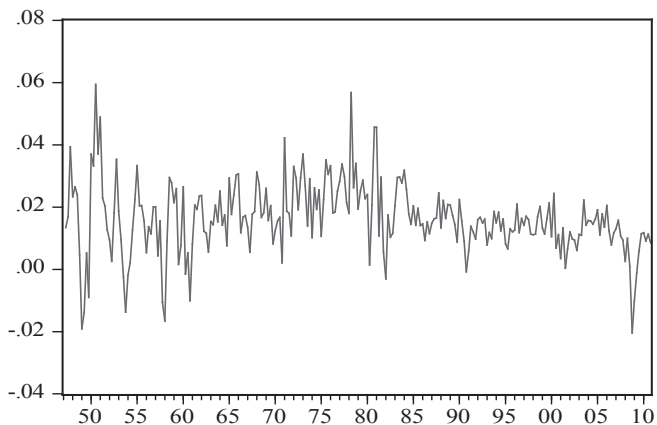


Fig. 1.5. Graphical Representation of a Stationary Process

An example of non-stationary process is presented on Fig. 1.6*.

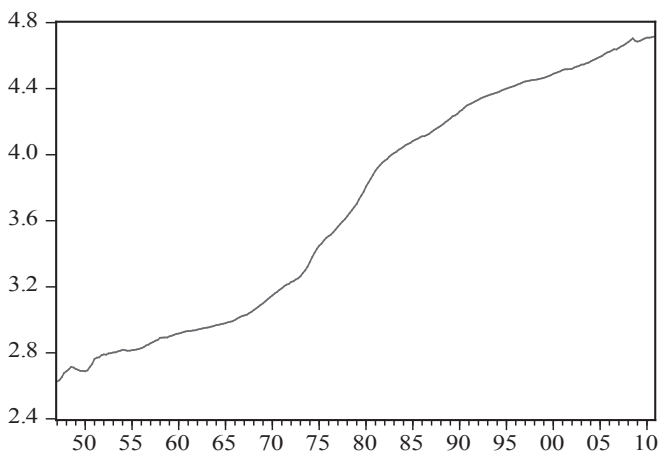


Fig. 1.6. Graphical Representation of a Random Walk Process

* American CPI was used here to construct the graph (yearly data, 1947–2010).
Source: www.bea.gov.

The non-stationary process can also look like data with any trend dynamics.

Any non-stationary data can be transformed into stationary process. Methods which allow us to do this transformation were presented above when we were speaking about trends. There are detrending and differencing.

Visual analysis can be informative for the first stage of any research, but it is necessary to give quantity and quality explanations to initial assumptions. For this purpose there are different unit root tests which help to determine stationarity or non-stationarity of the process. This methodology will be covered in the next chapter.

Questions

1. What main characteristics of time series do you know? Explain briefly what they mean.
2. List components which are included into time series. How can these parts interact with each other? Are there any ways to separate these components from the data?
3. What types of trend removing do you know? Briefly explain the idea behind each method.
4. Is it possible to correct seasonality in the data? List the ways to do it.
5. What does it mean stationary time series? Provide examples of stationary and non-stationary processes. Why cannot we apply standard statistical tests for non-stationary series?

Exercises

1. Find monthly data for any economic indicator and consider a 3-years period.
 - a) Present the data graphically. How can you describe this dynamics? Is there a trend? What can you say about seasonality?
 - b) If seasonality is presented in your data, what kind of models (additive or multiplicative) should you apply to correct it? Use seasonal indexes to do it.

- c) Present initial and adjusted data on one graph. Has a variance become lower after all transformations?
2. Show analytically that AR(1) process with $|\alpha_1| < 1$ has constant mean, variance and covariance.
 3. Show analytically that a white noise with a linear trend has a mean depending on time and constant variance.

Chapter 2

UNIT ROOT TESTS AND TESTS FOR STRUCURAL BREAKS

2.1. The Idea and the Dickey — Fuller Test

Stationarity is a desirable case for time series analysis as it was marked above. Non-stationary processes do not converge to the mean and cannot be estimated with standard statistical tools. It is possible to make some inferences about stationarity of the data looking at a graph. But we should use a criterion to formulate final conclusions, and for this purpose there are some parametric tests which allow us to test stationarity statistically.

Consider a standard AR(1) process without intercept:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t. \quad (2.1)$$

We know that this process will be stationary if the next condition satisfies:

$$|\alpha_1| < 1. \quad (2.2)$$

It means that for testing stationarity we should formulate the next hypothesis:

$$H_0 : |\alpha_1| = 1, \text{ the process is non-stationary;}$$

$$H_1 : |\alpha_1| < 1, \text{ the process is stationary.}$$

Formally this hypothesis can be tested. But if the initial data generate a non-stationary process, we must not use standard tests because the results will be false. In 1987 Phillips showed that non-stationary process did not converge to standard distribution constants. Moreover,

estimators present a stochastic process. It means that we should transform our equation. We subtract an element y_{t-1} from both sides of the equation:

$$\begin{aligned} y_t - y_{t-1} &= \alpha_1 y_{t-1} - y_{t-1} + \varepsilon_t, \\ \Delta y_t &= (1 - \alpha_1) y_{t-1} + \varepsilon_t, \\ \Delta y_t &= \gamma y_{t-1} + \varepsilon_t, \quad \gamma = 1 - \alpha_1. \end{aligned} \tag{2.3}$$

We have actually performed differencing and have obtained a stationary process. It is possible to rewrite the hypothesis:

$H_0 : \gamma = 0$, the process is non-stationary;

$H_1 : \gamma < 0$, the process is stationary.

Now we can apply a standard t-test to check stationarity. This test is known as the Dickey — Fuller test. It allows us to control stationarity. An observed value can be calculated as follow:

$$t_{DF} = \frac{\hat{\gamma}}{se(\hat{\gamma})}. \tag{2.4}$$

Accounting our alternative hypothesis*, a rejection rule will be written in the next way:

$$t_{DF}^{\text{observed}} < t_{DF}^{\text{critical}} \rightarrow H_0 \text{ should be rejected.}$$

For this comparison Dickey-Fuller critical values are used (see App., Tab. 1).

There are different specifications of the test. The first specification is called “none” (model A), and it corresponds with our initial equation:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t. \tag{2.5}$$

* In this case we have a left-side critical region because of the sign of the alternative hypothesis.

The second type is called “intercept” (model *B*):

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \varepsilon_t. \quad (2.6)$$

The third specification includes a trend component (model *C*):

$$\Delta y_t = \alpha_0 + \beta t + \gamma y_{t-1} + \varepsilon_t. \quad (2.7)$$

For all these specifications the testing hypothesis is kept, but critical values and observed statistics change. There are special tables with such critical values for different specifications and degrees of freedom.

The Dickey — Fuller test will work properly if the error term in the model is presented as a white noise. It means that we should correct autocorrelation and heteroskedasticity. To solve this problem the augmented Dickey — Fuller test is applied. All possible specifications can be presented as follow:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^k \varphi_i \Delta y_{t-i} + \varepsilon_t, \quad (2.8)$$

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \sum_{i=1}^k \varphi_i \Delta y_{t-i} + \varepsilon_t, \quad (2.9)$$

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \sum_{i=1}^k \varphi_i \Delta y_{t-i} + \varepsilon_t. \quad (2.10)$$

In these equations k is a number of lags which allow us to remove autocorrelation in the model. There is the next rule how to state a maximum k :

$$k_{\max} = \frac{T}{4}.$$

To choose a proper number of lags we should start from the maximum and then to decrease this number until autocorrelation disappear. The number of lags does not influence critical values, but it determines a observed statistics. It means that it is necessary to be attentive choosing the number k in the model. Choosing k , it is possible to consider the significance of the last lag on a 10 % level or account information criteria (they should be minimal in a proper specification).

If the data are stationary on all specification and critical values, there is no necessity to do any transformation for achieving stationarity. But often we have cases when data are not stationary on some specifications. The lack of the bases to reject the null hypothesis for any specification can help us to establish a source of non-stationarity. For example, if the null hypothesis was not rejected for the model C (“trend and intercept”), we can assume that non-stationarity is connected with the trend. To solve this problem it is recommended to do some transformations. To achieve stationarity we can apply differencing. For economic data stationarity is usually achieved after first differencing and rarer after second differencing. Higher orders are not applied because there are difficulties with their interpretation. Detrending also can be applied for achieving stationarity.

There are different cases when the range can include more than one unit root. For economic data it is possible to have two unit roots, cases with three and more roots are rare. In 1987 Dickey and Pantula suggested an algorithm which allowed them to test more than one unit root. They considered a top-down approach to compare different numbers of unit roots in the process. Authors showed that such correspondence worked only with such approach, and an opposite was false. Also there may be seasonal unit roots which also can be tested (the Hylleberg test will be considered below).

In general, the Dickey — Fuller test is quite widespread and standard instrument for testing stationarity. It is automatically realized in all applied statistical packages. But there are some problems with this test. This fact stimulated researchers to create new modifications for testing stationarity with autocorrelation and heteroskedasticity. Further we will consider other tests and problems associated with the Dickey — Fuller test.

2.2. The Phillips — Perron Test

This test has no significant distinctions from the Dickey-Fuller test in formulating initial hypotheses. This test also checks a single unit root, but it is constructed in such way to correct heteroskedasticity and

autocorrelation in the model. The test includes non-parametric correction to remove these negative things from the analysis and to obtain an adequate result.

Null and alternative hypotheses are formulated as follow:

$H_0 : \gamma = 0$, the process is non-stationary;

$H_1 : \gamma < 0$, the process is stationary.

But the test statistics differ from the Dickey — Fuller test:

$$t_{pp} = \frac{\hat{\sigma}_\varepsilon}{\hat{\sigma}} t_{DF} = \frac{1}{2} (\hat{\sigma}^2 - \hat{\sigma}_\varepsilon^2) \left(\hat{\sigma} \left(T^{-2} \sum_{t=2}^T (y_{t-1} - \hat{y}_{t-1})^2 \right)^{1/2} \right)^{-1}, \quad (2.11)$$

$$\hat{y}_{t-1} = (T-1)^{-1} \sum_{t=1}^T y_t,$$

$$\hat{\sigma}_\varepsilon^2 = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^2,$$

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^2 + 2T^{-1} \sum_{j=1}^K w_{ij} \sum_{t=i+j}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-j},$$

$$w_{ij} = 1 - \frac{j}{l+1}.$$

In this case $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}^2$ are consistent estimators for the variance in the error term of the model. A calculated value should be compared with Dickey — Fuller critical values. Then it is possible to speak about rejection of the null hypothesis.

The Phillips — Perron test is also performed in most statistical packages. As a rule, the results obtained with two considered tests correspond despite the differences in calculations and estimations. But there are problems generated by these tests. In 1982 Nelson and Plosser noticed that the Dickey — Fuller test often gave unit roots in the series where it should not. They studied macroeconomic data from 1860 to 1970 on the basis of autoregression analysis and got that only 1 range

from 14 was stationary. Authors showed that the Dickey — Fuller test gave non-stationarity if their studying coefficient was close to unity. Also the test does not account structural breaks. This observation allowed them to say that the test was not powerful. It means that the probability of type II errors is high for the Dickey — Fuller test, and often we accept false hypothesis as true one. This fact led to creation of new tests for stationarity which would be more accurate.

2.3. The Kwiatkowski — Phillips — Schmidt — Shin (KPSS) Test

As it was indicated above a standard Dickey — Fuller test has high probability of type II errors. It decreases the power of the test. To solve the problem and increase the power researches started to seek for new ways to improve the quality of estimation. As a result, the Kwiatkowski — Phillips — Schmidt — Shin (KPSS) test was created.

In standard tests the null hypothesis was about non-stationarity of the dynamic process. The KPSS test works with an opposite order. It states the next hypotheses:

$$H_0 : \text{stationary process,}$$

$$H_1 : \text{non-stationary process.}$$

Now we will describe the idea of the test. It is possible to present any time series as a mixture of three elements:

- Trend;
- Random walk;
- Stationary error term.

Formally this combination will look as follow:

$$y_t = \beta t + r_t + \varepsilon_t. \quad (2.12)$$

The first element in the right part presents a linear trend. The second component can be written as a random walk:

$$r_t = r_{t-1} + u_t, \quad (2.13)$$

$$u_t \sim iid(0, \sigma_u^2).$$

The last element in the initial decomposition is a random error term. In principle, it is possible to include any other deterministic components in the right part of the equation (constants etc.). Using facts presented above we can rewrite hypotheses in formal representation:

$$H_0 : \sigma_u^2 = 0,$$

$$H_1 : \sigma_u^2 \neq 0.$$

In other words, if the null hypothesis is true, r_t will not change in time. It means that such random walk will be transformed into constant because shocks do not take place in this particular process.

There are two possible specifications for further estimation. The first one is called level stationarity:

$$y_t = a_0 + \varepsilon_t. \quad (2.14)$$

The second type is named trend stationarity and can be written as follow:

$$y_t = a_0 + \beta t + \varepsilon_t. \quad (2.15)$$

This test assumes a standard OLS estimation and uses *LM*-test. Such statistics is calculated according to the next formula:

$$LM = T^{-2} \frac{\sum_{t=1}^T s_t^2}{s^2}, \quad (2.16)$$

$$s_t = \sum_{i=1}^t \varepsilon_i.$$

In these formulas s represents a consistent estimator of a long-run variance in ε_t calculated with the help of estimators from initial regression in levels or with a trend.

LM-test is built on the basis of χ^2 distribution. This criterion is characterized with right-side critical region. Then a rejection rule sounds

like that: if an observed value is more than a corresponding critical one (see App., Tab. 2), the null hypothesis should be rejected in favor of the alternative statement. It means that analyzing data generate a non-stationary process if we reject the null hypothesis.

In general, the KPSS test has helped to improve a quality of stationarity identification. As it was marked above, we can face other kinds of unit roots. For example, it is possible to catch a seasonal unit root or a unit root in panel data which also include dynamic component. Further we will consider tests which can determine these unit roots.

2.4. Seasonal Unit Root and Unit Roots in Panel Data

If we speak about seasonal unit roots, it is possible to remove seasonality from the data and provide standard unit root tests. However, sometimes it is necessary to keep information about seasonality because it is connected with a purpose of particular research. For this case there is a special test which allows us to catch seasonal unit roots without any previous transformations of the data.

This methodology was suggested by Hylleberg, Engle, Granger, Yoo (HEGY) in 1990. They considered a polynomial, $\Phi(L)$:

$$\Phi(L) = (1 - L^4).$$

Let's consider a range y_t . To test seasonal unit roots, they suggested estimating the next equation with the OLS:

$$\Phi(L)y_t = \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{3,t-1} + \varepsilon_t. \quad (2.17)$$

The labels presented on the right side of the equation perform the next values:

$$y_{1,t} = (1 + L + L^2 + L^3)y_t,$$

$$y_{2,t} = (1 - L + L^2 - L^3)y_t,$$

$$y_{3,t} = (1 - L^2)y_t.$$

If to consider the element $\Phi(L)$, we will get the next characteristic equation:

$$\Phi(L) = (1 - L^4) = (1 - L^2)(1 + L^2) = (1 - L)(1 + L)(1 - iL)(1 + iL). \quad (2.18)$$

Then we have the next possible values for π which correspond with roots of this characteristic equation:

$$\pi_1 = 1, \pi_2 = -1, \pi_3 = i, \pi_4 = -i. \quad (2.19)$$

If to match this result with possibility of seasonality, there will be the next hypotheses:

$$\pi_1 = 1, \text{ there is no seasonality;}$$

$$\pi_2 = -1, \text{ there are two cycles within a year;}$$

$$\pi_3 = i, \pi_4 = -i, \text{ there is one cycle within a year.}$$

Stationarity will take place only if all coefficients are not equal to zero simultaneously. If π_3 and π_4 are equal to zero, we have two unit roots. If π_1 and π_2 are zero, we obtain one unit root.

The analysis is built on t -test and F -test. It is also admissible to include a constant term and trends into analyzing regression. Thus, this test helps to consider data with seasonality and not to remove a significant part of information about time series.

Panel data also includes a dynamic component. As a result, there is a possibility of unit roots in such data. There is a test which was suggested by Im, Pesaran and Shin. They considered a panel as a set of n ranges. Every range includes t observations. Then the next equation can be estimated with the OLS:

$$\Delta y_{it} = a_{i0} + \gamma_i y_{i,t-1} + a_{i1}t + \sum_{j=1}^p \beta_{ij} \Delta y_{i,t-j} + \varepsilon_{i,t}. \quad (2.20)$$

This form corresponds with a standard Dickey — Fuller regression. Other specifications can also be used here. For example, we can exclude a trend from the analysis. Such regressions should be provided for all n ranges. Moreover, all regressions must be estimated in the same specifications.

For every regression an observed statistics can be calculated according to the next formula:

$$t_{i,DF} = \frac{\hat{\gamma}_i}{se(\hat{\gamma}_i)}. \quad (2.21)$$

In the end of calculation n such values will be available. To find a critical value, we should determine an arithmetical mean:

$$\bar{t} = \frac{\sum_{i=1}^n t_{i,DF}}{n}. \quad (2.22)$$

Null and alternative hypotheses will be formulated as follow:

$$H_0 : \gamma_i = 0 \forall i = 1, \dots, n \text{ (unit root),}$$

$$H_1 : \exists i \gamma_i < 0 \text{ (stationarity).}$$

A test statistics will be biased because we use a mean value from the sample. Accounting this fact, an observed statistics will be calculated as:

$$\bar{z}_i = \sqrt{n} \cdot \frac{\bar{t} - E(\bar{t})}{\sqrt{Var(\bar{t})}} \sim N(0,1). \quad (2.23)$$

Critical values will be determined with a number of groups (n), the inclusion of the trend and significance level. A rejection rule is formulated as for the Dickey — Fuller test.

But this test has several lacks. The first one is that we consider a lot of γ . It is possible that, for example, only two coefficients from a hundred will statistically differ from zero. In such situation the test will give us stationarity, but in fact the process will be opposite. It means that we will make a false conclusion about the existence of unit roots in panel data.

The other problem is that there may be correlation between error terms in different regressions:

$$Corr(\varepsilon_{it}, \varepsilon_{jt}) \neq 0.$$

In this case we have cross-sectional correlation and cannot make conclusions with the unit root test in presented form.

There are different tests to find cross-sectional correlation. We will consider the Breusch-Pagan test. Null and alternative hypotheses are presented as:

$$H_0 : Corr(\varepsilon_{it}, \varepsilon_{jt}) = 0 \forall t, \text{ there is no cross-sectional correlation};$$

$$H_1 : \exists t Corr(\varepsilon_{it}, \varepsilon_{jt}) \neq 0, \text{ there is cross-sectional correlation.}$$

The test statistics looks like:

$$\chi^2 = T \sum_{i=1}^{n-1} \sum_{j=i+1}^n \hat{\rho}_{ij}^2 \sim \chi^2 \left(\frac{n(n-1)}{2} \right), \quad (2.24)$$

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^T \hat{\varepsilon}_{it} \hat{\varepsilon}_{jt}}{\sqrt{\left(\sum_{t=1}^T \hat{\varepsilon}_{it}^2 \right) \left(\sum_{t=1}^T \hat{\varepsilon}_{jt}^2 \right)}}.$$

If an observed value more than critical one, we should reject the null hypothesis.

This modification of the test works properly for small n and large t . If these numbers are quite big, the statistics will be modified:

$$z = \sqrt{\frac{1}{n(n-1)}} \cdot (T \hat{\rho}_{ij}^2 - 1) \sim N(0, 1). \quad (2.25)$$

There is another distribution, and it is necessary to use other critical values. Thus, by testing cross-sectional correlation, we can understand perspectives of using unit root tests for panel data.

So, we have studied different modification of unit root tests. But sometimes the results obtained with the help of such tests can contradict

with initial assumptions. The reason is that not all tests are sensible to structural breaks in the data. To find such breaks special procedures are used, and below we consider them.

2.5. Tests for Structural Breaks

Often the data in dynamics are characterized by structural breaks. There is a special kind of models which allows researchers to analyze consequences of such breaks and shocks. In this paragraph we will deal only with tests for structural breaks, and models for intervention analysis will be covered above.

A structural break presents a change in the dynamics in response to exogenous shocks. It is important to find such breaks because they strongly influence an estimation quality. It is possible that the data have to be described with different models before and after the break. The existence of such event also should be accounted in forecasting.

To find potential structural breaks, it is useful to look at the graphical representation of the data. There are three possible types of breaks:

- Breaks in trends;
- Breaks in intercepts;
- Breaks in trends and intercepts.

If we are speaking about linear trends, breaks are connected with changes in slopes, but not in intercepts. Visually it looks like Fig. 2.1.

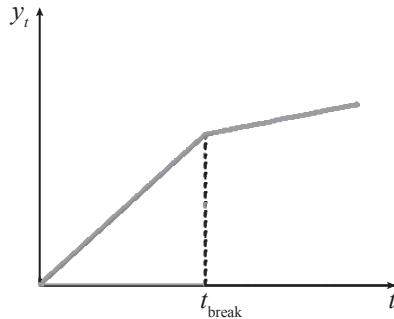


Fig. 2.1. Break in Trend

Breaks in intercepts lead to a parallel shift of initial graphs after the event (Fig. 2.2).

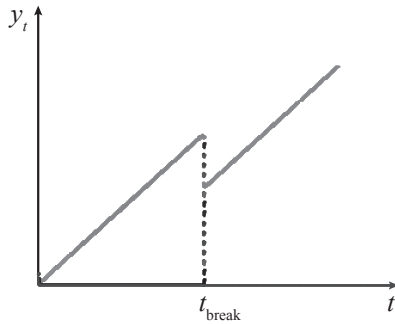


Fig. 2.2. Break in Intercept

Breaks in both deterministic components can be drawn as presented on the Fig. 2.3.

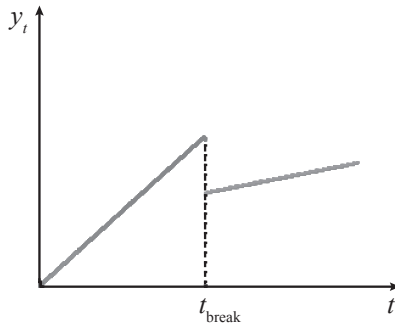


Fig. 2.3. Break in Trend and Intercept

All these breaks can be noticed on the graphs, but it is also necessary to provide some tests. Such tests help us to find unobserved visually structural breaks and check our assumptions about potential break points.

To illustrate these tests and principles of their work, we will use an example. There is a range, and it represents consumption per capita

for the U. S. economy. These data are quarterly and were collected from 1947 to 2010. Graphically presented data could be seen on the Fig. 2.4.

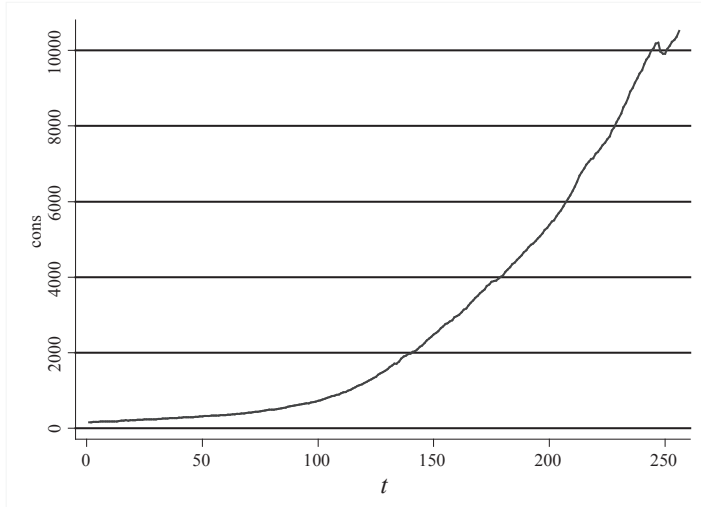


Fig. 2.4. Graphical Representation of Quarterly Consumption Per Capita for American Economy (1947–2010)

Let's make some assumptions about potential break points. There is an exponential trend in the data. In standard analysis it would be necessary to transform this one into a linear trend. But to take a logarithm means to do a monotonous transformation which does not change basic properties of the function. In other words, points with breaks will be kept at the same places, and it is not necessary to linearize the trend here. Thus, we will work with initial data without any additional transformations.

Now we will consider potential break points. The first break takes place at the moment #103 (the fourth quarter of 1971). This dynamics can be characterized as growth. Using some facts from the U. S. economic history, we can suppose that it was connected with overcoming of the recession in the end of 1960th years. This drop was stimulated

by differences in dynamics of real and nominal wages, deterioration of trade balance. As a result, consumption per capita did not demonstrate a tendency to grow. The situation had improved by 1971, and it led to growth in consumption per capita and a break in the trend.

The second break in trend and intercept takes place at the moment #243 (the fourth quarter of 2007). The potential source of such dynamics was the beginning of the crisis in the mortgage market. Also decreasing in consumption per capita is connected with dynamics of other macroeconomic indicators closely. This period was characterized with high prices for energy sources (oil, natural gas etc.), and this fact led to CPI growth and falling in consumption per capita. Thus, using visual representation and some facts about the U. S. economy, we have made assumptions about potential break points. But we should test statistically if our guesses are true. For this purpose there are several tests for structural breaks.

The first one is the Perron test. This test helps to find different types of structural breaks, and it is also connected with the unit root identification closely. To provide this test, we should have some assumptions about potential break points because we can check only one possible moment of time on every step.

To provide the test, we should create three new dummy variables:

$$D(t_{\text{break}}) = \begin{cases} 1, & \text{if } t = t_{\text{break}} + 1, \\ 0, & \text{otherwise;} \end{cases}$$

$$D(u) = \begin{cases} 1, & \text{if } t > t_{\text{break}}, \\ 0, & \text{otherwise;} \end{cases}$$

$$DT_t = \begin{cases} t - t_{\text{break}}, & \text{if } t > t_{\text{break}}, \\ 0, & \text{otherwise.} \end{cases}$$

These variables help to label moments of time after the break. The idea is to test if the variables are significant in the model or not.

There are three possible specifications, and they correspond with the number of break types. The first one presents the break in intercept (model *A*):

$$y_t = \mu + \beta t + dD(t_{\text{break}}) + \theta D(u) + \alpha y_{t-1} + \sum_{i=1}^K \rho_i \Delta y_{t-i} + \varepsilon_t. \quad (2.26)$$

The second model characterizes the break in trend (model *B*):

$$y_t = \mu + \beta t + \gamma DT_t + \alpha y_{t-1} + \sum_{i=1}^K \rho_i \Delta y_{t-i} + \varepsilon_t. \quad (2.27)$$

The third model incorporates these two cases (model *C*):

$$y_t = \mu + \beta t + dD(t_{\text{break}}) + \theta D(u) + \gamma DT_t + \alpha y_{t-1} + \sum_{i=1}^K \rho_i \Delta y_{t-i} + \varepsilon_t. \quad (2.28)$$

In all these specifications k is a number of lags including to correct autocorrelation. The last lag should be significant on 10 %. Also it is possible to choose k controlling information criteria (they are minimal in the best specification).

Initially we should choose one specification which we will test. In general, null and alternative hypotheses are formulated in the next way:

$$H_0 : \text{there is a unit root};$$

$$H_1 : \text{there is no unit root, and the specification is true.}$$

Let's write all possible null and alternative hypotheses for models. Model *A* has the next presentation for testing:

$$H_0 : \alpha = 1, \beta = 0, \theta = 0, d \neq 0,$$

$$H_1 : \alpha < 1, \beta \neq 0, \theta \neq 0, d = 0.$$

For model *B* hypotheses will be the next:

$$H_0 : \alpha = 1, \beta = 0, \gamma = 0,$$

$$H_1 : \alpha < 1, \beta \neq 0, \gamma \neq 0.$$

For model C we have these ones:

$$H_0 : \alpha = 1, \beta = 0, \theta \neq 0, d \neq 0, \gamma = 0,$$

$$H_1 : \alpha < 1, \beta \neq 0, \theta = 0, d = 0, \gamma \neq 0.$$

To test the range on stationarity, it is necessary to use special Perron's critical values. To find this number, we should calculate λ :

$$\lambda = \frac{t_{\text{break}}}{T}.$$

This value characterizes in what part of the whole observation period the break took place. There is a table with such critical values where λ and different probabilities are indicated. Such table is provided in the Appendix.

Now it is possible to consider our example. We supposed that the break in trend was at the moment #103. The model specification will correspond with the model B :

$$y_t = \mu + \beta t + \gamma DT_t + \alpha y_{t-1} + \sum_{i=1}^K \rho_i \Delta y_{t-i} + \varepsilon_t.$$

We should generate only DT_t variable here:

$$DT_t = \begin{cases} t - 103, & t > 103, \\ 0, & \text{otherwise.} \end{cases}$$

Null and alternative hypotheses have been considered above.

The hypothesis about stationarity can be checked with a standard t -test:

$$t = \frac{\hat{a} - 1}{se(\hat{a})}. \quad (2.29)$$

We should use Perron's critical values here (App., Tab. 3). Then we have a standard process for testing and making conclusions. For the second potential break point we should use model C , but the mechanism for testing will be the same.

Another test was suggested by Zivot and Andrews. The idea of the test is the same: we imply a null hypothesis about a unit root. There are also three possible specifications, but they differ a little from the Perron test:

$$y_t = \mu + \beta t + \theta D(u) + \alpha y_{t-1} + \sum_{i=1}^k \rho_i \Delta y_{t-i} + \varepsilon_t \text{ (model A, break in trend)}$$

$$y_t = \mu + \beta t + \gamma DT_t + \alpha y_{t-1} + \sum_{i=1}^k \rho_i \Delta y_{t-i} + \varepsilon_t \text{ (model B, break in intercept)}$$

$$y_t = \mu + \beta t + \theta D(u) + \gamma DT_t + \alpha y_{t-1} + \sum_{i=1}^k \rho_i \Delta y_{t-i} + \varepsilon_t \text{ (model C, break in trend and intercept)}.$$

This test assumes testing of a single structural break. In fact, null and alternative hypotheses are:

$$H_0 : \alpha = 1,$$

$$H_1 : \alpha < 1.$$

The statistics will look like:

$$t = \frac{\hat{a} - 1}{se(\hat{a})}.$$

The idea is that we test every moment of time (hypothetically from 1 to T) according to the chosen specification. As a result, we have got T statistical values for \hat{a} . Then we should choose the point with minimal t -statistics and continue the estimation here. There are also special critical values for the test. To find a proper value, it is necessary to know a type of the model (A , B , and C) and the level of significance.

According to these critical values, we can make conclusions about non-stationarity and structural breaks (see App., Tab. 4).

Another test is the Vogelsang test. It also indicates a single structural break and assumes consideration of all moment of time. The advantage of the test is that it estimates all possible types of breaks in one specification. The model for testing has the next representation:

$$\begin{aligned} \Delta y_t = & \beta_0 + \beta_1 t + \beta_2 t^2 + \delta_0 DU_t + \sum_{i=1}^p \delta_i DT_t^i + \pi y_{t-1} + \\ & + \sum_{i=1}^k \rho_i \Delta y_{t-i} + \varepsilon_t, \quad p=1,2. \end{aligned} \quad (2.30)$$

The next dummy variables should be generated to estimate the equation:

$$DU_t = \begin{cases} 1, & \text{if } t > t_{\text{break}}, \\ 0, & \text{otherwise;} \end{cases}$$

$$DT_t = \begin{cases} t - t_{\text{break}}, & \text{if } t > t_{\text{break}}, \\ 0, & \text{otherwise.} \end{cases}$$

In principle, it is possible to exclude a quadratic trend and provide the analysis without this component. Null and alternative hypotheses can be presented as follow:

$$H_0 : \delta_i = 0 \forall i,$$

$$H_1 : \exists i \delta_i \neq 0.$$

In other word, the null hypothesis assumes that there is a stable trend without breaks. The alternative one says that there is breaks in the trend. This test is built on F -statistics:

$$F = \frac{\frac{(SSR_r - SSR_{ur})}{j}}{\frac{SSR_{ur}}{(T - 2(p+1) + k + 1)}}. \quad (2.31)$$

In this formula j is a number of restrictions $(p + 1)$, and $2(p + 1) + k + 1$ reflects a quantity of parameters in the model.

The model should be estimated for all moment of time. Then the moment with a maximum F -statistics should be chosen and checked for breaks. Let's consider our example with consumption per capita.

There are special critical values for this test (see App., Tab. 5). They depend on a type of a break and other parameters which can be calculated for asymptotic distribution. We should find the next values:

$$mean(F) = \frac{(p + 1)F}{T},$$

$$exp(F) = \ln \left(\frac{e^{\frac{1}{2}(p+1)F}}{T} \right),$$

$$sup(F) = sup(p + 1)F.$$

If we know these values, it is possible to find critical values for every specification.

We have two assumptions about break points:

- At the moment #103 (break in trend);
- At the moment #243 (breaks in trend and intercept).

It is quite time-expensive to check all points, and we will study only the nearest ones. For the first moment we will take points #101, #102, #103, #104, #105, #106 and choose the value where F -statistics will be maximal. Our dynamics is characterized with non-linear trend, and we will keep a quadratic component in the regression equation:

$$\Delta y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \delta_0 DU_t + \sum_{i=1}^p \delta_i DT_t^i + \pi y_{t-1} + \sum_{i=1}^k \rho_i \Delta y_{t-i} + \varepsilon_t.$$

We included 6 lags of the dependent variable to remove autocorrelation ($k = 6$) and obtained the next results for F -statistics (Tab. 2.1):

Table 2.1

***F*-Statistics for the Regression in the Vogelsang Test
in the Neighborhood of 103**

The period	<i>F</i> -Statistics
$T=101$	38,7
$T=102$	38,7
$T=103$	38,7
$T=104$	38,71
$T=105$	38,71
$T=106$	38,71

Then we will investigate periods which are placed close to the moment #243: #241, #242, #243, #244, #245. We have the following results for the same regression presented in the Tab. 2.2.

Table 2.2

***F*-Statistics for the Regression in the Vogelsang Test
in the Neighborhood of 243**

The period	<i>F</i> -Statistics
$T=241$	55,92
$T=242$	57,48
$T=243$	59,74
$T=244$	59,64
$T=245$	58,71

According to this test, we have only one point where the break can be, and it is $T = 243$.

Now we should calculate a critical value. We considered $p = 2$ and $T = 256$, and we have the next values for all necessary statistics:

$$\text{mean}(F) = \frac{(2+1) \cdot 59,74}{256} = 0,7,$$

$$\exp(F) = \ln \left(\frac{e^{\frac{1}{2}(2+1) \times 59,74}}{256} \right) = 84,06,$$

$$\sup(F) = \sup(p+1)F = 3 \cdot 59,74 = 179,22.$$

We have a break in the last share of the sample. In the table with critical values we will look at $\lambda = 0,15$. We assumed the break and trend and intercept, and then we should take a critical value for model C . In principle, it is enough to find a critical value for $\sup(F)$. Using this number we can say if the null hypothesis has to be rejected or not. Our initial data are not stationary because there is an exponential trend. So, we must find $\sup(F)$ for non-stationary time series, model C , and $\lambda = 0,15$.

If we will consider $\alpha = 0,05$, this critical value will be equal to 17,88. There is a right-side critical region; an observed value is more than critical one. It means that we should reject the null hypothesis, and we actually have a break at the moment #243.

Thus, now we are familiar with unit root tests and tests for structural breaks. Also the conception of stationarity was presented above. It is possible to consider models of stationary time series, and it will be covered in the next chapter.

Questions

1. What is the main idea of unit root tests? Why cannot we apply standard tests to check the null hypothesis?
2. List unit root tests which you know. What are differences between them?
3. Does the Dickey — Fuller test have any disadvantages? How is it possible to solve the problems in such estimation if they take place?
4. Describe the idea of the unit root test for panel data. What restrictions exist for this test?

5. What instrument applies to test structural breaks? List tests for structural breaks, which you know, and the main ideas behind them.

Exercises

1. Take three macroeconomic time series with quarterly frequency (the minimal number of observations is 100).
 - a) Present the data graphically. What can you say about the existence of seasonality, trends, and stationarity?
 - b) Test every range for stationarity using a standard Dickey — Fuller test. What number of lags should you include to remove autocorrelation? If initial data are not stationary, apply some transformations and repeat the test until stationarity will be achieved.
 - c) Now test your data with the Phillips — Perron test. What is the main advantage of this test comparing with the Dickey — Fuller test? Do your results differ from the Dickey — Fuller test? If they do, what can be the reason for such difference?
 - d) Then use the KPSS test to check stationarity of these series. Compare your results with previous steps.
 - e) What test would you prefer and why?
2. For this task you can use the data from previous problem:
 - a) What can you say about the existence of structural breaks in the data? What kind of breaks do you suppose in any potential break point and why? Can you provide any economic explanation for the existence of the break in these points?
 - b) Provide the Perron test to check all potential break points. Were your initial guesses right? If they were not, can you give any econometric explanation to this fact?
 - c) Provide the Zivot and Andrews test. How many breaks can you check with this test? Do your results correspond with initial guesses? If they do not, try to explain such result using econometric toolkit.
 - d) Use the Vogelsang test. Compare the result with previous steps.

Chapter 3

MODELS OF STATIONARY TIME SERIES

3.1. ARMA Model

Earlier we have performed the concept of stationarity and different tests which allow us to test the process for unit roots. Now we will study different types of models using for stationary time series. The first class of such models is so-called ARMA models.

Let's consider different components of this process. The first part is autoregression process (AR). Formally it can be presented as follow:

$$y_t = \alpha_0 + \sum_{i=1}^p a_i y_{t-i} + \varepsilon_t. \quad (3.1)$$

The constant term is an exogenous parameter in the model. The number p should be chosen to remove autocorrelation form the model, and the model can be named as $AR(p)$. The error term in this model has to be a white noise (the mean is equal to zero, the variance is constant, autocorrelation does not take place).

The second part of ARMA model is a moving average process (MA):

$$y_t = \sum_{j=0}^q \beta_j \varepsilon_{t-j}. \quad (3.2)$$

The process does not include a constant. The number q also can be various and have to correspond the case when autocorrelation is absent in the model. Also the first coefficient, β_0 , usually is normalizes to unity.

ARMA model is a combination of presented processes with orders p and q accordingly:

$$y_t = \alpha_0 + \sum_{i=1}^p a_i y_{t-i} + \sum_{j=0}^q \beta_j \varepsilon_{t-j}. \quad (3.3)$$

It is ARMA(p, q) model. Numbers p and q can differ or correspond with each other. For example, we can write ARMA(3, 2):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}.$$

We have presented the model for initial data without any transformations. But it is known that often time series are non-stationary in levels. As a result, it is necessary to modify ranges to achieve stationarity. If we use differencing and build ARMA for a new range created by this way, we have ARIMA(p, k, q) model. The element “I” reflects the order of integration. In general, the range is integrated with the order k if it was generated as k differences from the initial series. For example:

$$\Delta y_t = y_t - y_{t-1} \sim I(1),$$

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} \sim I(2)$$

...

$$\Delta^k y_t = \Delta^{k-1} y_t - \Delta^{k-1} y_{t-1} \sim I(k).$$

Let's presents ARIMA(3, 1, 2) model:

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \alpha_3 \Delta y_{t-3} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}.$$

Stability condition can be presented for ARMA models:

$$\left| \sum_{i=1}^p \alpha_i \right| < 1. \quad (3.4)$$

This condition is not satisfied if the estimation was started before achieving stationarity. If stability condition is broken, the process does not converge. In this case we cannot find good estimators for coefficients of the model and have no ability to make forecasts. Also it is impossible to apply standard instruments for testing hypotheses. If the condition is satisfied as equality, we have a unit root in the process.

Speaking about ARMA process, it is necessary to tell about variance in the error term. The model assumes that the variance is constant and does not depend on time. In other words, conditional variance should

be equal to unconditional one. If it takes place, we have no problems in the model. But if the assumption is broken, we should apply a special model to estimate processes in the variance. This class of models will be discussed below.

Estimation of ARMA models usually follows the Box-Jenkins methodology. There are several steps:

1. The range has to be stationary:
 - Unit root tests are provided;
 - If the range is non-stationary, we should apply different transformations (detrending or differencing) and repeat the test until stationarity will be achieved;
2. We should estimate autocorrelation and partial autocorrelation functions and choose feasible lags for AR and MA processes;
3. Models, chosen on the second step, are being estimated. There have to be no autocorrelation in the error term. If the model has autocorrelation, it should be excluded from the analysis;
4. The best model is being chosen from all suitable variants according to minimal values of information criteria.

The first step was studied earlier. Now we should define autocorrelation (ACF) and partial autocorrelation (PACF) functions. ACF allows us to determine MA lags (the number q). Formally ACF looks like:

$$\rho_s = \frac{Cov(y_t, y_{t-s})}{Var(y_t)}. \quad (3.5)$$

This function includes correlation between all elements which are placed between y_t and y_{t-s} .

AR lags (the number p) can be supposed according to PACF function. The function includes the next elements:

$$\rho_r = Cov(y_t, y_{t-r}). \quad (3.6)$$

There is no additional covariance, this function accounts only connection between these two elements.

To identify autocorrelation we can also use special tests. For example, it is possible to use Box — Pierce Q -test. Null and alternative hypotheses look as follows:

$$H_0 : \rho_i = 0 \forall i = 1, \dots, k,$$

$$H_1 : \exists i \rho_i \neq 0, i = 1, \dots, k .$$

In this case ρ_i is an element of ACF. The null hypothesis says that there is no autocorrelation. The alternative hypothesis states the presence of autocorrelation. Statistics for the test is:

$$Q = T \sum_{i=1}^k \hat{\rho}_i^2 \sim \chi^2(k). \quad (3.7)$$

If the observed value is more than critical one, the null hypothesis must be rejected. Then we can say that there is autocorrelation on the lag k .

Another test is the Ljung — Box Q -test. The hypotheses and the rejection rule are the same, but the test allows us to control the size of the sample. Statistics will look like:

$$Q = T(T+2) \sum_{i=1}^k \frac{\hat{\rho}_i^2}{T-i} \sim \chi^2(k). \quad (3.8)$$

Also we can analyze ACF and PACF graphically. Let's consider Fig. 3.1 where these functions are presented.

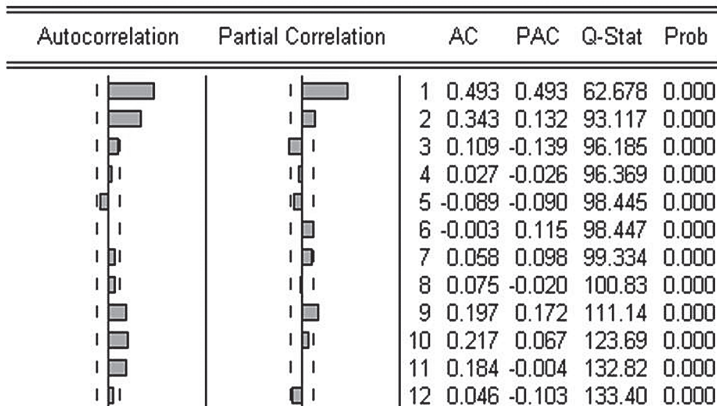


Fig. 3.1. Graphical Example for ACF and PACF functions

There are thresholds and bars. If the bar crosses the threshold, it gives a signal about the presence of autocorrelation on this lag. If there are doubts about the cross, it is necessary to look at the probability for Q -test and make an inference about rejection or acceptance of the null hypothesis (no autocorrelation). In our case we have the next potential processes: AR(1), AR(2), AR(3), AR(6), AR(9), MA(1), MA(2), MA(3), MA(9), MA(10), MA(11) and their possible combinations.

The maximal number of lags p and q also can be determined as $T/4$. Also there is the next tendency in time series: high frequency requires more lags.

On the third step models with all possible numbers of lags should be estimated. Every model must be tested for autocorrelation (it can be any test for autocorrelation). If the model has autocorrelation, it should be excluded from the analysis.

On the last step we choose the best model from all variants from the third step. The choice is based on minimal values of information criteria. There are three criteria which are often used in econometrics. The first is Akaike information criterion:

$$AIC = T \ln(SSR) + 2n, n = p + q + 1. \quad (3.9)$$

In this formula the letter n presents the number of parameters in the model.

The second criterion is Schwarz — Bayesian information criterion:

$$SBIC = T \ln(SSR) + n \ln T. \quad (3.10)$$

The third one is Hannan — Quinn information criterion:

$$HQIC = T \ln(SSR) + 2n \left(\ln \left(\ln T \right) \right). \quad (3.11)$$

It is possible when for different models we have opposite values of information criteria. For the first model AIC is minimal, but for the second model $SBIC$ or $HQIC$ is less. In such case we should choose the second model. It is connected with statistical power of information criteria. So, $SBIC$ and $HQIC$ are more powerful than AIC , and in our choice we should follow this rule.

It should be noted that ARMA models are not estimated with the OLS. As we have stability condition here, special methods with convergence are to be used here. It is iteration methods which work with special functions (likelihood functions). For example, the Newton approach for approximate calculations can be used to estimate ARMA processes. These methods are realized automatically in applied statistical packages.

ARMA and ARIMA models can be used for forecasting and presentation of common dynamics of the data. But there are cases when the assumption about constant variance in the error term is broken. In such case other models are applied, and further we will study their identification and estimation.

3.2. ARCH and GARCH models

As it was mentioned above, standard ARMA models work with some assumptions about a variance in the error term. Formally it can be written as:

$$Var(\varepsilon_t) = Var(\varepsilon_{t-s}) = Var(\varepsilon_t) \forall t, s.$$

It implies that the variance in the error term should be constant, and unconditional and conditional values are the same. But this assumption is not satisfied always. If it is not true, the estimation process requires other modifications of the model. As a result, models of autoregressive conditional heteroskedasticity (ARCH) and generalized conditional heteroskedasticity (GARCH) have to be used.

Let's consider a standard AR (1) process. It is possible to calculate unconditional and conditional means and variances and illustrate that these values will be different. Conditional mean and variance for AR(1) will look as follow:

$$E_t(y_t) = E(y_t | y_{t-1}) = E(\alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t) = \alpha_0 + \alpha_1 y_{t-1};$$

$$Var_t(y_t) = E((y_t - E y_t)^2 | y_{t-1}) = E((y_t - \alpha_0 - \alpha_1 y_{t-1})^2 | y_{t-1}) = E \varepsilon_t^2 = \sigma^2.$$

Unconditional values for the same process will be the next:

$$\begin{aligned} E(y_t) &= E(\alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t) = \alpha_0 + \alpha_1 E(\alpha_0 + \alpha_1 y_{t-2} + \varepsilon_t) = \\ &= \alpha_0 + \alpha_0 \alpha_1 + \alpha_1^2 E(\alpha_0 + \alpha_1 y_{t-2} + \varepsilon_t) = \dots = \frac{\alpha_0}{1 - \alpha_1} = \text{const} \end{aligned}$$

$$\begin{aligned} Var(y_t) &= Var(\alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t) = Var(\alpha_1 y_{t-1} + \varepsilon_t) = Var(\alpha_1 y_{t-1}) + \sigma^2 = \\ &= \alpha_1^2 (Var(\alpha_1 y_{t-2}) + \sigma^2) + \sigma^2 = \dots = \frac{\sigma^2}{1 - \alpha_1^2} = \text{const}. \end{aligned}$$

The last equation was written under assumption that there is no endogeneity problem (explaining variables are not correlated with the error term). ARMA processes can be presented with conditional homoskedasticity, but they cannot deal with conditional heteroskedasticity when σ^2 is not constant. In this case ARCH/GARCH models have to be applied.

Now we will consider ARCH and GARCH processes. It is clear that both these processes concern variance in the error term. In fact, ARCH process for the error term reminds a standard AR process for initial time series. Let's present the example of $AR(p)$ — $ARCH(m)$ process:

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \varepsilon_t, \quad (3.12)$$

$$\varepsilon_t = u_t \sqrt{\gamma_0 + \sum_{i=1}^m \gamma_i \varepsilon_t^2}, u_t \sim iid(0,1).$$

The expression under the square root is conditional variance for the model. The number m also should be chosen with the help of different procedures and test which will be discusses below. Also we can present conditions for non-negative values under the root:

$$\gamma_0 > 0, \gamma_i \geq 0 \forall i. \quad (3.13)$$

Stability condition can be presented in the next form:

$$\sum_{i=1}^m \gamma_i < 1. \quad (3.14)$$

If this condition is broken, we cannot estimate the process and coefficients properly. Moreover, it is impossible to use usual statistical tests to check our suppositions about parameters of the model. There can be not only AR(p)—ARCH(m) combination, but ARMA(p, q)—ARCH(m) too. The general representation of the model depends on the character of variance in the error term.

GARCH process reminds MA process for initial time series. Formally ARMA(p, q)—GARCH(m, n) process will look like:

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{j=1}^q \beta_j \varepsilon_{t-j} + \varepsilon_t. \quad (3.15)$$

$$\varepsilon_t = u_t \sqrt{\gamma_0 + \sum_{i=1}^m \gamma_i \varepsilon_t^2 + \sum_{j=1}^n \theta_j h_{t-j}}, u_t \sim iid(0,1),$$

$$h_t = \gamma_0 + \sum_{i=1}^m \gamma_i \varepsilon_t^2.$$

Conditions for non-negative values under the root are the same. Stability condition will have the next form:

$$\sum_{i=1}^m \gamma_i + \sum_{j=1}^n \theta_j < 1. \quad (3.16)$$

Now the formal presentation of the processes is clear. Now it is important to show the ways how conditional heteroskedasticity can be identified in the model. First of all we should estimate ARMA(p, q) model for which we will test the error term. Then it is necessary to save residuals from the model and generate their squares. Further we will work with the range from squares of the residuals, and all tests will be provided for this series. There are different methods to identify

conditional heteroskedasticity, and all of them are based on the analysis of the error term and correlation relations.

The first way is to generate the range of residuals from ARMA(p, q) model, create their squares and repeat the Box — Jenkins procedure for new range with all possible number of lags n and m .

The second method is supposed to analyze ACF and PACF functions. For this purpose Ljung — Box Q -test and LM -test can be used. Let's consider the first test.

Null and alternative hypotheses will be presented in the next form:

$H_0 : \rho_i = 0 \forall i = 1, \dots, k$ (there are no ARCH/GARCH processes);

$H_a : \exists i \rho_i \neq 0, i = 1, \dots, k$ (there are ARCH/GARCH processes).

Statistics for Ljung — Box Q -test is the same, but now we should estimate correlation in squares of the residuals from initial ARMA model:

$$Q = T(T+2) \sum_{i=1}^k \frac{\hat{\rho}_i^2}{T-i} \sim \chi^2(k), \quad (3.17)$$

$$\hat{\rho}_i = \frac{\sum_{i=1}^k (e_i^2 - \sigma^2)(e_{i-i}^2 - \sigma^2)}{\sum_{i=1}^k (e_i^2 - \sigma^2)^2},$$

$$\sigma^2 = \frac{\sum_{t=1}^T e_t^2}{T}.$$

In this case e_t presents residuals from initial ARMA model. If the observed value is more than critical one, we should reject the null hypotheses. It means that ARCH/GARCH processes are presented in the model, and they should be estimated.

LM -test helps to identify only ARCH(m) processes. To test the hypothesis about the existence of ARCH(m) process, we should estimate AR(m) model for the range of residuals in square:

$$e_t^2 = \gamma_0 + \sum_{i=1}^m \gamma_i e_{t-i}^2 + \Phi_t. \quad (3.18)$$

Null and alternative hypotheses will be the next:

$$H_0 : \gamma_i = 0 \forall i = 1, \dots, m \text{ (there is no ARCH}(m)\text{ process)};$$

$$H_1 : \exists i \gamma_i \neq 0, i = 1, \dots, m \text{ (there is ARCH}(m)\text{ process)}.$$

Statistics for the test has a form:

$$LM = TR^2 \sim \chi^2(m). \quad (3.19)$$

The number R^2 is taken from the regression for e_t^2 . If the observed value is more than critical one, we should reject the null hypothesis about the absence of ARCH(m) process.

If we have found ARCH process, now it is possible to catch GARCH one. For this purpose LM -test is also applied. On the first step it is required to estimate ARMA(p, q)—ARCH(m) model which was identified earlier. Then we will work with residuals (in particular with their squares) from this model. Let's denote this variable as l_t to prevent complications with previous labels. The next model should be estimated with the OLS:

$$\frac{l_t^2}{h_t} - 1 = \gamma_0 + \sum_{i=1}^m \gamma_i e_{t-i} + \varphi_t. \quad (3.20)$$

A fraction, which is places on the left side of the equation, is so-called standard residuals. The variable h_t can be calculated according to the formula indicated above. Null and alternative hypotheses will be the next:

$$H_0 : \text{ARCH}(m) \text{ model};$$

$$H_1 : \text{GARCH}(m, n) \text{ model}.$$

Statistics is calculated as earlier with R^2 from the regression presented for testing:

$$LM = TR^2 \sim \chi^2(m). \quad (3.21)$$

If we reject the null hypothesis, we can say about the existence of GARCH process in data.

Now we can present the procedure for estimation of ARCH/GARCH models:

1. Stationarity should be achieved in initial data (different transformations and unit root test are to be used);
2. A proper ARMA model is being chosen and estimated. Then the residuals from the model are saved. A new range from squares of these residuals are generated;
3. ARCH and GARCH processes are revealed with the help of different tests (Ljung — Box test, *LM*-test etc.);
4. When the process has been caught, lags are chosen (numbers m and n). For this purpose ACF and PACF functions for squares of the residuals should be studied (it is possible to use correlogram). The number m is connected with PACF, and n should be chosen according to ACF;
5. Then the model $\text{ARMA}(p, q)$ — $\text{GARCH}(m, n)$ is estimated with the maximum likelihood estimation method (MLE). All restrictions for the coefficients in the GARCH part must be satisfied:

$$\gamma_0 > 0, \gamma_i \geq 0,$$

$$\sum_{i=1}^m \gamma_i + \sum_{j=1}^n \theta_j < 1.$$

6. After the estimation standard residuals should be saved in the model. If there is no autocorrelation in the residuals, the model is good. It can be tested with *LM*-test, *Q*-tests etc. Sometimes it is necessary to add lags into ARMA to remove autocorrelation from the process.

There are different modifications of ARCH/GARCH models. For example, integrated GARCH (I-GARCH) process can be performed. Its main feature is:

$$\sum_{i=1}^m \gamma_i + \sum_{j=1}^n \theta_j = 1. \quad (3.22)$$

It means that the process can be non-stationary, and it will describe persistent variance. Also it is possible that such process has infinite variance. In such case the variance in the sample is not defined, and the process does not converge to a number.

Another example is GARCH in mean (GARCH-M) model. In this case conditional variance is used as a regressor in the ARMA part. There is an example for AR(2)—GARCH(1)-M process:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \mu h_t + \varepsilon_t, \quad (3.23)$$

$$\sqrt{h_t} = \sqrt{\gamma_0 + \gamma_1 \varepsilon_{t-1}^2}.$$

There are other types of GARCH models. For example, threshold (TGARCH) allows us to separate negative and positive shocks. Exponential GARCH is widespread in finance etc.

In general, ARCH/GARCH models are applied to data with high frequency where conditional heteroskedasticity is usual. This kind of models is popular in finance because the dynamics of data, using here, cannot be described with standard ARMA models properly. So, ARCH/GARCH models are quite useful for different researches especially if they concern high frequency data.

Thus, we considered models for stationary time series. But all presented processes work only with a single variable in dynamics. Often it is necessary to analyze different macroeconomic series together and observe the response of one variable to changes in another one. For this purposes intervention analysis is used. Further we will study main concepts and models in this field of applied econometrics for macroeconomics.

Questions

1. What kinds of models for stationary time series do you know?
List the differences between the models which you have indicated.
2. Describe the essence of AR and MA processes. For what purpose can they be used?
3. What are autocorrelation and partial autocorrelation functions?
How can they help to identify AR and MA processes?
4. What tests for autocorrelation do you know? State null and alternative hypotheses and a rejection rule.

5. Explain the Box — Jenkins methodology. How can we choose the best ARMA model?
6. What is the difference between ARMA and ARCH/GARCH models? Why should we use any additional corrections for ARMA models?
7. How can we catch ARCH/GARCH processes? What tests are used to do it? Describe the main principles of these tests.
8. In what researches ARCH/GARCH models can be applied?
9. List different types of ARCH/GARCH models. What are their key features and differences?

Exercises

1. For this problem you need two any macroeconomic series with quarterly or yearly frequency (the minimal required number of observations is 150).
 - a) Present you data graphically and check them for stationarity. If it is necessary provide all transformation to achieve stationarity,
 - b) Now you are asked to present ARMA model for your data. What potential combinations of AR and MA processes can you present? What properties does the best model have?
 - c) Estimate all potential ARMA models and choose the best one for both time series. What criteria have you used to find the best model?
2. To solve this task you can use data sets from the first problem.
 - a) Use the best ARMA models form the first problem. You should identify ARCH process in the data. What test will you use? Provide the test and report results for both time series,
 - b) If you have found ARCH process, now you are able to catch GARCH one. What test can you apply? Provide the test and report the result,
 - c) During the first two steps you have make inferences about the existence of ARCH/GARCH processes in the data. Now you should write and estimate proper models for these processes. Choose lags m and n and inspect the quality of your models.

Chapter 4

INTERVENTION ANALYSIS

4.1. General Facts about Intervention Analysis

We have already seen elements of intervention analysis in specifications of tests for structural breaks. The main idea of intervention analysis is to catch some events and test their impact on different variables. If we know a particular moment of time when the event happened, it is possible to create a new discrete variable to follow consequences of the event.

There are different ways to present such shocks in the model. The first type of representation is a pure jump. Choosing this one we assume that if the event took place, its influence will keep in the system for all next periods. Formally this variable can be performed as follow:

$$s_t = \begin{cases} 0, & \text{before the event,} \\ 1, & \text{at the moment and after the event.} \end{cases}$$

If you remember, structural breaks in intercept were constructed by this way.

Also the event can take place during a particular moment of time, and then its impact will disappear. It is a pulse function. For this representation the variable should be performed as:

$$s_t = \begin{cases} 0, & \text{before and after the event,} \\ 1, & \text{at the moment of the event.} \end{cases}$$

But it is not necessary that shocks will keep in the system or disappear at once. Sometimes the impact of the event can weak in time or, in opposite, increase. For the first case a prolong impulse function can be presented:

$$s_t = \begin{cases} 0, & \text{before the event,} \\ 1, & \text{at the moment of the event,} \\ 0,75, & \text{at the moment } t+1, \\ 0,5, & \text{at the moment } t+2, \\ 0,25, & \text{at the moment } t+3, \\ 0, & \text{at the moment } t+4 \text{ and later.} \end{cases}$$

For the case of gradual changes and accumulating impact the discrete variable can be constructed by the next way:

$$s_t = \begin{cases} 0, & \text{before the event,} \\ 0,25, & \text{at the moment of the event,} \\ 0,5, & \text{at the moment } t+1, \\ 0,75, & \text{at the moment } t+2, \\ 1, & \text{at the moment } t+3. \end{cases}$$

If you have your own research purposes, know some facts about the character of the event and need to construct a special variable, you are not restricted. But it is important to make a proper specification for your model and state the variable clearly.

Now it is important to state some facts about impulse and transfer functions. For this purpose we will consider the simplest case — a standard AR(1) process with one event:

$$s_t = \begin{cases} 0, & \text{before the event,} \\ 1, & \text{at the moment and after the event.} \end{cases}$$

Formally the process will look like:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \lambda s_t + \varepsilon_t. \quad (4.1)$$

It is obvious that before the event intercept is equal to α_0 . After the event we have an additional constant in the model, λ . Then new intercept will look like $(\alpha_0 + \lambda)$. If there is a difference in intercepts before and after the event, we can assume that long-run means will also

differ in response to the shock. Now we show that it is true, and the event changes a long-run mean. But one remark should be done here. We assume that the process is stationary (AR models are built only for stationary data). It implies that stability condition is satisfied:

$$|\alpha_1| < 1.$$

In opposite case we cannot estimate coefficients properly, and it will be impossible to calculate long-run means because the process will not converge.

The long-term mean (conditional) for standard AR(1) process was calculated above:

$$E(y_t | y_{t-1}) = \frac{\alpha_0}{1 - \alpha_1}. \quad (4.2)$$

If to provide the same procedure for the process after the event, we will obtain:

$$\begin{aligned} E(y_t) &= E(\alpha_0 + \alpha_1 y_{t-1} + \lambda + \varepsilon_t) = \alpha_0 + \lambda + \alpha_1 E(\alpha_0 + \alpha_1 y_{t-2} + \lambda + \varepsilon_t) = \\ &= \alpha_0 + \lambda + (\alpha_0 + \lambda) \alpha_1 + \alpha_1^2 E(\alpha_0 + \alpha_1 y_{t-2} + \lambda + \varepsilon_t) = \dots = \frac{\alpha_0 + \lambda}{1 - \alpha_1} = \text{const.} \end{aligned}$$

It means that the event has led to changes in the long-term mean. As a result, the dynamics of the process has changed too. It is important fact for forecasting because such changes require special models to control interventions.

In general, we want to know the response of our series to exogenous changes. These changes come to the model as different events and lead to shocks (positive or negative). In other words, we are interested in how current shocks influence future values of the variable.

To find an answer for this question, we will rewrite the process:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \lambda s_t + \varepsilon_t = \alpha_0 + \alpha_1 (\alpha_0 + \alpha_1 y_{t-2} + \lambda s_{t-1}) + \lambda s_t + \varepsilon_t \text{ etc.}$$

It is possible to continue this function, but all necessary moments are clear now. The function presented above is called impulse response function. It shows how current values react on shocks in previous

periods. If we want to know a particular value of changes, we should take a derivative. For example, we are interested how y_t responds to the event which took place in previous period. We can provide the next calculation:

$$\frac{dy_t}{ds_{t-1}} = \frac{dy_{t-1}}{ds_{t-1}} + \frac{dy_t}{ds_{t-1}} = \lambda + \lambda\alpha_1. \quad (4.3)$$

It should be remarked that it is necessary to account two elements in this derivative. The variable s_{t-1} influence not only current values, but also previous observations. If we want to estimate the reaction of y_t on changes in s_{t-2} , we should include three elements in the derivative:

$$\frac{dy_t}{ds_{t-2}} = \frac{dy_{t-2}}{ds_{t-2}} + \frac{dy_{t-1}}{ds_{t-2}} + \frac{dy_t}{ds_{t-2}} = \lambda + \lambda\alpha_1 + \lambda\alpha_1^2. \quad (4.4)$$

Now we can modify this presentation a little. We will present the reaction of future values on current changes in our discrete variable, s_t . In any moment of time $t + i$ the reaction on the event in period t can be calculated as:

$$\frac{dy_{t+i}}{ds_t} = \lambda(1 + \alpha_1 + \alpha_1^2 + \dots + \alpha_1^i). \quad (4.5)$$

Then it is possible to estimate this reaction on the infinite time horizon. If t converges to infinity, we get the next result:

$$\lim_{t \rightarrow \infty} \frac{dy_{t+i}}{ds_t} = \frac{\lambda}{1 - \alpha_1}. \quad (4.6)$$

It should be noted that this result was obtained with the assumption:

$$|\alpha_1| < 1.$$

If it is not satisfied, we cannot find a particular value for this limit. This fact also indicates the importance of stability condition.

Now we can show how this limit is connected with long-term means:

$$\frac{\lambda}{1 - \alpha_1} = \frac{\alpha_0 + \lambda}{1 - \alpha_1} - \frac{\alpha_0}{1 - \alpha_1}. \quad (4.7)$$

The first element on the right side of the equation is the mean after the event. The second element is the mean before the event. In fact, the value of the response is equal to the difference between long-term means.

To use and estimate the model with intervention properly, it is necessary to have enough data before and after the event. If this condition does not take place, we cannot use the estimators for forecasting and further analysis.

Also intervention analysis operates with the term “transfer function”. It describes the process of transferring from one moment of time to another one. This function includes an exogenous variable, and the variable produces potential shocks in the system. Using lag operators we can present the transfer function in the next form:

$$y_t = \alpha_0 + A(L)y_{t-1} + B(L)s_t + \varepsilon_t. \quad (4.8)$$

We assume that s_t is an exogenous variable. It means that it is not correlated with the error term and y_{t-1} . This condition is necessary to prevent endogeneity problem. But in fact we do not know if the variable s_t is exogenous or not. The source of endogeneity problem here is the existence of potential simultaneity. Often macroeconomic variables influence each other, and sometime these connections are quite complicated. If we ignore the existence of such interaction, we can obtain inconsistent and biased estimators because endogeneity problem. To prevent these negative facts it is possible to estimate a system of simultaneous equations. Such procedure also allows researchers to analyze shocks and use impulse response functions. The system of simultaneous equations for several time series with lags is called vector of autoregression (VAR). The concept of VAR and its different forms will be discussed in the next paragraph.

4.2. Vector Autoregression (VAR) model

This concept is quite convenient for the analysis. In particular, VAR allows researchers to observe responses of variables to changes in other parameters of the model. There are different forms of VAR

representation, but they have a common feature. So, we estimate not separate equations, but a whole system. Such approach helps to identify bilateral interactions and obtain good estimators for which consistency and unbiasedness are satisfied. Let's present a common view of VAR and consider its different forms.

Initially we must clarify what kind of data can be used in the VAR model. This model is an extension of standard autoregression models. As we saw earlier, these models are based on the assumption that the variable depends on its previous values and an exogenous parameter (usually it is a constant). A common principle is that time series have to be stationary. The second aspect is that ranges, which are being used in the estimation, have to be integrated with the same order. If initial data are stationary, we construct the VAR in levels. But if they are not, we have to provide some transformations for achieving stationarity.

When we use differencing for this purpose, it is possible to obtain stationary series with different orders of integration. For example, the first range is integrated with the order one, and the second series is stationary in levels (integrated with the order zero). But in such case it is also possible to construct the VAR model, but it should be in levels. For example, such approach is admissible when we want to estimate series integrated with different orders. Sometimes it is necessary for research purposes and can be done.

If time series have the same order of integration, there are two possibilities. The first one is that data are cointegrated. This concept will be considered in details in the next chapter. But the reason is that cointegration requires some changes in the VAR specification. If the series are cointegrated, we should include the error term as a regressor in the VAR. As a result, we will get an error correction model (ECM) which is constructed for cointegrated time series.

If our series are not cointegrated, it is enough to provide all necessary transformation for achieving stationarity. Then the VAR model can be built of the data have the same order of integration.

The VAR can be presented in two forms:

- Structural form;
- Reduced form.

For simplicity we will present the case with two variables in the system. The model will be written in levels. Then the VAR in structural form look like that:

$$\begin{cases} y_t = c_{1,0} + c_{1,1}z_t + \rho_{1,1}y_{t-1} + \rho_{1,2}z_{t-1} + \varepsilon_{1,t}, \\ z_t = c_{2,0} + c_{2,1}y_t + \rho_{2,1}y_{t-1} + \rho_{2,2}z_{t-1} + \varepsilon_{2,t}. \end{cases} \quad (4.9)$$

Above we have presented the VAR with the order one. This order is determined with the number of lags including in the model. For example, if we enter lags with a maximal number k , we will have the VAR with the order k . The system has to include exogenous parameters. In our case constants play this role in equations. Each variable depends on its lags and another series. Moreover, a current value of the variable is determined with the same one for another variable. As a result, we analyze interactions during present and previous moments of time. It is a key feature of the VAR in structural form. Using this model, we can represent and observe structural networks in different macroeconomic series. Also it is possible to analyze response functions and include shocks in the model*.

It is possible to present the VAR in the matrix form. For this we should enter the next labels:

$$A = \begin{pmatrix} 1 & -c_{1,1} \\ -c_{2,1} & 1 \end{pmatrix},$$

$$x_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix},$$

$$R_0 = \begin{pmatrix} c_{1,0} \\ c_{2,0} \end{pmatrix},$$

* Shocks can be presented in the error term, but it is possible to include another exogenous variable in the model. But in this case it is necessary to control if rank condition is satisfied or not to avoid overidentification.

$$R_1 = \begin{pmatrix} \rho_{1,1} & \rho_{1,2} \\ \rho_{2,1} & \rho_{2,2} \end{pmatrix},$$

$$e = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}.$$

Then we can rewrite the VAR with the order one in the matrix form:

$$Ax_t = R_0 + R_1 x_{t-1} + e. \quad (4.10)$$

As for other econometric models, the error term must be characterized with a zero mean, no autocorrelation and constant variance to provide the best estimators. These conditions can be violated in some cases, and often it depends on a chosen estimation method.

As it was marked above, the VAR presents a system of equations. As a result, it requires special methods for estimation. For example, it is possible to estimate these equations separately with the OLS, but in such case we lose effects which can be caught when we consider a system. Also if we ignore the system existence, we can face endogeneity problem. As a result, the estimators will be biased and inconsistent. In any case, it will be better to estimate a system if it takes place.

The system should be estimated simultaneously. For this purpose it is possible to use different modifications of the least squares method (for example, the GLS, 3SLS) or the maximum likelihood estimator. But there are several problems connected with the system identification. In particular, the rank condition should be satisfied. In other words, we must have enough instruments for endogenous variables. If this condition is violated, we have a set of possible values for parameters, and the system is not identified.

There are two conditions which help to determine if the equation is identified or not. The first one is an order condition. It says that any equation in the system is identified if the number of included

endogenous variables is less or equal to the number of excluded exogenous variables.

The second one is a rank condition. The rank condition says that the equation is identified if other equations include at least one significant exogenous variable which is excluded from the analyzing one. In other words, there are enough significant instruments for endogenous variables in the equation. The order condition is necessary for the rank condition, and the rank condition is necessary and sufficient for identification of the equation.

Let's present the example; we will present endogenous variables with the letter y , and exogenous ranges will be noted with z . The system includes three simultaneous equations which should be estimated:

$$\begin{cases} y_{1t} = y_{2t} + y_{3t} + z_{1t} + z_{3t} + \varepsilon_{1t}, \\ y_{2t} = y_{1t} + y_{3t} + y_{4t} + z_{2t} + z_{4t} + \varepsilon_{2t}, \\ y_{3t} = y_{2t} + z_{3t} + \varepsilon_{3t}. \end{cases} \quad (4.11)$$

Now we should find which equations are identified and can be estimated. We have four endogenous and exogenous variables. The equations have the next number of exogenous and endogenous variables on the right hand:

- The first equation: two included endogenous variables vs. two excluded exogenous variables;
- The second equation: three included endogenous variables vs. two excluded exogenous variables;
- The third equation: one included endogenous variable vs. three excluded exogenous variables.

To test the rank condition, we should know particular values for estimators and their statistics. Now we can check only the order condition. For the first equation this condition is satisfied because the number of included endogenous variables is equal to the quantity of excluded instruments. For the second equation the order condition is violated ($3 \text{ included} > 2 \text{ excluded}$). For the third equation this condition is true. But we have another problem here. There is an overidentification. For one endogenous variable we have an excessive number of instruments.

It is possible that some instruments are weak, and we can remove them from the IV regression. There are special tests for overidentification, and they help to catch this problem in the model*.

Sometimes the VAR in structural form is not identified because not all instruments (for example, lags) are exogenous. It means that we cannot find a single set of estimators. If we have this problem, it is possible to construct the VAR in reduced form. To derive this form, we will use the VAR in structural presentation with matrixes:

$$Ax_t = R_0 + R_1x_{t-1} + e.$$

Now we will multiple all elements on A^{-1} (an inverse matrix for A) at the left:

$$A^{-1}Ax_t = A^{-1}R_0 + A^{-1}R_1x_{t-1} + A^{-1}e. \quad (4.12)$$

Let's present the next labels:

$$A^{-1}R_0 = D_0,$$

$$A^{-1}R_1 = D_1,$$

$$A^{-1}e = \mu.$$

Now we can rewrite the equation in the next form:

$$x_t = D_0 + D_1x_{t-1} + \mu. \quad (4.13)$$

This is the VAR in reduced form. In fact, we removed values of the current period from the right side of the equation. It means that only previous periods influence present moment. Also one variable has no direct connections with other series, and we excluded structural components from the analysis. So, if we are speaking about economic researches, it is better to use the VAR in structural form not to lose hidden relations.

* This theme is not a key point for applied econometrics for macroeconomics. For the VAR concept we should know only sources of potential problems. If you want to study the system estimation in more details, you are recommended to use econometrics textbooks (for example: *Wooldridge J. M. Introductory Econometrics: A Modern Approach.*).

But if it is impossible because of problems with the system identification, you are recommended to estimate the VAR in reduced form.

In general, the VAR with the order k can be presented in matrixes. The model in structural form has the next view:

$$Ax_t = R_0 + \sum_{i=1}^k R_i x_{t-i} + e. \quad (4.14)$$

The VAR in reduced form will be written as:

$$x_t = D_0 + \sum_{i=1}^k D_i x_{t-i} + \mu. \quad (4.15)$$

As it was said earlier, the VAR model can include different number of lags k . It is necessary to choose the best model. To do it we can use different criteria and tests. In principle, we can use the same approach as for standard ARMA(p, q) models. So, it is possible to estimate models with different number of lags, to test the significance of the last one and also compare information criteria. The model with minimal values of these criteria will be the best one. Information criteria can be calculated according to the next formulas:

$$AIC = T \ln(SSR) + 2n, \quad n = p + q + 1,$$

$$SBIC = T \ln(SSR) + n \ln T,$$

$$HQIC = T \ln(SSR) + 2n \left(\ln \left(\ln T \right) \right).$$

In these formulas the letter n presents the number of parameters in the model.

Another approach is more powerful because it is based on asymptotical properties of estimators. It assumes the LR -test. Two models with different numbers of lags are compared. The model with more lags is called unrestricted (abbreviation UR), and another one is restricted (abbreviation R). To provide the test we must know determinants of estimated covariance matrixes of error terms from two models. This

determinant will be denoted as $|\Sigma|$. Null and alternative hypotheses are the next:

H_0 : the model with less lags is better
(lags with higher orders are not significant);

H_1 : the model with more lags is better
(at least one lag with higher order is significant).

Statistics for the test can be calculated according to the formula:

$$LR = T \left(\ln |\Sigma_R| - \ln |\Sigma_{UR}| \right) \sim \chi^2(m), \quad m = (k_{UR} - k_R)N^2. \quad (4.16)$$

The letter N indicates the number of variables in the VAR, k_{UR} and k_R reflect lags in unrestricted and restricted models accordingly.

There is a right-side critical region. If the observed value is more than critical one, the null hypothesis should be rejected. If the LR-test and information criteria give controversial results, we have to use the test to make the final inference. The reason is that the test is more statistically powerful than criteria, and the probability of type II error is less here.

The VAR model allows researchers to analyze different shocks and their consequences for the system. To provide such observations impulse response functions are used. Earlier we studied the mechanism how these function can be obtained. Now we will consider their implementation for the VAR model.

Impulse response functions show changes in one variable in response to exogenous shocks in another one. So, if we have three time series in the system, we can obtain nine impulse response functions. The first group of them will reflect the variance: the variable will react on its own shocks. There will be three such functions in the case of three time series. The second group will characterize relations between different variables. There will be six functions of such type.

Impulse response functions are built with the assumption that all variables are stable except the pair of analyzing series. The first series has a shock, and the second one demonstrated the reaction on this event in the system. Exogenous shocks can be accounted in the error term. The matter is that they are not generated in the system which is stable.

Initially we have said that the VAR model was built for stationary time series. It means that shocks will disappear in the system over time. Fluctuations, stimulated with any shocks, will converge to the constant mean for any time series*.

Performing impulse response functions for each variable we can observe its reaction on external events and make forecasts to predict the dynamics of studying series. All modern statistical packages can build impulse response functions automatically. We present the example for the U. S. economy. We analyze monthly data for inflation, interest rate and unemployment rate from 1955 to 2011. For the VAR model in the first differences impulse response functions are shown on the Fig. 4.1.

Usually impulse response functions are built for particular number of standard deviations. Innovation is a shock in the system, and its value is equal to one standard deviation in the considered example. There are ten time periods (months), but it is possible to widen this horizon**.

For example, on these graphs we can see that unemployment and inflation are characterized with negative relations. So, the growth in inflation leads to decreasing unemployment rate. But also we see that such interaction is not long, and the shock is neutralized by the end of the tenth period. Such dependence between inflation and unemployment is known as the Phillips curve, and it is not satisfied in the long-run period. We have illustrated this fact with the help of the VAR and impulse response functions.

Another useful consequence from the VAR model is variance decomposition. This instrument helps to present the variance of each variable as a result of different factors' interaction over time. More

* This fact will be satisfied if the data are stationary and stability conditions take place in the system.

** Cholesky approach was used here. It is sensible to the order of variables.

precisely, we can say how the innovation in one variable is important in the forecast for another time series. In fact, this approach helps to decompose the variance in forecast errors.

Response to Cholesky One S. D. Innovations

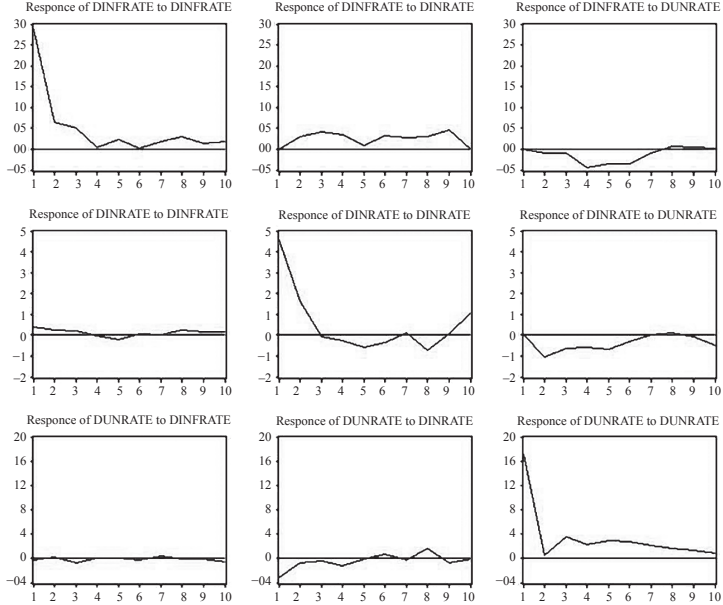


Fig. 4.1. Impulse Response Functions for the VAR model
in the First Differences
(the U. S. monthly data from 1955 to 2011)

We should find the mean squared error of the forecast which can be obtained with the help of this VAR. For standard models this value is calculated according to the next formula:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}. \quad (4.17)$$

So, we should know the forecast and the number of observations to calculate this value. In our case with the system it is more difficult

to find MSE because the structure of the model is more complicated. Applying to the VAR and impulse response functions, MSE presents a sum of innovation responses. Also we can estimate a gain of particular innovations into the error variance for each variable. In fact, variance decomposition is a relative measure which characterizes an importance of innovations in one variable for the forecast of another time series.

The variance decomposition can be presented as a set of graphs or as a table. We will show the table for our data on Fig. 4.2.

Variance Decomposition of DINFRATE:				
Period	S.E.	DINFRATE	DINRATE	DUNRATE
1	0.287142	100.0000	0.000000	0.000000
2	0.295799	98.92294	0.954218	0.122843
3	0.303038	96.94416	2.824147	0.231695
4	0.308371	93.64246	3.998944	2.358597
5	0.311352	92.38339	4.016581	3.600031
6	0.315024	90.24515	4.943524	4.811323
7	0.316930	89.48730	5.655926	4.856772
8	0.319760	88.72614	6.441085	4.832770
9	0.323374	86.92919	8.328564	4.742243
10	0.323911	86.96308	8.301121	4.735800

Fig. 4.2. Variance Decomposition for the First Difference of Inflation

We show this example only for one variable, but it is enough for our demonstration. We can observe how the gain of other factors to the variance of inflation increases in time. Initially all variance are explained by the variable itself. But by the end of the tenth period this share is only 86,9 %. Other percents are distributed between two variables in the system: interest rate (8,3 %) and unemployment (4,8 %). So, using this mechanism we can also observe the reaction of each variable for exogenous shocks.

Thus, using the VAR model we can identify different types of relationships between economic variables. Moreover, we can estimate consequences of shocks in the system and calculate gains of each factor in the common variance. This model is very important for macroeconomic analysis. But it is possible that relations between variables are not bilateral, and in such case we should use other models to identify this interaction.

4.3. Granger Causality

Sometimes we have cases when one variable does not determine another one, but it takes place in inverse order. In other words, we can separate a cause and an effect. As a result, it is possible to indicate the direction of such relationship. To find this interaction we should analyze causality for time series.

In fact, it is necessary to answer a question if lags of one variable enter in the equation for another one. Using causality approach we can also make inferences about exogeneity of one time series for any particular variable. This method was suggested by Granger, and now it is called Granger causality.

It should be remarked that Granger causality does not help to identify all types of interactions between different macroeconomic series. It is not consistent if the variables are driven by another process which is not included in the analysis. Granger causality can be provided for stationary and non-stationary series with some corrections. Then we will consider all basic cases which can take place in the process of Granger causality analysis.

Initially we will consider stationary series. In the simplest form we can use the standard F -test. Let's there are two variables: y_t and z_t . As the series are stationary, we can present the model in levels. We want to know if lags of z_t influence y_t or not. We can present unrestricted and restricted specifications for testing:

$$UR: y_t = c + \sum_{i=1}^k \rho_i y_{t-i} + \sum_{j=1}^m \pi_j z_{t-j} + \varepsilon_t. \quad (4.18)$$

$$R: y_t = c + \sum_{i=1}^k \rho_i y_{t-i} + \varepsilon_t. \quad (4.19)$$

Null and alternative hypotheses will look as follow:

$$H_0: z_t \text{ does not Granger cause } y_t,$$

$$\pi_1 = \dots = \pi_m;$$

$$H_1: z_t \text{ Granger causes } y_t,$$

$$\exists j \pi_j \neq 0, j=1, \dots, m.$$

Test statistics is calculated according to the next formula:

$$F = \frac{\frac{SSR_R - SSR_{UR}}{m}}{\frac{SSR_{UR}}{T - m - k - 1}} \sim F(m, T - m - k - 1). \quad (4.20)$$

There is a right-side critical region. Rejection rule will be formulated in the next way: if the observed value is more than critical one, we have to reject the null hypothesis. So, we can say that z_t Granger causes y_t .

There are three possible ways how the variables can interact with each other. The first one takes place when time series are not connected. In this case there is no Granger causality.

The second way is possible when the variables influence each other simultaneously. In other words lags of each variable are significant in the equation for another one. In this case k and m may be equal or not. If they are the same, we have a so-called near VAR case. But if these numbers differ, we should estimate a seemingly unrelated equations model.

The third possibility takes place when there is no simultaneous interaction. For example, y_t Granger causes z_t , and the opposite is not true. In this case z_t will be a cause, and y_t presents an effect. Then we can say that z_t is weakly exogenous for y_t , and it can be used as an instrument for the effect variable with some assumptions.

It is possible to present Granger causality in matrixes. Let's consider the VAR with the order p which includes n variables. The reduced form in matrix representation looks like:

$$\begin{pmatrix} x_{1t} \\ \dots \\ x_{nt} \end{pmatrix} = \begin{pmatrix} A_{10} \\ \dots \\ A_{n0} \end{pmatrix} + \begin{pmatrix} A_{11}(L) & \dots & A_{1n}(L) \\ \dots & \dots & \dots \\ A_{n1}(L) & \dots & A_{nn}(L) \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ \dots \\ x_{nt-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ \dots \\ e_{nt} \end{pmatrix}. \quad (4.21)$$

$A_{ij}(L)$ represents a polynomial of the lag operator L . It includes coefficients of lagged values of the variable j on the series i . If to formulate Granger causality in these terms, it can be written: the variable j does not Granger cause for the series i if all elements in $A_{ij}(L)$ are equal to zero.

In general, the idea of Granger causality is clear. Now it is possible to consider a practical application of this concept for the VAR model. Using Granger causality, we can determine if it is necessary to include an additional variable into the VAR structure or not. This test is known as the Block test.

We have the VAR in reduced form for two variables: y_t and z_t . We want to know if any new variable l_t should be included in the model. An unrestricted form for testing has the next view:

$$\begin{cases} y_t = c_{10} + \sum_{i=1}^k \rho_i y_{t-i} + \sum_{i=1}^k \gamma_i z_{t-i} + \sum_{i=1}^k \delta_i l_{t-i} + \varepsilon_{1t}, \\ z_t = c_{20} + \sum_{i=1}^k \rho_i y_{t-i} + \sum_{i=1}^k \gamma_i z_{t-i} + \sum_{i=1}^k \delta_i l_{t-i} + \varepsilon_{2t}, \\ l_t = c_{30} + \sum_{i=1}^k \rho_i y_{t-i} + \sum_{i=1}^k \gamma_i z_{t-i} + \sum_{i=1}^k \delta_i l_{t-i} + \varepsilon_{3t}. \end{cases} \quad (4.22)$$

The model with restrictions will be the next:

$$\begin{cases} y_t = c_{10} + \sum_{i=1}^k \rho_i y_{t-i} + \sum_{i=1}^k \gamma_i z_{t-i} + \varepsilon_{1t}, \\ z_t = c_{20} + \sum_{i=1}^k \rho_i y_{t-i} + \sum_{i=1}^k \gamma_i z_{t-i} + \varepsilon_{2t}, \\ l_t = c_{30} + \sum_{i=1}^k \rho_i y_{t-i} + \sum_{i=1}^k \gamma_i z_{t-i} + \sum_{i=1}^k \delta_i l_{t-i} + \varepsilon_{3t}. \end{cases} \quad (4.23)$$

Then null and alternative hypotheses can be written as follow:

$H_0 : \delta_i = 0 \forall i = 1, \dots, k$, the variable should not be included into the VAR;

$H_1 : \exists i \delta_i \neq 0, i = 1, \dots, k$, the variable should be included into the VAR.

To test the null hypothesis the *LR*-test is used. The procedure is the same as for the chose between different VAR specifications. To provide the test we should know determinants of covariance matrixes of error terms. Statistics is calculated according to the next formula:

$$LR = (T - 3k - 1) \left(\ln |\Sigma_R| - \ln |\Sigma_r| \right) \sim \chi^2(2k). \quad (4.24)$$

If the observed value is more than the critical one, the null hypothesis should be rejected. As a result, we can include an additional variable into the VAR and provide the analysis for three time series. If there are more than three variables, the testing procedure will be the same.

Earlier we have considered only the case for stationary time series. It is known that standard statistical tests do not work for non-stationary data because such series generate processes with the absence of convergence. Sims, Stock and Watson worked under causality tests for non-stationary data. We will consider two cases. As earlier, we have the next equation for y_t :

$$y_t = c + \sum_{i=1}^k \rho_i y_{t-i} + \sum_{j=1}^m \pi_j z_{t-j} + \varepsilon_t. \quad (4.25)$$

For simplicity and clarity the case with $k = m = 3$ will be considered:

$$y_t = c + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_3 y_{t-3} + \pi_1 z_{t-1} + \pi_2 z_{t-2} + \pi_3 z_{t-3} + \varepsilon_t. \quad (4.26)$$

Let y_t is not stationary in levels, but it is integrated with the order one. It means that the series is stationary in the first differences. The range z_t is integrated with the order zero, so it is stationary. Then we can use standard *t*-tests and *F*-tests to check the null hypothesis about the absence of Granger causality. In fact, we test coefficients for stationary variables, and standard estimation procedures give us true results.

But if we want to test hypotheses for y_t , we should transform the specification a little. Let's add and subtract from the right-hand side an element $\rho_2 y_{t-1}$. We will have the next result:

$$y_t = c + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_2 y_{t-1} - \rho_2 y_{t-1} + \rho_3 y_{t-3} + \pi_1 z_{t-1} + \pi_2 z_{t-2} + \pi_3 z_{t-3} + \varepsilon_t.$$

If to provide all necessary transformations, the equation will get the next view:

$$y_t = c + (\rho_1 + \rho_2) y_{t-1} - \rho_2 (y_{t-1} - y_{t-2}) + \rho_3 y_{t-3} + \pi_1 z_{t-1} + \pi_2 z_{t-2} + \pi_3 z_{t-3} + \varepsilon_t.$$

Let's introduce the next label:

$$\Psi = \rho_1 + \rho_2.$$

Then the equation can be rewritten in the next form:

$$y_t = c + \Psi y_{t-1} - \rho_2 \Delta y_{t-1} + \rho_3 y_{t-3} + \pi_1 z_{t-1} + \pi_2 z_{t-2} + \pi_3 z_{t-3} + \varepsilon_t.$$

As y_t is integrated with the order one, its first difference is stationary. As a result, we can test the coefficient before Δy_{t-1} with the help of standard procedures.

But if z_t is also integrated with the order one (the data are not stationary), we cannot use such approach. As a result, we have to change our model. Let's add and subtract from the right-hand side elements $\rho_2 y_{t-1}$ and $\pi_2 z_{t-1}$:

$$y_t = c + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_2 y_{t-1} - \rho_2 y_{t-1} + \rho_3 y_{t-3} + \pi_1 z_{t-1} + \pi_2 z_{t-2} + \\ + \pi_2 z_{t-1} - \pi_2 z_{t-1} + \pi_3 z_{t-3} + \varepsilon_t.$$

After all transformations we will obtain the next equation:

$$y_t = c + \Psi y_{t-1} - \rho_2 \Delta y_{t-1} + \rho_3 y_{t-3} + \theta z_{t-1} - \pi_2 \Delta z_{t-1} + \pi_3 z_{t-3} + \varepsilon_t, \quad (4.27)$$

$$\theta = \pi_1 + \pi_2.$$

Now it is possible to test restrictions for coefficients because the first differences are stationary. But there is an assumption: it is possible to test Granger causality only if the coefficient θ is equal to zero. This

assumption is connected with non-stationarity of the data. In fact, if we want to test Granger causality, we can state the next null and alternative hypotheses:

$$H_0 : \theta = \pi_2 = 0,$$

$$H_1 : \theta \neq 0, \text{ or } \pi_2 \neq 0, \text{ or both.}$$

Thus, tests for additional lags inclusion can be provided for stationary and non-stationary time series.

Granger causality helps to identify different types of interactions between variables, but there are some disadvantages in the method. For example, all tests work with the assumption that the error term is distributed normally and, as a result, symmetrically. This fact helps to find symmetric interactions, but often we face asymmetry in shocks and data distributions. Granger causality cannot catch this type of interactions, and we lose some information in the analysis. Also the tests do not work properly when analyzing data are driven by another process which is not accounted. Realization of these facts stimulated development of the causality concept, and new models and approaches for testing were appeared.

Thus, Granger causality allows us to identify a type of interaction between macroeconomic variables and decide if it is necessary to include additional lags into the model or not. Often this concept helps to find not obvious interrelations in the data and present their analysis. However, as a rule economic data demonstrate bilateral relations, and the interaction between data series can be quite strong. The concept of Cointegration can help to present mutual dynamics of time series. This part of applied econometrics for macroeconomics will be covered in the next chapter.

Questions

1. What is intervention analysis about?
2. How is it possible to present shocks in the model?

3. What are special features of the VAR model? What forms of this model do you know? Explain the main differences between each other.
4. What problems can appear in the process of system estimation? What method can be applied to estimate the system?
5. Formulate order and rank conditions. Describe the principle of their work. How can they help in the identification of system equations?
6. How will you choose the best specification for the VAR model? Describe two approaches to this problem. Which of them is more powerful and why?
7. What is Granger causality? What types of interaction does it help to identify?
8. What statistics are used to test Granger causality? Provide an example.
9. How can you test the necessity to include an additional variable into the VAR? Describe the idea of the test; formulate testing statistics and rejection rule.
10. Is it possible to test Granger causality for non-stationary data? If yes, what transformations should be provided? Are there any assumptions for this test?
11. Are there any disadvantages of Granger causality approach?

Exercises

1. For this task you need to use three macroeconomic series with quarterly frequency (minimum 100 observations).
 - a) You should choose the data and explain potential short-term and long-term interrelations which can be found here. To confirm your supposition you should use economic theory and your own experience;
 - b) Now you are asked to estimate two VAR models with these variables. Are there significant variables (lags) in every equation of both models?
 - c) You should choose the best model. Provide all necessary calculations and find the best specification.

- d) What does impulse response function show? How can it be derived? Present impulse response functions for your best model and interpret the result. Have your suppositions, made on the first steps, been true or not? How can you explain these results?
 - e) What is the variance decomposition? Present and interpret this decomposition for your best model.
2. To manage this task you need two any macroeconomic series with yearly frequency (minimum 50 observations).
- a) Present data dynamics graphically. What do you think about the direction of interactions between these variables?
 - b) Test the data for stationarity. If they are not stationary in levels, provide all necessary transformations;
 - c) Test Granger causality in two directions using facts about stationarity / non-stationarity. If your data are not stationary, do you have to provide any modifications in the model to test the hypotheses about causality? Present models with and without restrictions, test statistics and rejection rules. Make your conclusions about causality and its direction in the data.

Chapter 5

COINTEGRATION AND ERROR CORRECTION MODELS

5.1. Cointegrated Time Series

As we know not all time series are stationary. For example, they may be integrated with different orders. Usually if we two time series integrated with the same order, their linear combination will have this order too. The reason is that long-term dynamics of such series has particular patterns, and the combination will take these ones from both series and save their properties. But sometimes such combination can be integrated with the less order than initial time series. In this case we have so-called cointegration between two ranges.

This concept was suggested by Granger and then developed by Nelson, Plosser, Engle, Johansen and other researchers. In terms of systems cointegration shows the existence of long-run equilibrium. We can find enough examples of cointegrated time series in any economy because this system has sustainable relationships and equilibrium. Cointegrated time series demonstrate similar dynamics. Let's consider the next example. We have quarterly data from 1947 to 2010. There are time series for consumption per capita and GDP per capita for the U. S. economy. We present the data on a single graph. The result is reflected on the Fig. 5.1.

These series have the same dynamics: there are identical peaks and recession, similar exponential trend. Formally we also know that consumption and income (GDP per capita) have direct dependence: if GDP per capita grows, individual consumption will increase too. It is a typical example of cointegrated time series.

Now it is possible to present the definition of cointegration in formulas. Let's there are two time series: y_t and x_t . These data are integrated with the order one. Indicated time series will be cointegrated if

and only if their linear combination, z_p , is integrated with the order zero. Formally the next note will be done:

$$y_t \sim I(1), x_t \sim I(1),$$

$$z_t = \gamma_1 y_t + \gamma_2 x_t \sim I(0) .$$

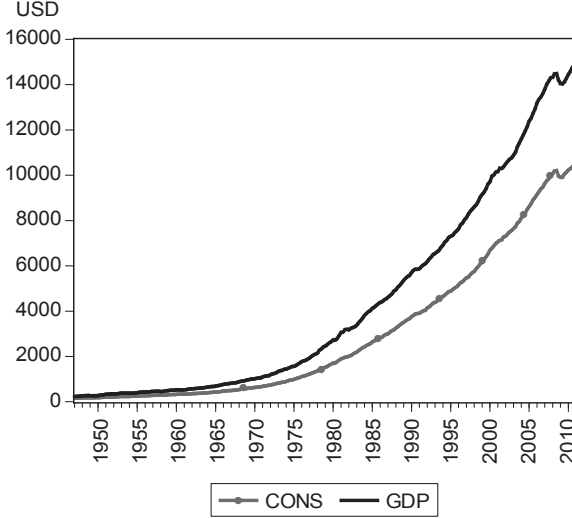


Fig. 5.1. Dynamics for the U. S. consumption and GDP per capita (quarterly data, 1947–2010)

Initial series have to be integrated with the same order. In opposite case it is impossible to speak about cointegration because a linear combination will inherit the highest order of the “parents”. Formally differencing is a linear combination of time series with the same order of integration. The result is cointegrated time series with the less order of integration:

$$y_t \sim I(1), y_{t-1} \sim I(1),$$

$$\Delta y_t = \gamma_1 y_t + \gamma_2 y_{t-1} \sim I(0), \gamma_1 = 1, \gamma_2 = -1 .$$

For our example we have $CI(1,1)$. The first number reflects the order of integration from initial series. The second one means the difference between orders of old series and a linear combination. In the example our series were integrated with the order one, and a new range had the order zero. It is not difficult to calculate that the difference is equal to one. Using these numbers we can present any cointegration in a short record.

Also there is a term “cointegration vector”, or CV . This vector is constructed with the help of coefficients from a linear combination:

$$CV = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}.$$

Often it is assumed that these coefficients are stable in time. In fact, this statement is not always satisfied. Also it is important to notice that there is a set of cointegration vectors. For example, we can multiple our CV on any positive number and the vector will change. To avoid this plurality it is a common step to normalize the first coefficient to unity. Then we have:

$$y_t \sim I(1), x_t \sim I(1),$$

$$z_t = y_t + \gamma_2 x_t \sim I(0),$$

$$CV = \begin{pmatrix} 1 \\ \gamma_2 \end{pmatrix}.$$

Earlier we have seen that cointegrated time series demonstrate the same dynamics (for example, in trend components). Now we can illustrate this fact using general facts about cointegration and time series decomposition*. In the first chapter we said that time series included deterministic and irregular parts. Then we can present our series,

* For more details you can follow: *Dolado J. J., Gonzalo J., Marmol F.* Lecture notes for Cointegration [Electronic resource]. URL: <http://www.eco.uc3m.es/~jgonzalo/cointegration.pdf>

integrated with the order one, as a linear combination of trend (T) and any stochastic process (ε). Formally it will look like:

$$y_t = T_{y,t} + \varepsilon_{y,t},$$

$$x_t = T_{x,t} + \varepsilon_{x,t},$$

$$y_t, x_t, T_{y,t}, T_{x,t} \sim I(1), \varepsilon_{y,t}, \varepsilon_{x,t} \sim I(0).$$

There are no any violations in assumptions about initial orders of integration. For instance, a linear combination of $I(1)$ and $I(0)$ always will be integrated with the order one. If initial series are cointegrated, the next fact takes place:

$$z_t = y_t + \gamma_2 x_t \sim I(0).$$

If we replace initial series with their decomposition, we will have the next result:

$$z_t = T_{y,t} + \varepsilon_{y,t} + \gamma_2 (T_{x,t} + \varepsilon_{x,t}) = T_{y,t} + \varepsilon_{y,t} + \gamma_2 T_{x,t} + \gamma_2 \varepsilon_{x,t} \sim I(0).$$

If a series is stationary, there must not be trend. It implies:

$$T_{y,t} + \gamma_2 T_{x,t} = 0,$$

$$T_{y,t} = -\gamma_2 T_{x,t}.$$

It means that cointegrated time series have to share the same trend dynamics up to a scalar. Of course, in general case we can find a set of such trends, but normalized coefficients allow us to determine this component in a single form.

Cointegration can be observed not only for two variables. There is a term multicointegration for this case. If the system includes more than two variables, it is possible to take series with different orders of integration. The reason is that such series can be combined in different order. For instance, we have two series with the order three, one range with the order two and the last series with the order one. Initially we will combine the first two series with the order three and obtain a new

range with $I(2)$. On the next step it is possible to create a combination from the new series and our initial range with the order two. As a result, we will get another series with $I(1)$. On the last step two $I(1)$ series will be mixed.

We can observe cointegration for series with the order higher than one. And now we will formulate a common definition for cointegration:

A vector of n time series, integrated with the order d , will be $CI(d, b)$ if their linear combination exists and integrated with the order $(d - b)$.

Formally we can write the next:

$$y_{1t} \sim I(d), \dots, y_{nt} \sim I(d) \forall i = 1, \dots, n, \quad (5.1)$$

$$z_t = \gamma_1 y_{1t} + \dots + \gamma_n y_{nt} \sim I(d - b),$$

$$CV = \begin{pmatrix} \gamma_1 \\ \dots \\ \gamma_n \end{pmatrix}.$$

Usually series integrated with high orders are not used for researches. It is quite difficult to interpret these data. Moreover, most economic series have the order not higher than two. Opposite cases take place rarely. But if it happens, you should think about interpretation of your future results or choose another approach to analyze the data*.

As it was said above, cointegration gives a signal about the existence of long-term equilibrium. In particular, the linear combination characterizes deviations from the long-term mean. If to consider an example where this variable is integrated with the order zero, we can say about stationarity of the linear combination. The long-term mean

* This case can appear in the data with high frequency where high orders are required for achieving stationarity. Also it is possible that such data have not a single unit root (this question was discussed in the second chapter). If it is true, the standard concept of cointegration will not work properly here. The reason is that it is built on the assumption about a single unit root. In opposite case different modifications and other procedures for identification should be applied.

should be equal to zero in this case. To illustrate this fact, we will use formulas and the case of two variables:

$$z_t = y_t + \gamma_2 x_t \sim I(0),$$

$$z_t = CV^T(y_t, x_t).$$

Long-term equilibrium; $z_t = 0$, as a result: $E(CV^T(y_t, x_t)) = 0$.

Using these facts, Granger formulated the next theorem (so-called representative theorem): if time series are cointegrated, they can be presented with the error correction model. The inverse theorem sounds as: if the error correction model exists, time series are cointegrated. This theorem says that for cointegrated time series it is possible to construct the model with the long-term mean equal to zero. In other words, this model allows us to correct a short-term dynamics for correspondence with long-term values. For this purpose cointegration vector and special weights are used. In fact, the error correction model (ECM) is a case of the VAR model which was studied above. In the next paragraph we will cover main feature of the ECM.

5.2. The Error Correction Model

The ECM is a case of the VAR model, but it helps to account the existence of cointegration between different time series. In fact, this model corrects short-run values and forecasts to provide the convergence to long-run equilibrium. Let's give a formal representation of the ECM.

As earlier, we will consider a case with two variables, integrated with the order one. Our time series are cointegrated, and their linear combination can be found according to the next formula:

$$z_t = y_t + \gamma_2 x_t \sim I(0).$$

Then the ECM will have the next representation:

$$\begin{cases} \Delta y_t = c_{10} + \beta_1 z_{t-1} + \sum_{i=1}^k \phi_i y_{t-i} + \sum_{j=1}^k \omega_j x_{t-j} + \varepsilon_{1t}, \\ \Delta x_t = c_{20} + \beta_2 z_{t-1} + \sum_{i=1}^k \phi_i y_{t-i} + \sum_{j=1}^k \omega_j x_{t-j} + \varepsilon_{2t}. \end{cases} \quad (5.2)$$

It is assumed that there is no autocorrelation in error terms of the model. If we rewrite the ECM and change z_p , we will see that β 's present correcting weights:

$$\begin{cases} \Delta y_t = c_{10} + \beta_1 y_{t-1} + \gamma_2 \beta_1 x_{t-1} + \sum_{i=1}^k \phi_i \Delta y_{t-i} + \sum_{j=1}^k \omega_j \Delta x_{t-j} + \varepsilon_{1t}, \\ \Delta x_t = c_{20} + \beta_2 y_{t-1} + \gamma_2 \beta_2 x_{t-1} + \sum_{i=1}^k \phi_i \Delta y_{t-i} + \sum_{j=1}^k \omega_j \Delta x_{t-j} + \varepsilon_{2t}. \end{cases} \quad (5.3)$$

So, these values correct coefficients of cointegration vector, and it helps to satisfy the condition about zero long-term mean. The model characterizes transitional dynamics to the long-run equilibrium. Also it was said that a linear combination of stationary time series showed deviations from this long-term mean. Moreover, the combination can be considered as residuals (an error term) from another regression. In fact, using the ECM we correct these residuals to satisfy the long-run mean.

We cannot apply the VAR model in its standard form because in case of cointegrated time series error terms are correlated with our regressors. It leads to appearance of endogeneity problem. To prevent this we should include the error term in the model as an explanatory variable. Then we will have ECM with consistent and unbiased estimators (if other conditions for a new error term are not violated).

If time series are cointegrated, at least one β coefficient must be significant. It is connected with causality which takes place in the case of cointegrated time series. A significant coefficient can indicate the direction of interaction between time series in dynamics.

Also it is possible to present the ECM in matrixes. Let's there is a vector x_t which includes n time series. There are p time series from this vector which are cointegrated. Then the ECM will be written as following:

$$\Delta y_t = c + \beta \gamma' y_{t-1} + \sum_{i=1}^k \phi_i \Delta y_{t-i} + \varepsilon_t, \quad (5.4)$$

$$y_t = (y_{1t} \dots y_{nt})',$$

$$\beta = (\beta_1 \dots \beta_n)',$$

$$\gamma = \begin{pmatrix} \gamma_{11} & \dots & \gamma_{1p} \\ \dots & \dots & \dots \\ \gamma_{n1} & \dots & \gamma_{np} \end{pmatrix},$$

$$\varepsilon_t = (\varepsilon_{1t} \dots \varepsilon_{nt}),$$

ϕ_i is a matrix.

Time series are not always cointegrated. Different possibilities can be presented here. For example, we have two time series. They can be integrated with the same order or not. Let's consider the first case. If our series have the same order of integration they can be cointegrated or not. In case of cointegrated series we have to estimate the ECM. If the data are not cointegrated, we should test Granger causality and try to reveal some patterns in these time series. In the second case when the data are not integrated with the same order, we can estimate the VAR model in levels and try to find interrelations here.

There are different ways how cointegration can be identified. However, two approaches are the most popular. The first one is Engle — Granger approach. The second way was suggested by Johansen on the basis of the ECM analysis. Both these approaches will be studied in the next paragraph.

5.3. Methods for Identification of Cointegration

As it was marked in the previous paragraph, there are two main methods to identify cointegration. The first one is the Engle — Granger methodology. This approach is quite simple and representative. It works properly in the case of two cointegrated variables, but it is not convenient if our vector includes more than two time series. Let's consider this approach more attentive.

We will consider an example with two time series as it has been done above. The procedure consists of several steps:

1. Initially we have to check the order of integration for our time series. To provide this step we can use different unit root tests. But the best results will be obtained with the Dickey — Fuller test and the KPSS test. As you remember, the Phillips — Perron test corrects statistics for heteroskedasticity and autocorrelation, but there is no necessity to provide this procedure for cointegrated series. All necessary lags and components should be included into the test specification.

Using these test inferences about the order of integration can be done. If time series are integrated with the same order, we can continue the identification. In opposite case we should use other methods to test interrelations (for example, the VAR model in levels);

2. Now it is possible to estimate long-run equilibrium in the model. For this purpose the standard OLS method is used. The next specification is suggested:

$$y_t = c + \gamma x_t + \varepsilon_t.$$

Time series are integrated with the same order. Intercept presents a deterministic component in the model because for non-stationary time series this part presents in decomposition. Also it is possible to include trend in the model. The linear combination can be calculated:

$$\hat{\varepsilon}_t = y_t - \hat{c} - \hat{\gamma}x_t.$$

3. As our time series are non-stationary, we cannot use standard statistics for checking the significance of coefficients. But in 1987 Stock showed that for cointegrated data OLS estimators were consistent. Also

he found that these estimators would converge to real values with higher speed. For this fact the term super consistency was introduced*. However, there is a problem connected with the identification of cointegration. So, we can obtain false cointegration which has the next features:

$$t_{\hat{\gamma}} \rightarrow \infty,$$

$$DW \rightarrow 0 \text{ if } t \rightarrow \infty.$$

To check if the regression is false or not, the cointegrating regression Durbin — Watson test can be applied. Null and alternative hypotheses have the next form:

$$H_0 : DW = 0,$$

$$H_1 : DW > 0.$$

Usually the Durbin — Watson test shows the presence of autocorrelation in residuals. The null hypothesis shows an absence of cointegration in the model (our linear combination has a unit root). In opposite, the alternative hypothesis gives a signal about a presence of cointegration (the linear combination is stationary).

Special critical values are used for this test. To find a particular value, we should know a number of potential cointegrated variables and a number of observations. If the value of Durbin — Watson statistics is more than critical one, the null hypothesis should be rejected. Critical values for the test can be found in the Appendix, Tab. 6.

But this test has one significant disadvantage. It works properly only if the linear combination can be presented as a standard AR(1) process. In other cases the test will not give us true results. The evidence is that often linear combinations are driven by more difficult and complicated processes. This limitation of the test does not allow us to use it for all possible cases.

* It works for large samples. If the sample is not enough big, super consistency will not take place.

Another test, which can be used for checking the order of integration in the linear combination, is the Davidson — MacKinnon test. In principle, it is the Augmented Dickey — Fuller test for the model without trend. But there is one modification. The reason is that we use estimated parameters to calculate the linear combination. It gives us additional variance in the variable. A standard Dickey — Fuller test does not account this fact. As a result, we have to use other critical values to provide the analysis. They are called MacKinnon critical values (see App., Tab. 7) and calculated for the case where additional variance takes place. Formally we estimate the next model for unit root testing:

$$\Delta \varepsilon_t = \alpha + \mu \varepsilon_{t-1} + \sum_{i=1}^k \delta_i \varepsilon_{t-i} + u_t.$$

For this model we calculate Dickey — Fuller statistics and compare it with the MacKinnon critical value. The table with these values is also included into the Appendix. Null and alternative hypotheses are the same as for the previous test. If the observed value is less than critical one (there is a left-side critical region), we have to reject the null hypothesis about an absence of cointegration. It means that time series are cointegrated.

4. If time series are cointegrated, we construct the ECM to correct the long-run mean:

$$\begin{cases} \Delta y_t = c_{10} + \beta_1 \varepsilon_{t-1} + \sum_{i=1}^k \varphi_i y_{t-i} + \sum_{j=1}^k \omega_j x_{t-j} + u_{1t}, \\ \Delta x_t = c_{20} + \beta_2 \varepsilon_{t-1} + \sum_{i=1}^k \varphi_i y_{t-i} + \sum_{j=1}^k \omega_j x_{t-j} + u_{2t}. \end{cases} \quad (5.5)$$

At least one coefficient β must be significant. The best specification can be chosen with the help of information criteria (they have to be minimal).

If time series are not cointegrated, we can test Granger causality and try to determine the direction of relationships in the system.

This method works well if we have only two time series. In practice we analyze more complicated systems and relationships. As a result, we

have to work with more than two variables. In this case it is convenient to use Johansen's methodology for identification of cointegration*.

The Johansen test gives estimators with good statistical properties, but it is built on the assumption about normal distribution of error terms. In fact, this statement can be violated. However, this test is more powerful than the Engle — Granger methodology. In particular, it does not require a single cointegration vector. Now we will study the main principles of the Johansen test.

Earlier we presented the ECM for the vector of n variables in matrixes:

$$\Delta y_t = c + \beta \gamma' y_{t-1} + \sum_{i=1}^k \varphi_i \Delta y_{t-i} + \varepsilon_t. \quad (5.6)$$

In fact, it is the VAR model in reduced form with the inclusion of our linear combination $\gamma' y_{t-1}$. Let's assume that all variables, presented in y_p , are integrated with the order m . We do not know how many cointegrated variables are here. It is possible that no one linear combination generates stationary time series. In opposite, all n ranges can be cointegrated with each other. We need a method to test the number of cointegrated time series (so-called cointegration restrictions).

Let's look at our ECM presented above. In fact, an element $\beta \gamma'$ is a matrix (a product of two vectors one of which is transposed) with n rows. The number of columns depends on the quantity of cointegrated time series. Let's say that there are r cointegrated series. Then our matrix will have the dimension $n \cdot r$.

What does this matrix present? It includes coefficients which determine our linear combinations. For example, if all variables are not cointegrated, there are only zero values in this matrix. In other words, our linear combinations are not integrated with the less order. But it is

* In more details you can see this concept here: *Dolado J. J., Gonzalo J., Marmol F.* Lecture notes for Cointegration [Electronic resource]. URL: <http://www.eco.uc3m.es/~jgonzalo/cointegration.pdf>; *Bo Sjö.* Testing for Unit Roots and Cointegration [Electronic resource]. URL: <http://www.iei.liu.se/nek/ekonometrisk-teori-7-5-hp-730a07/labbar/1.233753/dfdistrib7b.pdf>

possible that some coefficients are significant (not equal to zero), and then we can speak about the existence of cointegration in the model.

Actually we should find a rank of this matrix (a number of linearly independent rows and columns). To find the rank is equivalent to determine the number of significant eigenvalues. There are three possible combinations:

$rank(\beta\gamma') = r = 0$. It means that there are no cointegrated time series in the model. No one combination can generate a stationary range;

$rank(\beta\gamma') = r = n$. In this case all variables are cointegrated;

$0 < rank(\beta\gamma') = r < n$. There are r cointegrated time series in the model (r independent rows and columns).

In principle, the test works “down-top”. For example, on the first step the next null hypothesis can be formulated: there is no cointegration ($r = 0$). In opposite, the alternative hypothesis will sound like: there is at least one cointegration equation. If we reject the null hypothesis on this step, we can continue the analysis. Then the null hypothesis will be about the existence of a single cointegration relation. The alternative one will state that there is more than one cointegration equation. The test will continue until the null hypothesis is accepted.

In the first case it is possible to present the model in differences because no one combination generates stationary time series in levels. For the third case the situation can be the next. We may have the result that $r < n$ variables are cointegrated. But it does not mean that all r series are necessary to construct linear cointegrated with lower order of integration. These variables can generate necessary ranges in different combinations in each other. In such case it is useful to provide additional test and check if any variables can be excluded from the combination or not.

As the test is built on the assumption about normal distribution about error terms, we have to test residuals for autocorrelation and heteroskedasticity. In other words, the test is sensible to specification errors. As a result, we should be careful using this methodology. Also this test works properly on large samples (it is true for all parametric tests).

The Johansen test shows a number of cointegration equations, but it does not give us an ability to know what variables take part in

cointegration equation and how cointegration vector looks like. To determine a particular specification we should provide additional estimation procedures.

The test is based on a trace concept, and often it is called trace test. There are special critical values for the Johansen test (it is possible to include trend, and then you should use other critical values). So, this approach is convenient for the analysis of more than two time series and cointegration relationships.

Thus, we have studied the concept of cointegration and main methods how it can be caught. This concept has a wide application in macroeconomics and finance. It helps to analyze mutual dynamics and present equilibrium trajectories. Also it is possible to find different types of interactions between variables. The ECM can be used for long-term forecasting.

Now we have covered all basic concepts of applied econometrics for macroeconomics. The main purpose of all modern researches is to create a quality forecast. And below we will consider how these predictions can be made and their quality estimated.

Questions

1. What is cointegration? Present the concept briefly using graphical examples and formulas.
2. How can this concept be applied for economic theory? What type of interactions does it describe?
3. Show that cointegrated time series have similar trends up to a scalar. Use graphical examples and formulas to illustrate this fact.
4. What is the error correction model? How can it be applied for cointegrated time series? Describe the main idea of the model from the position of economic theory.
5. How can cointegration be identified in time series? List all approaches for testing cointegration.
6. Describe the Engle — Granger methodology. What advantages does this approach have?

7. Describe the idea of the Johansen test. When is it preferable to use this test for cointegration? List its advantages and disadvantages in the context of estimating techniques.

Exercises

For this exercise you have to use two pairs of time series. One pair should be cointegrated potentially, and the second couple should not. It is recommended to use quarterly data (if there is seasonality, you should remove it from time series). The minimal number of observations is 100.

- a) Illustrate your data with graphs. Can you provide any evidence from economic theory why your data may be cointegrated or not? You can use formulas and general theoretical statements.
- b) Check the order of integration for your time series. Is it possible to combine your series and check cointegration relationships? If your data are not cointegrated potentially, what kind of models can be used to reveal connections in the data?
- c) Check your first pair of time series for cointegration with the help of the Engel — Granger methodology. What test will you use? What assumptions can be done for testing procedure? Formulate null and alternative hypotheses, test statistics and rejection rules. What critical values have to be used for the test? Are your data cointegrated or not?
- d) Repeat the previous step for the second pair of time series. Are these data cointegrated or not?
- e) If your data are cointegrated, present the ECM for them. What can you say about the model? Choose the best specification.
- f) If your data are not cointegrated, provide Granger causality tests. What type of interactions can you identify here?
- g) Test cointegration with the help of the Johansen methodology. Do your results differ from previous steps?
- h) How do your inferences correspond with economic theory and your initial suppositions about cointegration? Explain your answer.

Chapter 6

FORECASTING

6.1. Main Principles of Forecasting

Earlier we have studied different types of econometric models and indicated main problems connected with their estimation. In the beginning of the book it was stated that the main purpose of applied econometrics for macroeconomics was to estimate dynamic processes and create forecasts. Now you know all basic methods of time series analysis and can use them for forecasting. It is necessary to present some methods to estimate the quality of forecasts and choose the best one.

In general, forecasts are built with the help of estimators which can be obtained during analysis of macroeconomic time series. There are different models, and not all of them can give us proper results. Thus, we should know some criteria which help us to identify the quality of any forecast.

It is known that we work with samples in econometrics. The sample includes only a part of possible data. Using this information we can provide various estimation procedures and obtain results. Also it is possible to calculate an error term which always exists in the model because the estimators are equal to their true values asymptotically (we use finite samples to predict necessary parameters, and obtained values can differ from true ones). Such estimation can be made for any sample.

But then we want to find future values, and there are no observations yet. In this case it is possible to use estimators, found for known samples, and calculate necessary data with the help of previous information. In other words, current and past observation will take part in constructing future values. To provide this calculation we should know estimators. And the quality of estimators will determine an accuracy of the forecast. That is why we always want to obtain consistent, unbiased and efficient estimators of unknown parameters.

So, for every sample we can obtain two forecasts: for current and previous data and for future values. The first one assumes the calculation of an error term. For the second part of our forecast we cannot find this term because we do not have a sample for future periods. In this case the forecast is built on previous values, and it is quite important to use all considered models (ARMA, ARCH/GARCH, VAR, cointegration analysis etc.) to make a good prophecy.

In principle, forecasting includes the next steps:

1. Different models are found and estimated for the given sample;
2. The error terms are calculated for all models estimated on the first step;
3. The best forecast (or a combination of different forecasts) is chosen;
4. Future values of the dependent variable are calculated.

If first two steps are familiar, the third one is not obvious. What indicators can show the forecast's quality? Now we will consider all possible criteria for choosing the best prediction.

The first one is a standard coefficient of determination and its different modifications. This number reflects the percent of explained variation in the data. For linear regression model with intercept this coefficient will look as follow*:

$$R^2 = \frac{SSE}{SST} = \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - \bar{y})^2}. \quad (6.1)$$

In this representation the value of the coefficient cannot be less than zero and not higher than unity. If the coefficient is closer to unity, the model approximates the sample well. As a result, we can generate good forecasts with the help of this model.

* There are different modifications of this coefficient. In particular, some of them allow negative values for the coefficient (it can be if squares of residuals do not take part in the calculation).

The next possible values of the coefficient can be presented:

- $R^2 = 0$;
- $0 < R^2 < 1$;
- $R^2 = 1$.

The first case assumes that all real values belong to the predicted regression line. It seems that this case is desirable. However, it is not a good sign in the estimation process. Often such values of the coefficient are connected with the presence of multicollinearity in the model. It is reasonable that not all real data will place on the regression line because there are shocks. Usually such values of the coefficient take place when independent variables are strongly correlated with each other. As a result, we have multicollinearity, and it leads to inconsistent and biased estimators. It is a signal that we have to change specification of the model or use other methods for estimation.

The second case, when $0 < R^2 < 1$, is usual. If the coefficient is close to unity, we can say about a good quality of the estimation. When this value is near to zero, it indicates that the chosen specification does not give proper results. To improve the quality it can be necessary to change the specification (include non-linear terms, lags etc.).

The third case when the coefficient is equal to zero says about bad results of the estimation. But there can be different reasons for this. The first one is that dependent and explanatory variables are not connected with each other. It means that we have to find other regressors and re-estimate the model.

The second reason is an error in the specification. For example, we consider linear relationships, but actually they are non-linear (quadratic etc.). In this case it is possible to change specification. If estimation results do not improve, you should think about causality and potential regressors in the model.

And the third case is possible if there are no correlation relationships between variables. But it does not mean that there are no any other interactions. In this case it is useful to analyze data more precisely and try other estimation procedures. For example, it is possible to construct

a theoretical model to reveal actual relationships between variables. Then using your quality analysis, it is possible to test found relationships with econometric tools.

One more disadvantage of this coefficient is connected with the number of regressors. So, the coefficient will grow when more variables are included in the right-hand side of the regression equation. It means that we cannot compare models with different number of regressors using this coefficient.

Also it is impossible to compare models with different specifications using such approach. For example, we cannot choose the best model with the help of the coefficient when we have linear and logarithmic representations. To provide such analysis we should use special procedures (for example, Box—Cox methodology). Thus, the coefficient of determination may be a good instrument for estimation of the forecast's quality. But it cannot be used to compare models with different specifications. As a result, we should use other indicators to determine the best forecast among all feasible opportunities.

There are other values which can help us to find the best forecast. All of them are built on the calculation of errors in different models. If we decide to find a mean error, we will get a zero value. To avoid such case it is necessary to provide some transformations with the error term. The first such indicator is a mean absolute deviation (*MAD*). It reflects a mean absolute error of the forecast and can be calculated according to the next formula:

$$MAD = \frac{\sum_{t=1}^T |y_t - \hat{y}_t|}{T}. \quad (6.2)$$

For the best forecast this value should be minimal.

The second indicator is a mean squared error (*MSE*) of the forecast. In fact, it is a variance of the error term in the model. This value calculates according to the next formula:

$$MSE = \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{T}. \quad (6.3)$$

Visually it is not the same as the variance because we do not subtract the mean. But the main assumption about the error term in econometric models is that its mean should be equal to zero. In fact, there are no any contradictions, and this value can be considered as a variance of the error term. The best forecast will have the least variance in the error term.

The third value, which helps to estimate the quality of made forecasts, is a mean absolute percentage error. It reflects a mean error ratio in initial data. This value can be found as follow:

$$MAPE = \frac{\sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{y_t}}{T}. \quad (6.4)$$

The forecast is good if this value is more than zero and less then 0,1. If you have several forecasts with satisfactory values of *MAPE*, you should choose the model with the minimal value.

The last indicator is a mean percentage error. It has the next formal representation:

$$MPE = \frac{\sum_{t=1}^T \frac{y_t - \hat{y}_t}{y_t}}{T}. \quad (6.5)$$

This value reflects a bias in the forecast. It can show if predicted values are higher or lower than actual numbers on the average. You can choose the forecast with the least bias.

In general, the best forecast can have the least values for all four indicators. But it is possible that these indicators will demonstrate opposite dynamics for different forecasts. In this case it is recommended to take the forecast with less *MSE* and *MPE*. Then it will be possible to predict future values of analyzing time series.

Also it is possible to use the *F*-test to compare different forecasts. For example, we want to know if our forecasts are equivalent asymptotically (it is possible) or not. This test is available only for two models. If you have more than two forecasts, you should construct all possible binary relations.

To provide the test we should calculate MSE for both forecasts. Then null and alternative hypotheses will be:

$$H_0 : MSE_1 = MSE_2 \text{ (forecasts are asymptotically equivalent);}$$

$$H_1 : MSE_1 \neq MSE_2 \text{ (forecasts are not asymptotically equivalent).}$$

Test statistics can be presented in the next form:

$$F = \frac{MSE_1}{MSE_2} = \frac{\frac{\sum_{t=1}^T (y_t - \hat{y}_{1t})^2}{T}}{\frac{\sum_{t=1}^T (y_t - \hat{y}_{2t})^2}{T}} \sim F(T, T). \quad (6.6)$$

Using this statistics we can reformulate null and alternative hypothesis:

$$H_0 : F = 1,$$

$$H_1 : F \neq 1.$$

If the observed value is more than critical one, we have to reject the null hypothesis. In this case we should choose the model with less MSE and forecast future values with its help.

All presented values can be calculated only for forecasts which were constructed with the help of the sample. If we find future values, we cannot estimate any statistical numbers because we do not know real data. But it is possible to check the forecast's quality post-factum and do necessary corrections to improve its quality.

Thus, there are different indicators which help to choose the best forecast. Now we will consider advantages and disadvantages of forecasts which can be made with the help of models presented earlier.

6.2. Forecasting With the Help of Different Models

Earlier we have studied different models which can be applied for time series analysis. And now it is necessary to understand which of them can be used for forecasting successfully and which cannot.

Let's start from different regular components. If we have data with high frequency, we face with seasonality problem. It is known that this component can be removed from the data, and different methods can be used for this purpose. For example, we can use seasonal dummy variables or moving averages to smooth the data. If we speak about forecasting, it is quite important to remove seasonality before we do any estimation. The reason is seasonality worsens estimation properties. As a result, we can obtain biased and inconsistent forecasts. So, it is necessary to remove seasonality on the first step of forecasting.

After this correction we can do forecasts with the help of trend dynamics. After removing seasonality we have “clean” trends. There can be different types of trend dynamics. For example, it is possible to face linear, quadratic, cubic, logarithmic, exponential and other trends. We can consider a quadratic trend. It can be presented that at a particular moment of time a variable has a minimal value. As a result, such trend will look like parabola:

$$y_t = \alpha + \beta t + \gamma t^2 + u_t. \quad (6.7)$$

To find parameters, we should create a variable which will reflect numbers of time periods. We can estimate this equation with the OLS and predict values for y_t :

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}t + \hat{\gamma}t^2. \quad (6.8)$$

If we do not know a particular trend function, we can estimate several models and choose the best one with the help of special indicators. But if we want to compare predicted values with initial data and if our forecast should include seasonality, we have to come back this deterministic component into the model after the estimation. In fact, we will do forecast using the trend equation, without seasonality and obtain good estimators, and then seasonal component will be accounted via inverse transformation. As a result, we will have predicted data which describe actual dynamics. This approach is quite simple, but it does not give good results. In fact, any trend is a mean value, and in most cases our predictions will not correspond with actual dynamics. However,

forecasting with the trend can help to present a common dynamics of the system and do some inferences in the first approximation.

Also we can use ARMA models for forecasting. This kind of models is quite convenient for predictions. To calculate future values we should know only previous values of the dependent variable and the error term. No additional data are required. Let's consider an example how a forecast function can be derived for an AR(1) process. Formally a regression equation has the next view:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t. \quad (6.9)$$

Then a future value of y_t will be presented as follow:

$$y_{t+1} = \alpha_0 + \alpha_1 y_t + \varepsilon_{t+1}. \quad (6.10)$$

An expected value for this variable will be:

$$Ey_{t+1} = \alpha_0 + \alpha_1 y_t = \hat{y}_{t+1}. \quad (6.11)$$

Then we can predict the next value:

$$Ey_{t+2} = \alpha_0 + \alpha_1 Ey_{t+1} = \alpha_0 + \alpha_1 (\alpha_0 + \alpha_1 y_t) = \hat{y}_{t+2}. \quad (6.12)$$

These calculations can be continued. Using presented results, we can write the forecast function for the AR(1) process and the moment of time k :

$$Ey_{t+k} = \alpha_0 (1 + \alpha_1 + \dots + \alpha_1^{k-1}) + \alpha_1^k y_t. \quad (6.13)$$

Using this function for any moment of time we can predict y_t . To do it we should know only coefficients (they can be obtained from the model) and a single value of y_t . It can be noticed that more complicated structures can be used to construct this function. As a result, it can differ from our presentation, but the main principle will be kept: to predict future values you should know only previous ones and coefficient from the estimated model.

You know the process how the best ARMA model can be chosen. This forecast is more accurate than trend functions, but it is also not so good. These models can be used for short-run forecasts, but they are weak if we want to predict long-run interactions and dynamic

processes. This kind of models works with one variable and does not account dynamics of other macroeconomic series. For example, if we have interventions and breaks in other variables, a standard ARMA model will not allow us to reflect this fact in the forecast. Potentially these shocks can be included into the error term, but it is not possible here.

ARCH/GARCH models can be used for forecasting data with high frequency. But as a rule, these data cannot be predicted accurately. We can follow common tendencies and pattern here, but all particular values will not be precise. However, if ARCH/GARCH processes take place in your ARMA model, they have to be accounted in the model. In opposite case estimators can have bad properties, and any short-run forecast, created with the help of such model, will be inconsistent. Thus, ARMA models can be used for short-run forecasts, and then they have to be revised and corrected for constructing new predictions.

Better forecasts can be obtained with the VAR model. The main advantage of this estimation mechanism is a system approach. It allows us to reveal mutual dynamics of different variables and create more accurate forecasts. If the VAR was estimated correctly, it is possible to follow all simultaneous changes in variables and present consequences of potential exogenous shocks (impulse response functions and variance decompositions can be used here). Moreover, it is possible to model different scenarios of various changes in variables forming the VAR. This model can be used for long-term forecasts.

In the case of cointegrated time series the best way will be to estimate the ECM. As it was discussed above, the ECM corrects short-run dynamics and describes long-term equilibrium. This model is the best candidate for forecasting in the case of long-term relations.

But the problem is that the ECM can be presented only for cointegrated time series. Moreover, if we have more than two cointegrated variables, we often know a little about cointegration vector and particular interactions in the model (this question was discussed when we studied the Johansen test). It means that the ECM is not always feasible, and in this case we should use the VAR for long-term forecasting and ARMA models for short-run prophecies.

Also it is possible to use causality relationships to create some forecasts, but in general this case will be equal to the VAR (if there are causality relationships) or the ARMA (if there is no causality).

Cointegrated time series have the same trend dynamics, and this fact can be used for general forecasts. For example, if we have enough information about one of cointegrated time series (it can be information which is not reflected in statistical data directly), we can predict its dynamics and forecast possible values for another series.

Thus, different models can be used for forecasting. But sometimes we have several quality predictions. In this case it is possible to construct combined forecasts which will be discussed in the next paragraph.

6.3. Combined Forecasts

Often it is impossible to find a single model which would describe the data perfectly well. But we need forecasts with maximal accuracy. Sometimes it is useful to integrate all proper forecasts and create a precise prediction. Below we will consider how it can be done in practice.

In general, a process of combining different forecasts can be presented in several steps:

1. On this step we should estimate several models for a particular sample;
2. We predict values of a dependent variable using models and estimators obtained on the first step;
3. Then we are to determine weights for every forecast in the final prediction with the help of fitted values and initial observations;
4. Now it is possible to calculate future values for a period where no sample is.

Now we will consider different approaches how such forecast can be found. The simplest way is to calculate an arithmetic mean for all found forecasts. It means that weights will be equal for all forecasts. But actually these predictions can enter with different shares into the final forecast. We will study cases when these weights can be not necessary equal.

The first method is quite simple, but it requires strict assumptions. Let we have a variable y_t and n forecasts for it (all forecasts will be denoted with the letter f). We want to construct a single forecast, and it can be presented as a linear combination of all existed predictions:

$$f_t = \varphi_1 f_{1t} + \dots + \varphi_n f_{nt}. \quad (6.14)$$

Then our variable can be written as follow:

$$y_t = f_t + \varepsilon_t. \quad (6.15)$$

As our final forecast is a linear combination (coefficients present weights in the model), the next assumption can be done:

$$\varphi_1 + \dots + \varphi_n = 1. \quad (6.16)$$

In fact, we do not know our final forecast. As a result, we cannot find these weights without additional transformations of the model. Using the assumption stated above we can provide the next transformation:

$$\varphi_n = 1 - \varphi_1 - \dots - \varphi_{n-1}. \quad (6.17)$$

Then we can put this expression into our equation for the final forecast and for the variable:

$$y_t = \varphi_1 f_{1t} + \dots + \varphi_{n-1} f_{n-1,t} + (1 - \varphi_1 - \dots - \varphi_{n-1}) f_{nt} + \varepsilon_t. \quad (6.18)$$

Providing transformations we can obtain the next equation:

$$y_t - f_{nt} = \varphi_1 (f_{1t} - f_{nt}) + \varphi_2 (f_{2t} - f_{nt}) + \dots + \varphi_{n-1} (f_{n-1,t} - f_{nt}) + u_t. \quad (6.19)$$

It is a regression equation, and it can be estimated with the OLS or other statistical methods. It is important that we can find these coefficients which present weights of different forecasts in the final prediction.

Also the next condition about the error term should be satisfied:

$$\bar{\varepsilon} = \frac{\sum_{t=1}^T \varepsilon_t}{T} = 0. \quad (6.20)$$

This fact can be checked for our model:

$$\begin{aligned}\hat{\varepsilon} &= \frac{\sum_{t=1}^T \varepsilon_t}{T} = \frac{\sum_{t=1}^T (y_t - \varphi_1 f_{1t} - \dots - \varphi_n f_{nt})}{T} = [y_t = f_{1t} = \dots = f_{nt}] = \\ &= \frac{(1 - \varphi_1 - \dots - \varphi_n) \sum_{t=1}^T y_t}{T} = 0.\end{aligned}$$

This condition will be satisfied only if the sum of all coefficients is equal to one. This assumption is quite strong, and often it is violated. Thus, it is necessary to find some ways to remove this assumption without losses in the forecast's quality.

If we simply remove the assumption without any changes in the model's specification, the condition about a zero mean of the error term can be false. In fact, the forecast will be biased. To solve this problem, Granger and Ramanathan suggested including intercept into the model of combined forecasts. This coefficient reflects a bias taking place when we remove the assumption about potential coefficient values. Then the model can be presented in the next form:

$$y_t = \varphi_0 + \varphi_1 f_{1t} + \dots + \varphi_n f_{nt} + \varepsilon_t. \quad (6.21)$$

This approach is the best way to combine forecasts because it does not require any assumptions for coefficients, but a zero mean condition is satisfied.

Thus, it is possible to create a proper forecast using several models. Such approach allows researchers to combine different methodologies and decrease errors in forecasts.

And one more remark can be done for this topic. There are no particular models and function which would generate perfect forecasts. Moreover, all modern predictions are not accurate enough. If you use complicated constructions and models, it does not guarantee that your forecast will be precise. One potential way to improve your forecasting is to construct a theoretical framework. It means that you should present a mathematical background for your future research. Initially you are recommended to reveal quality relationships and then to test them with statistical tools. Econometrics can help you to check interactions and

hypotheses, but it cannot formulate quality results. Thus, your research should have a strong theoretical framework and empirical confirmation.

All concepts, which were reflected in this book, can be estimated and tested in special statistical packages. Further we will show the ways how it can be done in applied programs such as Stata and Econometric Views (EViews).

Questions

1. What indicators can help to determine the forecast's quality? What is their statistical mention? What model will be the best according to these indicators?
2. List advantages and disadvantages of the coefficient of determination. How can the coefficient help to find the best forecast?
3. What kind of models can be used for forecasting? What are their advantages and disadvantages?
4. What model would you use to predict the data with high frequency in short-run? And in long run? Explain your answer.
5. List main steps for creating a combined forecast. Why is it useful to do such predictions?
6. What is the main principle of combined forecasts? What conditions should be satisfied? How is it possible to weaken these conditions?

Exercises

To do this exercise you are recommended to use quarterly data (one time series, a minimal number of observations is 100).

- a) Present the data graphically. What can you say about seasonality and trend? If there is seasonality, you have to remove it to make a proper forecast. To do it you can use any possible method;
- b) Now there is no seasonality in the data. What can you say about a trend function? Make assumptions about its general view and estimate unknown parameters. Predict fitted values and find *MSE*, *MAPE*, *MAD* and *MPE*. What can you say about a quality of your forecast? Using your estimation results you should predict the data for 20 periods (5 years in case of quarterly data).

Do not forget to come back seasonality after prediction. Present initial data and fitted values with forecasting results on a single graph;

- c) Now you should estimate two ARMA models for your data (do not forget to test stationarity and ARCH/GARCH processes!). After estimation you should choose the best model for forecasting using MSE , MAD , $MAPE$, MPE . When you choose the best model, please, do the forecast for nearest 20 periods;
- d) Now you have two forecasts: the first one, estimated with the help of trend equations, and the second one, obtained from ARMA models. Now you are asked to present a combined forecast for actual data and future 20 periods. What approach will you use to create this forecast? Estimate all necessary equations and present all made forecasts on a single graph.

Chapter 7

APPLIED ECONOMETRICS FOR MACROECONOMICS IN STATISTICAL PACKAGES

7.1. Adjustment Data in Statistical Packages

Modern econometric models can be very complicated for estimating. Moreover, if we work with many variables and specific functions, it may be impossible to provide all necessary estimation procedures without technical support. To solve this problem applied statistical packages were created. In particular, they can help us to do difficult calculations and estimations. Further we will consider how all models, indicated above, can be presented in applied packages.

There are different programs which can work with econometric models. Some of them require writing specific codes (for example, TSP). Other programs are more user-friendly. We will show estimation procedures for EViews and Stata. The first package is convenient for time series analysis. The second one also can solve the same tasks. In principle, it is a question of habits what package to use. Also you can apply other programs to solve exercises and practical questions.

To start your work in EViews you should create a so-called workfile. To do this you are to choose “File — New — Workfile”. In a new window you have to put a structure of your file. If your data are time series, you should indicate their frequency. Also it is possible to state “Start date” and “End date”. If the program indicated quarterly, monthly, weekly and other types of data separately, it is not necessary to write particular parts of a year (they will be counted from the first period automatically). But if you want to establish any specific date, you should use special syntax of the program. All parts of a year should be separated with “:”. For example, if you want to say “the fourth quarter of 1976”, you should write “1976:Q4” in EViews. The same note can be

done for months (in this case “Q” should be replaced to “M”), weeks, days etc. If you work with cross-sectional data, the option “unstructured” should be chosen. In case of panel data it is necessary to indicate “Balanced Panel”*. Thus, the program will now what type of data should be analyzed.

If you work with time series in Stata, you should know the next feature. The program does not separate time series, cross-sectional data and panels automatically. You must say what type of data will be studied. To do this you should state identifiers for time and objects. If you work with time series, it is enough to specify only the first identifier. In case of panel data both variables should be stated.

Let’s start from our initial point which was connected with data transformations and seasonality. It is always useful to present data graphically because it can help you to make inferences about potential problems in time series. It is quite simple to visualize your data in EViews. You can choose any variable and open it with double-click. In the appeared window you can see all data in the range and its frequency**. It is possible to study all descriptive statistics, characteristics of the distribution here. To plot a graph you should choose “View — Graph” There are different types of graphical representations. For the first approximation it is useful to study a line graph for your data. Also there are graphs which can help to analyze seasonality. They can be considered in further analysis.

Stata also allows us to create different graphs. In you command a graph type should be indicated. Also the graph has to include two variables. In opposite case the program will not work. For example, to plot a scatter graph for time series in Stata you need to state the time identifier and write a command “scatter variable t”. Instead of “variable” you can put any time series and follow its dynamics.

* In different versions of the program these names can be different, but the main principle is kept.

** Also it is possible to open several time series in a single window. To do this you should select all necessary variables with pressed Ctrl, click the right button of the mouse and choose “Open — As group”. Using this option you can analyze correlation relationships and identify potential sources of multicollinearity.

Using graphical representations you can make inferences about potential characteristics of your time series. For example, you can identify a trend and a function which can describe it. Also it is possible to find seasonality in the data using the graph. If you see periodical fluctuations in time series with high frequency, you can assume seasonality here. In EViews you can use several ways to remove it:

- Census X12;
- X11;
- Moving Average;
- Exponential Smoothing etc.

The first two methods present seasonal indexes as a main way to remove seasonality. The third method is a standard moving average which is calculated for particular frequency. Exponential smoothing is also quite popular and convenient for further analysis. In principle, you can choose any method if there are no restrictions in your research. All these methods remove seasonality successfully. In opposite case you should use a specific transformation to save main properties of the data. To remove seasonality in EViews you should open a necessary variable and choose “Proc — Seasonal Adjustment”. Then you can use any method presented in the program. The same can be done in Stata. Similar transformations can be done without special packages, but this way is more convenient and quicker.

Also it is possible to remove seasonality with the help of dummy variables. To do it you should generate a set of new variables. Each variable will reflect a particular time period within a year (quarter, month, week etc.). To remove seasonality you should simply run a regression of your variable on these dummy (do not forget about perfect multicollinearity when you estimate such equations with dummy's!). Then you can save residuals, and this new range will not contain seasonal components.

If you remove seasonality, it does not mean that the trend will disappear too. If you want to provide detrending, you can also include a trend function into your model with dummy variables. But it is possible to provide detrending separately. To do this you can assume a general form of the trend function. If you assume a linear trend, you can

estimate a corresponding regression model. The same is true for other trend functions.

To estimate this equation with the OLS in Stata you can use a variable which presents a time identifier. Let's assume that we have a linear trend in the data. Then the command can be written as "reg variable t". A command "reg" reflects the OLS and can be used for estimation of the trend equation. In Stata a constant term is included into the model automatically, and you do not need to indicate it separately. A variable "t" is a time identifier. If you save residuals from this model, you will get data without trend.

As you remember, we do not state any identifiers in EViews. Then we should generate a variable which will reflect a number for any period of time. To do this we need to specify "genr t = @trend". A command "genr" is standard for creating new variables. Then you can use this new variable and estimate a trend equation using a command "ls variable c t". A part "ls" shows that we use the OLS. If we include a constant term, we need to write it in the command. As earlier, residuals from the model should be saved, and a new series will not include trends.

Differencing also can be done in these applied packages. In EViews you need to specify a new variable (use the command "genr") and use a lag operator. For example, we want to find the first differences of any variable. Then we should write "dvar = d(variable)". In this case "dvar" is a name for a new series, and a part "variable" reflects initial data. Further this new series can be presented on the graph and tested for unit roots.

In Stata the command for lags will be different. A time identifier is also used for this operation. A command for creating new values is "gen" (or "g"). The next command will be written for the first differences: "gen dvar = variable — variable[n - 1]". In this case we specify the difference between present and previous values. If it is necessary to generate differences with higher orders, you can repeat this procedure for new time series.

Thus, all transformations and seasonal adjustments can be provided with the help of special statistical packages. Then we will consider how different unit root tests can be done in these programs.

7.2. Unit Root Tests in Applied Statistical Packages

A lot of tests and estimation procedures are realized in applied statistical packages. Unit root tests are not an exception. So, Dickey — Fuller, Phillips — Perron, KPSS, and other tests can be calculated in EViews. Stata allows us to present only the first two tests. Let's consider how it can be done in applied statistical packages.

Initially we present EViews possibilities. The Dickey-Fuller test is a standard procedure for estimation. To provide it you should open a variable which must be checked for stationarity. Then you need to choose “View — Unit Root Test”. The program contains the next methods for unit root testing:

- Augmented Dickey — Fuller test;
- Phillips — Perron test;
- KPSS test;
- Ng — Perron test;
- Elliott — Rothenberg — Stock Point-Optimal test.

You can choose any method, but now we will consider the first approach. The program allows you to test stationarity of the first and the second differences without creating new variables. So, you have three possibilities for testing: level (initial values of time series), the first differences, the second differences. Differences with higher order do not use in econometrics because their interpretation is complicated. But if you need to test such values, you always can generate corresponding variables.

Also you can choose a specification of the model which you will test. As it was marked in the second chapter, there are three alternatives: none, intercept, and trend and intercept. To establish a number of lags you can use two ways. The first one is to select lags automatically. The program suggests several criteria to provide this procedure. But the main principle is to minimize information criterion. Using your knowledge about these criteria it is possible to assume that Schwarz — Bayes and Hannan — Quinn criteria will be better than Akaike's approach because they are more powerful. The program does not realize the “top-down” approach when the significance of the last lag is tested. In this

case you can use the second alternative and state lags yourself. Choosing this number you can analyze significant lags or information criteria. The choice of lags is very important for test results, and you should provide this procedure carefully.

When all parameters of the test are chosen, you can study its results. They are reflected in the table. So, you can find observed and critical values for all levels of significance (MacKinnon critical values are presented here). Also it is possible to observe estimators for all parameters in the model and find general information about the regression (statistics, criteria etc.). The rejection rule says that if the observed value is less than critical one, the null hypothesis has to be rejected (there is no unit root, the data are stationary). If you have found a unit root, the test should be repeated for differences.

Also we can use the Phillips — Perron test to check the data for a unit root. You need to choose “View — Unit Root Test”, but now it is necessary to state “Phillips — Perron” in a dialog window. The test also can be presented for different levels and specifications. As you remember, the Phillips — Perron test allows us to correct autocorrelation and heteroskedasticity. To achieve this goal it includes non-parametric components in the test statistics. Other estimation methods must be applied to identify a unit root. The Bartlett kernel method is applied automatically for this test, but you can specify different methods. All these procedures help to correct standard errors (they are not estimated properly because of heteroskedasticity). Also it is possible to state if lags will be chosen automatically or not. Null and alternative hypotheses are the same as for the Augmented Dickey — Fuller test.

An observed statistics will be different from the previous test because it was calculated with corrections. But usually inferences about non-stationarity from these tests correspond. The rejection rule is the same: if the observed value less than critical one, the null hypothesis has to be rejected (time series is stationary).

The KPSS test also can be done automatically in EViews. There are two possible specifications: level and trend and intercept. Lags can be chosen automatically or specified by the user. The test also corrects heteroskedasticity and autocorrelation. As you remember, null and

alternative hypotheses were changed here. So, the null one states that the range is stationary. Also another distribution is used here*. As a result, there is another rejection rule (for the right-side critical region): if the observed value is more than critical one, the null hypothesis has to be rejected (it means that the data are not stationary). The KPSS test is more accurate, and its results can differ from the first two tests.

Now we will consider unit root tests in Stata. To provide the Augmented Dickey — Fuller test in this program you need to choose “Statistics → Time Series → Tests → Augmented Dickey—Fuller Unit Root Test”. In a dialog window you can specify the regression which will be estimated. For example, you can exclude a constant term (option “Suppress constant term in regression”) and obtain a standard none specification. Also it is possible to include a trend, and you will obtain other results. Lags have to be stated by a user. It is possible to use information criteria for choosing a proper number. But in this case it will be better to apply the “top-down” approach controlling the significance of the last lag and autocorrelation in the model. To provide such procedures it is convenient to state “Display regression table” in test’s parameters. The rejection rule for this test was presented above.

The Phillips — Perron test also can be done in Stata. You should follow “Statistics → Time Series → Tests → Phillips — Perron Unit Root Test”. There are two possibilities to choose lags. The first approach is based on the next formula:

$$k = \text{int} \left(\left(\frac{4T}{100} \right)^{\frac{2}{9}} \right). \quad (7.1)$$

The second variant supposes that a number of lags will be chosen by a user. There are no any additional parameters for the test which can be specified in Stata. The result has the same interpretation as in the Augmented Dickey — Fuller test. Stata also allows us to provide other

* This test is built on the *LM*-test. As a result, critical values have to be obtained with the χ^2 distribution.

tests for stationarity. It is tests for a white noise which also show if the process is stationary or not.

Also it is possible to analyze seasonal unit roots in applied statistical packages. For instance, it is possible to provide the Hylleberg — Engle — Granger — Yoo test in Stata (a command `hegy4`)*. Other tests also can be provided in statistical packages.

In general, EViews is more convenient for unit root tests because it allows us to change different parameters of specifications. Also a general report is more informative and user-friendly in EViews. But this program does not contain test for structural breaks in all possible modifications (they are not presented automatically). Speaking about this part of time series analysis it is reasonable to use Stata. Now we will consider how tests for structural breaks can be provided in this program.

7.3. Tests for Structural Breaks in Applied Statistical Programs

As it was mentioned above, it is convenient to provide test for structural breaks in Stata. Earlier we studied three tests for structural breaks: the Perron test, the Zivot and Andrews test and the Vogelsang test. Let's consider all of them.

The Perron test is not realized automatically in statistical packages, but it is not difficult to provide it manually. Initially we should present a graph and assume a particular moment where a break could be. Moreover, we should suppose a type of this break because it will determine our specification for testing the hypothesis about a unit root. In any case, it is necessary to create additional variables for the test:

$$D(t_{\text{break}}) = \begin{cases} 1, & \text{if } t = t_{\text{break}} + 1, \\ 0, & \text{otherwise,} \end{cases}$$

* Initially you must install this test in Stata. To do it you should use the command “`ssc install hegy4`”.

$$D(u) = \begin{cases} 1, & \text{if } t > t_{\text{break}}, \\ 0, & \text{otherwise,} \end{cases}$$

$$DT_t = \begin{cases} t - t_{\text{break}}, & \text{if } t > t_{\text{break}}, \\ 0, & \text{otherwise.} \end{cases}$$

Then we can write the next commands in Stata if we suppose a particular value of t_{break} . For the first variable it will be (names for new variables can be different):

```
gen dtb = 0
```

```
replace dtb = 1 if t = tbreak + 1
```

For the second variable we have the next commands:

```
gen du = 0
```

```
replace du = 1 if t > tbreak
```

The third variable will be the next:

```
gen dt = 0
```

```
replace dtb = t - tbreak if t > tbreak
```

Using these variables we can estimate a necessary regression (its common view depends on a type of the break) with the OLS. Lags have to be chosen with the help of the “top-down” approach (the last lag must be significant in the final specification). Then we can test null and alternative hypotheses (their formal presentation also depends on a type of the break). For particular variables we can estimate the significance of coefficients with standard statistical tests. If we want to check a unit root, we should use special critical values for the Perron test. Then it is possible to do inferences about the existence of structural breaks or unit roots in the model.

The Zivot and Andrews test can be done automatically in Stata*. This test shows a single break. It also can be provided in three possible modifications (break in trend, break in intercept or both). The test should be done separately for every modification. Also it is possible to present a graph with t -statistics after all estimations. To use the tests you need to write a command “zandrews variable, break(trend) graph”. In the brackets after “break” you should state a type of the break (trend, intercept or both). A minimal value of t -statistics will be reflected on the graph. Then this value has to be compared with critical one. The rejection rule says that if the observed value is less than critical one, the null hypothesis about a unit root has to be rejected in favor of the alternative hypothesis (it indicates that the break exists). Special critical values must be applied for this test too.

The third test suggested by Vogelsang also is not presented in Stata or EViews automatically. However, it is not a problem to provide all necessary estimation procedures manually. The Vogelsang test does not check a unit root in time series. It reflects only structural breaks. As you remember, we need to estimate the next specification:

$$\Delta y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \delta_0 DU_t + \sum_{i=1}^p \delta_i DT_t^i + \pi y_{t-1} + \sum_{i=1}^k \rho_i \Delta y_{t-i} + \varepsilon_t, \quad p = 1, 2. \quad (7.2)$$

If there is a linear trend in the data, we should choose $p = 1$. In case of non-linear trends (for example, a quadratic trend can exist) it will be reasonable to take $p = 2$. To find a potential break point, all F -statistics can be compared for the next null and alternative hypotheses:

$$H_0 : \delta_p = 0 \forall p = 0, 1, 2,$$

$$H_1 : \text{at least one coefficient is significant.}$$

* This test is not installed automatically. To do this you need to enter a command “ssc install zandrews”.

As it was considered in the example, it is quite difficult to check all points for breaks. So, we need to make an assumption about potential break points and test only their neighborhood. For each point the next variables must be created:

$$\begin{aligned} \text{gen } du &= 0 \\ \text{replace } du &= 1 \text{ if } t > t_{\text{break}} \\ \text{gen } dt &= 0 \\ \text{replace } dt &= t - t_{\text{break}} \text{ if } t > t_{\text{break}} \end{aligned}$$

After estimating each regression we should provide the F -test and remember observed values of statistics. Then the point with the maximal value has to be chosen. It means that the break can be found at this moment of time. Further it is necessary to compare an observed value in the break point with critical one for the Vogelsang test. To do it additional calculations of $mean(F)$, $exp(F)$, $sup(F)$ must be done. If these calculated values are more than critical points, the null hypothesis about a stable trend must be rejected in favor of the break existence.

Thus, it is possible to estimate structural breaks in applied statistical packages. Not all tests can be done automatically, but some of them are not complicated technically. Different models also can be presented in programs, and further we will consider models for stationary time series.

7.4. Models for Stationary Time Series

As you know, there are particular types of econometric models which can be applied only for stationary time series. In opposite case stability condition will be violated, and estimators will have bad properties. Moreover, they will not converge to particular numbers. It means that initially you have to check the data for stationarity, do all necessary transformations (if they are required for achieving stationarity). Only after all these procedures you can apply special models.

The first type of models, which will be considered in applied statistical packages, is ARMA models (or ARIMA if differencing was applied

for making the data stationary). Let's present this estimation in EViews. Initially it is useful to look at correlogram. To do it you should open a necessary variable and choose "View — Correlogram". The program will ask you to state the number of lags, and you can use a common rule to do it: the maximal number of lags can be found as the number of observations divided into four. Then you can remove insignificant lags.

Also it is possible to state for what variable you want to present this correlogram. So, you can choose initial values (levels) or differences for non-stationary time series. Using this graph and statistics you can assume potential AR and MA processes (to identify AR lags PACF has to be used; ACF indicates MA lags). A graphical example was provided in the chapter devoted to models of stationary time series. The correlogram also includes Q -statistics and probabilities for each lag. They can be interpreted as follow: if any probability is less than the stated critical level, the null hypothesis has to be rejected. In case of autocorrelation the null hypothesis says that all k lags of the autocorrelation function are not significant. If we reject this statement, it means that there is autocorrelation on some lags.

Then you must specify the equation using a special syntax. For example, ARMA(2, 3) can be written in EViews as:

ls variable c ar(1) ar(2) ma(1) ma(2) ma(3).

Also it is possible to specify equations using "Quick — Estimate Equation". The result will be the same. Then you have to test the model for autocorrelation. It is possible to look at the Durbin — Watson statistics (it is reflected in the window with results). But other tests also can be realized automatically. In the window with results you must choose "View — Residual Tests". It is possible to construct correlograms for residuals or use the LM -test. Null and alternative hypotheses will be the same. If there is no autocorrelation in the models, you have to remember values of information criteria to make a final comparison (these numbers are contained in the table with results of ARMA estimation). Similar actions are applied for other ARMA specifications, and then you can determine the best model.

It is also possible to estimate ARMA models in Stata. Moreover, it is more convenient because this program allows us not to write all ARMA lags manually. We must indicate only a total number of lags for AR and MA processes and specify the order of differencing*. To present correlograms in this program you can use a command “corrgram variable”. To provide the estimation you should use ‘Statistics — Time Series — ARIMA and ARMAX models’. In a dialog window you should reflect a variable for which the model must be constructed. Also it is possible to include exogenous variables into the regression (you should not write a constant term separately because it will be included automatically). Then you must establish lags for your processes (they also can be indicated separately). If you use only MA processes, you also have to choose “Suppress a constant term”.

After estimation you also have to test autocorrelation. In case of its absence you should remember information criteria. To find these values in Stata you must specify “estat ic”. The choice of the best model is based on minimal values of information criteria.

Now we will consider ARCH/GARCH models. To find these processes it is necessary to determine a proper ARMA models and work with its residuals. Let's start from EViews. Initially we must store residuals from a corresponding ARMA model. The program saves residuals of the last regression in a variable “resid”. So, we simply should generate a new series which will be equal to “resid” in square (it is necessary to analyze squares of residuals because we estimate process for a variance in the error term). Then we can provide tests for autocorrelation (the *Q*-test, the *LM*-test). Also the program allows us to estimate the presence of ARCH processes without generating new variables. After estimating the best ARMA model we can choose “View — Residual Tests — ARCH LM Test”. As you remember, the null hypothesis says that there is no ARCH process in residuals. If an observed value is more than critical one, this statement must be rejected. In this case we can say that ARCH processes take place in the data.

* Also the program can estimate seasonal ARMA models.

To catch GARCH processes we should use other tests which are not realized automatically in EViews. To provide this testing it is necessary to find standard residuals. It can be done with the help of the regression which was used for ARCH identification and the ARMA/ARCH model. After generating this variable we must find if there is autocorrelation or not (standard tests, the *LM*-test)*. If it is true, there is a GARCH process. In opposite case we will use only ARCH processes.

Potential lags can be found with the help of correlograms. To estimate such model in EViews we should choose “Quick — Estimate Equation”. In the row “Methods” ARCH must be chosen. The ARMA model, for which this estimation is being provided, is to be specified in a dialog window (do not forget to type a constant term in the ARMA specification). Also it is possible to estimate different modifications of ARCH/GARCH models (for instance, ARCH-M, EGARCH etc.). Then we should state lags for ARCH and GARCH processes. If there is no GARCH, this lag will be equal to zero. In case of TARCH a threshold level also must be presented. The program allows us to choose different specifications for error terms. After all these procedures the estimation can be done.

In the table with results the program shows coefficients for ARMA and ARCH/GARCH equations. A proper ARCH/GARCH model must not contain autocorrelation. Then we have to find standard residuals** and test them for autocorrelation. Sometimes additional lags are to be included into the ARMA model to remove autocorrelation from ARCH/GARCH processes.

Similar estimation procedures can be provided in Stata. This program contains more modifications of ARCH/GARCH models than EViews. For example, nonlinear and asymmetric models can be

* Also it is possible to use the BDS-test for these residuals to identify GARCH processes. This test checks if residuals are independently identically distributed or not. The BDS-test can be done in EViews automatically. To provide it you should open these residuals and choose “View — BDS Independence Test”. The null hypothesis states that the residuals are i.i.d.

** In case of GARCH processes we cannot test residuals which are not corrected for conditional variance. So, we can work only with standard residuals which can be obtained with the help of estimated coefficients.

estimated here. The process of stating specification parameters looks like for ARMA models. So, lags of ARCH and GARCH processes can be presented. Also a constant term can be excluded from a final regression equation. Results will be reported in a single table. Information criteria can be displayed with the command “estat ic”.

Thus, models of stationary time series can be successfully estimated in applied statistical packages. Other types of models for macro-economic time series also can be presented in Stata, EViews and other packages.

7.5. Intervention Analysis in Applied Statistical Packages

Earlier we have studied models which can be applied for the analysis of different interactions between variables. It can be models with events, the vector of autoregression or Granger causality relationships. Now their usage in applied statistical packages will be covered.

It is convenient to analyze consequences of events in Stata. For example, if we want to create a variable which would reflect an impulse or other types of interventions, commands “gen” and “replace” can be used. Such actions can be followed in the block devoted to tests for structural breaks in applied packages. Including such variables in regression equations it is possible to evaluate their significance and present impulse response functions.

The VAR and Granger causality are realized in programs too. Let's start from EViews. We can assume that there are three variables which are integrated with the same order. We want to estimate the VAR model, present impulse response functions and variance decompositions. Then we should select all necessary variables, click with the right button and choose “Open — As VAR”. A dialog window reflects two alternatives. The first one assumes that we can construct an unrestricted VAR. The second alternative allows us to estimate the error correction model (ECM). The ECM will be discussed later, now we consider only the unrestricted VAR model. Also we can specify exogenous variables. If there are no additional exogenous components or special variables,

we should indicate only a constant term (it is necessary to identify parameters of the system). The last thing is to state a number of lags including into the model. You can use autocorrelation functions for variables to choose a potential number of lags. You can assume what time is necessary to neutralize shocks in the system and take corresponding lags. For instance, if you suppose that in case of quarterly data chocks disappear during three years, you should try twelve lags. The final specification will be chosen with the help of information criteria and the *LM*-test.

The model will be estimated in the reduced form to find particular values for coefficients*. The model is estimated as a system. In other words, parameters in all equations are found simultaneously. There are no probabilities for coefficients, but the table includes standard errors and *t*-statistics for every parameter. Using this information, it is possible to say if a variable is significant or not. In the bottom of the table there is general information about the estimated system. If to remember about post estimation procedures, we are to be interested in the determinant of residual covariance matrix and information criteria. These values are important for a comparison of different specifications in the VAR model.

Other specifications are estimated with the same actions. Then the best model must be found. Initially we can assume that the model with minimal values of information criteria is the best one. However, for more accurate inferences we have to provide the *LR*-test. The program does not calculate all necessary statistics automatically. As a result, this test must be done manually. It does not require difficult calculations and special critical values (they are standard and taken from χ^2 distribution). When the test is finished, it is possible to make forecasts with the help of the chosen model.

When the proper model is chosen, impulse response functions and variance decompositions can be presented. To draw impulse response functions it is enough to click the button “Impulse” in the table with estimation results. A user can choose a type of displaying these functions

* In opposite case the model may be not identified (see additional notations in chapters above).

(table, multiple or combined graphs). Also a method for standard errors calculation can be changed. User can also specify a number of periods. Using presented graphs consequences of shocks can be observed.

To find variance decompositions you should follow “View — Variance Decomposition”. There are different forms of presenting results and methods for standard errors calculations. It is convenient to present variance decompositions in tables. Then gains of other factors into the variance of a particular variable can be analyzed.

Stata also has automatic procedures for the VAR analysis. Moreover, there is an opportunity to construct the VAR in a structural form. To provide the estimation you should choose “Statistics — Multivariate Time Series”. Then different forms of the VAR can be used. It is also possible to exclude a constant term from the analysis. Impulse response functions and variance decompositions can be found in “Statistics — Multivariate Time Series — IRF and FEVD Analysis”.

Both programs allow us to present Granger causality relationships with automatic tools. In EViews you can find corresponding tests in “Quick—Group Statistics — Granger Causality Test”. In a dialog window it is enough to indicate variables for which you want to test causality. For a particular number of lags the program gives a table with results. There are possible null hypotheses (if we analyze more than two variables, several tests can be provided automatically), test statistics and probabilities. If any probability is less than a critical level, the null hypothesis about an absence of Granger causality has to be rejected in favor of alternative one.

Granger causality also can be tested in Stata. To find this option you should use “Statistics — Multiple Time series — VAR Diagnostics and Tests — Granger Causality Tests”. It is possible to provide the test using the VAR model which was estimated earlier. Also the program can check Granger causality for any time series without previous system estimations.

Thus, intervention analysis can be done automatically. It is quite convenient to use special programs because this part of applied econometrics for macroeconomics is built on systems’ estimation. It requires complicated calculations and matrix analysis, so the programs help us

to solve these problems easier. Cointegration relationships can be presented as the ECM which is a special case of the VAR model. Such analysis can be provided automatically too.

7.6. Cointegration

Earlier we discussed two different approaches to the question of cointegration identification. The first method was suggested by Engle and Granger, and it works properly for the case of two time series. But if the number of ranges is more than two, such analysis becomes complicated, and it is reasonable to use the second approach which is called the Johansen test.

The Engle—Granger methodology can be presented manually without any difficulties. It is not presented in applied packages. The procedure is quite simple: we run the regression of one potentially cointegrated variable to another range (these series have to be integrated with the same order), save residuals and check them for stationarity. Also we are to control if this regression is false or not. When the residuals are integrated with the less order, we can say that time series are cointegrated. In opposite case they are not cointegrated, and we can test the data for Granger causality (this algorithm was considered above).

The Johansen test is not so simple for manual calculations, and it can be done automatically in EViews and Stata. To provide the test in EViews you should choose necessary variables and open them as a group. Then you should follow “View — Cointegration Test”. The program suggests different specifications which can be checked. You are to say what deterministic components will be included into cointegration equations (an abbreviation CE in the program). You can exclude trends (linear or quadratic), intercepts and combine these deterministic parts in the system. Lags also must be specified by a user.

In the table with results you can see different rows. The row “None” means the null hypothesis that there are no cointegration equations (time series are not cointegrated). The row “At most one” indicates the hypothesis that there is a single cointegration equation (a pair of cointegrated variables) etc. If the null hypothesis is rejected, the

program checks more cointegration restrictions. In case of two variables the rejection of “At most one” means that there are two cointegration restrictions, and time series are cointegrated.

In tables below it is possible to find coefficients for cointegration equations. Using results from the ECM it is possible to make forecasts for long-run periods. Thus, this methodology can be useful for forecasting and estimation of long-term interactions between different time series.

The same procedure can be done in Stata. To present the ECM you should follow “Statistics → Multivariate Time Series → Vector Error-Correction Model (VECM)”. All parameters of the model (dependent variables, deterministic components, number of included lags) must be specified in a dialog window. If you want to find a number of cointegration equations in the model, the next way should be used: “Statistics — Multivariate time series — Cointegrating Rank of a VECM”. Using this option it can be found what number of variables is cointegrated with each other. But we do not know what particular series form these relationships. To find cointegrating groups additional tests and estimations must be provided.

Thus, cointegration can be tested with the help of applied statistical packages. The Engle—Granger methodology can be provided manually. Main statistical programs do the Johansen test automatically. It helps to find a number of cointegration equations and determine coefficients in the ECM. Then these data and results can be used for forecasting.

7.7. Forecasting

The main purpose of time series analysis is to create a proper forecast. All presented concepts can be used for predictions. Some of them are good for short-run forecasts; other can be applied for long-term periods. To make predictions it is necessary to know estimators of unknown parameters in the model. Then these values can be used for calculation of future data with the help of previous information. There are different indicators which can help to find the best forecast. As main types of

time series models can be estimated in applied statistical packages, it is also possible to determine and evaluate forecasts here.

Let's start from EViews. We can do retrospective forecasts (predict future values using information about previous periods) with the help of ARMA, ARCH/GARCH, VAR, ECM models. To do the forecast the model must be estimated properly. In fact, we have to provide a correct in-sample forecast. It means that all necessary stability conditions, assumptions about error terms have to be satisfied. In other words, estimators must have good statistical properties.

You know how these models can be estimated in EViews. If we estimated the VAR model, it is possible to make an in-sample forecast and find indicators of its quality. To provide such prediction, the model has to be estimated in the program. In the window with final results you should choose "Proc — Make Model". A system will be presented here. Then it is necessary to click "Solve" to find fitted values. Additional parameters for solution must be stated here. In the first part (usually it is called "Basic Options") you can choose solution types, a kind of the forecast, a sample range etc. Usually we are interested in deterministic parts of time series; in such case this option must be chosen. Also the VAR model presents a dynamic system, and we want to know future values. Then the forecast will be dynamic. It is possible to create predictions only for some parts of the sample. If you need fitted values for several periods (not for a whole range), start and final data must be indicated. In the part "Solver" different solution methods can be chosen. Also it is useful to switch on an option "Stop solving for missing data". When all solvers' parameters are indicated, the process can be started.

As a result, a set of new variables will be created. These time series present fitted values of initial data estimated with the help of the VAR model. In principle, it is possible not to use the system for the forecast (an option "Fit" should be chosen in the part "Basic Options" for a solver). So, separate equations from the VAR can take part in calculations. But if we are interested in the system and interactions between time series, it is reasonable to provide corresponding methods of forecasting.

Initial and new time series can be presented on a single graph. It can help you to understand if the forecast approximates the real data

well or not. To estimate the forecast's quality more precisely special indicators must be calculated. This procedure is not presented automatically in EViews. So, if you want to find *MAD*, *MSE*, *MAPE*, *MPE*, it must be done manually (for instance, Excel can be used for such operations). It makes sense to compare several forecasts and choose the best one where these indicators are minimal*.

When in-sample forecasts were done and the best model was chosen, an out-of-sample prediction can be made. When you started your work in EViews, the size of the sample had to be stated. To create additional observations for out-of-sample forecasts we need to widen the sample range. This parameter can be changed in your work-file. In earlier versions of EViews it was necessary to choose "Proc — Change Work-file Range". In later editions it sounds like "Proc — Structure / Resize Current Page". Using this option you can widen the sample on necessary number of observations. Then you should repeat all actions which were done for forecasting with the VAR model. The same method can be used if you want to provide a forecast with the help of the ECM (in case of cointegrated time series).

As you remember, the VAR model often is used for long-run forecasts. To predict short-term dynamics it can be useful to find a proper ARMA model. You know how this model must be specified and estimated. Usually there are several ARMA models without autocorrelation. It means that they can be used for forecasting. To find the best model *MAD*, *MSE*, *MAPE*, *MPE* should be calculated. To find these values new fitted variables must be found. It can be done in the window with estimation results. You should click the button "Forecast" and specify all necessary parameters (a type of the forecast, displaying of the graph etc.). Then actual and predicted values can be compared; and all indicators of the forecast's quality must be calculated. The best forecast has minimal values of *MAD*, *MSE*, *MAPE*, *MPE*. The out-of-sample forecast can be done with the help widening the sample's range.

* It is possible that for different forecasts these indicators demonstrate opposite dynamics. In this case you are recommended to use MSE and MPE as main signals for the best model (we are interested in the forecast with minimal variance in the error term and the least bias).

Stata also can be used for forecasting. If you use a standard ARMA or ARCH/GARCH model, it is enough to use a command “predict variable”*. Then actual and fitted values can be compared. If you use multivariate time series, you should follow “Statistics — Time Series — Multivariate Time Series — Dynamic Forecasts”. Using this option, predictions can be calculated in the system. It is important if we want to observe structural interactions and relationships**.

Thus, forecasting can be provided in applied statistical packages. To do this it is necessary to widen the sample’s range. The programs do not calculate *MAD*, *MSE*, *MAPE*, *MPE* automatically, but these actions can be done manually without any difficulties.

* Instead of “variable” a name for new time series must be inserted.

** If you want to provide an out-of-sample forecast in Stata, you should widen the sample manually, generate new observations and insert them into the data set for further calculations.

CONCLUSION

We studied main types of models which can be used for time series analysis. The basic concept in time series analysis is stationarity. It means that the process converges to a constant mean. In other words, shocks do not accumulate in the system. Stationarity can be tested with the help of special procedures. Sometimes the data have structural breaks in trend, or in intercept, or both. These dynamic changes also can be tested with econometric tools.

There are special models which can be applied for stationary time series. If there is no conditional heteroskedasticity, ARMA models can be used successfully to present dynamics. But if this assumption is violated, ARCH/GARCH models must be presented.

Also it is possible to provide intervention analysis. We can study how different event determine time series. The event (shock) can be included in the model as a discrete variable. The result of such analysis is an impulse response function. It shows how current events influence future values of time series.

Sometimes variables are connected with each other. To follow their mutual dynamics the VAR (vector of autoregression) model can be constructed. This model presents a system which can be estimated simultaneously. To study mutual dynamics and consequences of shocks impulse response functions and variance decompositions can be used.

It is possible that time series have the same dynamics. It can be observed in deterministic parts of time series (trends, seasonal fluctuations etc.). If such data are integrated with the same order and their linear combination is integrated with the less one, we can say about cointegration. Cointegrated time series can be presented in terms of the ECM (error correction model). This model presents long-run equilibrium and correct short-term dynamics.

The main goal of time series analysis is to create proper forecasts. To provide it we should use econometric models with good statistical

properties of estimators. There are in-sample and out-of-sample forecasts. Using the first type we can find the best model for forecasting. To make a correct choice special statistical indicators must be calculated (*MAD*, *MSE*, *MAPE*, *MPE*). The best model has to be characterized with minimal values of these indicators. In some cases it is reasonable to combine different models. Then the final specification will be used for out-of-sample forecasts.

Most estimation procedures can be done in applied statistical packages. But you have to know how all these methods work to avoid mistakes in the process of forecasting. So, you should know which model is better for short-run predictions and which one generated proper long-term forecasts.

Today applied econometrics for macroeconomics is developing. New tests and approaches are appearing; different types of known models are being suggested. The reason for further studies is that we do not know how to construct the best predication. Any dynamics cannot be presented perfectly. One part of this error may be explained through stochastic processes. But another part is connected with imperfect estimation of time series. It means that further researches in this sphere are quite prospective.

APPENDIX

Table 1

Dickey — Fuller Critical Values

Time Span	Critical Level			
Model <i>A</i>	0,10	0,05	0,025	0,01
$T = 25$	-1,60	-1,95	-2,26	-2,66
$T = 50$	-1,61	-1,95	-2,25	-2,62
$T = 100$	-1,61	-1,95	-2,24	-2,60
$T = 250$	-1,62	-1,95	-2,23	-2,58
$T = 300$	-1,62	-1,95	-2,23	-2,58
$T = \infty$	-1,62	-1,95	-2,23	-2,58
Model <i>B</i>	0,10	0,05	0,025	0,01
$T = 25$	-2,62	-3,00	-3,33	-3,75
$T = 50$	-2,60	-2,93	-3,22	-3,58
$T = 100$	-2,58	-2,89	-3,17	-3,51
$T = 250$	-2,57	-2,88	-3,14	-3,46
$T = 500$	-2,57	-2,87	-3,13	-3,44
$T = \infty$	-2,57	-2,86	-3,12	-3,43
Model <i>C</i>	0,10	0,05	0,025	0,01
$T = 25$	-3,24	-3,60	-3,95	-4,38
$T = 50$	-3,18	-3,50	-3,80	-4,15
$T = 100$	-3,15	-3,45	-3,73	-4,04
$T = 250$	-3,13	-3,43	-3,69	-3,99
$T = 500$	-3,13	-3,42	-3,68	-3,98
$T = \infty$	-3,12	-3,41	-3,66	-3,96

Table 2

Critical Values for KPSS test

Critical Level	Critical Value	
	For η_u	For η_τ
0,10	0,347	0,119
0,05	0,463	0,146
0,025	0,574	0,176
0,01	0,739	0,216

Table 3

Perron's Critical Values

Pre-break Fraction, Λ	Critical Level			
Model A	0,10	0,05	0,025	0,01
$\lambda = 0,1$	-3,40	-3,68	-3,93	-4,30
$\lambda = 0,2$	-3,47	-3,77	-4,08	-4,39
$\lambda = 0,3$	-3,46	-3,76	-4,03	-4,39
$\lambda = 0,4$	-3,44	-3,72	-4,01	-4,34
$\lambda = 0,5$	-3,46	-3,76	-4,09	-4,45
$\lambda = 0,6$	-3,47	-3,76	-4,09	-4,45
$\lambda = 0,7$	-3,51	-3,80	-4,07	-4,42
$\lambda = 0,8$	-3,46	-3,75	-3,99	-4,33
$\lambda = 0,9$	-3,38	-3,69	-3,97	-4,27
Model B				
$\lambda = 0,1$	-3,36	-3,65	-3,94	-4,27
$\lambda = 0,2$	-3,49	-3,80	-4,08	-4,41
$\lambda = 0,3$	-3,58	-3,87	-4,17	-4,51
$\lambda = 0,4$	-3,66	-3,94	-4,20	-4,55
$\lambda = 0,5$	-3,68	-3,96	-4,26	-4,56
$\lambda = 0,6$	-3,66	-3,95	-4,20	-4,57

End of tab. 3

Pre-break Fraction, Λ	Critical Level			
$\lambda = 0,7$	-3,57	-3,85	-4,13	-4,51
$\lambda = 0,8$	-3,50	-3,82	-4,07	-4,38
$\lambda = 0,9$	-3,35	-3,68	-3,96	-4,26
Model C				
$\lambda = 0,1$	-3,45	-3,75	-4,01	-4,38
$\lambda = 0,2$	-3,66	-3,99	-4,32	-4,65
$\lambda = 0,3$	-3,87	-4,17	-4,46	-4,78
$\lambda = 0,4$	-3,95	-4,22	-4,48	-4,81
$\lambda = 0,5$	-3,96	-4,24	-4,53	-4,90
$\lambda = 0,6$	-3,95	-4,24	-4,49	-4,88
$\lambda = 0,7$	-3,86	-4,18	-4,44	-4,75
$\lambda = 0,8$	-3,96	-4,04	-4,31	-4,70
$\lambda = 0,9$	-3,46	-3,80	-4,10	-4,41

Table 4

Zivot and Andrews Critical Values

	Critical Level			
	0,10	0,05	0,025	0,01
Model A	-4,58	-4,80	-5,02	-5,34
Model B	-4,11	-4,42	-4,67	-4,93
Model C	-4,82	-5,08	-5,30	-5,57

Table 5

Vogelsangs's Critical Values

Stationary case, $\lambda = 0,01$									
Critical Level	$p = 0$			$p = 1$			$p = 2$		
	Mean(F)	Exp(F)	Sup(F)	Mean(F)	Exp(F)	Sup(F)	Mean(F)	Exp(F)	Sup(F)
0,10	2,00	1,59	9,24	3,49	2,76	13,62	4,74	3,70	16,06
0,05	2,66	2,20	10,85	4,42	3,52	15,44	5,65	4,41	17,89
0,025	3,34	2,80	12,46	5,36	4,18	17,26	6,69	5,22	19,57
0,01	4,21	3,63	14,49	6,64	5,24	19,90	8,14	6,24	21,65
Stationary case, $\lambda = 0,15$									
Critical Level	$p = 0$			$p = 1$			$p = 2$		
	Mean(F)	Exp(F)	Sup(F)	Mean(F)	Exp(F)	Sup(F)	Mean(F)	Exp(F)	Sup(F)
0,10	1,58	1,23	7,32	2,70	2,33	11,25	3,58	3,18	13,96
0,05	2,20	1,89	9,00	3,50	3,13	13,29	4,41	3,98	15,48
0,025	2,85	2,53	10,69	4,35	3,88	15,12	5,25	4,68	17,61
0,01	3,70	3,46	13,02	5,55	5,05	17,51	6,47	5,78	19,90
Unit root case, $\lambda = 0,01$									
Critical Level	$p = 0$			$p = 1$			$p = 2$		
	Mean(F)	Exp(F)	Sup(F)	Mean(F)	Exp(F)	Sup(F)	Mean(F)	Exp(F)	Sup(F)
0,10	3,32	4,02	16,14	7,14	6,98	22,60	10,18	9,58	28,11
0,05	3,91	4,84	18,20	8,22	8,18	25,27	11,74	11,09	31,35
0,025	4,53	5,68	20,23	9,29	9,27	27,76	13,17	12,50	34,45
0,01	5,35	6,69	22,64	10,54	10,56	30,44	14,80	14,42	38,43
Unit root case, $\lambda = 0,15$									
Critical Level	$p = 0$			$p = 1$			$p = 2$		
	Mean(F)	Exp(F)	Sup(F)	Mean(F)	Exp(F)	Sup(F)	Mean(F)	Exp(F)	Sup(F)
0,10	2,28	3,87	15,78	6,12	6,90	22,29	8,65	9,54	27,99
0,05	3,43	4,71	17,88	7,19	8,12	25,10	10,00	11,07	31,29
0,025	3,99	5,57	20,08	8,07	9,24	27,56	11,32	12,47	34,39
0,01	4,65	6,60	22,48	9,17	10,54	30,36	13,02	14,34	38,35

Table 6

**Critical Values for the Cointegrating Regression
Durbin — Watson Test**

Number of Variables	Sample Size		
	50	100	200
2	0,72	0,38	0,2
3	0,89	0,48	0,25
4	1,05	0,58	0,3
5	1,19	0,68	0,35

Table 7

MacKinnon Critical Values

Number of Variables	Critical Level		
	0,01	0,05	0,1
2	-3,9	-3,34	-3,04
3	-4,24	-3,74	-3,45
4	-4,64	-4,1	-3,81
5	-4,96	-4,42	-4,13

REFERENCES

- Amisano A.* Topics in Structural VAR Econometrics / A. Amisano, C. Giannini. 2nd ed. Berlin : Springer-Verlag, 1997.
- Banerjee A.* Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data / A. Banerjee, J. Dolardo, J. Galbraith, D. Hendry. Oxford : Oxf. Univ. Press, 1993.
- Bauwens L.* Bayesian Inference in Dynamic Econometric Models / L. Bauwens, M. Lubrano, J.-F. Richard. Oxford : Oxf. Univ. Press, 1999.
- Campbell J. Y.* The Econometrics of Financial Markets / J. Y. Campbell, W. Andrew, A. C. MacKinlay. Princeton : Princ. Univ. Press, 1997.
- Clements M. P.* Forecasting Economic Time Series / M. P. Clements, D. F. Hendry. Cambridge : Cambr. Univ. Press, 1998.
- Dickey D.* Distribution of the Estimators for Autoregressive Time Series with a Unit Root / D. Dickey, W. Fuller // J. of the American Statistical Association. 1979. Vol. 74. P. 427–431.
- Enders W.* Applied Econometric Time Series / W. Enders. 3rd ed. New Jersey : John Wiley & Sons, 2010.
- Engle R.F.* Co-integration and Error Correction: Representation, Estimation and Testing / R. F. Engle, C. Granger // Econometrica. 1987. Vol. 55(2). P. 251–276.
- Favero C.* Applied Macroeconometrics / C. Favero. Oxford : Oxf. Univ. Press, 2001.
- Fuller W.* Introduction to Statistical Time Series / W. Fuller. 2nd ed. N. Y. : John Wiley & Sons, 1996.
- Granger C.* Some Properties of Time Series Data and Their Use in Econometric Model Specification / C. Granger // J. of Econometrics. 1981. Vol. 16. P. 121–130.
- Hamilton J.* Time Series Analysis / J. Hamilton. Princeton : Princ. Univ. Press, 1994.
- Harvey A.* Forecasting, Structural Time Series Models and the Kalman Filter / A. Harvey. Cambridge : Camb. Univ. Press, 1989.
- Hatemi A.* Multivariate Tests for Autocorrelation in the Stable and Unstable VAR Models / A. Hatemi // Economic Modelling. 2004. Vol. 21(4). P. 661–683.

- Johansen S.* Likelihood-Based Inference in Cointegrated Vector Autoregressive Models / S. Johansen. Oxford : Oxf. Univ. Press, 1995.
- Kocenda E.* Elements of Time Series Econometrics: An Applied Approach / E. Kocenda, A. Cerny. [S. l.], 2007.
- Lutkepohl H.* Applied Time Series Econometrics / H. Lutkepohl, M. Kratzig. N. Y. : Cambr. Univ. Press, 2004.
- Maddala G.* Unit Roots, Cointegration and Structural Changes / G. Maddala, I. Kim. Cambridge : Cambr. Univ. Press, 1998.
- Makridakis S.* Forecasting: Methods and Applications / S. Makridakis, S. Wheelwright, R. J. Hyndman. N. Y. : John Wiley & Sons, 1998.
- Mills T. C.* The Econometric Modeling of Financial Time Series / T. C. Mills. N. Y. : Cambr. Univ. Press, 2008.
- Nelson Ch.* Trends and Random Walks in Macroeconomic Time Series / Ch. Nelson, Ch. Plosser // J. of Monetary Economics. 1982. Vol. 10. P. 139–162.
- Ng S.* Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag / S. Ng, P. Perron // J. of the American Statistical Association. 1995. Vol. 90. P. 268–281.
- Pesaran M. H.* Wickens Handbook of Applied Econometrics: Macroeconomics / M. H. Pesaran. Wiley — Blackwell, [S. a.].
- Phillips P. C. B.* Testing for Unit Roots in Time Series Regression / P. C. B. Phillips, P. Perron // Biometrika. 1988. Vol. 75. P. 335–346.
- Stock J. H.* Units Roots, Structural Breaks and Trends / J. H. Stock // Handbook of Econometrics. N. Y., 1994. Vol. 4.
- Tsay R. S.* Analysis of Financial Time Series / R. S. Tsay. New Jersey : John Wiley & Sons, 2010.
- Verbek J.* A guide to Modern Econometrics / J. Verbek. New Jersey : John Wiley & Sons, 2008.
- Wooldridge J. M.* Introductory Econometrics: A Modern Approach / J. M. Wooldridge. 2nd ed. South-Western College Pub., 2003.

GLOSSARY

Alternative hypothesis — a general statement which is opposite to the null hypothesis. In fact, it is a negation of the null hypothesis. Usually it is formulated in terms “there are differences between objects”. The next expressions can be used in the alternative hypothesis: “is not equal”, “is more than”, “is less than”, “at least one” etc.

Autoregression process (AR) — a dynamics based only on previous values of the variable. It means that current time series are determined only by their history. The process is estimated only for stationary time series (stability condition must be satisfied). This process has memory and can be caught with the help of autocorrelation tests. Potential lags for this process can be assumed using correlogram.

ARCH/GARCH processes — ARCH (autoregressive conditional heteroskedasticity) and GARCH (generalized autoregressive conditional heteroskedasticity) processes take place when the assumption about unconditional mean and variance are violated in the ARMA model. These processes are to be constructed only for stationary time series. The ARCH process looks like the AR process estimated for the error term (from the ARMA model). The GARCH process can be compared with the MA process applied for the error term.

Asymptotic properties — properties appear on infinitely large sample. They are connected with assumptions about distributions of error terms and parameters in the model. Asymptotic properties are very important for estimating the model's quality.

Autocorrelation — dependence between values of different time periods. So, current observations can be connected with previous ones. In terms of Gauss — Markov assumptions autocorrelation is a violation of the fifth condition (error terms for different objects or time periods must not be correlated). The presence of autocorrelation in residuals means that there are problems with the model's specification. Autocorrelation leads to inconsistent and biased estimators. Autocorrelation can be identified with the help of special statistical tests (the Q-test, the LM-test etc.).

Autocorrelation function (ACF) — a measure of dependence between values from different moments of time. It also includes correlations for all period placed between stated time targets.

Cointegration — a presence of long-term interactions between time series.

Cointegration can be observed only for series integrated with the same order. Time series are cointegrated if their linear combination is integrated with the less order than initial data. Cointegrated time series have similar deterministic components (trends, seasonal patterns etc.). Cointegrated time series can be presented via the error correction model (ECM), which characterizes long-run equilibrium.

Combined forecast — a mixture of different models using for predictions.

This forecast is a linear combination (various assumptions about the sum of weights can be done) of models. Such forecast can be used for achieving more accurate results. Weights for each component in combines forecast can be estimated with the help of regression equation.

Conditional mean — a mean calculated with respect to some distribution function. As usual, conditional mean is not equal to unconditional one.

Conditional variance — a variance calculated with respect to some distribution function. As usual, conditional variance is not equal to unconditional one.

Consistency — one of main statistical properties of estimators. It states that an estimator must converge (in probability) to a real value asymptotically. If it is true, the estimator is consistent. It means that results for large samples have to be more accurate than for small numbers of observations. This property works asymptotically, and it is a difference between consistency and unbiasedness.

Constant term — a deterministic part of a regression equation. Also it is called intercept. This parameter can be considered as an exogenous variable (for instance, this approach is applied in VAR models). Usually this term is included into all specifications.

Correlation — a connection between different variables. Regression analysis is built on the assumption that dependent and explanatory variables must be correlated. Usually the strength of correlation is measured with the help of a special coefficient. If its value is close to unity, there are strong correlation relations. If the coefficient is close to zero, variables have weak connections. In some cases the presence of correlation can become a problem. For example, the fourth Gauss — Markov condition says that regressors must not be correlated with the error term. In opposite case we have endogeneity problem.

Critical level — a probability established for constructing a confidence interval.

This interval is built with the assumption about a particular distribution.

For different critical levels these intervals will be different. Critical level is necessary for seeking for critical points and testing hypotheses.

Critical region — a part of numeric line where the null hypothesis has to be rejected. Critical region lies outside of the confidence interval. There are two-side and one-side critical regions. Two side critical regions are applied for symmetric distributions and alternative hypotheses in terms of “not equal”. One-side critical regions are used in other cases (asymmetric distribution and other terms for alternative hypotheses).

Critical value — a calculated value for particular distribution, critical level and degrees of freedom. This value is necessary for testing hypotheses. Observed value for statistics have to be compared with the critical value. Then it is possible to determine if the null hypothesis must be rejected or not.

Detrending — a possible way to exclude deterministic components from time series. It assumes that a trend function must be specified and a corresponding equation has to be estimated. Then residuals from this model should be saved, and a new range without trend is made. As usual, data after detrending are stationary and can be used for ARMA and ARCH/GARCH estimation.

Difference — a variance between two objects or periods. For time series difference is calculated for current and previous values. Differencing is the second way to achieve stationarity. It is possible to take differences finite number of times and create stationary time series.

Distribution of a random variable is a special order which reflects probabilities for all possible outcomes.

Dummy variable — a variable which can accept only two possible values. Usually they are denoted as 0 and 1. Dummy variables are a special case of discrete variables. They are used to formalize quality features and characteristics.

Endogeneity — a violation of the fourth Gauss — Markov assumption: regressors have not to be correlated with error terms. Endogeneity leads to biased and inconsistent estimators. There are four main sources of endogeneity: omitted variables, specification errors, measurement errors, simultaneity.

Error term — residuals in regression models. In fact, it is a difference between actual and fitted values.

Forecasting — a process of prophecies creation. Possible values of current and future observations can be found with the help of regression models

and statistical estimators. There are different types of forecasts. It is possible to present dynamic and static forecasts.

Granger causality — a case when lags of one variable determine dynamics for other time series. Granger causality includes not all types of interactions. There are three possibilities: variables are not connected; one variable determines another one, and inverse statement is not true; bilateral interaction. Granger causality can be caught with special tests.

Heteroskedasticity — a violation of the sixth Gauss — Markov assumption. It means that a variance in the error term is not constant. In fact, it is a function of regressors.

Impulse response function — a function, which reflects how one variable changes in response to shocks in another one.

Information criteria — a special statistical value which indicate the model's quality. These criteria are calculated with the help of SSR (sum of squared residuals) and a total number of observations. The best model has least values of information criteria. In general, criteria have less statistical power than asymptotic tests.

Innovation — a shock in a variable. If there is a dynamic system (or equation), such innovation can lead to changes in other variables. Mutual changes can be analyzed via impulse response functions.

Lag operator — a short method to present lags in the model. Usually it is noted with the letter “ L ”. For any time series L^0 is a unity and means y_t . Then y_{t-1} will be $L y_t$, y_{t-2} will become $L^2 y_t$ etc.

Long-run equilibrium — a steady state in the system. It exists if variables are connected with each other. The model has to generate a process converging to long-run equilibrium.

MA process — a process, which can be described as dependence of current values from actual and previous error terms. MA process can be constructed only for stationary time series. In general view it does not include any constant terms (a mean for error terms has to be equal to zero). Potential lags for this process can be chosen with correlogram (PACF) and statistical tests.

Maximum likelihood estimator (MLE) — one of possible statistical estimators. It works with likelihood function which replicates the sample with the highest probability. In case of *MLE* and right specification of a likelihood function we have consistent, unbiased (asymptotically) and efficient (asymptotically) estimators.

Mean absolute deviation (MAD) — one of possible indicators of the forecast's quality. It reflects a mean absolute difference between actual and predicted values:

$$MAD = \frac{\sum_{t=1}^T |y_t - \bar{y}_t|}{T}.$$

The best forecast has minimal *MAD*.

Mean absolute percentage error (MAPE) — one of possible indicators of the forecast's quality. It calculates a mean absolute difference between actual and predicted values in percents:

$$MAPE = \frac{\sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{y_t}}{T}.$$

The best forecast has minimal *MAPE*.

Mean percentage error (MPE) — one of possible indicators of the forecast's quality. It calculates a mean difference between actual and predicted values in percents:

$$MPE = \frac{\sum_{t=1}^T \frac{y_t - \hat{y}_t}{y_t}}{T}.$$

This value reflects the presence of biases in the forecasts. If it is not equal to zero, the forecast generates higher or lower values than actual observations (the direction of this bias can be found with the sign of *MPE*).

Mean squared error (MSE) — one of possible indicators of the forecast's quality. It calculates a mean square of differences between actual and predicted values:

$$MSE = \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{T}.$$

In fact, it is a variance of these differences. The best forecast will have the least variance in the error term.

Multicollinearity — a case when explanatory variables are strongly correlated with each other. It leads to inconsistent and biased estimators.

Null hypothesis — a general statement which has to be tested. Usually it formulates in terms “there is no difference” (“is equal to”).

- Ordinary least squares (OLS)** — a statistical estimator. It chooses model parameters to minimize squares of residuals. When all Gauss — Markov assumptions are satisfied, the OLS is the best linear unbiased estimator.
- Partial autocorrelation function (PACF)** — indicates correlation relations between observations from different periods. This function does not include additional correlations from moments placed between studying data. PACF helps to find potential MA lags.
- Random walk** — a stochastic process which cannot be predicted. Changes in future values are determined only by shocks presented in the error term.
- Reduced form (of the VAR model)** — does not include current values in the regressors' list. This transformation helps to avoid problems in the estimation process (identification problems etc.).
- Rejection rule** — an order stating when the null hypothesis has to be rejected. In general sense it can be written as follow: when the observed value is in the critical region, the null hypothesis must be rejected.
- Residuals** — a difference between actual and fitted values in the model.
- Seasonality** — a deterministic component in time series. It presents a pattern determining dynamics for various moments within a particular period (for instance, a year). It leads to additional fluctuations and a higher variance in the data.
- Seemingly unrelated models (SUR)** — a set of connected equations which have different explanatory (in particular, exogenous) variables. These equations can be estimated separately, but better results will be obtained in case of system estimation. If to ignore this fact, it is possible to face endogeneity.
- Simultaneous equation models (SEM)** — a system of equations including the same variables. These equations have to be estimated simultaneously.
- Specification** — a general view of the model. It includes a functional form and a set of regressors.
- Stability condition** — a restriction for possible coefficients. It must be satisfied to guarantee convergence in the process and find a single set of estimators.
- Standard residuals** — residuals corrected for conditional heteroskedasticity.
- Stationarity** — a key concept in time series analysis. It means that time series have constant mean, variance and correlation between observations on the same distance. In economic sense stationarity characterizes a system where shocks do not accumulate.
- Structural break** — a change in deterministic components of time series (trend, or intercept, or both).

Structural form (of the VAR model) — indicates structural interactions between time series. It includes current values of variables into the right-side part of the equation. Sometimes it is impossible to find a single set of estimators for this model. Then the reduced form must be applied.

Trend — a deterministic component in time series. It characterizes long-term dynamics of time series, which is possible to find a trend function, estimate it and decompose the data.

Unbiasedness — one of main statistical properties. It states that an expected value of the estimator is equal to its real one.

Unit root — a feature of non-stationary processes. It can cause problems with estimation procedures. As a result, time series with unit roots have to be transformed or estimated with special methods (for example, cointegration analysis can be applied for non-stationary series).

White noise — a special kind of distribution which is desirable in the error term. It requires a zero mean, a constant variance and independent identical distribution. If this assumption is violated, it becomes impossible to test hypotheses, and statistical properties of estimators worsen.

Учебное издание

Мариев Олег Святославович
Анцыгина Анастасия Леонидовна

ПРИКЛАДНАЯ ЭКОНОМЕТРИКА
ДЛЯ МАКРОЭКОНОМИКИ
APPLIED ECONOMETRICS
FOR MACROECONOMICS

Учебное пособие

На английском языке

Зав. редакцией *М. А. Овечкина*
Редактор *В. И. Первухина*
Корректор *В. И. Первухина*
Компьютерная верстка *Н. Ю. Михайлов*

План выпуска 2014 г. Подписано в печать 29.10.2014.
Формат 60×84 ¹/₁₆. Бумага офсетная. Гарнитура Times.
Уч.-изд. л. 8,5. Усл. печ. л. 8,8. Тираж 150 экз. Заказ № 1618.

Издательство Уральского университета
620000, Екатеринбург, ул. Тургенева, 4
Отпечатано в Издательско-полиграфическом центре УрФУ.
620000, г. Екатеринбург, ул. Тургенева, 4.
Тел.: +7 (343) 350-56-64, 350-90-13.
Факс: +7 (343) 358-93-06.
E-mail: press-urfu@mail.ru

