INSTITUTE OF AGRICULTURAL
AND FOOD ECONOMICS
NATIONAL RESEARCH INSTITUTE

Essential econometric methods of forecasting agricultural commodity prices

no 90.1
Warsaw 2013

MULTI-ANNUAL PROGRAMME
2011 - 2014

COMPETITIVENESS OF THE POLISH FOOD ECONOMY UNDER THE CONDITIONS OF GLOBALIZATION AND EUROPEAN INTEGRATION

Essential econometric methods of forecasting agricultural commodity prices

INSTITUTE OF AGRICULTURAL
AND FOOD ECONOMICS
NATIONAL RESEARCH INSTITUTE

Essential econometric methods of forecasting agricultural commodity prices

Editor:
Mariusz Hamulczuk, PhD, IAFE-NRI

Authors:
Mariusz Hamulczuk, PhD, IAFE-NRI
Sylwia Grudkowska, MSc, NBP
Stanisław Gędek, PhD, Professor of Rzeszow University of Technology
Cezary Klimkowski, MSc, IAFE-NRI
Stanisław Stańko, PhD, Professor of IAFE-NRI

MULTI-ANNUAL PROGRAMME 2011 - 2014

COMPETITIVENESS OF THE POLISH FOOD ECONOMY UNDER THE CONDITIONS OF GLOBALIZATION AND EUROPEAN INTEGRATION

Warsaw 2013

This publication was prepared as a contribution to the research on the following subject
Economic modelling in the analysis of competitive growth of agri-food sector
within the framework of the research task *The forecasting system aimed at increasing the competitiveness of the agri-food sector*

The aim of the study was to present selected econometric models which can be applied for forecasting of agricultural commodity prices.

Reviewer
Jacek Bednarz, PhD, John Paul II Catholic University of Lublin (KUL)

Prepared for print by
Mariusz Hamulczuk

Proofreader
Joanna Gozdera

Cover Project
AKME Projekty Sp. z o.o.

ISBN 978-83-7658-422-5

*Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej
– Państwowy Instytut Badawczy
ul. Świętokrzyska 20, 00-002 Warszawa
tel.: (22) 50 54 444
faks: (22) 50 54 636
e-mail: dw@ierigz.waw.pl
http://www.ierigz.waw.pl*

Contents

Introduction

Experience has shown that in forecasting of agri-food sector the statistical and econometric tools are relatively rarely used. Most of the analyses and forecasts for this sector are based on experience of experts. We do not claim that the methodology used is inadequate, and that the effects are of poor quality. The problem is that it is unclear what conditions and assumptions form the basis of the conclusions. Therefore, subjective opinions should be complemented with knowledge about certain objective regularities.

The aim of the book is to present possibilities of analysis and forecasting agricultural commodity prices on the basis of quantitative methods. On the one hand it can be done with the use of time series models on the basis of the patterns observed in the past prices. On the other hand model base on causal relationships with other phenomena in the economic system can be employed. Practical application of above methods raises a lot of questions concerning data availability, identification of the most important factors determining the level of agricultural prices, the choice of suitable model or extrapolation of the patterns and relationships into the future. So the advantages and limitations of commonly used forecasting models have to be known.

The study was carried out within the Multi-Annual Programme: "Competitiveness of the Polish food economy in the conditions of globalization and European integration" in the task "The forecasting system aimed at increasing the competitiveness of the agri-food sector". Its objective is to provide methodological basis for the use of quantitative methods in forecasting agricultural commodity prices. The book is addressed to practitioners who deal with forecasting in agribusiness: market analysts, policy makers or market agents. As the readers or potential forecasters have different levels of knowledge in statistics and econometrics, the nature of methods and applications presented here has a varying degree of sophistication.

The material in this text is divided into five chapters. Chapter 1 presents fundamentals of agricultural commodity prices formation. It describes demand and supply conditions, the role of global markets in domestic price determination, the impact of economic policies on the price level and the nature of the relationships between agricultural markets.

Chapter 2 contains background and method of time series forecasting. It presents general characteristics of patterns existing in agricultural commodity time series, ways to measure them, conditions for extrapolation of these patterns into the future and a general introduction to time series methods. The selected time series methods, like econometric models, exponential smoothing models, ARIMA models, were presented in a concise way. Some attention was paid also to application issues.

Chapter 3 covers issues of measuring patterns occurring in time series using the X-12-ARIMA and TRAMO/SEATS methods. It develops the considerations of chapter 2 with modern analysis and forecasting tools included in DEMETRA+ software. The last part of this chapter focuses on empirical analysis of agricultural commodity prices with the use of seasonal adjustment methods including forecasting with the use of RegARIMA models.

The fourth chapter discusses methodological aspects of the analysis and forecasting of short-term price changes with the use of models with explanatory variables. It shows the basic specification of static and dynamic models, methods of testing the properties of stochastic processes as well as vector autoregression models and vector error correction models. At the end of the chapter the problems of application selected models for short-term forecasting of agricultural prices are discussed.

The last chapter focuses on partial equilibrium models, which provide the basis for forecasting and simulation of the agricultural sector. It presents the theoretical foundations of equilibrium models, pros and cons of these models and their structure. Three partial equilibrium models are discussed in details: AGLINK-COSIMO, FAPRI and AGMEMOD model. The chapter ends with discussion on medium- and long-term forecasting of domestic prices on the basis of world price projections.

The authors would like to thank Jacek Bednarz, PhD. for his valuable comments and suggestions that were given to the original text.

1. Factors determining agricultural commodity prices

Forecasting of agricultural commodity prices seems impossible without the knowledge of economic factors underlying price formation. Knowledge of economic theory provides guidance in reconsidering both the data collection and the analysis performing. Agricultural markets are subject to the law of supply and demand, yet there are certain specificities of the agricultural market resulting in greater price volatility than on markets of other products. On the basis of literature we can conclude that the agricultural commodity prices result from [Ferris 2005, Hill 1990, Ritson 1977]:

- law of supply and demand,
- biological and technical nature of agricultural production,
- indirect links between farmers and the consumers,
- cross-commodity markets links,
- linkage of domestic prices with global prices,
- macroeconomic factors – especially agricultural and trade policy measures.

The most important rule in price formation is the law of supply and demand. The price is the result of the equilibrium between market demand and the market supply for a given commodity. In this system, prices play the role of a regulator of market processes. When we are dealing with demand that is not fully met, prices increase and when production (supply) surpluses are evident – prices decrease. Under pure competitive condition the price is determined by market situation. Along with the growing importance of factors of institutional nature the role of market as processes' regulator is weakening.

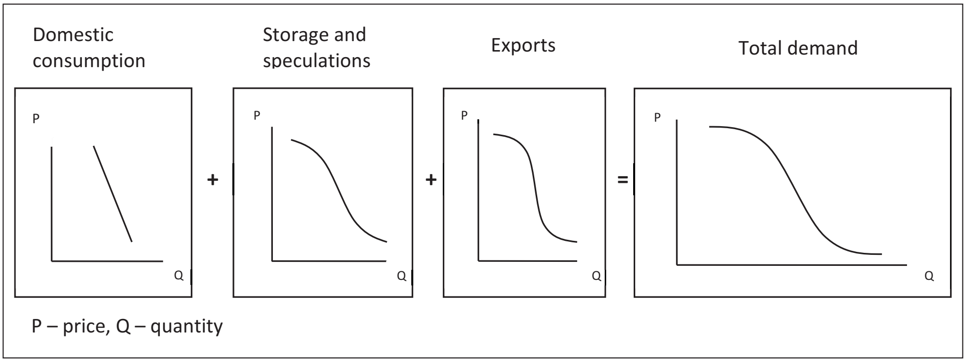
1.1. Demand shifts

Demand is defined as a quantities of a given good that buyers are willing to purchase at alternative prices during given period of time, *ceteris paribus*. *Ceteris paribus* assumption states that everything else remains constant. Demand for agricultural products can be analysed from the point of view of the whole sector or demand may be considered at a more dis-aggregated level, i.e. for a given commodity.

Demand components

Demand for a given commodity exists when people have desire for the commodity coupled with the willingness and ability to purchase. Market demand may have different sources depending on the direction of the use of the given agricultural commodity (Fig. 1.1). It can be assumed that domestic demand can be decomposed into demand for: domestic consumption, storage, speculation and export (losses have not been taken into account) [Ferris 2005]. These different elements constituting the total demand for agricultural raw materials indicate the factors responsible for demand formation.

Figure 1.1. Demand components for agricultural products



Source: elaboration based on Ferris 2005, p. 9.

The domestic consumption usually has the largest share in the total demand for agricultural products. This group is heterogeneous, however, and includes demand of different nature: food use, feed use or industrial use. The food utilisation constitutes the largest share of the domestic consumption of food especially in the case of fruits and vegetables, meat and milk products. For some products the high share constitutes feed utilization (grains, oilseeds). Along with economic development more and more agricultural products become an important raw material in industrial processing. Agricultural products may be used for production of ethanol, dextrin, glues and other technical preparations used in textile and paper industry as well as metallurgy.

Quite a large part of demand for agricultural commodity may result from storage and market traders' activity. For some products demand for storage and speculation may even exceed the ultimate demand for consumption. On highly perishable goods the consumers may predominate but on storable and seasonal products (feedstuff) storage may be more important. Certain entrepreneurs may temporarily increase demand by storing commodities, predicting future price increase. This type of demand bases on anticipation of future changes affecting the current prices of commodities. This is why market agents' expectations play an important role in short-term pricing.

In the open economy demand for domestic commodities depends on demand and supply relationship (and prices) on foreign markets. If domestic prices are lower than world prices there is a tendency to the increase of export of commodities. A reverse situation (world prices lower than domestic ones) may lead to an increase of commodities' import. The competitiveness of domestic products in foreign markets depends however on exchange rates' changes.

Demand conditions

Generally, it may be assumed that aggregated domestic demand for the agricultural products may be determined by factors influencing particular components of the total demand: domestic consumption, storage and speculation, and exports. The most important ones are as follows [Ritson 1977, Ferris 2005]:

- current and expected prices of a given commodity,
- availability, and current and expected prices of substitute and complementary commodities,
- domestic institutional arrangements,
- possibility of storage and its costs,
- world market prices and transportation costs,
- trade barriers,
- income and its distribution,
- population and its demographic structure,
- consumer habits, tastes, and preferences.

Demand elasticity and flexibility

The price dynamics and the process of demand and supply adjustments are determined by the shape and slope of demand and supply curves. This is reflected by the so-called price elasticity of demand and supply. It measures the response of demand (supply) to a change in its own price. It gives the percentage change in quantity demanded in response to a one percent change in price, *ceteris paribus*. Price elasticity may be measured both as a point elasticity or arc elasticity.

Price elasticity of demand for goods depends on several factors. These factors are somewhat related to factors influencing different components of the total demand and the *ceteris paribus* conditions mentioned when price elasticity was defined. In general, products that are difficult to substitute by others are characterised by low price elasticity of demand. If there are no close substitutes, the substitution effect will be small and the demand inelastic. More elastic is the demand for products having close substitutes (e.g. substitution of certain fruit with others or certain vegetables with others). There is also the problem of definition connected with the level of aggregation of the variable being analysed. Domestic demand for food may be extremely price-inelastic whereas domestic demand for particular products is characterized by relatively higher price elasticity. This is connected with necessity and existence of substitutes. The more necessary a good is, the lower the price elasticity. There is no substitutes for food and people will attempt to buy it no matter the price whereas specific food products have some substitutes.

The second factor influencing price elasticity of demand is a share of the consumer's income that the product's price represents. When the given product represents only a small portion of the budget the income effect will be insignificant and demand inelastic. The price (and income) elasticity of demand in developing countries is higher than in the developed ones. This pattern is resulting from the Engel's law. Along with the increase in income the share of expenditure on food reduces. Even if some part of income increase is spent on food, this does not so much concern the increase of quantity but rather the increase of demand for more processed food. When identifying demand for agricultural products, a consumer, their needs, preferences, tastes, budget limitations, etc. should be the starting point. With the increase of income of the population not only the share of expenditure on food in total expenditure decreases, but the demand structure changes as well. Generally, demand for food is inelastic as compared to income and changes in long- rather than short-term.

The third factor affecting the price elasticity of demand is a market level. Demand for agricultural products is dual in nature: direct and indirect. Direct demand (final demand) originates mainly from households purchasing products for consumption, e.g. bread, pasta, cured meat, cheese. Agricultural products may also be used to satisfy direct food needs, e.g. fresh potatoes, fruit, vegetables or flowers. However, along with the economic development of countries the importance of direct consumption falls. The main part of demand for agricultural products therefore is the derived (indirect) demand. Consumers rarely buy food directly from farmers. Instead farmers sell raw materials to marketing service providers, who store, process, transport, and otherwise add utility, and who sell it to the consumers. As a result the price elasticity of demand is different at the farm gate, processing plant or retail store level [Drummond, Goodwin 2004; Heijman *et al.* 1997].

For example, demand for raw milk is less elastic than demand for cheese or yogurt. Before raw milk becomes fit for consumption, it is a subject to multiple transformations. Processing of agricultural products requires utilisation of additional production factors so raw milk is only a part of the final product (yogurt). The difference between retail value and farm value of the product is called marketing margin. Generally, along with the increase of the degree of food processing (marketing margin) the price elasticity increases.

In theory of demand it is important to distinguish between the short- and the long-term price elasticity of demand. Generally, demand for agricultural products is relatively more elastic for periods within the storage life of a product than for periods than exceed the storage life. Moreover, demand for agricultural commodities is more elastic over long-term perspective than over shorter periods that still exceed the storage life of products [Goodwin 1994]. As a result, monthly demand for a given product may be relatively more price elastic than annual demand. What is interesting is the fact that a multiyear demand also tends to be more price elastic than annual demand. Demand for storage (short term) and consumers' adjustment possibilities (important in multiyear horizon) play very important role here.

Price elasticity of demand is concerned with the responsiveness of quantity demanded by buyers (consumers) to the prices changes of a given commodity. Forecaster, however, is concerned with the variability that might be expected in the price as a result of change in the quantity of the product [Goodwin 1994]. The inverse of elasticity would measure the sensitivity of price to the quantity change and is called a price flexibility. The more inelastic price elasticity of demand, the more volatile prices. Prices of most agricultural products are highly volatile because their price elasticity is relatively low.

Derived demand, price spreads and price transmission

As mentioned above, demand for agricultural commodities can be considered as derived demand from retail demand. Therefore, retail demand may be considered as composite demand for agricultural raw materials and marketing services and materials. The difference between retail and farm prices is called as price spreads. They are a function of the marketing materials and services. The higher farm-retail spreads, the less elastic farm-level demand for domestic food and the higher agricultural price flexibility in respect to the domestic demand [Drummond, Goodwin 2004; Tomek, Robinson 2003].

The question is, where the prices of agricultural product are formed: at the farm-level or at the consumer-level. Waught [1964] stresses that in the short-term the retail price movements should reflect the price changes on the farm level. However, price response (transmission) at retail or wholesale level is not immediate and complete. Deviations caused by various factors may be noted here. Generally, shorter period of reaction is noted on the markets with large share price-spreads or when the form of final product differs insignificantly from the raw material obtained in agricultural holdings (e.g. fresh fruit and vegetables). When final products contains a large share of components related to processing and trade, the reaction may be quite weak.

Waught [1964] indicates that in the long-run the consumer demand dominates over supply. Therefore, farm prices in the long-term are determined at the retail market. Results of empirical research show that there is no immediate and straight rule of price formation in the marketing chain. This results from simultaneous, but also different in nature movements of agricultural prices, non-agricultural raw materials used in production, trade and processing and food products. They may vary depending on the shift in retail demand and supply of inputs of agricultural or non-agricultural origin [Gardner 1975].

When forecasting prices in food marketing chain different approaches are needed in respect of the forecasting horizon. Forecasting monthly data (short-term horizon) we should start with farm prices and include costs of marketing materials and services to calculate retail price forecast. When forecasting in the long term the farm price is a retail price minus all marketing costs.

1.2. Supply shifts

The supply of a commodity can be defined as the quantity that producers are willing and are able to offer at alternative prices in a given time period, *ceteris paribus*. We can say that the supply is a relationship between prices and offered quantities. A key difference between demand and supply analysis is the distinction between current prices (demand) and expected prices (supply). Due to biological lag in an agricultural production expected prices predominate in demand analysis [Ferris 2005].

Supply conditions

Based on literature [i.e. Ritson 1977; Tomek, Robinson 2003; Varian 1999; Ferris 2005] the following factors influencing the supply of agricultural products can be mentioned:

- price of a given product,
- prices of other products that can be produced (supplied),
- state of production technology,
- prices of factors of production,
- level of fixed resources,
- the weather,
- the objectives of producers (firms),

- number of producers,
- institutional arrangements, etc.

The quantity of specific agricultural products offered for sale depends on expected prices of the product being considered and the range of *ceteris paribus* conditions. The supply of agricultural products is generally more volatile than demand as the price and cost expectations of market agent play a crucial role in estimating the future supply. Also *ceteris paribus* conditions of supply are subject to higher variability than those of demand. Therefore, crucial to forecasting of agricultural prices is an anticipation of the impact of changes in the *ceteris paribus* conditions and future supply based on expected prices.

Price elasticity of supply and flexibility of prices

The responsiveness of producers in terms of output to changes in the prices of their product is measured by the price elasticity of supply E_s. It demonstrates the percentage change in quantity supplied of a given commodity to one-percent price change of that product.

Impact of the change of quantity (supply or demand Q) on prices P is reflected by the price flexibility coefficient F_i [Tomek, Robinson 2003]:

$$F_i = \frac{\Delta P}{\Delta Q} * \frac{Q}{P}. \tag{1.4}$$

where:

Q – quantity,

ΔQ – change in quantity,

P – commodity price,

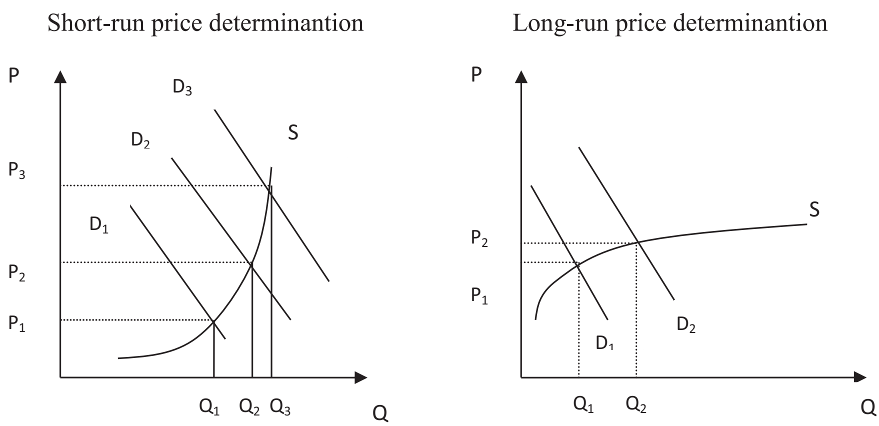
ΔP – price change.

It is an approximation of the reverse price elasticity coefficient and expresses the percentage change of prices due to changes in quantity by 1 percent, applying the *ceteris paribus* rule. High flexibility of prices in respect of quantity supplied is one of the most distinguishing characteristics of the agricultural commodity market.

Short- and long-run elasticity of supply

Price elasticity of supply of agricultural products depends on the time frame being considered. In the short-term, i.e. one production cycle, price elasticity of supply is close to zero, which results from the specific nature of production. Fixed costs in agriculture have a high share in total costs and, once investments are made, the producer cannot limit costs by decreasing supply. Even under conditions of prices decreasing below the total costs, the producer can increase production if variable costs are lower than prices. It is rational because it leads to a reduction of losses as the prices are higher than variable costs. This type of a response is so-called reverse reaction effect and must be treated as short-term and exceptional. In the case of arable production, when plants are nearly ready for harvest, change of the product price affects supply to a small extent since most costs of cultivation have already been incurred and producers will continue the production till harvest time.

Figure 1.2. Short- and long-run price determination



Source: elaboration based on M. Radetzky 2010, p. 58.

Possibilities to increase supply of the agricultural raw material in a given production cycle are limited by the production potential. In plant production it will result from the surface area of crops and harvest. After harvest of plants, price increases may result only in a slight increase of supply resulting from changes of harvest allocation. Its size will be limited by harvest and stocks from the previous years. These conditions result from the fact that the supply of agricultural commodities in a given production cycle has a specific limit, the supply curve is vertical. Further increase of prices will not lead to an increase of supply. This affects the behaviour of balance prices in the short run. Small changes of supply with a given demand result in relatively large changes of short-run equilibrium prices (Fig. 1.2). In general, market equilibrium prices result from different factors in short- and in long-run.

Changes of supply of agricultural raw materials in longer perspective are possible through adjusting the potential to market signals. In the long-run, changes may be introduced in the production technology or production structure by adjusting it to market needs. In plant production the volume of production is determined by crop and harvest structure. High prices, e.g. of wheat may be a signal for producers to increase the area of crops and to intensify production. This may lead to the increase of production and supply in subsequent periods. Such processes make the supply curve become "flat" in long-run (greater than the production cycle). The longer the time frame, the greater the supply elasticity [Hill 1990, Radetzky 2010].

Supply and price movement over time

From the forecast point of view, it is important to know the dynamic aspects of price variations. In the long-term the forecasting of a trend is the most important issue. Trends in prices are caused by gradual changes in demand (mostly) and supply conditions (e.g. policy, incomes, and consumer preferences, prices of agricultural inputs or technology).

Prices of agricultural commodities cyclically and seasonally fluctuate around long-term trends. Such movements are of a great importance for medium- and short-term forecasting. Their source is mainly in supply side conditions. The main sources of cyclical fluctuations in the prices of agricultural raw materials include biological constraints, weather conditions, interactions between markets, market psychology or economic expectations. The length of the cycle depends on biological process of production in crop and animal production. In annual crop products the minimum length would be 2 years. The minimum length of cycle in livestock inventories is about four times longer than the time from the birth to the first reproduction. For pig production it is at least 3 years [Goodwin 1994]. The factual length could be considerably modified by weather conditions, market intervention or farmers reaction (expectations).

Due to the specific nature of agricultural production, a certain period needs to elapse from the moment of making a decision to launch production to the moment when the commodities appear on the market. Production is planned on the basis of the current situation (naive expectations), past and current condition (adaptive expectations) or on the basis of past, current and future (predicted) situation (model of rational expectations). Market supply is therefore always, to a greater extent, a function of the past rather than the current prices. This mechanism (known as a cobweb model) results in cyclical fluctuations of production and prices of agricultural commodities.

According to this mechanism, a decision to increase agricultural production is a response to high prices in a period of shortages. To expand production (herds or area sown) some amount of current production is withdrawn from the market causing further increase of prices. High commodity prices encourage farmers for further expansion of production but some period will elapse from launching production to the moment when production reaches the market. Increasing supply leads to a fall in prices as it exceeds consumer demand. Reaction of farmers is often to increase production in order to achieve profits by increasing the scale of production. These decisions, which are rational from the microeconomic point of view, cause a further decline of market prices and producers are recording losses. As a result, the farmers decide to reduce production, which in subsequent periods will result in the increase in market prices [Tomek, Robinson 2003].

Cyclical fluctuations in the supply of agricultural commodities cause high volatility of their prices. Despite greater knowledge of mechanisms among agricultural producers, greater availability of market information and stabilisation measures, these fluctuations are still present. Therefore, they should be accepted and knowledge of these mechanisms should be used for forecasting.

Another type of fluctuation typical for agricultural prices is a seasonality which is a regular movement observed within production period. Price seasonality is a consequence of the variation of labour intensity, supply and market trade. Seasonal nature of supply, and thus prices, has its origin in biological process of production which is strictly related to the temperature. Most crops are planted in the spring and harvested in the fall. But the product has to be available between harvests. So the more limited possibility to store (and higher costs) the commodity the higher seasonal variation. Seasonal patterns are noted not only in plant prices but also and in livestock products prices. For example, the lowest live pig prices are quoted in the 1st quarter of the year, whereas the highest ones in 3rd quarter. Such fluctuations are caused mainly by seasonality of the costs of production (energy, fodders, etc.).

1.3. Global conditions

Integration and globalisation processes make price formation of agricultural commodities in a given open economy more complex than in the closed economy. For most countries domestic commodity prices are a result not only of the supply and demand relation in the country, but also of the impact of the situation on the so-called world markets. The notion of world market shall mean a country or region having a considerable share in international trade and therefore the supply and demand situation on that market has significant impact on the formation of global prices.

Attention should be paid to the fact that not all the commodities produced in various countries are subject to international trade to the same extent. In general, raw materials and products which can be stored and those which are more processed are subject to wider international trade than perishable products. So the impact of the world markets on domestic one can vary considerably among the products.

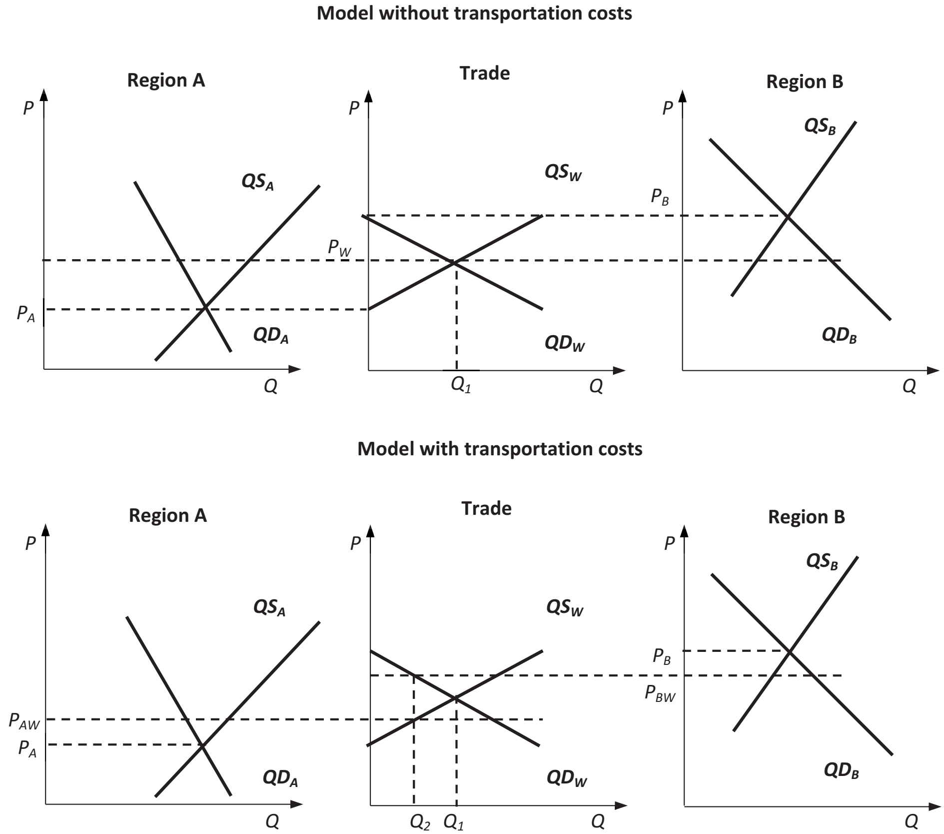
Spatial equilibrium model

A useful analytical structure used frequently to present the price formation mechanism under conditions of international trade is the spatial equilibrium model. Spatial equilibrium model can be defined as the model solving the simultaneous equilibria of plural regional markets under the assumption of existence of cost of arbitrage (transportation costs among them) between two regions. It focuses on the notion of competitive equilibrium and Pareto efficiency manifested in zero marginal profits to arbitrage. At the heart of most analyses of market integration lays the Enke-Samuleson-Takayama-Judge (ESTJ) spatial equilibrium model dated to Enke [1951], Samuelson [1952], and Takayama and Judge [1971]. The model is also used to assess the impact of the change of trade policy instruments on market prices and welfare.

Figure 1.3 shows how the market equilibrium of two markets (countries, regions) is determined when one is characteristic of surpluses and the other of shortages. Linear functions of supply and demand were adopted for demonstrational purposes. When there is no trade, equilibrium prices on these markets are P_A and P_B respectively. Therefore, there is significant difference between prices for products in the two markets. This has certain consequences both for producers and consumers. Producers in region A obtain less money than in region B for one unit of the same product sold but simultaneously consumers pay less than consumers in region B.

As far as trade is concerned (the lack of trade barriers between these spatially integrated markets is assumed), the equilibrium price is established on the new level of P_W. We assume that the world price is determined by these two markets (A and B). If there are no transportation costs (transport, loading, unloading), equilibrium price will be set at the level at which the global demand equals global supply. The price is determined by the amount of surplus of supply in the country with a lower domestic price (QS_W=QS_A-QD_A) and the surplus of demand in the country with a higher price (QD_W=QD_B-QS_B) in the respective countries. In such a case Q_1 quantities of a given commodity will be traded.

Figure 1.3. Spatial equilibrium model (two regions, assumption of a large country)



Source: elaboration based on Figiel 2002.

Law of One Price

Assuming the lack of transportation costs (and other barriers), it should be expected that the world price P_W will be accepted by producers and consumers in both countries. Determination whether the price of a given product is at the same level in the whole world is called Law of One Price (LOP) in economic theory. For it to really work in a strong form, lack of transportation costs, lack of state interference, lack of transaction costs and full information for market participants should be assumed. The LOP defines the extend in which spatial commodity market are integrated into a single market [Ardeni 1989, Figiel 2002].

International trade will result in changes in producer and consumer surpluses. It will lead to a decrease of consumer surplus and to an increase in producer surplus in region A. Simultaneously, consumer surplus increases and producer surplus decreases in region B. However, theory of international trade demonstrates that national price convergence with world prices signifies the increase of social welfare.

Equilibrium prices do not equal the theoretical world price P_W in a more realistic case. This results from the fact that quantity exported by country A to country B (imported by country B from country A) will be ultimately determined by transportation costs. Their level contains a certain fixed component related to transportation of commodity and a certain variable component related to the distance between the two markets. This means that the costs increase with distance. In the presence of transaction costs domestic prices will be established on P_{BW} and P_{AW} (Fig. 1.3). The transportation cost equals the difference between P_{BW} and P_{AW}. Transaction costs make the exported (imported) quantity determined at the level of P_W decrease from the level Q_1 to the level Q_2. Increase of welfare will also be smaller as compared to the hypothetical lack of such costs. If the transportation costs exceed the price difference on A and B countries' markets, trade will not take place.

In fact, price differences in separately integrated market may be higher (more often) or lower than transportation costs. This may be caused by the following factors [Tomek, Robinson 2003]:

- lack of complete information on prices and quantities offered and desired in the world,
- preferring commodities from a specific area (e.g. as a result of long-term agreements, maintaining old supply channels or out of habit),
- institutional and legal limitations to trade (for more details, see Chapter 1.4).

Theoretical analyses and empirical research make an assumption as regards the size of a country (region) determining the impact on global prices. In the case of a "large country", its export and import on international markets has considerable share in the world market of a given commodity (greater level of Q in Figure 1.3). The supply and demand situation on a given large economy market and agricultural or trade policy may affect the level of global prices in this case (countries are price makers).

The changes of the supply and demand situation as well as the economic policy have no significant impact on global prices in the case of a "small country" assumption. This means that consumers and producers take the world prices as domestic ones. This is the case of perfect competition market, when the economy is open, free from trade barriers and when the lack of transaction and transport costs is assumed.

Price transmission

The above-mentioned Law of One Price is a useful theoretical construction for analysing relationships between spatially separated commodity markets. Weak form of the Law of One Price is manifested in the relationships according to which observed price differences in regions may diverge from transportation costs, but spatial arbitrage will cause the difference between the two prices to move towards the transportation cost [Baulch 1997; Tomek, Robinson 2003].

Global prices and exchange rates are the basic parameters determining directly the commodity prices in the small open economy. The two elements may formally be expressed as [Ardeni 1989]:

$$P_t^{ij} = \alpha \cdot E_t^{ij} \cdot fP_t^{ij}, \tag{1.2}$$

where:

t, i, j – refer to, respectively: time, country to which prices are compared and the commodity in question;

P_t^{ij} – domestic prices of a commodity expressed in domestic currency;

α – transmission parameter, *1-α* expresses departure from the LOP and impact of transfer costs, long-run difference between domestic and global prices;

E_t^{ij} – exchange rate as a value of foreign currency of country *i* expressed in domestic currency;

fP_t^{ij} – world prices of commodity *j* expressed in the currency of country *i*.

Due to a relatively small production and consumption potential, Poland is regarded as small economy. Consequently, Poland is assumed to be the world price taker. This is confirmed by empirical research [Rembeza 2010]. This means that changes in the domestic supply and demand situation have relatively small impact on the level of domestic prices of most agricultural commodities in Poland. In consequence, there is a need for careful observation of conditions underlying global and European prices of agricultural commodities. From the forecast point of view, it would be more justified on numerous commodity markets to forecast domestic prices as functions of world prices and the exchange rates.

The concept of price transmission can be thought of as being based on three components [Rapsomanikis, Hallam, Conforti 2003]:

- co-movement and completeness of adjustment which implies that changes in prices in one market are fully transmitted to the other at all points of time;
- dynamics and pace of adjustment which implies the process by, and rate at which, changes in prices in one market are transmitted to the other market;

- asymmetry of response which implies that upward and downward movements in the price in one market are symmetrically or asymmetrically transmitted to the other.

The complete price transmission between two spatially separated markets is defined as a situation where changes in one price are completely and instantaneously transmitted to the other price. If price changes are not passed-through instantaneously, but after some time, price transmission is incomplete in the short run, but can be complete in the long run. Changes in the price at one market may need some time to be transmitted to other markets for various reasons, such as policies, the number of stages in marketing and the corresponding contractual arrangements between economic agents, storage and inventory holding, delays caused in transportation or processing. In the short run asymmetric price transmission may also occur for different reasons resulting in different reaction of the domestic price on increase and different reaction on decrease of world prices [Rapsomanikis, Hallam, Conforti 2003].

Changes of the exchange rates also affect the income situation of agricultural producers, profitability of not only enterprises importing and exporting their commodities, but also enterprises carrying out activity solely on their domestic market. Agricultural producers in Poland are exposed to the exchange rate risk and thus the income risk at least in three domains. Firstly, the current Common Agricultural Policy of the European Union, namely direct payments to agricultural holdings, makes the level of support obtained dependant on the exchange rate the European Central Bank used to convert payments. Secondly, the exchange rate affects the level of prices obtained for products sold. Thirdly, the exchange rate determines the level of costs.

One of the results of the LOP is the reduction of impact of internal shocks on domestic prices. Possibilities to restock shortages or to sell off market surpluses on foreign markets decrease variability of domestic commodity prices and raise the effectiveness of economic entities. On the other hand, the impact of global disturbances is higher. It seems that consumers are the main beneficiaries of trade, while producers may be exposed to extension of the range of risk factors in certain periods. This follows from that fact that in the open economy the natural hedging, manifesting itself in the decrease of prices in the period of overproduction and vice versa, does not work to the extent possible in a closed economy.

1.4. Government policies

Government and agriculture

Market prices result not only from pure supply and demand game in the country, region or the world. The agri-food sector is one of those sectors that are subject to strongest regulatory processes. Implementation of different agricultural and trade, as well as macroeconomic policies is justified by economic and political factors. In concept, government interventions work to facilitate market competition and to help the market to achieve the national policy objectives. The argument, which is most often raised when opting for intervention, involves market failures. The economic theory of market failure seeks to account for inefficient outcomes in markets that otherwise conform to the assumptions about markets held by neoclassical economics. When failure happens, less welfare is created than could be created given the available resources. Many social welfare programs find their theoretical justification in market failure or in other violations of the neoclassical market assumptions.

Among the reasons of state intervention are: assuring market information flows, combating externalities, providing public goods, controlling non-competitive behaviour or changing of income distribution. It is stressed that without public intervention free market would lead to socially inacceptable income disparities and thus to weakening of processes related to the development of agriculture as compared to other sectors [Kowalski, Rembisz 2005]. The aim of market intervention in the area of agriculture and food economy is also to: ensure sustainable growth (social and environmental issues, external effects and public goods), assure self-sufficiency in production, limit price and income volatility or ensure reasonable food prices for consumers [Tomek, Robinson 2003].

Under market intervention, we are dealing with the competitiveness of the state against the market in regulatory functions. Disturbances in free market mechanisms in the form of state intervention lead to the change of allocation of factors of production, their valuation (e.g. income) and affect the level of market prices. Producers receive other market signals than in a free market economy, which leads to alternation in production decisions. The fundamental question concerns the relation of market failures and imperfection of governmental regulation [Rembisz 2010].

Criticism of the market failure notion and of using government to remedy market failure's effects is articulated in the public choice school of economics. Public choice theory is often used to explain how political decision-making results in outcomes that conflict with the preferences of the general public. It is stressed that politicians and interest groups influence government policy to benefit themselves or to achieve their goals (rent-seeking). This leads to higher government failures than market failure. It is generally recommended that governments should play a facilitating rather than a direct role in markets. For more information see: Gardner [1990] or Kowalski and Rembisz [2005].

Policy instruments may be divided into ones that contribute to the increase of prices (majority) or ones that lead to the decrease of prices. This classification is presented by Tomek and Robinson [2003]. The macroeconomic, trade and agricultural policy instruments may have direct or indirect impact on prices of agricultural and food products. Classification into income-related instruments (payments are not linked with production in the current CAP) and market instruments would provide another classification of policy tools. Tools that belong to the first group (direct payments, LFA payments, etc.) have small (negative), indirect and long-term impact on prices. Under conditions of income support, pressure on the increase of agricultural prices decreases. Income support leads also to the decrease of risk exposure of agricultural producers (the considerable share of income is almost market risk-free), so the pressure on the use of market intervention instruments decreases [see: Gardner 1990, Moschini, Hennessy 2000].

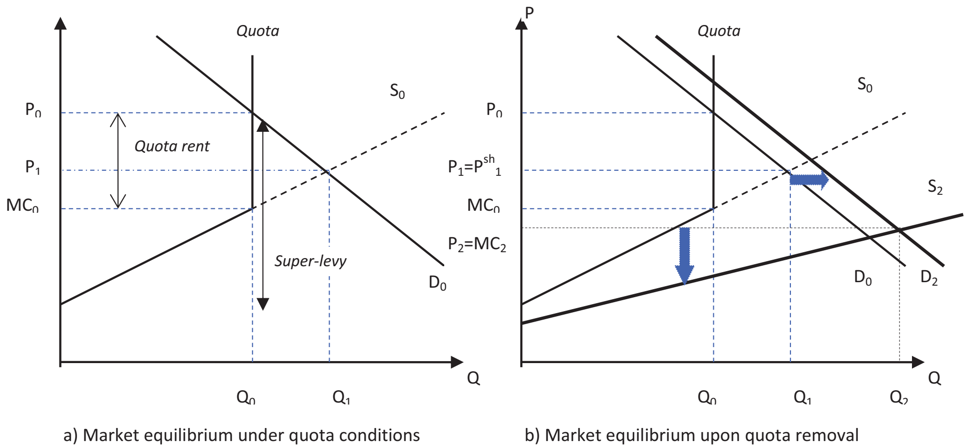
Various market intervention instruments have considerable impact on the level and volatility of agricultural commodity prices. The aim of these tools is a direct control of market supply of commodities (production quotas, import quotas, intervention purchases, non-tariff barriers, etc.) or direct control of their supply (e.g. burden import duties, export subsidies). On the demand side there are also instruments affecting the agricultural commodity prices. In recent years, such an example is the mandatory blending into transport fuels [Gardner 1990, Tyner 2010].

Production constraints

Mechanisms of selected instruments shall be presented below, simultaneously indicating their impact on prices. Figure 1.4 shows the impact of production limits on market equilibrium and market prices on the example of milk quotas. Land limitations or setting aside have similar impact. It should be stressed that the impact presented below is possible only in a closed economy or with other trade limitations.

Quantitative restriction of supply means that the market equilibrium is established at the level different than the one that would be reached in a free market economy. Under quota regime, a difference is noted between the P_0 price and the marginal cost MC_0, the so-called quota rent. Control of market supply by quota makes it possible for farmers to obtain a price for product (P_0) that is higher than the equilibrium price without this limitation. Milk quotas are the assets – they can be bought and sold, and there is a market for them. A production quota becomes "valuable", which is reflected by the difference between P_0 and MC_0.

Figure 1.4. Effect of abolishment of milk quota system



Source: own compilation based on: Gardner 1990; Réquillart et al. 2008; Hamulczuk, Stańko 2009.

Functioning of the production quota measures is possible due to penalties paid for exceeding the quota. The amount of levy is strictly related to milk prices and production costs (Figure 1.4a). The levy is established at such a level that income on additionally produced and sold product unit decreased by the amount of levy fails below marginal costs, which in turn leads to production constraint [Réquillart *et al.* 2008].

A shadow price P^{sh}_0 may be determined basing on prices for milk and quota rent. It should be at least equal to the marginal cost of milk production $MC_0 = P^{sh}_0$. It is a minimum market price at which farmers will produce milk in milk quota system. When market price is lower than marginal costs, production is not economically profitable for farmers and quota is not binding. This results in the limitation of production, which will be lower than the quota Q_0. When market prices are higher than MC_0, production equals quota Q_0 due to administrative limitations and levies for exceeding the quota [Hamulczuk, Stańko 2009; Réquillart *et al.* 2008].

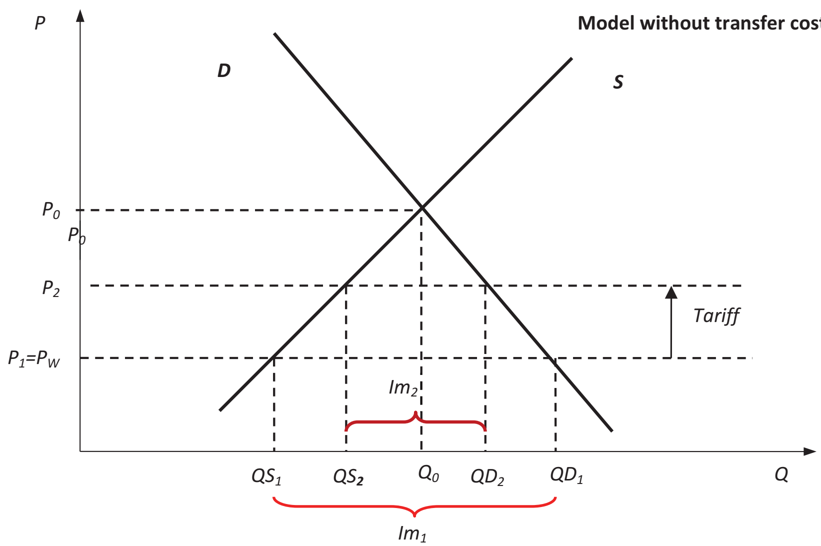
Production implications of quota abolishment may be examined in a short and long term. The short term means that producers are unable to lower their costs. The long term means qualitative changes consisting in the possibility to adjust production costs to the changing market situation. In the short term, removal of quotas results in a new market equilibrium which is established at the point of balance of demand and supply (P_1-Q_1). Once quotas are raised (or removed), artificial difference between the market price and the shadow price $P_1 = P^{sh}_1$ is not present.

Due to quota removal, in the long term market will experience additional mechanism resulting in shifting down the aggregated supply curve S_0. New supply curve (S_2) reflects the increase of effectiveness of the sector as a whole. Improvement of sector effectiveness, drop of prices, and increase of competitiveness of processors in the long term may affect the changes of consumer preferences, resulting in the shift or change of the demand curve. Possible changes are presented in Figure 1.4b in the form of curve D_2. Therefore, a new equilibrium is established in a point P_2-Q_2, where: $P_2 < P_1$ and $Q_2 > Q_1$. In this case, economic optimum is achieved, since the equilibrium price equals marginal cost $P_2 = MC_2$ [Hamulczuk, Stańko 2009].

Trade protection

Another way to maintain prices of agricultural commodities results from trade protection. Protectionism may be of various nature. The most commonly applied internal market protection instruments for counteracting cheap import include various types of import barriers, custom duties among them. The use of such a measures prevents foreign products from competing with products manufactured in a given country. Such limitations may be either continuous or introduced depending on the relation between the domestic price and the global price. Supply limitation on the domestic market of foreign products is also introduced by administrative tools and various non-tariff barriers. Quantitative limitations (quotas) in import are similar to customs duties because they both limit the supply of domestic raw materials. The supply from foreign markets may also be limited by various forms of administrative barriers hindering access to the internal market and decreasing total supply. These are specific technical standards or quality standards, procedures of their approval, sanitary standards, commodity packaging rules, etc., which must be complied with by foreign commodities [Gardner 1990; Hill 1990; Drummon, Goodwin 2004].

Figure 1.5. Impact of tariff barriers on market equilibrium in a small economy



Source: own elaboration based on literature.

Impact of tariff barriers on price formation is presented in Figure 1.5. Under conditions of small closed economy, market equilibrium is established in point P_0-Q_0. However, when global price P_W is lower, under conditions of open economy (and assuming the lack of transfer costs), domestic market price will reach the level of the global price $P_1=P_W$. Import then equals $Im_1=QD_1-QS_1$. Import shifts right through the supply curve and balance the unmet domestic demand resulting from high domestic prices on the initial level of P_0.

When tariffs barriers are imposed on imported good, the price of imported commodity in the local market increases. This means that the price of commodity imported at the border equals $P_W+Tariff=P_2$. Assuming the lack of transfer costs, the price becomes the domestic price since there is the $P_2<P_1$ relation. Consequently, the import of commodity Im_2 upon the introduction of imports tariff will be lower than under conditions of lack of such limitations $Im_1>Im_2$. This means that consumers will be able to purchase fewer commodities than in the case of lack of tariff limitations, but simultaneously more than in closed economy (lack of trade). In the case of high customs duties $Tariff>P_0-P_W$, the domestic price is similar to that of the closed economy. Similar mechanism is noted for quantitative limitations introduced in import (e.g. import quotas) in terms of the general rule and impact.

Tariff barriers have different impact in large economies. There, customs duty has driven a wedge between internal prices and global prices (similarly to the small economy). However, in the small economy import (or its limitation in the case of customs duties) is small enough not to affect global prices. In the large economy tariff barriers increase domestic prices (to a greater extent) and simultaneously decrease global prices (to a lesser extent). If prices in exporting countries are to be assumed global prices, tariffs (in importing countries) result in the decrease of prices in exporting countries.

Export refunds (export subsidies) are an important means of supporting agricultural prices. When prices in the world markets are lower than domestic prices, exporters may apply for export subsidies for exported market surpluses. This allows export of market surpluses to the world markets and guarantees price competition. This is a mechanism opposite to the one presented in Figure 1.5.

In order to protect consumer interests, for food safety reasons or due to political factors, export limitations may be introduced periodically. For example, when global prices are higher than domestic ones, export payments (export tax) may be introduced not to allow excessive export. Such measures stop the increase of domestic prices which would take place in the case of open economy. In recent years, such measures have been introduced in Ukraine and Russia in the cereal market. They affect not only the domestic market, but also the world market due to the fact that these countries are significant world exporters. Such a policy should not be excluded as a speculative measure in the interest of the main exporters.

Stabilizing markets

One of the main policy objectives is to stabilise the market for agricultural commodities and to prevent effects of changes in farm incomes from year to year. Incomes instability is caused mostly by the high price volatility. In order to alleviate price fluctuations, a number of intervention instruments may be used as the need arises. The most commonly used instruments include intervention purchases of market surpluses in a period of low prices and their sale in periods of shortages. Purchases like these are carried out when market prices are below the target or intervention prices. This results in the artificial increase of demand for products, which leads to the price increases. In the periods of shortages, the stored surpluses are sold off (artificial increase of supply), thus prices fail to achieve levels like in the case of lack of such interventions.

Similar impact is observed in the case of private storage aid schemes. The aim of these schemes is to offer financial support at times when the prices of certain products are low. In other words, these schemes are not available at all times but they work by temporarily removing surplus product from the market. Intervention schemes can be seasonal or in response to exceptional market conditions. Subsidising private storage costs releases the market from surplus of supply and stabilises prices [Gardner 1990; Hill 1990; Tomek, Robinson 2003].

Biofuel policy

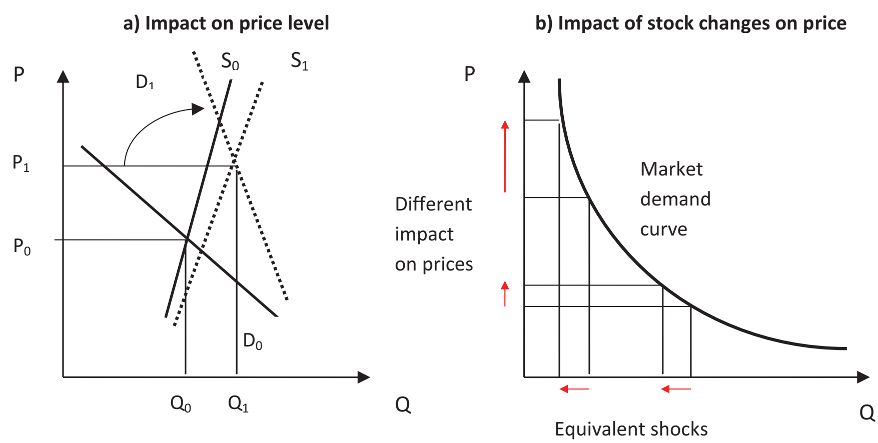
In recent years, the energy policy has had a growing impact on agricultural markets. Introduction of minimum limits on bio-component content in transportation fuels in numerous countries resulted in an increase of world demand for selected agricultural commodities and led to establishment of a stronger link between prices of agricultural products (and their variability) and prices of crude oil [Abbott *et al.* 2008; Serra, Zilberman 2011; Hamulczuk, Klimkowski 2012]. Direct impact of the policy relates to the greatest extent to such plants as rape, maize, sugar cane or coconut palm. On the other hand, this policy indirectly affects other crop and livestock markets.

The use of agricultural commodities for biofuel production is affected by a combination of several factors (including policy instruments). First of all, it results from the ratio of crude oil prices and raw materials prices. For example, Tyner and Taheripour [2008] demonstrate the ratio of oil to maize price at which biofuel production is profitable. When the oil to maize price ratio is high, biofuel production is economically rational, and not – when the ratio is low. Therefore, the demand for agricultural raw materials increases when oil prices are high. However, the activity leading to the increase of share of biofuel production (in particular with unfavourable price ratios), is backed with financial incentives (tax reliefs, payments). In such case biofuel production is profitable even under lower price ratios. The question is when to introduce incentives, in what form and at what level [Tyner 2010].

However, the most important instrument includes minimum limits of content of the bio-component in transportation fuel – so called *mandatory blending.* Due to the introduction of limits, the curve of demand for a given agricultural commodity shifts right, irrespective of the ratio: oil price/agricultural raw materials price, which leads to the increase of commodity prices. Therefore, the impact of this tool on the level of agricultural raw materials' prices depends on the pace of increase of limits as compared to technological progress. However, it should be borne in mind that even though crude oil and agricultural commodities are considered substitutes, there are technical limitations in substituting them (with the state of new technologies).

The theoretical foundation of the impact of use of agricultural commodities for biofuel production is presented in Figure 1.6. Enhanced biofuel production increases the demand for a given commodity what shifts the demand curve to the right. However, the demand for a given agricultural commodity is more inelastic (shift of demand curve from D_0 to D_1). Due to mandatory blending a market demand is much less price responsive than without biofuel policy [Wright, Cafiero 2011]. Higher prices of commodities lead to an increase of production which is shown in shifting of the supply line to the right (form S_0 to S_1). Since an increase in demand is higher than in supply price rises from P_0 to P_1 (Figure 1.6a).

Figure 1.6. Theoretical foundation of biofuel introduction on price level and volatility



Source: own elaboration based on Abbot et all. 2011, Wright & Cafiero 2011.

An increase of biofuel use results in the decrease of supply of these products for traditional purposes (food, feed use, stocks). At the markets in cereals, oilseeds crops or sugar cane a decrease of the stock to use ratio has been observed in recent years, which means that there are less physically available commodities on the market. It has to be emphasized that under higher price regime stocks are lower and prices become more sensitive to shocks in supply. This is where the additional problem arises, namely, the increase in volatility of world agricultural prices. The rate at which biofuel limits were introduced in some countries exceeded the production efficiency growth rate, which is conditioned by yield growth potential. This led to the above-mentioned reduction of stock levels (Figure 1.6b). Obviously, if the stock levels are low, prices are extremely sensitive to the changes in supply-demand relations [Abbott *et al*. 2011]. Fuel demand is highly rigid. In the conditions of high yield variability resulting from climatic conditions, prices become extremely volatile. For example, on the basis of simulations, Tyner et al. [2012] observed that severe drought accompanied by rigid indices of blending biofuels with liquid conventional fuels may contribute even to a 60% growth of maize prices. Price reactions become less strong as the flexibility of indicative objectives concerning mandatory blending levels increases.

All above mentioned state intervention measures weaken and distort natural market processes and, therefore, market prices in the long and short term perform differently than they would do in the absence of such institutional arrangements. The forecasting problem here is that most of these activities are of periodic nature and it is quite difficult to estimate the net impact of each intervention.

1.5. Cross-market interactions

When analysing the agricultural sector, an expert should take into account a number of interactions between various agricultural branches and interaction between the agriculture and the entire economy. The branches of the agricultural sector may be determined by the type of farming, such as dairy, poultry, crop, etc. It can also be divided by the type of harvest, such as agronomy and animal husbandry. From the microeconomic point of view, this includes adjustment of production decisions to the conditions and opportunities, in which agricultural holding (enterprise) operates. In the whole process the key role is played by relative prices. On the basis of the relations between the prices of the factors of production, prices of products sold and relations between the prices of products sold and products purchased, each producer makes both operational and investment decisions.

Complementary and competitive relationships

There are basically two types of relationships between agricultural branches and products: competitive and complementary [Varian 1999]. These relations may be seen as the links between sectors of industry, mainly between crop production and animal production, and the interdependence within respective sectors.

Theoretically, all agricultural branches compete for the same limited resources used as the factors of production in agriculture (in agricultural holdings). However, competition occurs when an increase in production in one branch leads to a reduction in production in another. This occurs when the branches are competing at the same time for the same factors of production. Let's take as an example the competition for arable land. The land allocation between the various crops at given yields determines the production and supply of agricultural crops. This is particularly important in the short- and medium-term perspective. Increasing the sown area of one cereal may only occur at the expense of reducing the cultivation of another plant. At the same time not only cereals but also oilseeds and root plants compete for arable land. Production and supply of agricultural products are determined by relationship between prices of crops that can be grown. Similar correlation can be distinguished in other types of agricultural activity.

Complementarity between branches of production is based on the principle that those branches complement each other. Branches of crop production are associated with the branches of livestock production mainly through fodder for animals and animal fertilizers. Economic policy (environmental) may increase the compliance with the relevant standards (requirements) or higher payments for those entities which carry out sustainable production.

Pork market example

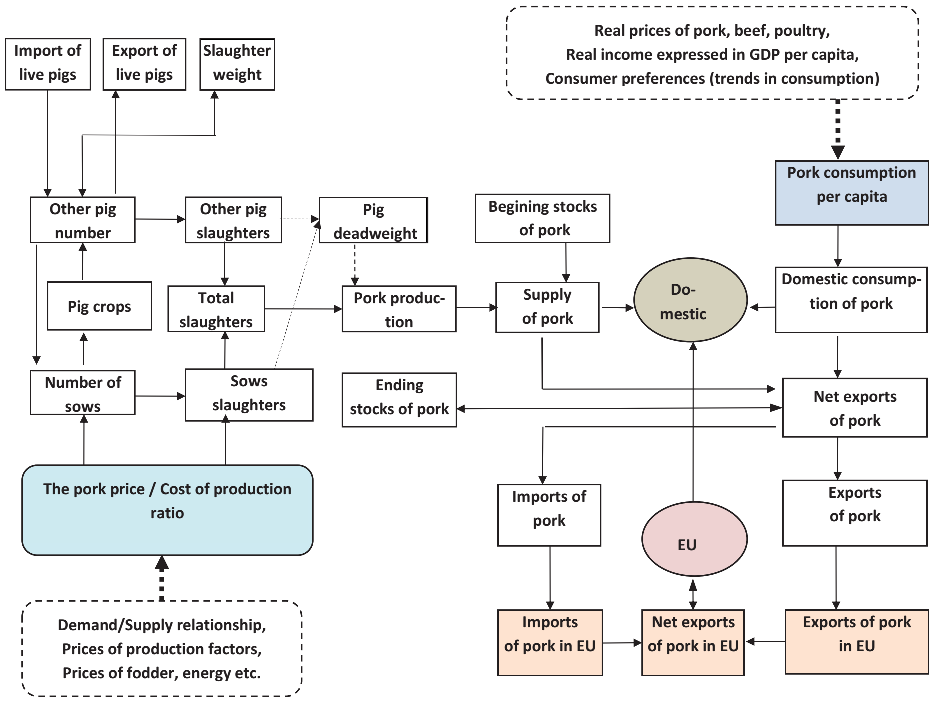
An example of the links between different categories of pork production and links between pork market and other markets is shown in Figure 1.7, where the analytical framework for modelling with the use of AGMEMOD[1] partial equilibrium model is presented. For consumers who have limited budget, the demand for a product (e.g. pork) depends on various factors, prices of the product and of substitute goods including, e.g. different types of meat (beef, chicken, mutton), etc. Thus, consumption of pork per capita depends on the relationship between its prices and the prices of other types of meat, of people's income and preferences. It is a relation between one kind of meat and others, through prices that are the result of supply-demand situation, which determines the formation of price relations. An excessive increase in prices in one market will decrease demand for the commodity, as a result of changes in consumer preferences (the substitution effect). The change of consumers' preferences does not take effect immediately because of the stocks held or habits of consumers. In an open economy, demand for specific commodities results not only from domestic prices, but also depends on the prices in other countries. Lower domestic prices than prices in other countries may encourage exports (see Figure 1.7). This means that in forecasting of the demand and supply situation on the domestic market, international trade should also be taken into account.

Another factor influencing the market equilibrium involves supply. The profitability is regarded as the most important cause affecting supply. For example, the profitability of pig breeding depends on prices received for livestock and

[1] More on AGMEMOD model can be found in a chapter 5.4.

the production costs. Among production costs the prices of fodder, which is based on cereals, play an important role here. When profitability increases, pig producers decide to increase the scale of breeding, in anticipation of the growth in revenue in the future. Thus, the demand for piglets is growing and the number of farrowing sows is increasing. At the same time a growing number of livestock creates demand for grains influencing the increase in their prices. At the peak of livestock development the profitability of pig production falls as a result of decrease of pork prices and increase of fodder prices.

Figure 1.7. Flow diagram for pig livestock and pork meat in AGMEMOD model



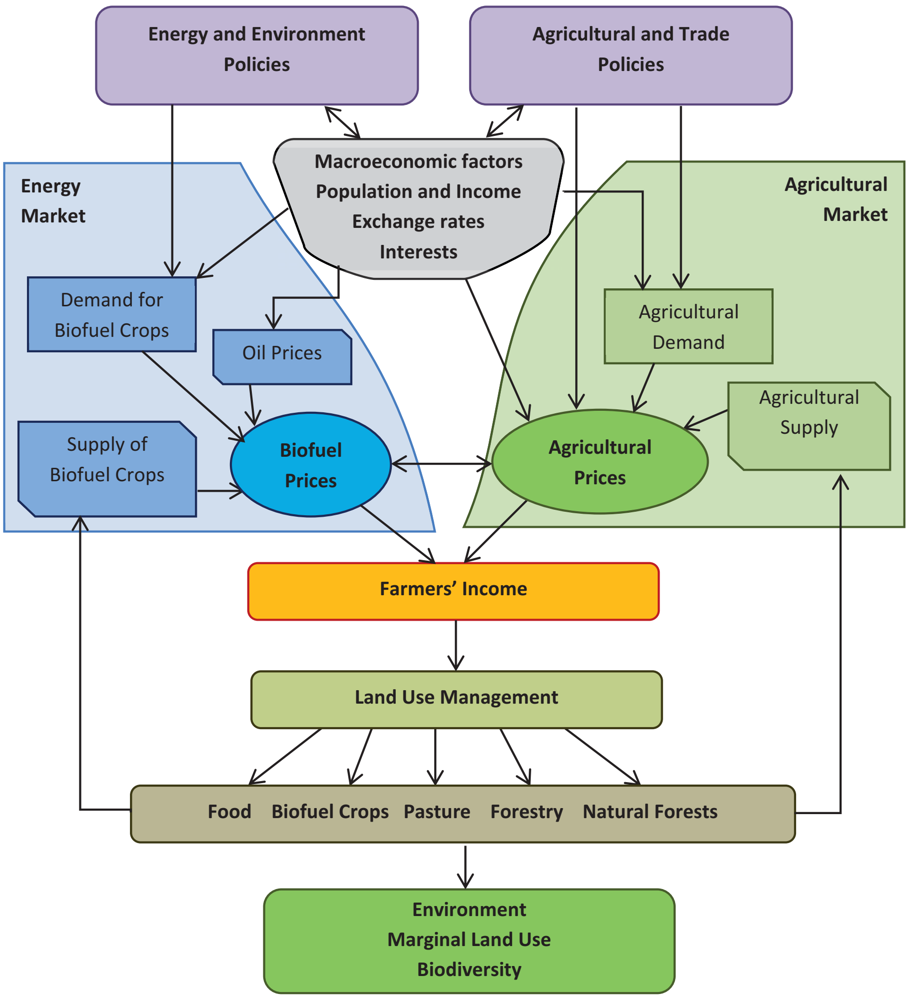
Source: own compilation based on AGMEMOD Partnership [2005].

Due to the biological and technical nature of production, delays between the time of decision-making and demand occur, which means that current production and demand depend on past prices. Thus, market forecasts should take into account such time lag. Domestic prices depend also on processes in the world markets, in European Union on EU prices (Fig. 1.7).

Biofuel policy impact example

Another good example of market interactions is presented in Figure 1.8. There are various channels biofuel policy affects the market through market equilibrium. Implementation of the biofuel policies leads directly or indirectly to the effects on land use, crop production, food consumption, feed use, demand for factor of production, etc. [Tyner 2010, Rathman *et al*. 2010].

Figure 1.8. Analytical framework for analysis of the biofuel-food market interactions



Source: authors' own compilation based on Rathman et al. [2010].

In the centre here is the competition for the land use. Increase of crops for biofuel production (maize, wheat, rapeseeds, etc.) results in the decrease of land use for other plants. This is reflected in the prices of other plant products. Higher prices for crops mean increased feed costs which is later reflected in the prices of meat and milk products. Higher prices of food concerns consumer decision – so there is the expecting decrease of food demand under high prices.

Different issues are important when forecasting agricultural prices in short and in the long term. Under the conditions of market relations it is important to determine price transmission mechanism between markets and delays between such impulses. The knowledge of these regularities allows for the correct specification of the forecasting model and for evaluation of the forecasts.

2. Time series forecasting

Forecasting employing time series models is based on conclusions drawn from observation of the patterns occurring in a historical period. This chapter addresses the issue of specifying the nature of such patterns, as well as the problem of extrapolating these patterns into future. It also presents the basic statistical models and conditions of their application.

2.1. The nature of time series forecasting

2.1.1. The concept of time series forecasting

The basis for forecasting prices of agricultural commodities is knowledge about market patterns occurring in agriculture and its environment, referred to as "agribusiness". Agricultural commodity market is characterised by specific features that cause different market performance, compared to the markets of industrial products and services. In particular, this refers to greater price volatility and uncertainty than in markets of other products. Forecasting is one of many ways which allow to reduce market uncertainty. The questions are: to what extent are we able to predict the future agricultural prices and how to do it?

Economic processes do not occur in a completely random and chaotic manner over time. They are characterised by certain patterns like trend, cycle or seasonality. Effective forecasting of economic phenomena is possible when we properly identify such patterns. The patterns exiting in the data reflect the influence of various factors. These factors have already been extensively discussed in literature on the subject which was summarised in Chapter 1. However, identifying all factors which influence a given phenomenon, e.g. agricultural prices, is not always possible or necessary. It is often reasonable to consider only the effects of these causes which are reflected in patterns shown by a given phenomenon over time.

The data used for forecasting are mostly time series, which means that they consist of a sequence of observations over time. Time series models belong to quantitative methods which can be applied for forecasting purposes only if the following conditions are fulfilled [Makridakis *et al.* 1998]:

- information about past of the variable being forecasted is available,
- such information can be quantified in the form of numerical data,
- it can be assumed that past patterns or relationships observed in the past will continue in the future.

Thus the time series forecasting is based on the existence of statistical data concerning variables being predicted. There are price data of different frequency: yearly, quarterly, monthly, weekly, daily or others. Depending on the frequency of data we can capture different combinations of patterns. For example, when analysing yearly data it is impossible to capture intra-year fluctuations like seasonality. Investigating the cyclical fluctuations is also limited then. Forecasting horizon also depends on the frequency of the data. There is no use to predict for one year on the basis of daily data, whereas the horizon of forecasts calculated with yearly data is usually a few years.

The key premise underlying the prediction based on quantitative methods, time series models among them, is the assumption of continuity. It is assumed that certain regularities of data, like trend or seasonal fluctuations, will not change in the horizon of the forecast. This assumption is not always true, so a forecaster should assess the stability of the system over time. We know that nothing remains exactly the same, however, sometimes history repeats itself.

Time series models look at the past patterns of data and attempt to predict the future based upon the underlying patterns contained within these data. Time series forecasting treats the system (phenomenon) as a black box and makes no attempt to discover the factors influencing its behaviour. However, knowing the factors underling different patterns makes verification of the quality and reasonableness of the generated forecast much easier.

The essence of forecasting on the basis of time series consists in the assumption that a forecast of a given phenomenon reflects all of the factors that have impact thereon. There is a widely known saying that prices reflect all known and relevant information. Assuming its accuracy, there is no need to study the causes of price changes. Prices can be treated as a "black box" and it is unnecessary to go deeper into the causes and relationships with other phenomena [Box, Jenkins 1970].

Time series models are relevant when:

- the system is not understandable or it is difficult to measure relationships that govern its behaviour,
- we are interested in predicting what will happen in the future, but not in explaining why it happens,
- data limitation and high cost of use of the alternative methods exist.

Human history from its very beginnings has been interested in the idea of foreseeing the future. The question about the future has been asked throughout such disciplines as economics, statistics, econometrics and philosophy. As Adam Smith pointed out, if millions of greedy and selfish individuals who strive to achieve their goals and, in most cases, are not controlled by the state do not cause anarchy, it is most probable to foresee the behaviour of market participants. The question is, to what extent the economic processes can be foreseen.

The basis for forecasting is the use of certain forecasting methods. The forecasting approach consists of two phases: diagnosis of the future and determination of the future [Cieślak 2005]. They are related to modelling and extrapolation of regularities, respectively.

The first phase is related to looking for regularities in the past and an attempt to present them in a model. The model constitutes a simplified description of the reality, which overlooks irrelevant aspects to explain the internal activity, form or construction of a more complex mechanism. Quantitative models may be based on relationships between the forecasted variable with other variables or on relationships in a given time series.

Since this chapter concerns forecasting based on time series, it will be limited only to dynamic relationships in a single time series. The model approach to a time series assumes more or less clear (direct) distribution of regularities for historical periods. The approach distinguishing between a trend, cyclical, seasonal and incidental fluctuations is one of the propositions facilitating understanding of the nature of time series modelling. In most cases, patterns occurring in a time series are hidden in the data. Time series methods treat the system as black box and make no attempt to discover the factors influencing its behaviour [Box, Jenkins 1970].

Attractiveness of time series methods results from several issues. First of all, as a rule the only source of information concerning the future progress of a phenomenon is the past data on the forecasted variable. There is no need to gather and analyse vast amount of information from different sources. This may be justified by the general saying that "the price reflects all available information". From this perspective, the use of time series method is relatively cheap and time-saving in comparison to other models based on large amount of data. The quality of such forecasts is no poorer than the forecasts based upon more complex models. In many cases it is enough to answer the question "how will it happen" and not "why will it happen". One of the most important reasons to use time series models is that there is no need to make assumptions about the values of explanatory variables in the forecasted period. It is evident that the forecast might be close to reality only if explanatory variables would be determined properly.

2.1.2. What does a time series pattern reflect?

Literature on the subject assumes that a time series may consist of the following mutually independent components (patterns): trend, cyclical fluctuations, seasonal fluctuations and incidental fluctuations. Some authors do not distinguish between trend and cycle, treating them as one component. Below we will briefly present the economic mechanisms underlying the above patterns, which is important to understand the idea of time series forecasting of prices.

A trend is defined as a long-term propensity to one-directional changes of the value of a given variable over time. Trends observed in agriculture and the food economy are caused by technical innovation, changes in preferences and tastes of consumers, or the general level of inflation. The trend does not include variability which is crucial for the price risk. Each market participant has time to adapt to long-term changes through technology change, concentration of production or lower costs.

Different deviations take place around the trend, including cyclical fluctuations in the form of more or less regular fluctuations around the trend. Cyclical fluctuations cover medium- and long-term changes. The length of cyclical fluctuations measured between the two successive upper or two successive lower turning points is longer than one year. Distinguishing between cyclical fluctuations and the trend is quite problematic when it comes to the methodology. Both types of fluctuations are in fact changes in the long term and are often not separated, but treated as a long-term trend (trend-cycle). This approach is justified by the fact that in economic reality there are cycles of different lengths, often overlapping. What seems to be a trend for a short time series may appear to be a good approximation of a cycle, easily noticeable for a longer time series.

Factors underlying cyclical fluctuations in agri-food economy are of different nature: endogenous or exogenous. Biological and technological constraints are the main factors causing cyclical fluctuations in the prices of agricultural raw materials. Due to the specificity of agricultural production, a certain period of time needs to elapse from the moment of making a decision on launching a given production to the moment when the commodities appear on the market. Due to biological and technological limitations, agricultural production is highly inelastic and does not respond to price changes in a short-term perspective. Therefore, small change in supply (mostly) and demand for agricultural raw materials may cause very considerable fluctuations in their prices. Reaction of farmers to high prices – in periods of shortages – is to increase production. When we are dealing with plant or livestock production, a few months to a few years will pass from the moment of making the decision to enter the market. Increasing supply leads then to a fall in prices. Therefore farmers' decisions, which are rational from the microeconomic point of view, cause a further decline in prices at which agricultural producers are getting higher losses. As a result, they reduce production, which in subsequent periods will result in an increase in market prices. This describes the endogenous mechanism of typical commodity cycle.

The length of the cycle and its persistence depends on the technological constrains (production cycle) and the speed of producers' reaction to market signals. Production can be planned based on different types of expectations: naïve (cob-web model), adaptive or rational. Therefore, producers make decisions taking into account past, current and/or expected future situation. Theoretically, assuming rational expectations some might expect that commodity cycles should disappear. Empirical evidence from the past decade does not prove that.

By analysing cyclical changes in agricultural markets in Poland, we cannot limit ourselves solely to domestic reasons. We need to bear in mind that in the conditions of relatively open trade, behaviour of commodity prices in a given country stems not only from domestic demand-supply relations, but also from the impact of the situation on what is referred to as European or world markets. In fact, changes in the domestic supply-demand have a relatively small impact on the prices of most agricultural raw materials in Poland.

Seasonality is the best recognized volatility factor in agricultural prices, due to the dependence of the production on weather conditions. Seasonal fluctuations are revealed as periodical changes lasting one year (we ignore daily or weekly seasonal cycles). Seasonal fluctuation in prices is due to the variability of costs, supply and market turnover. For example, in the cereals market, the seasonal effect is manifested by a sharp fall in prices after the harvest period (high supply and turnovers) and steady increase in prices later (till the next harvest period) due to increase in storage costs. The more limited ability to store raw materials, the more obvious effects of seasonality. Seasonality in prices is nothing more than an example of the impact of the general law of supply and demand. Demand, on a relatively stable level, encounters supply, which varies over time, what in turn leads to price changes.

The last type of variability observed in time series is an irregular fluctuation, which is responsible for the influence of all incidental factors and factors which are impossible to predict. Among irregular changes one may distinguish the effects caused by random factors, such as calamities, sudden policy changes or strikes.

To the above four components one can add components that describe untypical behaviour in time series, such as outliers and structural changes. Structural breaks are evident in many economic phenomena. Agricultural prices that are affected by numerous factors, such as weather conditions, animal diseases or agricultural policies, are recognised as those frequently affected by structural changes [Wang, Tomek 2007]. This results in permanent or temporary shift of a given variable. A good example would be a time series referring to beef prices in Poland, which noted a clear sudden increase after Poland's accession to the European Union [Hamulczuk 2012].

2.2. Time series patterns identification – classical decomposition approach

The identification of patterns existing in a series is one of the most significant issues in time series forecasting. It can be done by analysing price graphs or with the use of statistical tools. One of such tools is classical time series decomposition method (known as Census I).

2.2.1. Decomposition models

Dividing time series into individual components is called decomposition of time series. As indicated in the previous chapter, it is assumed that a price level (Y) is a combination of a trend (T) and cyclical (C), seasonal (S) and irregular (I) fluctuations. The two principal decomposition models of time series include additive or multiplicative models. Due to mathematical reasons when multiplicative model is chosen all calculations are performed via log-additive model (based on natural logarithms of data). Formulas for these models are as follow:

$$Y_t=T_t+C_t+S_t+I_t, \text{ (additive model)}, \quad (2.1)$$

$$Y_t=T_t \cdot C_t \cdot S_t \cdot I_t, \text{ (multiplicative model)}, \quad (2.2)$$

$$lnY_t=lnT_t+lnC_t+lnS_t+lnI_t, \text{ (log-additive model)}. \quad (2.3)$$

The difference between the additive model and the multiplicative model is based on diverse relations between their components. The easiest way to illustrate it is on the example of the seasonal component. In the additive model, seasonality is not linked to the level of the phenomenon over time. In case of multiplicative model, seasonal effects are fixed in relative terms, i.e. the greater the values of a phenomenon resulting from a trend, the greater the amplitude of seasonal fluctuations. In empirical analyses more the multiplicative approach is more frequently applied.

2.2.2. Classical decomposition – an empirical illustration

There are many procedures of time series decomposition. This chapter presents the simplest of them, called the classical decomposition on the basis of monthly pig price series (Fig. 2.1, 2.2). This constitutes an introduction to a more advanced approaches, like the seasonal adjustment method X-12--ARIMA presented in Chapter 3.

Before making the calculations, the type of model must be considered. Price plot in the Figure 2.1 does not clearly show which model is more appropriate, so we choose the additive one as it is the simpler one. Therefore, we assume interactions between variables in the following form: $Y_t=T_t+C_t+S_t+I_t$. However, we should remember that most of price series act according to multiplicative models.

The general stages of time series decomposition according to classical procedure involve the following steps:

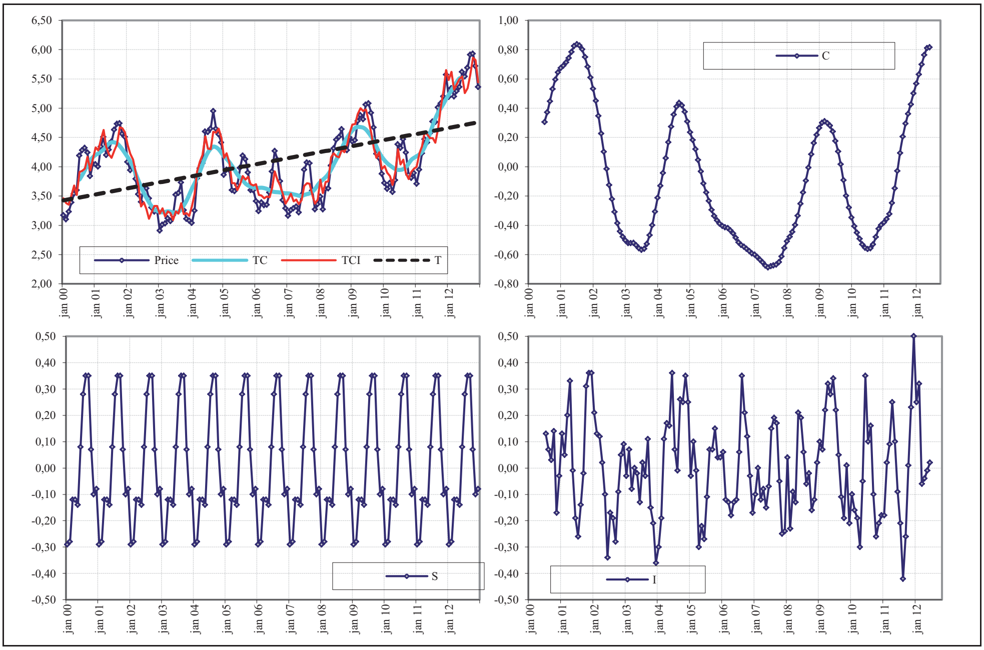
- Calculation of the central moving average, which reflects the trend-cycle TC_t,
- Adjustment of data representing the long-term trend Y_t-$(CT_t)=S_t+I_t$,
- Calculation of the seasonal component on the basis of Y_t-CT_t ,
- Extraction of original series from seasonal component Y_t-S_t,
- Division of the trend-cycle into the trend and cyclical fluctuations by fitting trend line model to TC_t,
- Calculation of irregular component as a residual value.

In the first step, we calculate the central moving average. The number of total observations based on which we calculate such an average equals the number of seasons in the year (12 for monthly time series). The values of the calculated average are attributed to the central observation. If the number of seasons is an even number, we add one observation to the average, and we give the weight of ½ to the last and the first of them. For example, the trend-cycle in 7th period is calculated as follows: $TC_7 = (0{,}5 * Y_1 + Y_2 + Y_3 + ... + Y_{12} + 0{,}5 * Y_{13}) / 12$, where Y_1 is the price in the first analysed month (January 2000), Y_2 – price in the second month (February 2000), etc. The subsequent terms of the moving average are calculated in a similar way, only based on a different range of data (we calculate TC_8 on the basis of the data from observations 2 to 14). The moving average thus calculated is presented in Figure 2.1.

In order to calculate the seasonal component, we must first remove the trend-cycle from the actual data: Y_t-$TC_t=S_t+I_t$. Calculation of the seasonal component comes down to averaging the S_t+I_t series to observations from the same seasons. If the sum of all the average indicators for individual months is equal to zero, we recognise them as final seasonal indicators. If it is not, we need to adjust them, to fulfil the above-mentioned condition. In the case of multiplicative model, the sum should equal to the number of seasons in the year.

These indicators represent average deviations of pig prices, in various seasons, from the long-term trend in the analysed period. According to the performed analysis the lowest pig prices in Poland are in January (seasonal indicator is -0.29) and the highest are in August and September (seasonal indicators are 0.35). In other words, in January pig prices are on average lower by 0.29 PLN/kilo from the long-term trend. In August and September pig prices are higher by 0.35 PLN/kilo than the prices resulting from the long-term trend. So the variation of pig prices due to seasonality is 0.64 PLN/kilo.

Figure 2.1. Decomposition of time series for prices of pig livestock (PLN/kg)



Source: authors' own compilation based on the data of the Central Statistical Office.

The next step is to eliminate seasonality from the original data. In the case of the additive model, we subtract the seasonal effect from the original data. Seasonally adjusted data (T_t+C_t+I_t) are presented in Fig. 2.1.

We can then separate the trend from the cycle, which requires the calculation of the trend. To this end, we use analytical functions of the trend. As a rule, we apply simple ones (linear or exponential forms of the trend). The estimated linear trend equation (for T_t+C_t+I_t series) is as follows: T_t=0.0077·*time*+3.466. Based thereon, we will calculate the individual values of the trend. Cyclical component is calculated by subtracting the trend component from trend-cycle: $C_t=(TC_t)-T_t$. The results are presented in Fig. 2.1. Graphical insight leads to the conclusion that cyclical fluctuations have greater share in price volatility than seasonal fluctuations. The last component, which is the irregular component I_t, is calculated as a residual value. We can use e.g. the formula $I_t=Y_t-TC_t-S_t$.

The final result of the above operations can be seen in Figure 2.1. Time series of prices was decomposed into a trend, seasonal and irregular fluctuations. We can see that cyclical fluctuations have the greatest share in price variability and their amplitude exceeds 1.40 PLN/kilo. Seasonal fluctuations are of smaller importance (amplitude is less than 0.70 PLN/kilo). Irregular fluctuations have a similar proportion in total variability to seasonal variation.

The disadvantage of the presented approach consists in shortening of the data in the beginning and in the end, due to the use of moving averages. It is possible to cope with this problem, but the methods to do so would have a negative impact on the clarity of the example used to illustrate the idea of decomposition of time series.

2.3. Basic time series models

In this chapter basic time series models are presented. The first part is dedicated to the econometric models, the second one to adaptive methods. We start from describing trend line models, and then the discussion is extended to cover also the models with seasonal and autoregressive components. When dealing with adaptive models we begin with simple exponential smoothing method. We focus then on Holt's and Holt-Winters models in classical approach as well as with damping factor.

2.3.1. Econometric models

Trend line models

One of the simplest and the most intuitive methods of time series forecasting is trend extrapolation. When using this technique we assume that the trend detected in past observations will continue into the future. Trend line is calculated as a regression where time or its transformations constitute a set of independent variables. The basic equation of trend line model is given by:

$$Y_t = f(t) + \varepsilon_t\,, \tag{2.4}$$

where:

Y_t – observed value of analysed time series at time t,

$f(t)$ – trend function,

ε_t – random error at time t.

When trend function is identified, the forecast equals the value of trend function in the period of forecast horizon $F_{t+h}=f(t+h)$. There is a wide variety of trend functions that can be applied to the equation (2.4). However, choosing the most appropriate trend function can be a quite complicated procedure. An experienced forecaster does not only care for good fitting of model to past data. *A priori* expectations about future changes based on the knowledge about the nature of analysed process are also very important.

Examples of trend functions that are most often used in practice are listed below:

$$f(t) = \beta_0 + \beta_1 \cdot t \,, \tag{2.5}$$

$$f(t) = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 \,, \tag{2.6}$$

$$f(t) = e^{(\beta_0 + \beta_1 \cdot t)} \,, \tag{2.7}$$

$$f(t) = \beta_0 + \beta_1 \cdot 1/t \,, \tag{2.8}$$

$$f(t) = \beta_0 + \beta_1 \cdot \ln(t) \,, \tag{2.9}$$

$$f(t) = \beta_0 \cdot \beta_1^{\beta_2^t} \,. \tag{2.10}$$

If there is a presumption that the direction of level change will remain steady (in absolute terms) and trend will continue, it is possible to use a simple linear function (equation 2.5). Parameter β_0 denotes the intercept that represents the value at time zero, and β_1 determines the slope of the trend line. However, there are many examples of economic processes that cannot be properly forecasted by linear trend. If there is a premise that the growth observed in the past will accelerate, quadratic function (equation 2.6) can be applied. When accelerating growth is the most expected the future trend exponential function (2.7) can also be used. It should be noted, that these functions are appropriate mostly for short-term forecasts. If used for long-term forecasts risk of significant errors is substantial. When there are factors indicating that the trend will decrease its rate and approach horizontal asymptote, we can use i.e. reciprocal function (2.8) or logarithmic function (2.9). In fact, we can use any other function that will fit to observed data and our assumptions referring to future trend. There are large numbers of composite functions that can to fit to supposed future changes of econometric processes. For instance, the Gompertz function (2.10) is used when forecasting trend is assumed to have initially increasing and then decreasing rate of growth.

After choosing an appropriate function, we have to estimate the unknown parameters β_0, β_1,…, β_n. As it was mentioned before, as for forecasting good fitting to past observation is not the only objective, it is still very important. There are many various estimation techniques, however, the Least-Square Estimation (LSE) method is the most popular. Although there are many iterative methods of estimating non-linear functions, most forecasters transform non-linear functions into linear ones. The most convenient method is the logarithmic transformation. Using this transformation we can change the exponential function (2.7) into the linear one as given by:

$$\log(f(t)) = \beta_0 + \beta_1 \cdot t, \tag{2.11}$$

where: log – natural logarithm. This transformation allows for using LSE on the equation 2.11.

Trend line with seasonal dummies

Simplicity of trend line models is a big advantage. However, if there is a seasonal variation in the analysed time series data, there is a need to expand this method to catch the seasonal component. Seasonality – repetitive and predictable movements around the trend line in constant periods – is the common characteristic of many time series. Seasonal dummies allow to include the variation into forecasts. The equation of a model with trend and seasonal dummies is as follows [Kufel 2007]:

$$Y_t = f(t) + \sum_{i=1}^{r-1} \alpha_i S_{it} + \varepsilon_t, \tag{2.12}$$

where:

r – number of seasons in a year,

α_i – seasonal coefficient at the season i,

S_{it} – seasonal dummy variable that takes the value of one in the i-th season and 0 otherwise.

To avoid the so-called dummy variable trap arising from perfect co--linearity of seasonal dummy variables with seasonal intercepts, different strategies can be used. The first one requires to drop one of seasonal dummies, as a rule the last one is omitted, as presented in the equation 2.12. In this strategy, the intercept β_0 can be treated as the intercept calculated for this last season. Intercepts for other seasons equal $\beta_0+\alpha_{i.}$ The second solution is to drop the intercept β_0 in trend line function and use all seasonal dummies.

After identifying the trend and seasonal dummies forecasts are calculated on the assumption that the trend will continue and the seasonal indicators for corresponding seasons in the forecast time horizon remain the same as in the past.

Models with autoregressive component

The trend and seasonal fluctuations are not the only patterns that can occur in time series. If there are any significant short-term deviations from the long-term trend model given by the equations 2.4 or 2.12, they cannot be considered as adequate forecasting tools. These short-term fluctuations disturb the characteristics of random errors ε_t in such a way that it cannot be treated as a white noise process. Autoregressive models (see in details Chapter 2.4.1) are appropriate to capture short-term fluctuations. The equation of the trend line model with the autoregressive component is as follows [Falk, Roy 2005; Kufel 2007]:

$$\begin{cases} Y_t = f(t) + u_t & (2.13) \\ u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \ldots + \phi_p u_{t-p} + \varepsilon_t & (2.14) \end{cases}$$

where:

u_t – residuals which follow the autoregressive process of order p,

ϕ_j – j-th autoregressive parameter.

The transformation of equation 2.12 allowing to capture autoregressive component in the trend line with seasonal dummies model is given by [Falk, Roy 2005; Kufel 2007]:

$$\begin{cases} Y_t = f(t) + \sum_{i=1}^{r-1} \alpha_i S_{it} + u_t & (2.15) \\ u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \ldots + \phi_p u_{t-p} + \varepsilon_t & (2.16) \end{cases}$$

Various procedures to estimate this model can be applied. The simplest one is the two-step OLS method. In the first step, we estimate the equations 2.13 or 2.15, obtaining the trend, seasonal estimates as well as residuals u_t. Then the autoregressive model for residuals u_t is applied to obtain the autoregressive parameters ϕ_j. The forecast is equal to the sum of forecasts from the two aforementioned equations.

We can also estimate parameters in the models with autoregressive component using the Feasible Generalized LS (FGLS) method. Applying this technique we first estimate the residual equation using the Ordinary LS (OLS) method. Then the estimators of trend and seasonal component are modified according to the nature of estimated autoregression. The generalised Cochrane-Orcutt iterative procedure is most widely used practical tool using FGLS estimators. It helps to solve the problem of biased estimation when random errors are serially correlated by modelling the residuals and transforming the model by taking quasi-differences of the analysed time series. The value of this quasi-difference depends on the estimated value of the autoregressive parameter. For details see [Falk, Roy 2005].

Another possibility is to collapse the equations 2.13 and 2.14 or 2.15 and 2.16 into one single equation. The combined transformed equation which follows from collapsing equations 2.15 and 2.16 is given by [Osińska ed. 2007]:

$$Y_t = f(t) + \sum_{i=1}^{r-1} \alpha_i S_{it} + \sum_{j=1}^{p} \phi_j Y_{t-j} + \varepsilon_t . \tag{2.17}$$

It is now possible to estimate equation 2.17 by using the OLS method. The number of lags equals the order of residual autoregressive process and it should guarantee the white noise properties of residual ε_t.

It should also be mentioned that all of the above-discussed models can be extended by adding some structural variables such as level shift or additive outliers, which is described in detail in Chapter 3.

2.3.2. Exponential smoothing methods

If the structure of forecasted variable is subject to temporary changes, the exponential smoothing models can be used to calculate the short-term forecasts. Models belonging to this group are characterised by the fact that no fixed analytical form of a mechanism explaining changes of the analysed time series is assumed, but changes at irregular intervals are acceptable. These models allow for instable economic structure and changes of parameters over time. High flexibility of the exponential models makes them useful in the short-term forecasts.

Simple exponential smoothing model (SES)

Although the exponential smoothing methods were introduced almost half a century ago, their simplicity and practical utility make them very popular among forecasters even today [Pedregal Young 2005]. The simplest of these methods is the one called simple exponential smoothing model (SES). Its forecasting mechanism is a combination of naïve method, where all forecasts equal last observed value, and the average method assuming all future values are simple averages of observed data. Like both of the above-mentioned techniques SES has a flat forecast function. This characteristic of SES is suitable only for stochastic series with no trend or seasonal pattern.

The basic equation for SES is [Stańko ed. 2013]:

$$F_{t+h} = \alpha Y_t + (1-\alpha) F_t , \tag{2.18}$$

where:

F_{t+h} – a forecast made at time t for h period ahead,

Y_t – an observed value at the time t,

α – a smoothing constant.

The SES model allows us to calculate the level of time series that describes the recent changes of the series in a way that adjusts smoothly over time. Alternative representation of the SES basic equation is the one emphasizing adaptive character of forecast creation and is given by:

$$F_{t+h} = F_t + \alpha(Y_t - F_{t-1}) = F_t + \alpha e_t \,, \tag{2.19}$$

where:

e_t – an one-step forecast error at the time t.

The equation (2.18) can be also rearranged to a form that accentuates the role of past observation in determining the present level of the value:

$$F_{t+h} = \alpha Y_t + (1-\alpha)(\alpha Y_{t-1} + (1-\alpha)F_{t-1}) = \alpha(Y + (1-\alpha)Y_{t-1}) + (1-\alpha)^2 F_{t-1} \,. \tag{2.20}$$

If we continue to substitute earlier levels by observed values backwards, we can rewrite the equation (2.20) in the following form:

$$F_{t+h} = \alpha(Y_t + (1-\alpha)Y_{t-1} + (1-\alpha)^2 Y_{t-2} + ... + (1-\alpha)^t Y_1) + (1-\alpha)^{t+1} F_1 \,. \tag{2.21}$$

The equation (2.21) shows the weights given to the past values of series and allows us to understand the meaning of "exponential smoothing" in SES. The only problem left is the initialisation of the smoothing process, since we need to specify an initial value F_0. The most common approach is to set $F_1 = Y_1$. However, we can also set initial value as average of first k observation $F_1 = 1/k\sum_{i=1}^{k} Y_i$ or use backforecasting technique.

It is also worth mentioning that according to the equation (2.21), the forecast depends on initial value. When time series is relatively long and the smoothing constant α is far from 0, the weight attached to the F_1 value is relatively small.

Still, the most important task is to properly define the smoothing constant value. By convention, this parameter is limited to the range of $0<\alpha<1$. The smaller α, the more weight is given to the observation from the more distant past and to the initial value as well. The larger α, the more important are recent observations. If α equals 1, SES forecasts are the same as the ones obtained on the basis of the naïve method. In practice forecasters use different rules of thumb – i.e. we can use small values of α if time series is relatively stable over time. However, more objective way to obtain smoothing constant value is to estimate them from observed data. Using iterative methods we can choose the value of α so that it will minimise the sum of the squared errors (SSE).

Holt's model

To overcome the limitations of SES, Holt [1957] extended this method to forecast data with trends. Holt's forecasts are calculated as a sum of two values

– the level and the trend of time series. The level and the trend as well as forecasts for next h periods can be written in the following form:

$$\begin{cases} L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) & (2.22) \\ T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} & (2.23) \\ F_{t+h} = L_t + hT_t & (2.24) \end{cases}$$

where:

L_t – a level (weighted average) at the time t,

α – a level smoothing constant,

T_t – a trend at the time t,

β – a trend smoothing constant.

The first component of forecast referring to the time series level at the time t is a sum of the weighed recent value and the weighted previous level increased by previously estimated trend. The latter component is the weighted average of previous trend estimation and the recent levels of difference.

There are obvious similarities between SES and Holt's model. Likewise in SES, the crucial points in adapting Holt's model are setting α and β values as well as the initial values of level and trend. As with SES, smoothing parameters are limited to the range from 0 to 1 and can be specified using some rules of thumb or through optimization process minimising errors. The example of the first one is to set α and β close to 1, when there is a strong irregular component of time series. Ord and Fildes [2013] suggest the following values: $0.05>\alpha>0.3$ and $0.05>\beta>0.15$.

The initialization process for Holt's model requires two estimates – the starting values of trend and level. Different methods can be applied here. The convenient method frequently present in literature on the subject states that starting values should be calculated on the basis of the first three observations. Then the initial values will be as follows [Ord, Fildes 2013]:

$$\begin{cases} T_1 = (Y_3 - Y_1)/2 & (2.25) \\ L_1 = (Y_1 + Y_2 + Y_3)/3 - T_1 & (2.26) \end{cases}$$

These values correspond to fitting a straight line to the first three observations. However, some other optimization methods can be applied to setting the initial value with the use of backforecasting techniques.

Additive Holt's model with a damping factor

There is a great number of economic processes for which forecasts assuming the constant linear trend seems to be inappropriate, especially for longer time horizons. Such trend series as the ones referring to the volume of export or import, value of sales or price levels cannot keep the steady growth or decline in long term, since life-cycle effects are common phenomena in economics. As a result, empirical evidence indicates that SES as well as Holt's model tend to over- or underforecast. To solve this problem Gardner and MacKenzie [1985] introduced a method allowing for slowly changing the forecast trend into a flat line.

If we decide to expand Holt's model on the damping factor ϕ, then the equations (2.22, 2.23, 2.24) become [Hyndman *et.al*. 2008]:

$$\begin{cases} L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + \phi T_{t-1}) & (2.27) \\ T_t = \beta(L_t - L_{t-1}) + (1-\beta)\phi T_{t-1} & (2.28) \\ F_{t+h} = L_t + \sum_{i=1}^{h} \phi^i T_t & (2.29) \end{cases}$$

where: ϕ – damping factor with a value between 0 and 1.

Under the Holt's model with additive damped trend the forecasts for next h steps can be rewritten in the following form:

$$F_{t+h} = L_t + (\phi + \phi^2 + \phi^3 + \ldots + \phi^h)T_t . \quad (2.30)$$

In this model the forecast values tend to recede from the trend, approaching the cut-off value $L_t+T_t/(1-\phi)$. According to the equation (2.30), it is obvious that that the larger damping factor, the weaker trend damping. If $\phi=1$ the investigated model is equivalent to classical Holt's model. We can also set $\phi>1$. Then the factor ϕ instead of damping the trend will affect forecasts in opposite way than the accelerating factor.

Holt-Winters additive model

Both previously considered methods cannot be used in the case of data characterised by seasonality. If we try to make forecasts for time series with distinct seasonal component, Holt-Winters model should be employed. It is based on three smoothing equations – the additional one comparing to Holt's model deals with seasonality. There are two variations of Holt-Winters models – additive and multiplicative one. The difference between them refers to the nature of seasonal component. The additive model expresses seasonal changes in absolute terms, when in the multiplicative model seasonal component is expressed in relative terms. Additive version of Holt-Winters model is as follows [Ord, Fildes 2013]:

$$\begin{cases} L_t = \alpha(Y_t - S_{t-r}) + (1-\alpha)(L_{t-1} + T_{t-1}) & (2.31) \\ T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} & (2.32) \\ S_t = \gamma(Y_t - L_t - T_{t-1}) + (1-\gamma)S_{t-r} & (2.33) \\ F_{t+h} = L_t + hT_t + S_{t+h-r} & (2.34) \end{cases}$$

where:

S_t – a seasonal component at the time t,

γ – a seasonal smoothing constant,

r – a length of seasonality,

the rest notations as in Holt's model.

Since Holt-Winters model consists of three equations, setting initial values of smoothing coefficient is more complicated than in Holt's model. There is a wide variety of different methods of setting those values. As Ord and Fildes [2013] suggest, it is convenient to define starting values in the following way:

$$\begin{cases} L_r = (Y_1 + Y_2 + \ldots + Y_r)/r + T_r(r-1)/2 & (2.35) \\ T_r = \left[\frac{1}{r}(Y_{r+1} + Y_{r+2} + \ldots + Y_{2r}) - \frac{1}{r}(Y_1 + Y_2 + \ldots + Y_r)\right]/r & (2.36) \\ S_{(i)} = (Y_i + Y_{i+r})/2 - (Y_1 + Y_2 + \ldots + Y_{2r})/2r & (2.37) \end{cases}$$

where: $i = 1,2,\ldots,r$.

The above-mentioned initial values are calculated on the basis of observations from the first two years. If the cycle is one year long, the level value is defined as average from the first year plus the estimated trend for half a year. The trend value is estimated as a difference between average level values in the second and the first year divided by the number of seasons in a year. Seasonal assessments are estimated as a sum of given months from the first two years minus the average value for the whole two years. As with Holt's model, smoothing constants are limited to the range from 0 to 1, and can be specified using some rules of thumb or through optimization process minimizing errors.

Holt-Winters additive model with damping factor

Likewise in Holt's model it is also possible to expand the Holt-Winter model and add the damping factor thereto. Holt-Winters additive model with damping factor is described by four equations [Hyndman *et al.* 2008]:

$$\begin{cases} L_t = \alpha(Y_t - S_{t-r}) + (1-\alpha)(L_{t-1} + \phi T_{t-1}) & (2.38) \\ T_t = \beta(L_t - L_{t-1}) + (1-\beta)\phi T_{t-1} & (2.39) \\ S_t = \gamma(Y_t - L_t - \phi T_{t-1}) + (1-\gamma)S_{t-r} & (2.40) \\ F_{t+h} = L_t + \sum_i^h \phi^i T_t + S_{t-r+h} & (2.41) \end{cases}$$

The possible damping factor values and their impact on forecasts are the same as in Holt's model with damping factor explained above. Initial values can be assumed in the same manner as in case of the classical Holt-Winters model.

Pegel classification

Although every exponential smoothing method has its own characteristics, they all have quite similar structure. So differences between them can be presented in a concise table showing Pegels' classification [Pegels 1969]. Original Pegels' framework has two-way classification. Nine methods were divided per trend type (none, additive, multiplicative) and seasonal component (none, additive, multiplicative). It was later extended several times to incorporate models with damping factor and multiplicative error [Hyndman *et al.* 2008].

2.4. ARIMA models

ARIMA stands for the AutoRegressive Integrated Moving Average models. This term covers both stationary and non-stationary data models, as well as seasonal and non-seasonal ones. In the literature on the subject these are sometimes referred to as Box-Jenkins models, named after the authors who popularized this approach: George Box and Gwilym Jenkins [1970].

The ARIMA model belongs to time series forecasting methods. It assumes that the time series is a realization of stochastic process $\{Y_t\}$. Stochastic process is a sequence of random variables (Y_t) indexes by time (t). The goal of time series modelling is to describe the probabilistic behaviour of the underlying stochastic process that is believed to have generated the observed data in a concise way. The ARIMA models may be used for forecasting different phenomena: characterised both by a stable level (stationary), trend, cyclical fluctuations and seasonal fluctuations. These models are regarded as a tool for short-term forecasting of various economic processes worth recommendation.

2.4.1. Stationary time series processes

The basic autoregressive and moving average models describe the so-called stationary series (stationary stochastic processes). The stochastic process is stationary, in a broad sense (strict stationary), if the joint distribution of random variables in a strictly stationary stochastic process is time invariant. The stochastic process is weakly stationary (covariance stationary) if its mean, variance and auto-covariance do not change over time [Box, Jenkins 1970; Brockwell, Davis 2002]. In other words, stationary processes are characterised by constant variance, and their values in particular moments are formed around a relatively constant level (average). The value of co-variance is not time-dependent, but depends only on the interval between two observation moments. In a non-stationary process, one or more of these assumptions are not satisfied.

Autoregressive model AR(p)

In many cases the description of the investigated variable is possible through the assumption that the current value of forecasted variable depends on its last p values. Such a model is called linear autoregressive model of order p [Box, Jenkins 1970]:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t = \phi_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t \,. \tag{2.42}$$

where:

Y_{t-i},– analysed variable in period t-i, where $i=1, 2, \ldots, p$,

p – autoregressive order meaning maximum delay of the endogenous variable,

ϕ_0 – constant,

ϕ_i – i-th autoregressive parameter,

ε_t – the error term at time t, the white noise process.

Very useful notation device is the backshift (lag operator) operator B, which allows replacing Y_{t-1} by BY_t. Its general formula of is as follows: $B^i Y_t = Y_{t-i}$. Hence, the AR(p) model could be rewritten as follows [Gruszczyński *et al.*, eds. 2009]:

$$Y_t = \phi_0 + \phi_1 BY_t + \phi_2 B^2 Y_t + \ldots + \phi_p B^p Y_t + \varepsilon_t \cdot \tag{2.43}$$

Using the lag polynomial notation $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ and moving the summation term to the left side of equation, autoregressive model can be concisely specified as follows:

$$\phi(B) Y_t = \phi_0 + \varepsilon_t \,. \tag{2.44}$$

Moving average model MA(q)

Moving average model of order q involves only error terms and can be written as follows [Gruszczyński *et al.*, eds. 2009]:

$$Y_t = \theta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} = \theta_0 + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \,, \tag{2.45}$$

where:

q – order of moving average,

θ_0 – constant,

θ_i – i-th moving average parameter.

With the use of the moving average backshift operator $\theta(B)=1+\theta_1 B+\theta_2 B^2+...+\theta_q B^q$ the MA(q) model can be transformed as follows [Gruszczyński *et al.*, eds. 2009]:

$$\begin{aligned} Y_t &= \theta_0+\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+...+\theta_q\varepsilon_{t-q} \\ &= \theta_0+\varepsilon_t+\theta_1 B\varepsilon_t+\theta_2 B^2\varepsilon_t+...+\theta_q B^q\varepsilon_t \\ &= \theta_0+(1+\theta_1 B+\theta_2 B^2+...+\theta_q B^q)\varepsilon_t \\ &= \theta_0+\theta(B)\varepsilon_t \cdot \end{aligned} \tag{2.46}$$

Autoregressive and moving average model ARMA(p,q)

The ARMA (p,q) model constitutes a combination of the AR(p) and the MA(q) models and can be written [Enders 2010, Tsay 2010] as follows:

$$Y_t=\phi_0+\sum_{i=1}^{p}\phi_i Y_{t-i}+\varepsilon_t+\sum_{i=1}^{q}\theta_i\varepsilon_{t-i}\,, \tag{2.47}$$

where: symbols as above.

Using backshift notation the ARMA(p,q) model is concisely written:

$$\phi(B)Y_t=\phi_0+\theta(B)\varepsilon_t. \tag{2.48}$$

2.4.2. Non-stationary and seasonal time series processes

The ARIMA(p,d,q) models

ARMA (p,q) models are suitable only for a stationary series. If the series is non-stationary then it should be transformed. In the simplest case, it might be obtained by subtracting the deterministic trend (on the basis of the trend line fitted to the data, see eq. 2.13-2.14). This approach is justified when fluctuations around such a trend line are stationary. Such series are referred to as trend-stationary. In many cases time series are non-stationary with respect to their variance. It means that the correct method of trend elimination is a differencing of the series. It constitutes a classical Box-Jenkins approach to the time series modelling. We define the differenced series as: $\Delta Y_t = Y_t - Y_{t-1}$. Also in this case, we often use the formula with differencing operators and lag (backshift) operators. These are co-dependent: $\Delta Y_t = Y_t - Y_{t-1} = Y_t - BY_t = (1-B)Y_t$. It transpires from the above that $\Delta=(1-B)$ and $\Delta^d=(1-B)^d$. The transcription of a *d*-times differenced time series can be written as $\Delta^d Y_t=(1-B)^d Y_t$.

The forecasting models, based upon differenced data, are known as integrated autoregressive and moving average models. They can be written by means of the ARIMA(p,q,d) notation, where p stands for autoregressive order, d – number of differentiations, q – the moving average order.

Therefore with the use of backshift operator the ARIMA(p,d,q) model can be written as follows [Box, Jenkins 1970]:

$$\phi(B)(1-B)^d Y_t = \phi_0 + \theta(B)\varepsilon_t . \tag{2.49}$$

For instance, the ARIMA(2,1,1) model, with two autoregressive and one moving average terms and based on differenced series, has the following form: $(1-\phi_1 B-\phi_1 B^2)(1-B)^1 Y_t = \phi_0 + (1+\theta_1 B)\varepsilon_t$.

The SARIMA(p,d,q)(P,D,Q)$_S$ models

If the phenomenon is characterised by seasonality, the model should be extended with seasonal parameters. Such models are often referred to as SARMA (for stationary time series) or SARIMA (for non-stationary time series). The latter model can be expressed by shorthand notation: SARIMA(p,d,q)(P,D,Q)$_S$. Symbols P, D, Q stand for: the seasonal autoregressive order, seasonal differencing number, and seasonal moving average order, respectively. The S value indicates the number of periods per year in a given series.

If a time series shows a strong seasonal component with a season of S length then a seasonal difference of the form $\Delta_S Y_t = Y_t - Y_{t-S}$ can be used to generate stationary time series. It consists in calculating the differences between time series values from analogous periods of subsequent years. Also in this case seasonal differential operators $\Delta_S Y_t = Y_t - Y_{t-S} = Y_t - B^S Y_t = (1-B^S)Y_t$ can be applied.

Non-stationarity can also result from both the occurrence of a trend and seasonal variations. A time series differenced with seasonal and non-seasonal length shall be transcribed as a product of two operators $(1-B)^d(1-B^S)^D Y_t$. Hence the general transcription of the SARIMA(p,d,q)(P,D,Q)$_S$ model is the following [Brockwell, Davis 2002; Makridakis *et al.* 1998]:

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D Y_t = \phi_0 + \theta(B)\Theta(B^S)\varepsilon_t \tag{2.50}$$

where: $\Phi(B^S)$, $\Theta(B^S)$ are seasonal autoregressive and moving average operators respectively, representing polynomials in the backshift operators of seasonal part of the model. For instance, the SARIMA(2,1,1)(0,1,1)$_{12}$ model without constant can be concisely written with the following mathematical form:

$$(1-\phi_1 B-\phi_1 B^2)(1-B)(1-B^{12})Y_t = (1+\theta_1 B)(1+\Theta_1 B^{12})\varepsilon_t .$$

2.4.3. The ARIMA model step-by-step

In practice, forecasting based on ARIMA models requires finding ARIMA approximation of the process that actually generates the data. This is not an easy task as real economic processes are quite complex. Despite the flexibility of ARIMA models, the final result depends on automation of the procedure, *a priori* assumption and the experience of the forecaster. Hence different forecasters dealing with the same phenomenon, may select different models and receive different forecasts (though forecasts for competitive models are often comparable). It should be emphasised that, in general, construction of good ARIMA models requires more experience than the commonly used statistical methods.

A practical guide for building an appropriate model might be an iterative Box-Jenkins approach which is summarised in Figure 2.2. The Box-Jenkins methodology includes three phases: identification, estimation and testing, as well as application. If in the diagnostics phase it turned out that the model failed to meet the necessary conditions, then the model should be revised and whole procedure repeated.

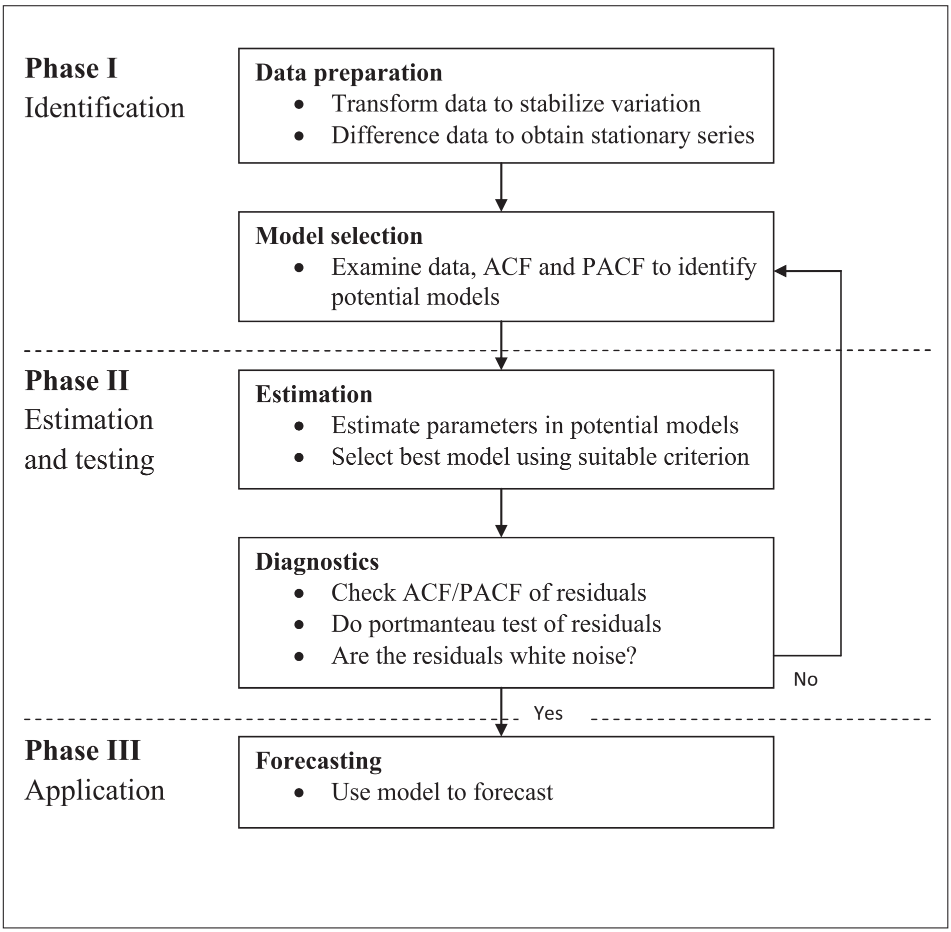
Data preparation

The first step in developing ARIMA model is to determine if the series is stationary (mean and variance) and if there is any significant seasonality that needs to be modelled. The identification procedure starts with graphical analysis of a time series aimed at initial determination of the occurrence of patterns. This concerns mainly the structure of the analysed time series (trend, seasonality) and stability of variance over time (additive or multiplicative model). If a time series structure is multiplicative, the log transformation of data should be considered for stabilising the variance [Lütkepohl, Krättzig 2007; Makridakis *et al.* 1998].

To assess the stationarity of time series, we might use both graphical analysis of a time series, graphical analysis of autocorrelation (ACF) and partial autocorrelation (PACF), and statistical tests. Graphical analysis of a time series is limited to the assessment whether a phenomenon shows a trend and seasonality. Therefore, if a trend is present, one might assume that the series is a non-stationary one. Non-stationarity may also result from seasonal fluctuations, which is confirmed by seasonal cycles.

In practice, unequivocal determination of the non-stationary type is controversial. However, the choice of trend removal is significant and apparently to a large extent influences forecasting results. If we assume price stationarity around a deterministic trend, such a trend constitutes a certain state of equilibrium, at which the prices converge in a long term. In other words, shock occurring in the system has transitory effect on it. Stochastic trend is present in the data when shocks, policy or otherwise, exercise permanent effects which do not decay as they would if the process was stationary. So prices may diverge significantly from the level of the deterministic long-term trend (or constant level). In that case we can say that a price series has a unit root. Then an appropriate order of differencing (regular and seasonal) is needed to stationarise the series and remove the gross features of seasonality.

Figure 2.2. Schematic representation of the Box-Jenkins methodology for time series modelling



Source: Makridakis et al. [1998, p. 314].

A wide range of unit root tests may be applied to assess stationarity of price series. A unit root test is a statistical test for the assumption that in an autoregressive statistical model of a time series the autoregressive parameter equals one. The most widely used tests for non-seasonal unit root include: Dickey-Fuller (DF and ADF) test; the Kwiatkowski, Phillips, Schmidt and Shine (KPSS) test; Philips-Perron (PP) test. For testing the seasonal integration the Dickey, Hasha and Fuller (DHF) test or Hylleberg, Engle, Granger and Yoo (HEGY) test can be applied [Enders 2010; Maddala, Kim 1998][2].

Model selection

Once we had determined the order of regular and seasonal differences, the next step is to identify the ARMA structure. In other words, we need to decide how many autoregressive and moving average (regular and seasonal) parameters should be included. In practice the number of model parameters rarely exceed three.

In a pure AR and MA models identification of the order p and q is carried out by comparing the estimated simple (ACF) and partial (PACF) autocorrelation functions (of initial series or the series resulting from transformation) with the theoretical functions of the ARMA process. The theoretical autocorrelation functions for different types of ARMA models can be found in various textbooks or can be simulated. Since the input data constitute only a limited sample of the series, the sample ACF and PACF computed from the series only approximate the true autocorrelation function of the process that generates the series. The sample pattern of ACF and PACF indicates a possible model. If the series is a white noise (a purely random process), then there is no need to fit a model.

The above approach is rather efficient for pure AR and MA models. In the case of mixed models and models with seasonal part the task is more difficult. We can try to extend the pure model suggested by sample ACF and PACF to a mixed model [Makridakis *et al.* 1998]. However, more than one plausible model may be identified and appropriate model selection measure is needed.

In a time series practice the information criteria based on likelihood are used. They include two components: one describing fitting of the model to the data (likelihood), and another, determining model complexity (penalty term). If the extra term does not improve the given information criterion, there is no use to adding it. The most popular information criteria are Akaike's Information Criterion (AIC), Bayesian Schwartz Information Criterion (BIC or SBIC) or Hannan-Quinn Information Criterion (HQ). The model with the smallest information criteria's value is preferred. However, the AIC tends to over-parameterize the models, so the BIC criterion (or modified AIC) is usually preferred because it tends to select simpler and more parsimonious models[3]. Box and Jenkins argue

[2] Some of them are discussed in a chapter 4.2.1.
[3] AIC, modified AIC and BIC criteria are discussed in a section 3.2.1.

that parsimonious models produce a better forecast than over-parameterized models [Lütkepohl, Krättzig 2007; Bisgaard, Kulahci 2011].

Apart from the above parameters, it should be decided whether a model includes a constant. The constant is usually present in models for non-differenced series. It is possible (but not obligatory) to include a constant in a model with one order of total differencing. Models with two orders of total differencing are constructed without a constant. If a constant is included in the model, the number of parameters is increased which influences the value of a given information criterion.

Estimation and testing

Estimation is a determination of parameters' values through software application. To estimate ARIMA parameters least squares or maximum likelihood methods are applied. There are a lot of numerical algorithms (mostly nonlinear) and the choice of the method (algorithm) influences estimated coefficients. The methods do not always converge successfully for a given set of data and in such cases it is recommended to simplify the model or modify the starting values [see for example: Evans 2003].

Diagnostic of ARIMA models is, in principle, similar to validation of other econometric models. A well-estimated model should [Enders 2010]:

- be parsimonious (significant coefficients, not too many parameters),
- fit the data well (this criterion is somehow contradictory to parsimony),
- have appropriate distribution of residuals (residuals should act as a white noise process, normal distribution),
- have coefficients that imply stationarity and invertibility (model being fit has to be stable),
- have coefficients that do not change over sample period (may be analysed with the use of recursive estimation),
- have a good out-of-sample forecasts (multi-step forecast errors for a validation period).

If the model is invalid then there is a need to repeat the steps of identification, estimation and diagnostics. Only accepted model can be applied to predict the future.

2.5. Application issues

The key issue regarding the application of time series forecasting concerns selection of the appropriate model. There are a lot of factors influencing the choice of the method applied, including:

- components of the forecasted time series,
- skills of the forecaster and technical possibilities (software),
- forecast horizon and number of past data,
- extrapolation condition.

Time series patterns and models

The method is selected on the basis of identification of regularities (components of a time series). Therefore, methods should reflect regularities and enable their extrapolation beyond the sample using of a more or less formal approach. The measurement of these regularities (decomposition) is not necessary for forecasting, but enables better understanding of the phenomenon, and, as a result, a fact-based evaluation of rationality of the estimated forecasts.

For stationary time series data moving average models, SES model or ARMA models can be used. We need to bear in mind that ARMA model can be used if the number of data in the time series is sufficient (at least 30).

For data with trend (without seasonality/or seasonally adjusted or without very strong cyclical pattern), trend line models, exponential smoothing models (Holt's) or ARIMA models are recommended. If the trend is deterministic the trend line model is sufficient, but if the trend is stochastic (with time-varying slope/direction) the adaptive models, like Holts or LES and ARIMA model, are suggested. Trend line models might be appropriate mostly for forecasting yearly series. It is rather impossible to effectively predict monthly or weekly price series (often with cyclical and seasonal fluctuations) according to trend line models. Adaptive models and ARIMA are quite efficient for a short-term forecasting and can be used for all data frequencies.

When the time series contains a seasonal component, the model applied should allow for extrapolation of seasonal regularities as well. Such series are usually of quarterly, monthly or weekly frequency. One can directly apply some models and one can separately predict seasonality and trend. For a direct prediction of time series SARIMA model, Holt-Winter's model or trend line model with seasonal dummies (and autoregressive lags if needed) can be appropriate.

Less formal approach is based on the decomposition of the time series. Very frequently the data are seasonally adjusted (with the use of seasonal dummies, classical decomposition or more sophisticated method like TRA-

MO/SEATS, X-12-ARIMA or X-13ARIMA/SEATS[4]) and the remaining part of variability is modelled and forecasted according to the methods mentioned in the last two paragraphs. If seasonally adjusted component is the time-varying trend, one can apply ARIMA model or Holt's model. The final forecast constitutes a sum of forecast of the seasonal component and the trend component. This allows for a better insight into the phenomena [Armstrong *et al.* 2001].

The most demanding time series for forecasting are those with a cyclical component. Such series usually have weekly, monthly or quarterly frequency. As it was said before, trend and cycle might be treated as one component called a trend-cycle or a stochastic trend. In such case the models mentioned in the previous paragraph can be applied. SARIMA model is particularly recommended in the literature of the subject. Another possibility is the use of a model with time variable, seasonal dummies and lagged endogenous variables (autoregressive part). Lags and other components of the model can be selected with the use of methods based on general-to-specific modelling (Gets)[5].

Holt-Winter's model is less advised for data with cyclical variation. It can track the data quite well, but on the other hand it is not appropriate for forecasting processes, where there is a high risk that the turning points of the cycle will occur. One of the solution to overcome this problem is an extension of the model with the damping factor. However, in all methods the main problem is forecasting of the turning points. The wrong answer to the question about the moment in time when the agricultural prices can change their direction, seems to be the main source of foresting errors.

Chapters 2.2 and 2.3 discuss forecasting by time series decomposition. This is a multi-level procedure frequently used by practitioners because it provides opportunities for analysts to use their market knowledge. The final forecast constitutes, in such cases, the sum of partial forecasts of the trend (made by a trend line), the seasonality and cyclical components. The forecast of the cyclical component is usually done with the use of qualitative method. The forecast of cyclical fluctuations for the future is based on many factors simultaneously: analogies to former cycles, and the use of other market information (e.g. leading indicators).

The choice of the right model

The structure of the time series is, inevitably, a key factor underlying the choice of forecasting model. However, often the same phenomenon can be predicted using different models: simple or advanced, econometric or based on data

[4] They are discussed method is discussed in a chapter 3.
[5] More about Gets can be found in a chapter 4.3.3.

smoothing, of varying degrees of formality. The choice is not so easy as the models belong to various classes of methods.

Evans [2003] emphasises that forecasting is an art, not a science, so the forecaster's choice has a strong influence on the results obtained. Therefore, the knowledge (in the field of statistics and market) as well as skills and experience play an important role in the process. Farmers or other market agents usually apply simple models. Simple models (smoothing, trend line, decomposition) can be used by individuals with relatively limited knowledge of statistics and econometrics, as well as by the experienced analysts. There is no evidence that forecast calculated with the use of simple models are worse that those obtained in line with more sophisticated procedures. The advantage of simple models lies in their relatively easy to understand algorithms that is used to calculate the forecast. Specialists in large consulting organisations and researchers have more skills, hence they more often use advanced statistical tools.

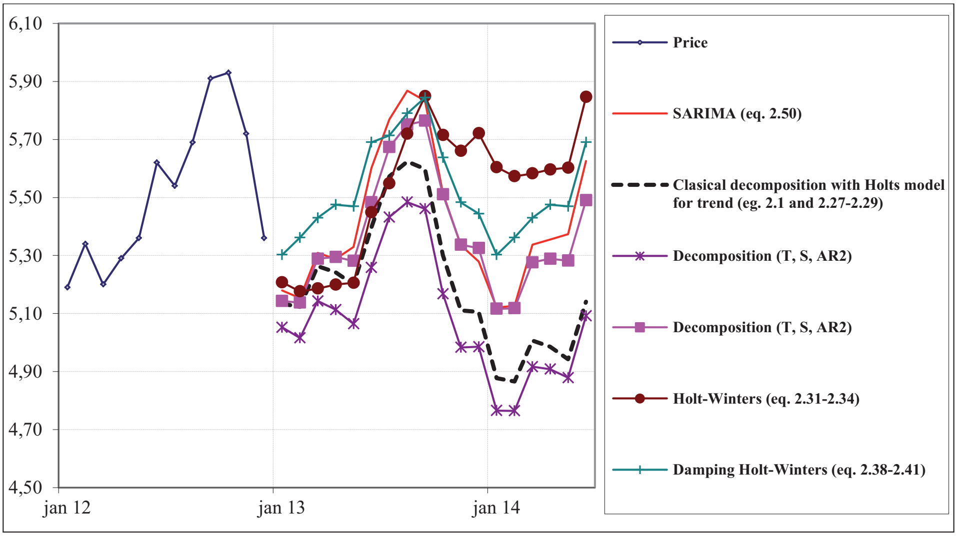
Access to statistical software is another crucial factor limiting the scope of models that can be chosen. A spread sheet is sufficient to calculate forecasts based on simple models. The more advanced model, the more advanced specialised software has to be used. Forecasting procedures are implemented in such commercial tools as EViews, Statistica, SPSS or Autobox. There are also free and Open Source software as R, Gretl, Zaitun and other.

The minimization of the future errors should be the main criterion when choosing the model and setting the parameters. It does not always means using maximization of goodness-of-fit statistics over the sample period as criterion of the model selection [Evans 2003]. There are some models which fit the data very well (e.g. exponential smoothing model with high smoothing constants), but they do not perform well in a long-term forecasting. The choice should be based on *ex post* performance of particular model in forecasting of a given price series for specific time horizon.

The typical problem of the right choice of model is presented in Figure 2.4. For monthly pig prices data in Poland for 2000-2012 (Figure 2.1) a few quantitative models of the types discussed above (which are suitable to such time series patterns) were fitted and forecasts were calculated. Even though all of them can be regarded as appropriate for such data, the discrepancies between obtained forecasts is remarkable.

Thus the key problem is the choice of the best model and the best forecast. It is widely recommended in the literature of the subject to use a few models for forecasting [Armstrong *et al.* 2001] since they can present the possible range in which real values might be contained. The final forecast can be calculated as weighted average of individual forecasts. Weights given to particular models (forecasts) may be assigned by the forecaster due to, for instance: fitting statistics or *ex post* errors. The weights may vary depending on the forecast horizon.

Figure 2.3. Forecasts of pig livestock prices in Poland from Figure 2.1 (PLN/kg)



Source: authors' own calculations.

In our case the percentage discrepancies between fitted values and real data vary from 3.63% (decomposition with Holts model) to 4.54% (Holt-Winters model). Consequently, the solution might be to give more weight for the forecast from the better models and lower weights for forecasts from the poorer models. However, the models used belong to different classes of models (econometric, exponential, more or less formal like decomposition) so fitting statistics are not comparable. As a result, a forecaster decides how to average the forecasts.

Rule-based forecasting

The selection of forecasting model is not an easy task. There are usually a few alternative models which can be applied for predicting a given phenomenon. In most cases the decision about the choice of a model is based on fitting statistics or errors of particular models estimated with the use of statistical software. However, such automatic selection may lead to choosing a model that certainly is best fitted, but not necessarily generates the reliable forecasts.

Practitioners suggest to incorporate a priori assumption based on forecaster knowledge into forecasting process. It is stressed very frequently that judgment and statistical forecasting methods should be integrated. One of the best known combination of judgment and extrapolation of forecasting techniques is Rule-Based Forecasting (RBF). RBF constitutes a strategy for forecasting time series based on validation of *ex post* forecasts [Collopy, Armstrong 1992; Adya *et al.* 2001; Armstrong *et al.* 2001]. They concerned such models as random walk, SES, LES, Holts model and trend line based on regression. Empirical results indicate that forecasts based on RBF procedures are statistically more accurate than forecasts obtained from traditional extrapolation techniques.

Rules associated with RBF can be divided into five groups [Collopy, Armstrong 1992]:

- using features of the time series to establish weight for combining (averaging) forecasts,
- using heuristics to establish parameters for exponential smoothing models,
- using separate models for short- and long-term extrapolations,
- damping the trend under certain conditions,
- incorporating domain knowledge in an extrapolation.

RBF is based on the premise that the features of time series can be reliably identified by examining the series plots. Collopy and Armstrong [1992] identified 18 features of time series in the following areas: type and significance of trend, uncertainty, types of data, instability, functional form (additive, multiplicative) or an existence of the cycle. The knowledge about underlying process is also important in identification of some time series characteristics. Among the most important time series features are those associated with instability of series.

One of RBF recommendations is to forecast separately for short- and for long-term horizon [Armstrong *et al.* 2001]. The longer the forecast horizon is, the more damping and smoothing the data needs. It means that recent changes in trend or level are of key importance in the short-term. When extending forecasting horizon a basic trend seems to play a more important role. Combining forecasts should be based also on judgment. The weights should depend on time horizon, type of model and features of time series. In general, knowledge and experience play crucial role in this method and strongly influence the forecast via imposing limits on parameters (or its calibration), the choice of models, weights of forecast in combining process.

3. X-12-ARIMA and TRAMO/SEATS procedures

The chapter presents X-12-ARIMA and TRAMO/SEATS procedures. They include RegARIMA model which is an extension of ARIMA framework discussed in the Chapter 2.4. Both of them have potential to be used for analysis of agricultural price series as well as for predicting them.

3.1. Introduction

According to Shishin, "a principal purpose of studying economic indicators is to determine the stage of the business cycle at which the economy stands. Such knowledge helps in forecasting subsequent cyclical movements and provides a factual basis for taking steps to moderate the amplitude and scope of the business cycle" [Shiskin 1957]. The question is, which tools are the most appropriate for this task. The quality of ARIMA class models, including the generated forecasts, is usually satisfactory. However, despite its flexibility in terms of parameter selection, such models cannot fully reflect the complexity of the processes that generate the analysed time series. As time series shows the evolution of economic phenomena, it usually contains observations that do not follow a simple ARIMA model. In fact, an observed time series is a combination of several distinctly different unobserved components, each representing the impact of certain types of real world events on the data. The characteristic pattern of each component is not constant in time and can be disrupted by unexpected events.

These problems have been handled by seasonal adjustment methods which can be successfully applied to the monthly and quarterly time series. The most popular of them are X-12-ARIMA[6] (X-13ARIMA-SEATS[7]) developed at

[6] X-12-ARIMA is a seasonal adjustment program that belongs to the X-11 family developed and supported by the U.S. Census Bureau. It includes all the capabilities of the X-11 program, which estimates trend and seasonal component using moving averages. X-12-ARIMA offers useful enhancements including: extension of the time series with forecasts and backcasts from ARIMA models prior to seasonal adjustment, adjustment for effects estimated with user-defined regressors, additional seasonal and trend filter options, alternative seasonal-trend-irregular decomposition, additional diagnostics of the quality and stability of the adjustments, extensive time series modelling and model selection capabilities for linear regression models with ARIMA errors. For basic information on the X-12-ARIMA program see "X-12-ARIMA Reference Manual" [2011]. More information on X-12-ARIMA can be found in http://www.census.gov.

[7] X-13ARIMA-SEATS is the newest seasonal adjustment program developed and supported by the U.S. Census Bureau that contains two seasonal adjustment modules: the enhanced

the U.S. Census Bureau [Ladiray, Quenneville 2001; Findley *et al.* 1998] and TRAMO/SEATS[8] developed by Victor Gómez and Agustín Maravall, from the Bank of Spain. Both methods apply ARIMA methodology[9] to the algorithms they use. Not only they extract different types of movements from the time series but also they calculate the forecast in a more reliable way than simple SARIMA models. Apart from modelling seasonal fluctuations in time series, seasonal adjustment has other important aims [ESS Guidelines…, 2009]. One of them is to facilitate short-term forecasting of nonseasonal movements in time series [Bell, Sotiris 2010]. Therefore, seasonal adjustment methods can be used for prediction of the short-term time series development.

The abovementioned methods enable to extract respective components from time series, i.e.:

- trend-cycle (TC), reflecting long-term movements and cyclical fluctuations having a periodicity longer than one year,
- seasonal fluctuations (S), which are cyclical movements repeating within one year, caused by the climate, institutional conditions and short-term cyclical changes in economic activity,
- irregular variations (I) composed of random fluctuations.

Time series may be thus understood as a sum of variations of various frequencies of occurrences. Its decomposition aims to eliminate the seasonal component from the series and to separate trend-cycle from the irregular variations. For seasonal adjustment purposes there is no need to divide the trend-cycle (TC) into trend (T) and cyclical variations (C), although it is possible for sufficiently long time series.

X-11 seasonal adjustment procedure and ARIMA model based seasonal adjustment procedure from the SEATS seasonal adjustment program developed by Victor Gomez and Agustin Maravall [2013]. For basic information on the X-13ARIMA-SEATS program see "X-13ARIMA-SEATS Reference Manual" [2013]. More information on X-12-ARIMA can be found in http://www.census.gov.

[8] TRAMO/SEATS is a model-based seasonal adjustment method developed by Victor Gomez and Agustin Maravall (Bank of Spain). It consists of two linked programs: TRAMO and SEATS. TRAMO ("Time Series Regression with ARIMA Noise, Missing Observations, and Outliers") performs estimation, forecasting, and interpolation of regression models with missing observations and ARIMA errors, in the presence of possibly several types of outliers. SEATS ("Signal Extraction in ARIMA Time Series") performs an ARIMA-based decomposition of an observed time series into unobserved components. Both programs are supported by Bank of Spain. For basic information on the TRAMO/SEATS see G. Caporello and A. Maravall [2004]. More information on TRAMO/SEATS can be found in www.bde.es.

[9] Both abovementioned methods are recommended by Eurostat and European Central Bank for seasonal adjustment of reporting data by national statistical institutions and central banks.

Both X-12-ARIMA and TRAMO/SEATS methods are distinctly divided into two parts: the stage of preliminary modelling of a time series aimed at e.g. clearing the time series from the impact of shocks and the stage where the actual decomposition of the time series, forecast calculation and model quality assessment are carried out. The use of predicted values reduces the size of the revisions of the seasonal adjustment, in particular at the end of the series. Forecasting also improves quality of the end adjustments.

The preliminary modelling of a time series, not included in ARIMA methodology, is of primary importance for the quality of the obtained results. It consists of identification of various types of outliers that had an impact on the series in the period under analysis, as well as of modelling of the non-seasonal calendar-related movements and other external factors. Such factors cause non-linearities in the data, which result in a poor quality of the model. Therefore these effects need to be identified, estimated and removed from time series before the actual estimation starts. This process is referred to as the linearization of a time series. The comparison of original and linearized time series is presented in a chapter 3.4.

The previously removed components are added to the corresponding components of the time series or are disclosed under separate categories after the decomposition is complete. Moreover, at the preliminary stage the type of relationship between the components of the time series is determined and the forecasts used at the further stage of estimation are calculated. Detailed description of the algorithms of X-12-ARIMA and TRAMO/SEATS is presented below.

3.2. X-12-ARIMA

X-12-ARIMA method has been developed on an empirical basis, without explicitly defined statistical decomposition model. It widely uses moving average filters, which are defined as:

$$M(Y_t) = \sum_{k=-p}^{+f} \theta_k Y_{t+k} \, , \tag{3.1}$$

where:

Y_t – time series,

θ_k – parameters of the moving average model.

The quantity $p+f+1$ is referred to as the moving average order. Application of the abovementioned filters leads to the smoothing of a time series since moving averages replace the original time series by weighted averages, the current values, p observations preceding the current observation and f observations following the current observation. When p is equal to f, the moving average is said to be centred. A symmetric moving average is a centred moving average for which $\theta_k = \theta_{-k}$ for any k. A symmetric moving average can be presented in the form of equation 3.2 [Planas 1998]:

$$M(Y_t) = \left(\theta_0 + \sum_{k=1}^{p} \theta_k (B^k + B^{-k}) \right) Y_t, \tag{3.2}$$

where B is backshift operator (see Chapter 2.4.)

To estimate the trend and the seasonal variations, the composite moving averages are used. They are obtained by composing a simple moving average of order P, whose coefficients are all equal to *1/P* and a simple moving average of order Q, whose coefficients are all equal to *1/Q*. The order of a composite moving average is denoted as $P \times Q$. For odd values of P and Q a composite moving average $M_{P \times Q}(Y_t)$ is expressed as [Grudkowska, Paśnicka 2007]:

$$M_{(2k+1)\times(2n+1)}(Y_t) = \frac{1}{2k+1} \sum_{j=-k}^{k} S_{t+sj}^{2n+1}, \tag{3.3}$$

where:

$S_t^{2n+1} = \frac{1}{2n+1} \sum_{j=-n}^{n} Y_{t+sj}$,

$2k+1 = P$,

$2n+1 = Q$,

s – number of observations in one year.

For even values of P and Q the formula 3.3 is as follows:

$$M_{2k \times 2n}(Y_t) = \frac{1}{2k} \sum_{j=-k+1}^{k-1} S_{t+sj}^{2n}, \tag{3.4}$$

where:

$S_t^{2n} = \frac{1}{2n} \sum_{j=-n+1}^{n-1} Y_{t+sj}$,

$2k = P$,

$2n = Q$,

s – number of observations in one year.

For example, coefficients of a moving average 3 x 5 are {1,2,3,3,3,2,1} and coefficients of a moving average 2 x 4 are {1,2,2,2,1}.

Symmetric moving averages can eliminate certain frequencies from a time series and do not result in a phase effect [Grudkowska 2013]. Therefore, selected symmetric moving averages can be used to extract the seasonal component from a time series. The features can be analysed for example for a time series Y_t of the frequency ω, amplitude R and phase ϕ:

$$Y_t = R\sin(\omega t + \phi). \tag{3.5}$$

The moving average filter applied to the time series results in:

$$M(Y_t) = M(R\sin(\omega t + \phi)) = G(\omega)R\sin(\omega t + \phi + \Gamma(\omega)), \tag{3.6}$$

where:

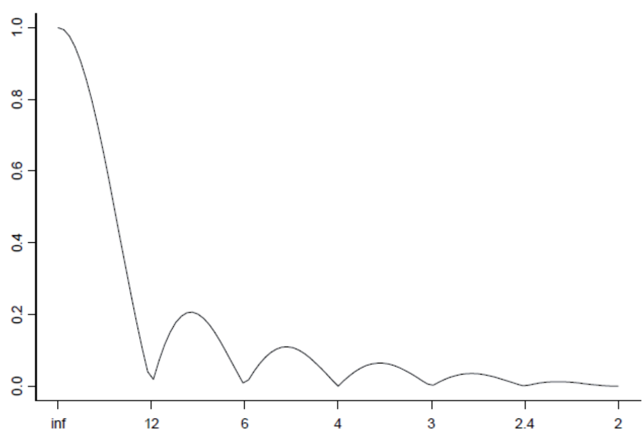
$|G(\omega)|$ – gain function expressed as $G(\omega) = \theta_0 + 2\sum_{k=1}^{r}\theta_k \cos k\omega$, presenting the frequencies eliminated or preserved by the moving average [Ladiray, Quenneville 1999],

$\Gamma(\omega)$ – the phase shift function. For a symmetric centred moving average $\Gamma(\omega)$ equals zero.

Ideally, the filter should cancel out the periodicities below specific threshold value from a time series while retaining the remaining variations unchanged, which can be written as follows:

$$G(\omega) = \begin{cases} 1 \ for \ \omega \le \omega_0 \\ 0 \ for \ \omega > \omega_0 \end{cases}. \tag{3.7}$$

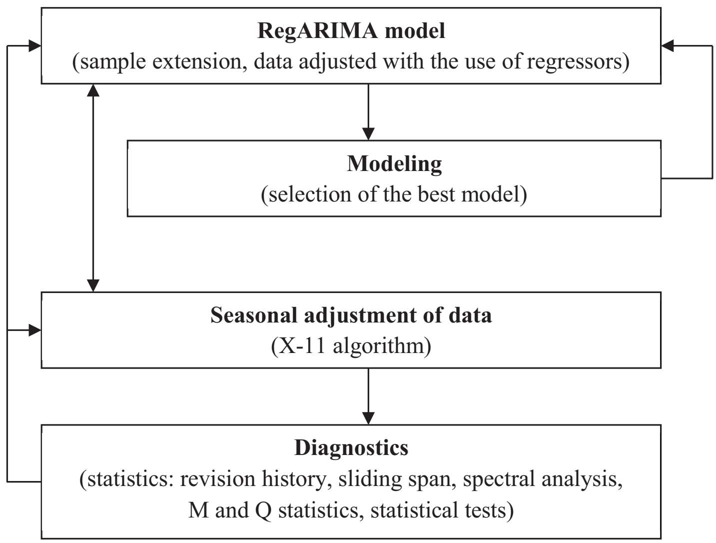
Figure 3.1. Gain function of a 2 x 12 filter



Source: [An Introductory… 2005].

The moving average filters do not fully satisfy this condition. For example, gain function of a 2×12 filter presented in Figure 3.1 removes seasonal frequencies (i.e.: $\frac{\pi}{6},\frac{\pi}{3},\frac{5\pi}{6},\frac{\pi}{2},\frac{2\pi}{3},\pi$) from the time series, but at the same time it reduces the amplitude of variations of periods less than a year, which is undesirable. Therefore, the moving average filters do not prove an efficient tool for seasonal adjustment of data in a frequency domain, but they permit to obtain a smoothed time series in a time domain. As a consequence, the decomposition of a time series by X-12-ARIMA method is carried out in time domain.[10]

Figure 3.2. Diagram for Seasonal Adjustment with X-12-ARIMA



Source: [Findley et al. 1998].

The X-12-ARIMA method is composed of two the stages: RegARIMA, aimed at linearization of a time series, and X-11, where a time series decomposition is performed using selected moving averages. This algorithm, completed with model diagnostics, is presented in Figure 3.2.

3.2.1. The RegARIMA model

RegARIMA (Regression model with ARIMA errors) is an automatic selection procedure implemented in the X-12-ARIMA method to improve the

[10] Division of the time series into components in the frequency domain is done under the TRAMO/SEATS algorithm discussed later in the study.

quality of the estimates. This model enables to complete two distinctively different tasks:

- to extend the time series with forecasts and backcasts, resulting in improvement of the X-11 filter output and reduction of revisions at the beginning and at the end of the time series;
- to estimate and remove the calendar effects, outliers and other effects from the time series, which lead to more stable and reliable seasonal adjustment estimates.

In fact the RegARIMA model combines two models: an ARIMA model and a regression model. The general regression model estimated at the RegARIMA stage is as follows [Findley *et al.* 1998]:

$$Y_t = \sum_i \beta_i X_{i,t} + Z_t , \tag{3.8}$$

where:

Y_t – original time series,

β_i – i^{th} regression coefficient,

$X_{i,t}$ – i^{th} regression variable (e.g. trading (working) days variable, leap year effect, outlier, Easter effect, ramp, intervention variable, user-defined variable such as special dummy variable or external explanatory variable),

Z_t – term that follows the general SARIMA(p,d,q)(P,D,Q) process:

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D Z_t = \theta(B)\Theta(B^S)\varepsilon_t , \tag{3.9}$$

where designations are identical with those used in Chapter 2.4.

Thus, RegARIMA fits to the original time series the regression model in which the error term follows a SARIMA process. After substitution of the formula (3.8) with the formula (3.9) the equation is as follows:

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D\left(Y_t - \sum_i \beta_i X_{i,t}\right) = \theta(B)\Theta(B^S)\varepsilon_t , \tag{3.10}$$

and it can be treated as a generalisation of the SARIMA model (formula 2.50).

At the RegARIMA stage the type of relationship between time series components is determined. This decision is crucial for the quality of seasonal adjustment as the choice of decomposition scheme aims both at obtaining the stationary time series and the most stable seasonal component [ESS Guidelines… 2009]. The choice of a decomposition scheme has also a great impact on the forecasts. Many types of relationship between the components may be considered. For the vast majority of a time series either additive or multiplicative type of relationship is an adequate one.[11]

The additive model

An additive model assumes that the value of a time series is a sum of its components, thus the original time series can be presented as follows:

$$Y_t = TC_t + S_t + I_t + D_t + E_t \, , \tag{3.11}$$

where:

t – number of observations, $t=1,\dots,T$,

TC – the trend-cycle component,

S_t – the seasonal component,

I_t – the irregular component,

D_t – the calendar effects,

E_t – the Easter effect.

An additive model is most suitable for series in which the behaviour of the irregular fluctuations, as well as the seasonal and calendar effects are independent of the trend-cycle. It means that it should be applied when series level does not influence on the fluctuations overlapping the trend-cycle component.

The multiplicative model

A time series is presented as a product of its components in the case of a multiplicative model. Therefore, this decomposition scheme implies that the seasonal and irregular variations change proportionately to the trend-cycle. In the case of a multiplicative model, the original time series is presented as follows:

$$Y_t = TC_t \cdot S_t \cdot I_t \cdot D_t \cdot E_t \, , \tag{3.12}$$

where designations are identical with those used in 3.11. The characteristic feature of the multiplicative model is that cyclical, seasonal and irregular variations can be described as relative deviations from trend-cycle.

Other models

Apart from two basic decomposition schemes presented above, there are several other models that can be used to describe the relationship between its components. These models include:

- the log-additive model, where the time series components are related in the following manner:

[11] Formulas of respective models are based on [Ladiray, Quenneville 2001].

$$\ln(Y_t) = \ln(TC_t + S_t + I_t + D_t + E_t)\,, \tag{3.13}$$

- the pseudo-additive model:

$$Y_t = TC_t(S_t + I_t + D_t + E_t - 1)\,, \tag{3.14}$$

designated for time series that include zero values.

The choice of a decomposition scheme

A multiplicative model (either pure multiplicative or log-additive) is used most often due to the fact that for most economic time series the amplitude of seasonal variations is proportional to the level of that series. The multiplicative model is particularly useful when the seasonal and irregular fluctuations change in a specific manner, as a result of the behaviour of the trend. In the case of a multiplicative relationship the amplitude of the seasonal fluctuations increase (decrease) with an increasing (decreasing) trend-cycle. On the contrary, in the additive case, the components are not independent from each other. When trend in both the mean and the variance is present, the log-additive decomposition is recommended [ESS Guidelines…, 2009]. In general, the inappropriate choice of decomposition scheme has an adverse effect on the seasonal adjustment results.

Regressors

The regressors estimated by RegARIMA include: the constant, which corresponds to the parameter ϕ_0 in equation 2.50, outliers, seasonal dummy variables, calendar effects and other regression effects (e.g. user-defined regression variables).

Outliers referred to include additive outliers (AO), level shifts (LS), temporary changes (TC), ramp effects (RP), seasonal level shifts (SLS), and reallocation outliers (RO).[12] The nature of those outliers is presented below (and in Figure 3.3).

Additive outliers

An additive outlier is a variable for a point outlier which occurred in a given date t_0. It is marked as AO and modelled by the regression variable[13]:

$$AO_t^{t_0} = \begin{cases} 1 \; for \; t = t_0 \\ 0 \; for \; t \neq t_0 \end{cases}. \tag{3.15}$$

[12] Definition of the AO, LS, TC and RP type outliers presented later in this study is based on [*X-12-ARIMA*…, 2011].
[13] Definition from [Grudkowska 2013].

This regressor is used for the modelling of an impact of a single impulse on a time series. Such outliers can result from human errors (mistakes during writing data) or simply through natural deviations in populations. Examples of such events are weather anomalies or strikes.

Level shifts

A level shift (LS) is a variable for a constant level shift beginning on the given date t_0. It is modelled by the regression variable:

$$LS_t^{t_0} = \begin{cases} 0 \; for \; t < t_0 \\ 1 \; for \; t \geq t_0 \end{cases}. \tag{3.16}$$

Examples of such events are changes in economic policy: tax rates or levies of customs duties.

Temporary changes

A temporary change (TC[14]) is a variable for a level change beginning on the given date t_0 and decaying exponentially over the following periods. A rate of decay back to the previous level is denoted as α $(0 < \alpha < 1)$. It is modelled by the following regression variable:

$$TC_t^{t_0} = \begin{cases} 0 \; for \; t < t_0 \\ \alpha^{t-t_0} \; for \; t \geq t_0 \end{cases}. \tag{3.17}$$

The TC variable is used for the modelling of events whose impact on a time series trend decreases in time. This may include for example the effect of extreme weather conditions sustaining for several periods (such as severe drought or flood).

Ramp effects

A ramp effect (*RP*) is a linear increase or decrease in the level of the series over a specified time interval t_0 to t_1. It is modelled by the variable:

$$RP_t^{(t_0,t_1)} = \begin{cases} -1 \; for \; t \leq t_0 \\ (t-t_0)/(t_1-t_0)-1 \; for \; t_0 \leq t < t_1 \\ 0 \; for \; t \geq t_1 \end{cases}. \tag{3.18}$$

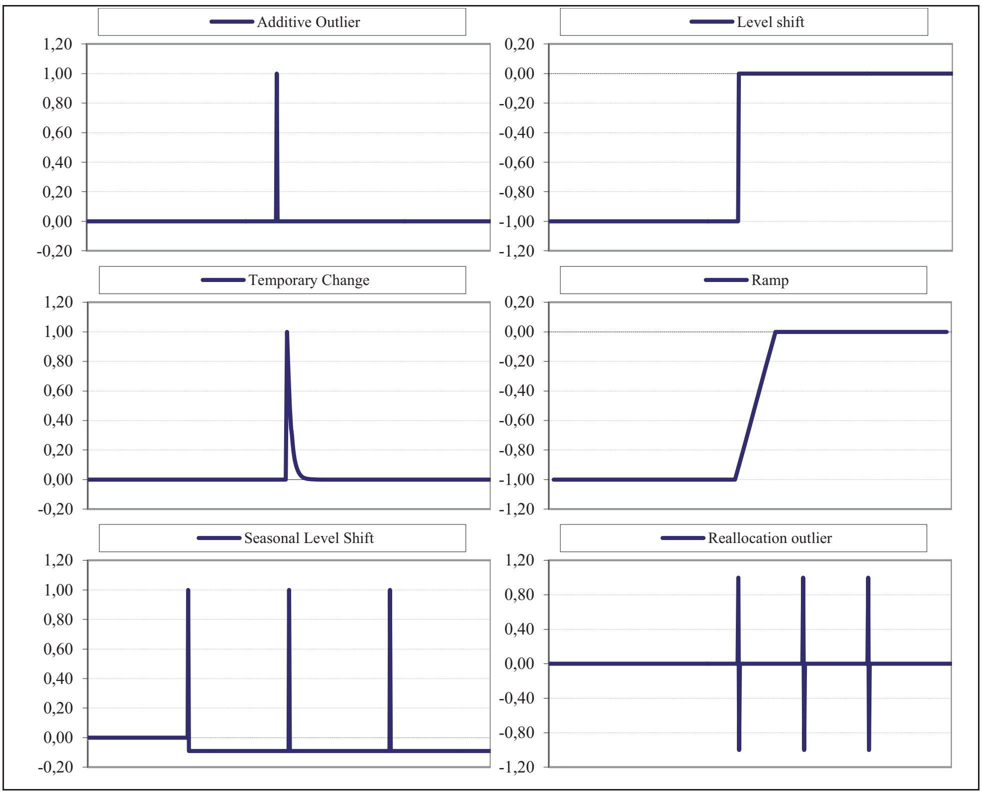
[14] As a rule, in the present study TC refers to long-term trend. Here, by way of exception, we refer to the transitory change of the time series, due to the designations adopted in the X-12-ARIMA methodology.

This regressor is used to describe the same events as LS. However, contrary to LS, it assumes that the change of a trend level is spanned over more than one period.

Reallocation outliers

In the case of a reallocation outlier (RO) the impact of the regressor on observations from the neighbouring periods is opposed and equal in terms of absolute value [Wu *et al.* 1993]. Fluctuations are modelled as a composition of two AO variables. This type of regressor is used for the modelling of changes in the seasonal pattern occurring in the year (e.g. a shift of payment of annual bonuses from December to January).

Figure 3.3. Outliers' types



Source: author's own compilation.

Seasonal level shifts

Seasonal level shifts (SLS) is a variable for a permanent change in the seasonal pattern of a time series beginning at time t_0. This regressor captures an abrupt change in the seasonal pattern, and maintains the level of the series with a contrasting change spread over the remaining months or quarters.[15] SLS is modelled by the following variable:

$$SLS_t^{t_0} = \begin{cases} 0 \; for \; t < t_0 \\ 1 \; for \; t \geq t_0, t \; same \; period \; as \; t_0 \\ \dfrac{-1}{(s-1)} \; otherwise. \end{cases} \tag{3.19}$$

where s is the frequency of the time series being modelled (12 for monthly series, 4 for quarterly series). An example of a situation where SLS variable can be used is the introduction of annual bonuses from the period t_0, payable to the employees each year.

Calendar effects

The relationship between the value of observation and the number of the work-free days in a given period can be observed in the case of certain time series. It is caused by the calendar effects. They result from the differences in the number and pattern of the working and non-working days in respective periods in consecutive years. The calendar effects cause regular movements in some series. These movements may be caused, e.g. by a leap year effect (because of the extra day in February every four years) and by a presence of a moving holidays.

The calendar effects can be separated into a seasonal part and a non-seasonal part. The seasonal part arises from the properties of the calendar that recur each year (e.g. the number of working days of months with 31 calendar days is on average larger than that of months with 30 calendar days). The non-seasonal part of the calendar effect which includes e.g. the leap year effect (it does not occur every year) is called “the calendar component”. The time series are corrected for it in a pre-adjustment step by including the appropriate regression variables. The examples of such variables are presented in RegARIMA model description.

The seasonal adjustment methods correct time series for several types of calendar effects [Findley *et al.* 1998] including:

- Working day effect, that distinguishes two types of days: working days and non-working days. This effect is applied for time series, which values depend on the number of working days and in can be assumed that there are two

[15] Definition from Monsell: *www.fcsm.gov/07papers/Monsell.II-B.pdf*

states of the level of the economic activity: one of them is characteristic for each working day and the second one is characteristic for non-working days;
- Trading days effect, that arises from systematic effects in a time series related to changes in the day-of-week composition of each period;
- The Length-of-Month effect, that indicates deviation of the number of days in a given period from the long-term average;
- Leap Year effect, which measures the impact of an additional day in the year (in the case of Gregorian calendar it is inserted in February every four years except from the years that are evenly divisible by 100, on the value of a time series);
- Moving holidays effect, that relate to the holidays that fall in different days and months from year to year. The examples are: Easter, Pentecost, Corpus Christi and Chinese New Year. Their presence influence on the economic activity not only in the day they fall, but also before and/or after it.

The trading and working days effects are rarely identified the case of price-related time series because price level rarely depends on the specific day of the week. However, it is probable, at least in the case of products that their price changes before holidays. Therefore, the significant moving holiday effect may appear in some price-related time series. In Poland the moving holiday that has the strongest impact on the economic time series is Easter.

The choice of the RegARIMA model

The process of determining the RegARIMA model (3.10) is a multilevel one.[16] The SARIMA $(0,1,1)(0,1,1)_s$ model is adopted for the time series and it is used for testing of significance of calendar effects and other regressors belonging to vector $X_{i,t}$, except for the automatically detected outliers, in the first stage of this procedure. The significance of those variables is verified with the use of the modified Akaike information criterion (AICC)[17]. The significance of the

[16] Description of the procedure is based on *X-12-ARIMA Reference Manual.* The described algorithm is based on the procedure used by the TRAMO program. In the case of X-12--ARIMA it is also possible to apply other methods for determining the form of the autoregressive model.

[17] The formula for Akaike information criterion is a follows: $AIC_N = -2L_N + 2n_p$. AICC is AIC with a greater penalty for extra parameters. The statistics of the modified Akaike information criterion is as follows: $AICC_N = -2L_N + 2n_p\left(1-\frac{n_p+1}{N}\right)^{-1}$, where: N – number of observations, n_p – number of model parameters, L_N – logarithm of model reliability function [X-12-ARIMA Reference Manual].

constant in the model (3.10) is verified with Student's t-test. In the next stage, outliers are automatically identified, after which the significance of calendar effects and the constant are verified again. The model residuals are diagnosed (the Ljung-Box test on residuals and standard deviation of residuals are applied) and disruptions modelled by the identified regressors are eliminated from the series in the last stage.

The aim of the second step of the algorithm is to identify the regular and seasonal differencing orders of the target SARIMA model, i.e. determination of the values of parameters *d* and *D*. This is carried out with the use of the Hannan-Rissanen method.[18] Once they are determined in the model, the constant equal to the average from the differentiated time series is included and its significance is verified.

The next stage is an identification of parameters of the seasonal part of the SARIMA model, i.e. *P* and *Q*, for a stationary time series. This is carried out through comparison of the value of Bayesian information criterion (BIC)[19] for various specifications of the SARIMA model in the form: $(3,d,0)(P,D,Q)$. The model with the lowest value of information criterion is preferable. Then on the basis of BIC the values of parameters *p* and *q* are determined in the model (p,d,q)(P,D,Q), where the values of *P* and *Q* have been calculated in the previous step. After that parameters *P* and *Q* are selected again with the use of BIC [X-12-ARIMA… 2011].

At the fourth stage of this procedure the model determined in the previous step is compared with the default model $(0,1,1)(0,1,1)_S$ and one of them is selected.[20] The properties of the chosen model are checked, including the residuals analysis and checking the significance of parameters of the ARIMA model and regressors.

The model selected by RegARIMA procedure is used both for forecasting and backcasting. Linear forecasts Y_t obtained from the current and the past observations of the time series reduce MMSE (Minimum Mean Square Error), assuming that the structure of the ARIMA model and regressor selection are correct. In their calculation it is assumed that in the period covered by the forecast

[18] The Hannan-Rissanen method is a two-step procedure for selection of orders of the autoregressive process and the moving average process in the ARIMA model. Detailed description can be found in [*X*-12-ARIMA Reference Manual*,* 2011].

[19] The statistics of Bayesian information criterion is as follows: $BIC_N = -2L_N + n_p \log N$, where: N – number of observations, n_p – number of model parameters, L_N – logarithm of model reliability function.

[20] Model selection criteria have been described in [X-12-ARIMA… 2011].

no outliers occur. Application of the forecasted values permits the X-11 algorithm to apply symmetrical filters over a whole time series span [X-12-ARIMA… 2011].

3.2.2. The X-11 algorithm[21]

The X-11 algorithm developed by U.S. Census Bureau is a decomposition procedure using local filters that enable to minimise the revision in the centre of the time series span. It assumes that each time series can be decomposed into the trend-cycle, the seasonality and the irregular component. The algorithm is based on an iterative principle of estimation of different components using appropriate moving averages at each step of the algorithm, taking into account the possible presence of extreme observations. The algorithm proceeds iteratively: estimation of components, search for disruptive effects in the irregular component, estimation of components over a corrected series, search for disruptive effects in the irregular component, and so on. The results from each step are saved in appropriate table.

The description below applies to the additive model; other models require prior application of appropriate transformations to obtain the additive form.[22] The original X-11 program consists of four processing stages and three diagnostic parts:

- Part A: Pre-adjustment;
- Part B: First automatic correction of the series;
- Part C: Second automatic correction of the series;
- Part D: Seasonal adjustment;
- Part E, F and G: Statistics and Charts.

Before RegARIMA model has been adopted for linearization of a time series, the correction of the series for trading days effect, statutory holidays, and known outliers was performed in part A. In X-12-ARIMA this step is skipped.

In stages B, C and D the basic algorithm is applied to gradually improve estimates of the components. Part B leads to the first estimation of the components, detection and automatic correction of extreme observations. Part C results in better estimation of the components and correction for the extreme observations. Finally, Part D results in the final estimates of the components. In Part E the components are adjusted for large extreme values. Part F includes seasonal adjustment quality measures and Part G presents some graphics.

[21] Description of the algorithm is based on [Ladiray, Quenneville 2001].
[22] For example, for the multiplicative model the appropriate operation is logarithmic.

The basic algorithms, that is used in stages B, C and D consists of eight steps. They are listed below.

1. Estimation of Trend-Cycle by 2x12 moving average:

In the first part the trend-cycle component is estimated by applying the 2×12 MA of coefficients $\{\frac{1}{24},\frac{2}{24},\frac{2}{24},\frac{2}{24},\frac{2}{24},\frac{2}{24},\frac{2}{24},\frac{2}{24},\frac{2}{24},\frac{2}{24},\frac{2}{24},\frac{2}{24},\frac{1}{24}\}$ to the original time series:

$$TC_t^{(1)} = M_{2\times12}(Z_t), \tag{3.20}$$

where:

TC_t – trend-cycle,

$^{(1)}$ – iteration number,

Z_t – linearised time series from the equation (3.8).

As it can be seen in Figure 3.1, this filter removes seasonal variations that occur with the frequencies of once, twice, three times, four times, five times and six times a year. This filter has a little impact on long-term variations and at the same time it reduces the amplitude of high-frequency variations that correspond to the irregular component.

2. Estimation of the Seasonal-Irregular component:

Then the sum of the seasonal and irregular components (Seasonal-Irregular component SI) is calculated by removing the trend-cycle from the time series:

$$(S_t + I_t)^{(1)} = Z_t - TC_t^{(1)}, \tag{3.21}$$

where other designations identical to those above. The exemplary chart of the SI component is discussed in Chapter 3.4.

3. Estimation of the Seasonal component by 3x3 moving average over each month:

The seasonal and irregular components are separated through application of a moving average 3×3 to the SI component, as a result of which it is smoothed, which is a preliminary estimation of the seasonal factor:

$$S_t^{(1)} = M_{3\times3}[(S_t + I_t)^{(1)}]. \tag{3.22}$$

The moving average used here is a 3 × 3 moving average over 5 terms, of coefficients $\frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \frac{2}{9}, \frac{1}{9}$. The filter is applied to the seasonal-irregular ratios for each period, separately, over 5 years.

In the next step the seasonal factors are normalised using a centered 12-term moving average, so that the seasonal effects over the whole 12-month period are approximately cancelled out[23].

$$\tilde{S}_t^{(1)} = S_t^{(1)} - M_{2\times12}(S_t^{(1)}). \tag{3.23}$$

4. Estimation of the seasonally adjusted series:

The estimation of the seasonally adjusted series is done by removing from the starting series (Table B1) the first estimate of the seasonal component (Table B5):

$$SA_t^{(1)}=(TC_t+I_t)_t^{(1)}=X_t-\tilde{S}_t^{(1)}. \tag{3.24}$$

5. Estimation of the Trend-Cycle by 13-term Henderson moving average:

The second estimation of trend-cycle is obtained from seasonally adjusted series, calculated on step 4, smoothed with 13-term Henderson filter.

$$TC_t^{(2)}=H_{13}\left(SA_t^{(1)}\right). \tag{3.25}$$

6. Estimation of the Seasonal-Irregular component:

The second estimate of the seasonal-irregular component is calculated by subtracting the trend-cycle from the linearised time series.

$$(S_t+I_t)^{(2)}=X_t-TC_t^{(2)}. \tag{3.26}$$

7. Estimation of the Seasonal component by 3x5 moving average over each month:

The second estimation of the Seasonal-Irregular component is done by removing the trend-cycle from the time series using *3x5* moving average over 7 terms, of coefficients $\frac{1}{15}, \frac{2}{15}, \frac{3}{15}, \frac{3}{15}, \frac{3}{15}, \frac{2}{15}, \frac{1}{15}$:

$$S_t^{(2)}=M_{3\times5}\left[(S_t+I_t)^{(2)}\right]. \tag{3.27}$$

The coefficients are then normalized such that their sum over the whole 12-month period is approximately cancelled out.

$$\tilde{S}_t^{(2)}=S_t^{(2)}-M_{2\times12}\left(S_t^{(2)}\right). \tag{3.28}$$

[23] The 2x12 filter, similarly to the 3x3 filter, maintains the trend. Normalisation with the use of the filter consists in elimination of the trend from the seasonal component. In the case of a multiplicative model, the sum of seasonal components in one year is equal to the number of periods in the year (12 for monthly data, 4 for quarterly data).

8. Estimation of the seasonally adjusted series:

The estimation of the seasonally adjusted series is done by removing from the linearised series the second estimate of the seasonal component:

$$SA_t^{(2)}=(TC_t+I_t)^{(2)}=X_t\text{-}\tilde{S}_t^{(2)}. \tag{3.29}$$

To run the algorithm the choice of the moving averages used for the estimation of the trend-cycle in steps 1 and 5, and for the estimation of the seasonal component in steps 3 and 5 need to be done. The method of choosing these values is presented in [Ladiray, Quenneville 1999].

3.2.3. Model validation

In the final step of the X-12-ARIMA method the model is subject to diagnostics. The presence of seasonality in a time series and its characteristic are determined on the basis of seasonality tests. Other tools, like *M* and *Q* statistics as well as sliding spans diagnostic checks the quality of adjustment. The size of revision of the trend and the seasonal component is calculated with the use of revision history.[24] In more complex versions of this algorithm, filter length selection depends on the characteristic features of the components [X-12--ARIMA… 2011].

As X-12-ARIMA is a heuristic seasonal adjustment method based on adhoc filters rather than econometric model, its results should not be tested using standard statistical tests. The validation of its results is based on criteria that arise from the analysis of the seasonal adjustment of typical series. Therefore, in the case of X-12-ARIMA, the model validation results indicate potential problems, but do not necessarily approve or reject the outcome. The diagnostics include:[25]

- seasonality tests,
- spectral analysis of the seasonally adjusted data,
- M and Q quality statistics,
- analysis of SI indicators,
- revision size assessment,
- stability analysis.

[24] Seasonality tests, periodogram, M and Q statistics, sliding spans and revision history are discussed in detail in Chapter 3.4. For description of revision history see *X-12-ARIMA Reference Manual* [2007].

[25] Description of respective statistics and measures of the quality of the seasonal adjustment of data is available in [Findley *et al.* 1998], [Ladiray, Quenneville 2001] and [Gomez, Maravall 2001].

Seasonal adjustment of data should not be carried out in case of series for which the seasonal component is not significant, because it leads to artificial introduction of seasonal components into the data. The nature of seasonality is also important. A seasonal pattern that changes too fast is difficult to model and requires precise data analysis. Comprehensive assessment of seasonal variations comprises the stable seasonality test, Kruskall-Wallis test and test for the presence of seasonality assuming stability, as well as the moving seasonality test, identifiable seasonality test and combined seasonality test. A detail description of the tests is available in [Ladiray, Quenneville 2001].

These tests are performed on different stages of the algorithm. They test the presence and the characteristic features of the seasonal movements in the time series. The tests concentrate on detecting the stable seasonality, as it is a necessary prerequisite for seasonal adjustment, so rejection of the hypothesis that the stable seasonality exists, implies that the adjustment should not be performed.

M and Q statistics

The compliance of the relationships between the obtained components and the expectations is verified with the use of the X-12-ARIMA method on the basis of M and Q statistics.[26] The set of M statistics includes:[27]

- M1, which measures the share of the irregular component variance in the series variance;
- M2, which measures the volume of the irregular component with regard to the linear trend. The value of M2 statistics may lead to erroneous conclusions if the trend is not (approximately) linear. Therefore, it is often omitted in the analysis;
- M3, measuring the relationship between the value of the irregular component and the trend at the first stage of estimation. High value of M3 indicates strong irregularity in the time series which might affect the course of decomposition;
- M4, verifying the randomness of the irregular component on the basis of the autocorrelation test. Due to the fact that lack of autocorrelation of the seasonal component is not required to obtain high quality of the seasonal adjustment, M4 statistics is not considered key quality measure;

[26] The assumptions adopted under the TRAMO/SEATS method are verifiable with the standard assessment of component correlation coefficients. Considering the common knowledge of this type of analysis, its description has been omitted in this study.
[27] Detailed description of M statistics including the appropriate formulas can be found in [Lothian, Morry 1978].

- M5, indicating the number of periods necessary for the trend-cycle variance to exceed the irregular component variance;
- M6, testing whether the volume of changes in the irregular component measured year-on-year is appropriate for the application of the *3x5* filter to *SI* estimation. The excessively high value of M6 indicates the need to apply a shorter filter for *SI* estimation [Guide to… 2007];
- M7, which is the identifiable seasonality test;
- M8, verifying the volume of short-term, quasi-random disruptions;
- M9, testing the existence of long-term fluctuations in the seasonal component;
- M10, which is defined analogically to M8 and which is calculated for the observations from the last three years in the sample;
- M11, which is defined analogically to M9 and which is calculated for the observations from the last three years in the sample.

M8, M9, M10 and M11 statistics are useful for detection of occurrences of seasonal pattern breaks, which are undesirable from the point of view of the seasonal adjustment of data and which require individual analysis.

Each of the M1 to M11 statistics may have values from the range [0,3], whereas acceptable values are the ones below 1.

The abovementioned statistics can be combined into a general measure of the quality Q, which is defined as follows:

$$Q = \frac{10M1 + 11M2 + 10M3 + 8M4 + 11M5 + 10M6 + 18M7 + 7M8 + 7M9 + 4M10 + 4M11}{100}. \quad (3.30)$$

If the time series is shorter than 6 years, M8, M9, M10 and M11 statistics cannot be calculated and the value of Q is expressed as follows:

$$Q = \frac{14M1 + 15M2 + 10M3 + 8M4 + 11M5 + 10M6 + 32M7 + 0M8 + 0M9 + 0M10 + 0M11}{100}. \quad (3.31)$$

The quality of the seasonal adjustment results is considered acceptable if the values of respective M statistics and Q measure are below 1.

A model of seasonal adjustment of data should generate stable results, i.e. the results which are not subject to material changes as the time series is extended by subsequent observations. This feature is verified through sliding spans diagnostic.

Sliding spans

Sliding spans is a diagnostic tool designed to examine the stability of seasonal adjustment outcome using so called spans. The span is defined as a range of data between two dates. Spans are chosen from a time series in a way that one observation belongs to several spans. Each span is seasonally adjusted separately. The tool compares the outcomes of seasonal adjustment for each observation that belongs to the more than one span and checks if the difference between the values are above or below the threshold value. The investigation can be conducted for:

- seasonal component;
- trading days effect (if it is present in the time series);
- seasonally and trading day (if this effect is present) adjusted time series.

Apart from assessing the stability of seasonal adjustment, the tool is also useful for detecting significant changes in the original time series, like seasonal breaks, large number of outliers and fast moving seasonality.

3.3. TRAMO/SEATS

TRAMO/SEATS is a seasonal adjustment method where components are separated from the time series on the basis of appropriately selected ARIMA models. It is composed of two stages. At the first stage – TRAMO – an estimation and forecasting of regression model with possibly non-stationary (ARIMA) errors and any sequence of missing values is performed. TRAMO interpolates missing values, identifies and corrects for several types of outliers, estimates calendar effects and different types of intervention variables. At the second stage – SEATS – an estimation of unobserved components in time series takes place following the so-called ARIMA-model-based method.

The time series components: trend-cycle, seasonal, irregular, and transitory, are estimated and forecasted with signal extraction methodology applied to ARIMA models. SEATS validates the model by checking its statistical properties and calculates standard errors of the estimates and forecasts. For validation TRAMO/SEATS procedure the tests presented in section 3.2.3 can be applied.

3.3.1. The TRAMO procedure

The key objectives of this procedure (*Time Series Regression with ARIMA Noise, Missing Observations and Outliers*) include: interpolation of the time series in the possible presence of outliers, estimation of the regression model with errors described with the use of ARIMA model and forecasting of the time series on the basis of the estimated model.

The TRAMO algorithm fits the following regression model to the original time series z_t [Maravall 2006]:

$$z_t = y_t\beta + x_t . \qquad (3.32)$$

where:

$\beta = (\beta_1, ...\beta_n)$ – vector of regression coefficients;

$y_t = (y_{1t}, ..., y_{nt})$ – n regression variables (trading days variables, the leap year effect, outliers, the Easter effect, ramps, intervention variables, user-defined variables);[28]

x_t – term that follows the general ARIMA process:

$$\phi(B)\delta(B)x_t = \theta(B)a_t, \tag{3.33}$$

where:

B – the backshift operator,

$\delta(B) = (1-B)^d(1-B^s)^D$ – polynomial in B including roots related to the order of regular and seasonal differentiation of the time series,

$\varphi(B) = (1+\varphi_1 B + ... + \varphi_p B^p)(1+\varphi_1 B^s + ... + \varphi_P B^{sxP})$ – polynomial in B including autoregressive process roots,

$\theta(B) = (1+\theta_1 B + ... + \theta_p B^q)(1+\theta_1 B^s + ... + \theta_Q B^{sxQ})$ – an invertible[29] moving average (MA) polynomial in B,

a_t – a white noise process[30], $a_t \sim N(0, V(a))$.

The procedure for estimation of equation 3.32 is almost identical to the estimation made by RegARIMA. Linearisation of the series by RegARIMA follows the solutions developed for TRAMO, therefore, these two programs usually produce the same results. The ARIMA model estimated by TRAMO, including the calculated forecasts, is used by the SEATS program.

3.3.2. The SEATS procedure

The SEATS procedure (*Signal Extraction in ARIMA Time Series*) consists in decomposition of the time series described by the ARIMA model into the unobserved component ARIMA models: the trend-cycle, the seasonal fluctuations, the transitory component and the irregular movements. For the purposes of the estimation, SEATS uses the ARIMA model selected by TRAMO.[31] The decom-

[28] The constant is equal to the average from the differentiated series $\delta(B)z_t$. Regressors analysed by TRAMO are defined identically as it is the case under RegARIMA.

[29] The moving average process is reversible if all results of the formula $\varphi(B) = 0$ are higher than 1 in terms of module.

[30] Awhite noise process has been designated according to the convention applied in the studies on the subject relating to the TRAMO/SEATS method. Under the X-12-ARIMA method the white noise process is marked as ε_t.

[31] In the case when the ARIMA model selected by TRAMO cannot be decomposed, SEATS carries out a new identification of the ARIMA model.

position may have the additive or the multiplicative form, whereas the latter may be transformed to the additive form with the use of logarithmic transformation. In the case of an additive form, the time series x_t is presented as a sum of the components:[32]

$$x_t = \sum_{i=1}^{k} x_{it} \,, \tag{3.34}$$

where each i^{th} component is the realisation of the ARIMA process:

$$\delta_i(B)x_{it} = \psi_i(B)a_{it} \,, \tag{3.35}$$

where:

i – trend-cycle, seasonal, transitory and irregular components, respectively,[33]

$\psi_i(B) = \dfrac{\theta_i(B)}{\varphi_i(B)}$, where the polynomials $\theta_i(B)$, $\phi_i(B)$ and $\delta_i(B)$ are of finite order,

$a_{it} \sim WN(0, V(a_i))$ – white noise process with zero mean and constant variance, $V(a_i)$ referred to as innovation of the i^{th} component,

a_{it} is the error estimator in the single-period forecast of the i^{th} component.

The number of possibilities in which the ARIMA model for observed time series can be decomposed into ARIMA models for the components is infinite. Therefore some assumptions are made. They assure that from a statistical point of view the chosen decomposition is the most appropriate one.

First of all, it is assumed that in the estimation process the time series components are orthogonal to each other.[34] This condition is verified through analysis of the correlations between the components obtained through estimation and the appropriate theoretical estimators. The correlation is considered negligible if the significance level is higher than the assumed significance level of 5%. Apart from that, it required that models for components should not contain the

[32] In the multiplicative case the formula is $x_t = \prod_{i=1}^{k} x_{it}$, and after calculating logarithms on both sides $\left(\log(x_t) = \sum_{i=1}^{k} \log(x_{it}) \right)$ it can be analysed identically to the additive case.

[33] The irregular component, which, as a rule, is a white noise process, is always presented as ARIMA (0,0,0)(0,0,0).

[34] This assumption means that the course of respective components is caused by various, independent reasons. For example, seasonal and calendar factors result in a creation of the seasonal component, while the trend-cycle results e.g. from a specific production method, selected technology and macroeconomic stimuli. The orthogonality assumption permits unequivocal attribution of specific frequencies of the input series to one of the components and further independent analysis of the course of each component.

same unit roots of the autoregressive process $\varphi(B)$. Finally, the components should be independent of each other. The compliance of the results with these assumptions is verified with appropriate test.

The decomposition is performed in the frequency domain and consists in the division of the spectral frequency function of the series x_t into the spectral frequency functions of respective components. The trend-cycle includes the values accumulated around the zero-level spectral frequency. The seasonal component comprises the values of the spectral frequency function within specific range around the seasonal frequencies. The transitory component consists of cyclical fluctuations whose change period is more than one year. The irregular component is a white noise process.

It is expected that the estimated trend-cycle and seasonal components are as stable as possible. Because of that in the estimation process the variance of the irregular component is maximised, thus separation of a white noise from other components is impossible.[35]

The properties of the model applied by SEATS result in the theoretical estimator of a given component having lower variance than such component [Maravall 1993]. Moreover, variance of the theoretical estimator should be close to the variance of the component obtained as a result of the estimation. If, for a given component, the variance of the theoretical estimator is significantly higher than the variance of the component, it means that that component is overestimated in the estimation process. The reverse implies underestimation of that component [Grudkowska 2013].

Auxiliary tools for the analysis of the results include spectral frequency functions and squared gain of the component filter. Spectral frequency function presents the decomposition of the spectrum for the linearised time series to the spectra for the components. This tool informs in a graphical way which frequencies exist in a time series and how they have been distributed among particular components.

The squared gain of the component filter controls how a movement of particular amplitude at a frequency ω is delivered to the output series. It reveals how the variance of the series contributes to the variance of the component for each frequency. When squared gain is zero in some span, it implies that the given series is free of movements in this range of frequencies [Planas 1998]. On the contrary, if for some ω the square gain is 1, then all variation from linearised time series is passed on to the component estimator.

[35] It is a so-called canonical decomposition, which assumes that apart from the irregular component none of the time series components includes the white noise process.

3.4. Application of X-12-ARIMA and TRAMO/SEATS to analysis and forecasting sheep prices in Poland

This chapter discusses the outcome of the seasonal adjustment using exemplary time series. It focuses on the pre-adjustment part of the processing, as this step is crucial to obtaining reliable results and high quality of the forecasts. This section also explains how to assess the quality of the adjustment. The example used here is monthly time series that presents live sheep prices in Poland from January 1996 to June 2013 (Figure 3.4). To check the quality of estimated forecast the estimation span is shortened to June 2012. Estimations were carried out in *JDemetra+* software[36].

Automatic identification

The automatic model identification was performed at first. The procedure chooses the model that passes validation tests and has the lowest information criteria. Although the general procedure for picking the model is similar in both leading seasonal adjustment methods, there are some differences that may result in different results. This is the case of the analysed time series. The RegARIMA model chosen by TRAMO/SEATS is more complex than the one estimated by X-12-ARIMA. Apart from that, the model decomposition is not the same in both cases. Therefore, the user is expected to investigate the results and improve them manually. Table 3.1 reveals the results from automatic model identification procedure.

Table 3.1. Estimation of RegARIMA model within TRAMO/SEATS and X-12-ARIMA

Parameters	RegARIMA model					
	TRAMO/SEATS method			X-12-ARIMA method		
	Multiplicative decomposition			Additive decomposition		
	Coefficients	T-Stat	P[|T| > t]	Coefficients	T-Stat	P[|T| > t]
AR (1)	-0.6651	-11.97	0.0000	-0.6753	-12.55	0.0000
MA(1)	0.2586	2.92	0.0039	0.2331	2.60	0.0101
SMA(1)	-0.5620	-7.62	0.0000	-0.6252	-8.84	0.0000
Easter [15 days]	0.1563	4.31	0.0000	0.9343	4.37	0.0000
LS[02.1998]	-0.3067	-4.10	0.0001	-	-	-

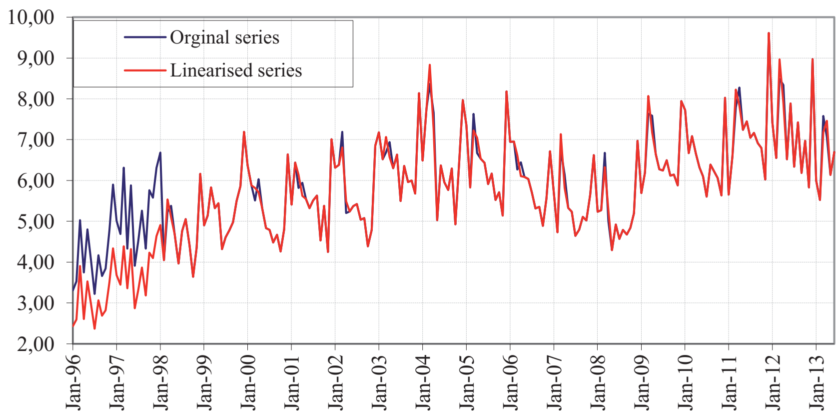
Source: author's own compilation, prepared with JDemetra+.

[36] *JDemetra+* is a new open source tool for seasonal adjustment that enables the implementation of ESS Guidelines on SA. The software is promoted by Eurostat and can be downloaded from http://www.cros-portal.eu. For more information about *JDemetra+* see [Grudkowska *et al.* 2013].

The decomposition type is the most crucial difference between the models (Table 3.1). ESS Guidelines on Seasonal Adjustment recommends visual inspection of the time series and manual choice on problematic cases. The original time series is presented on Figure 3.4. The time series is highly seasonal, although the relationship between trend-cycle and seasonal fluctuations is difficult to assess.

As a result, the chart does not provide a clear answer concerning the most appropriate decomposition scheme. However, one can notice the impact of the decomposition type to the quality of RegARIMA residuals presented in the Table 3.2, which includes a set of standard diagnostics[37]. The values that are lower than 0.05 indicate the statistics for which the relevant test failed, which imply that the residuals do not manifest the expected feature. The comparison of the respective values reveals that the quality is higher in the case of results from X-12-ARIMA than TRAMO/SEATS. In particular, the residuals from this model can be considered normal, which is not the case of residuals from the model estimated using the X-12-ARIMA method. Therefore, it seems that the additive decomposition is more accurate for the analysed time series.

Figure 3.4. The comparison of original live sheep price series and series corrected by RegARIMA model – the results from TRAMO/SEATS method (PLN/kg)



Source: author's own compilation based on data of the Central Statistical Office, prepared with JDemetra+.

The level shift in 1998 is detected when the multiplicative decomposition using the TRAMO/SEATS method is applied. Although there is no clear interpretation of this event, this outlier enhances the quality of the RegARIMA model chosen by TRAMO/SEATS. Extraction of this effect from the original time

[37] The description of the tests is available in [Grudkowska 2013].

series results in decreasing the level of time series before February 1998, while the correction for the Easter effect is visible as small changes in values each March and April (Figure 3.4).

In comparison with the SARIMA model, the RegARIMA model enables to take into account the calendar effects and the outliers. These regression effects are often present in time series so RegARIMA model seems to be more useful to reflect the behaviour of lamb price series as well as other economic time series.

Table 3.2. Quality of the residuals from RegARIMA model within TRAMO/SEATS and X-12-ARIMA procedures

Test	TRAMO/SEATS	X-12-ARIMA
Normality of the residuals		
Mean	0.1773	0.4895
Skewness	0.0115	0.0120
Kurtosis	0.5586	0.1792
Normality	0.0355	0.0430
Independence of the residuals		
Ljung-Box (24)	0.8494	0.9002
Box-Pierce (24)	0.8969	0.9418
Ljung-Box on seasonality (2)	0.3039	0.2495
Box-Pierce on seasonality (3)	0.3217	0.1611
Randomness of the residuals		
Runs around the mean: number	0.6678	0.6045
Runs around the mean: length	0.0271	1.0000
Up and Down runs: number	0.0271	1.0000
Up and Down runs: length	0.0271	1.0000
Linearity of the residuals		
Ljung-Box on squared residuals (24)	0.2911	0.441
Box-Pierce on squared residuals (24)	0.3832	0.5458

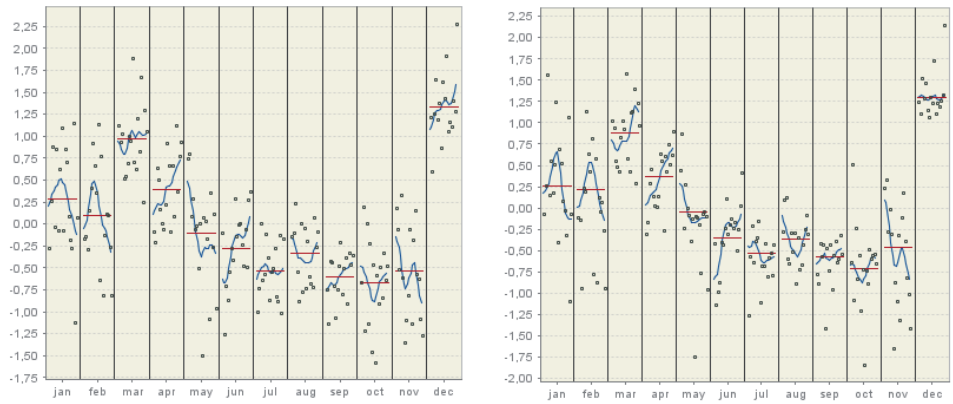
Source: author's own compilation, prepared with JDemetra+.

Additive decomposition, seasonal component

Once the decomposition has been changed to the additive one, the outlier has no effect on the residuals and can be removed from RegARIMA model. The final model chosen for both methods is the RegARIMA model (0,1,1)(1,1,1) with the Easter effect. Such model is a recommended forecasting tool for live sheep prices series in Poland.

The yearly average of seasonal factors is equal to 0 as the seasonal effect within a calendar year cancels out in the case of an additive model. One can assess the impact of the seasonal fluctuations on the time series within a time series span. It is expected that for a given calendar period the influence of the seasonal-related factors is similar from year to year, although it can gradually change in time. The empirical results reveals that both methods produce similar seasonal factors (Figure 3.5). The figure shows that for both models the prices are higher than yearly average in the first part of the year, especially in March and April. Then they tend to decline, except from December, when then are higher than average by around 1.25 PLN.

Figure 3.5. The comparison of SI factors from the RegARIMA model (left column - TRAMO/SEATS, right column - X-12-ARIMA)

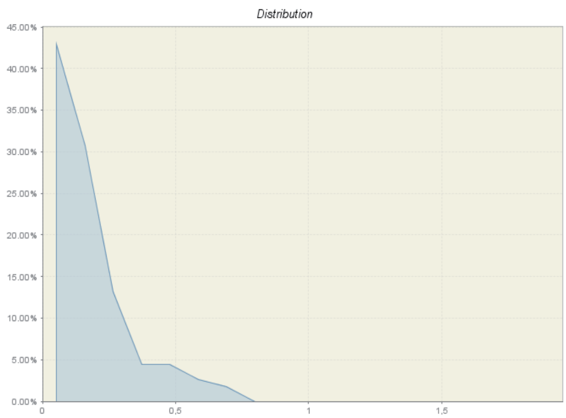


Source: author's own compilation, prepared with JDemetra+.

As mutton is not a popular meat in Poland, this yearly fluctuations mostly result from timing of the most important holidays (Easter, Christmas), as in these times the consumption of mutton increases. The seasonal pattern related to the production cycle over year is also evident. The gradual changes in SI ratios are visible for both models. In some cases, e.g. May, the change from below to above the overall mean can be observed, which indicates that over the time series span the impact of seasonality on the time series changes from positive to negative. In general, such movements are not welcomed and alternative specifications that produce more stable SI ratios should be tested. However, in this case, other models do not provide better results. It imply that the changing impact of seasons on mutton price is a characteristic feature of this time series and therefore it is visible in the respective figure.

Despite visually observed movements in SI ratios the seasonal factors seem to be stable enough as sliding spans statistics show that only 1.8% of observations have been marked as unstable (Figure 3.6). Therefore, the seasonal component can be regarded as stable because the series with stable seasonal adjustment are defined as those with less than 15% of unstable observations. The chart on the left presents the cumulative frequency distribution of the sliding spans statistics using so called frequency polygon. The sliding spans statistics are shown on the horizontal axis, while vertical axis presents the frequency (in percentages) of each class interval. The analysis of chart points out that the majority of sliding spans statistics are below 0.05. None of these statistics is higher than 0.08.

Figure 3.6. The results of sliding spans diagnostic



Abnormal values : 1,8%

Breakdowns of unstable factors and Average Maximum Differences across spans

Period	*Breakdowns*	*Average*
January	0	0,2
February	**2**	**0,4**
March	0	0,2
April	0	0,1
May	0	0,1
June	0	0,1
July	0	0,1
August	0	0,1
September	0	0,1
October	0	0,2
November	**0**	**0,4**
December	0	0,1

Year	*Breakdowns*	*Average*
2003	0	0,2
2004	**1**	**0,2**
2005	**1**	**0,2**
2006	0	0,2
2007	0	0,1
2008	0	0,1
2009	0	0,2
2010	0	0,2
2011	0	0,2
2012	0	0,1

Source: author's own compilation, prepared with JDemetra+.

The table on the right present the breakdowns of unstable periods. It intends to highlight months and years that are particularly unstable. In the analysed case the number of breakdowns is low for the all months. Therefore, seasonal breaks, fast moving seasonality or undetected outliers are not suspected. Also the breakdowns do not cluster in any particular year. It means that it can be thought that the seasonal component is stable.

Model validation

The quality of the results from the X-12-ARIMA method have been checked using M and Q statistics (Table 3.3). The M statistics reveal that the relation of the trend-cycle component to the irregular fluctuations deviates from the one which is observed for the typical time series (M-1, M-3 and M-5 measures). This observation however does not determine that the seasonal adjustment results are poor. On the contrary, the overall assessment is good (Q below 1). This result proves that the X-12-ARIMA procedure may generate the acceptable results even if the series is not totally typical one.

In the case of the TRAMO/SEATS method the primary aim of the quality diagnostic is to check the initial assumption concerning the components and their estimates. The correlation test checks if the components' estimators are not correlated with each other (Table 3.4). The same test is performed for the components' estimates. The table containing these correlations is presented below. The last column (P-Value) in the table contains the results of the test for no correlations between the components. The P-Value higher than 0.05 means that that correlations are negligible. Therefore, for analysed time series, for each component, there is no sign of correlation between the components.

Table 3.3. Quality measures for X-12-ARIMA method

Quality measure	Value	Quality measure	Value
M1	1.194	M7	0.408
M2	0.566	M8	0.773
M3	1.155	M9	0.349
M4	0.088	M10	0.837
M5	1.383	M11	0.834
M6	0.318	Q	0.707

Source: author's own compilation, prepared with JDemetra+.

Table 3.4. Result of the correlation test for live sheep price series

Cross-correlation	Estimator	Estimate	P-Value
Trend/Seasonal	-0.1233	-0.1183	0.9672
Trend/Irregular	-0.3234	-0.3085	0.8638
Seasonal/Irregular	0.0635	0.0757	0.8183

Source: author's own compilation, calculated with JDemetra+.

The variance test compares the variance of the stationary transformation of the components innovation (second column of Table 3.5) with the variance of their theoretical estimators and the variance of their empirical (actually obtained estimate) [Grudkowska 2013]. For all the components estimator's variance is higher than estimate's variance but the discrepancies are small enough (P-Value higher than 0.05) to assume that there is no underestimation of the component variance.

Table 3.5. Analysis of variance of components and their estimators for live sheep prices

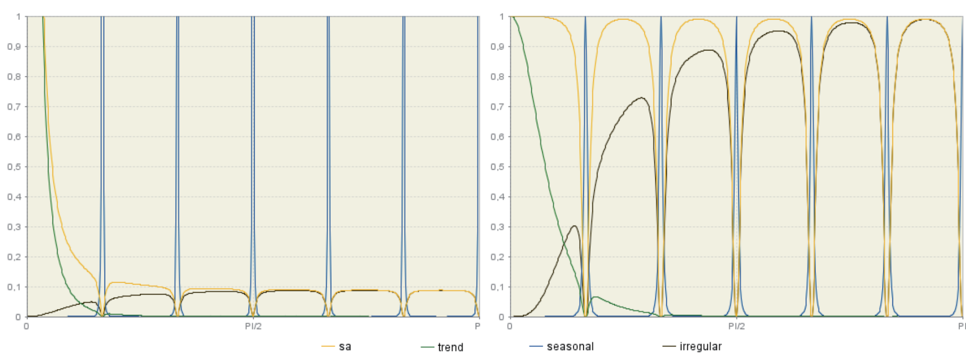
Cross-correlation	Component	Estimator	Estimate	P-Value
Trend	0.0223	0.0008	0.0007	0.5843
Seasonally adjusted	2.8671	2.4837	2.3484	0.7348
Seasonal	0.1018	0.0199	0.0140	0.3692
Irregular	0.6037	0.4346	0.4064	0.5967

Source: author's own compilation, calculated with JDemetra+.

The shape of the spectrum of the final estimators is shown in the first graph (Figure 3.7) on the left. It is clear that the seasonal component estimator includes only frequencies that are very close to the seasonal one (the spectral holes in seasonally adjusted series are very narrow). The narrow seasonal bands imply that the seasonality in the analysed time series is deterministic rather than stochastic.

The graph on the right (Figure 3.7) presents the squared gain functions. It points out that seasonal frequencies are assigned to the seasonal component while the seasonally adjusted series captures all the variance that result from the non-seasonal part of time series. The squared gain of seasonally adjusted data is nearly zero for seasonal frequencies, which means that the estimator of the seasonally adjusted series does not contain the seasonal frequencies. For other frequencies the squared gain of seasonally adjusted data is close to one, so the estimator do not disturb the non-seasonal frequencies.

Figure 3.7. Spectrum of estimators and squared gain of components filter



Source: author's own compilation, prepared with JDemetra+ .

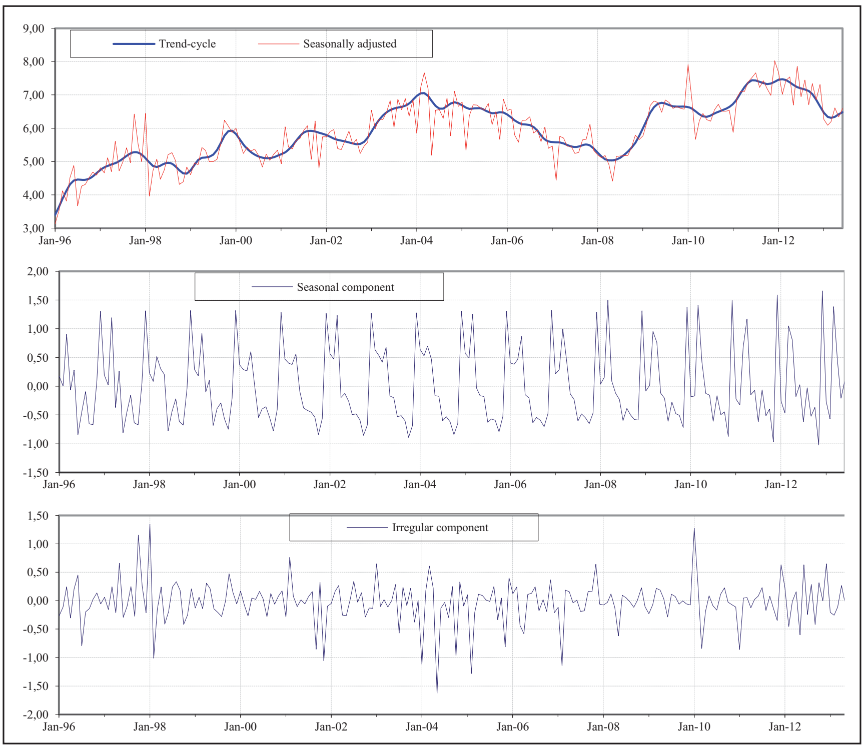
To sum up, the satisfactory results were obtained for both methods. Both methods indicate that Easter has a positive impact on mutton prices. No unusual events that disturb the evolution of a time series have been detected. The impact of the seasons on the mutton price seems to evolve in time. Nevertheless, the quality diagnostic shows that the estimates are relatively stable (the X-12-

-ARIMA method) and all model assumptions have been fulfilled (the TRAMO/SEATS method).

Decomposition and forecasting

Decomposition of the original time series (Fig. 3.4) into different components is presented in Figure 3.8. The rising trend is observed in analysed period with time varying cyclical pattern around it. The graph presents also time varying seasonality with amplitudes around 2-2.5 PLN. This variability is comparable with variability due to cycle. The highest prices of live sheep are in December and March what is connected with demand conditions (Christmas, Easter). Information about patterns in the sheep prices may be used by analysts who prepare short-term forecasts of price development.

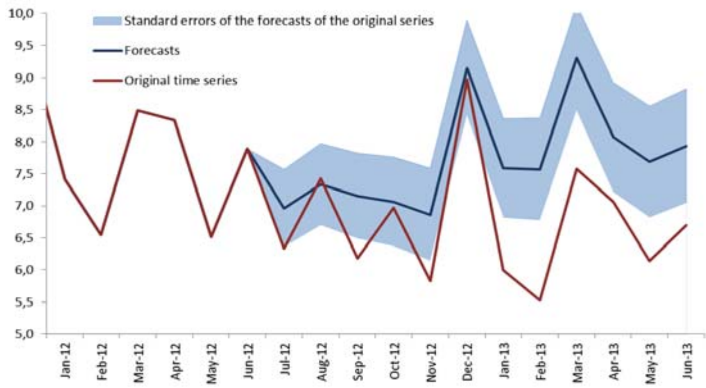
Figure 3.8. Decomposition of time series for mutton price series (PLN/kg) according to X-12-ARIMA additive procedure



Source: authors' own compilation based on the data of the Central Statistical Office.

The applied procedures are quite flexible in capturing regularities that occur in the time series. But this does not definitively decide on their forecasting capabilities. We calculated ex-post forecasts for a period July 2012–June 2013 in our example. The forecasts for the last 12 months obtained using both methods are very similar to each other, so the Figure 3.9 displays only one forecast with its upper and lower 95% confidence levels. The calculated forecasts generally tend to underestimate the actual values (Figure 3.8). Most of forecasts, especially in last six month, are out of confidence range. Errors for ex-post forecast (average for 12 months is 18.0%) were higher than errors for ex-ante forecasts (11.5%).

Figure 3.9. Ex-post forecasts and actual values (PLN/kg) for sheep price in Poland



Source: author's own compilation, prepared with JDemetra+.

4. Causal forecasting methods

Prices of agricultural products show significant variability. This results in a high level of risk of agricultural production. One of the basic tasks of agricultural economics is to explain the sources for this variability to reduce that risk. The causal analysis is one of the approaches to this problem. The causal analysis is based on the assumption that price variability is a result of some exogenous factors variability that influence the price level. The price forecast is constructed on the assumptions that the future values of exogenous factors are known or can be forecasted.

This chapter describes details of the methodology outlined above. In the following sections dynamic one- and multi-equation econometric models are explained. The methods shown are particularly useful in the situation where data are gathered on the weekly, monthly or quarterly basis. The problems of analysing the properties of stochastic processes, specifying causal models (congruent among them) and applying them for analysis and forecasting agricultural process are raised as well.

4.1. Introduction to causal dynamic models

The econometric model is a basic tool for exploration of the interdependence of economic variables. It is a formal description of the stochastic interactions of the phenomenon (phenomena) or the economic process (processes) on the factors that influence them, presented as a single equation or a system of equations. The structure of each equation of the econometric model is defined by endogenous variable, exogenous variables, structural parameters of the model, residuals and the functional relationship between the endogenous variable, exogenous variables and residuals.

Two main types of data are used in econometrics (not necessarily mutually exclusive): a cross-sectional data and a time series data. All these different data types require specific econometric and statistical techniques and models for data analysis. Cross-sectional data are collected by observing many entities (such as farms, companies, countries, individuals etc.) at the same point of time, or without considering the differences in time. This type of data is used in case of construction of the static econometric models to compare the differences among the entities. Time series data is a sequence of data points, measured typically at successive times spaced at uniform time intervals, i.e. annual, quarterly, monthly, daily and so on. This type of data has then a natural temporal ordering and a frequency and is used to specify dynamic econometric models. The dynamic models allows to study causality between two or more variables, not just correlation as in the case of the static models. For this reason the dynamic models are used for the agricultural price forecasting.

4.1.1. Model types

Natural conception of forecasting of economic phenomena is to assume that the variable (or variables) forecasted is a function of other variables. If the change in explanatory variable in one point in time has an impact on one or more explained variables such a model is called a causal model. If there is no cause-effect relationship between variables and the model is based only on correlation between variables then we have descriptive (symptomatic) models. Both types of relationship can be used for forecasting agricultural prices. Taxonomies of econometric models based on causal relationships the most frequently take into account the cognitive properties of the model, the number of explanatory variables, the number of equations, the functional form of relationships between variables, the existence of lagged variables and deterministic components. Some of these issues will be discussed below.

The number of equations

The analysis of the interdependence of economic variables very often uses a single-equation linear multiple regression econometric model which is given by the formula [see for example Goldberger 1998, Chap. 19.1]:

$$Y_i = f(X_{1i}, \ldots, X_{ki}, \varepsilon_i). \tag{4.1}$$

Assuming linear analytical form of equation (4.1) it can be rewritten in notation with sigma operator:

$$Y_i = \sum_{j=0}^{k} \beta_j X_{ji} + \varepsilon_i \tag{4.2}$$

or in the matrix notation:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{4.3}$$

where:

Y_i – i-th value of endogenous variable (i = 1, 2, ..., n),

X_{1i}, ... X_{ki} – i-th value of j-th exogenous variable (j = 0, 1, ..., k), values of x_{0i} (called the constant regressor) are equal to 1,

ε_i – i-th error term (residual),

β_j – j-th structural parameter of the model (j = 0, 1, ..., k),

Y – n dimension vector of values of endogenous variable,

$\mathbf{X}$ – $n \times k$ matrix of values of exogenous variables,

$\boldsymbol{\beta}$ – k dimension vector of structural parameters of the model,

$\boldsymbol{\varepsilon}$ – n dimension vector of residuals of the model.

A multi-equation econometric model consists of many single-equation econometric models and can be written as follows:

$$\begin{aligned} Y_{1i} &= f(X_{1i}, \ldots, X_{ki}, Y_{2i}, Y_{3i}, \ldots, Y_{mi}, \varepsilon_{1i}) \\ Y_{2i} &= f(X_{1i}, \ldots, X_{ki}, Y_{1i}, Y_{3i}, \ldots, Y_{mi}, \varepsilon_{2i}) \\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ Y_{mi} &= f(X_{1i}, \ldots, X_{ki}, Y_{1i}, Y_{2i}, \ldots, Y_{m-1i}, \varepsilon_{mi}) \end{aligned} \tag{4.4}$$

where both endogenous variables, exogenous variables (regressors), as well as an error term (residuals) can be connected to various types of relations.

Lagged variables

The distinction between models based on instantaneous (immediate) relationships and models with time-lagged variables is a very important one. The model without lagged variables sometimes is referred to as a "static" regression model based on time series. One-equation model without lagged variables can be written as follows:

$$Y_t = f(X_{1t}, \ldots, X_{kt}, \varepsilon_t), \tag{4.5}$$

where:

Y_t – endogenous variable,

$X_{1t}, \ldots X_{kt}$ – exogenous variables in time t (regressors),

ε_t – error term in time t (residual).

The application of models in the form of 4.5 for time series is rather difficult in economic reality. Usually the change in one variable is transmitted to other variable through many moments of time. The fundamentals of such relationships in agricultural economics have been shown in Chapter 1.

Dynamic models can contain lagged forecasted variable, lagged explanatory variables or both of them. It means that the commodity price is a function of the current and lagged values of exogenous variables and the lagged values of endogenous variables. The model with one explanatory variable X and lags in endogenous and exogenous variables can be written as follows:

$$Y_t = f(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}, X_t, X_{t-1}, X_{t-2}, \ldots, X_{t-q}, \varepsilon_t), \tag{4.6}$$

where: Y_{t-p} – lagged endogenous variable in time $t\text{-}p$, i=1,2,..,p, X_{t-q} – lagged exogenous variables in time $t\text{-}q$, i=1,2,..,q.

Models with lags are preferable in practical applications of commodity price forecasting. The chance for specification a good model without any lags might be only in the case of yearly data. This is caused by propagation effect through time and persistency of economic phenomena. The best situation might be if we find such an explanatory variable which would serve as a leading variable for a forecasted one.

Transformations of variables

Models can be constructed on the basis of original data or transformed data. The logarithmic transformation is the most commonly used data transformation. Another useful transformation of variables changing their properties is the Box-Cook transformation. This type of transformation alters the properties of time series from additive to multiplicative and helps to get good properties of the regression model. This is an effective way to stabilize the variance across time. It is possible to transform either the independent (explanatory) or dependent variables or both types of variables. The logarithmic transformation changes also the model analytical form.

Another transformation may be the removal of deterministic components from the set of variables. For example, before the model 4.4 parameters are estimated, the trend can be removed (eq. 2.1-2.10, Chap. 2), as well as seasonal fluctuations (eq. 2.12, Chap. 2; X-12-ARIMA or TRAMO/SEATS, Chap. 3).

The differencing is used to transform data in the case of non-stationarity of variables caused by non-constant variance. The first difference of a time series is the series of changes from one period to the next (Chap. 2.4.2). Most commonly both the explained and explanatory variables are differenced. However, it is possible to difference only the selected variables after a preliminary analysis of their properties. There are also models (e.g. error correction models) that utilize the levels of variables and first differences as well.

Deterministic component

Dynamic econometric models can also be expanded by deterministic variables to include information about internal structure of a modeled time series. A common practice is an inclusion of time variables to the model in order to obtain the parameters consistent with the economic theory. By including time variable we can eliminate linear trends from all the variables in a model which means that this model is estimated as in the case of de-trended data. Thus, there is a greater chance to eliminate spurious relationships.

In case of estimation of models based on quarterly or monthly data (as well as weekly data) a set of explanatory variables can be extended for artificial dummy variables for seasonal variation. Such variables are taken into account when at least one of the variables (explained or explanatory) contains seasonality. Time series models with trend and seasonal dummies are demonstrated in Chapter 2.3.1.

One of the most important assumptions about models is stability of their parameters over time. This assumption may be violated if there are structural changes in the analysed economic phenomena. It can be assumed that a structural break would occur if at least one of models' parameters changed at the break date within the sample period. The issues of outliers and structural changes in a time series are discussed in Chapter 3.2.1.

Agricultural prices that are affected by numerous factors, such as weather conditions, animal diseases or changes in agricultural policies, are regarded as those in which the structural changes may occur frequently [Wang and Tomek 2007]. To eliminate negative consequences of structural breaks some additional dummy variables can be added to models. Structural breaks may be limited to the level shift (LS), trend change (TC) or regime change (RC) when there is allowance for change of structural parameters in different regimes [Perron 2005].

4.1.2. Model specification and forecasting

Forecasting steps

Specifying and the use of the econometric model for forecasting requires solving a number of problems. These problems concern the following issues:

- the selection of the model's analytical form (linear or non-linear),
- the selection of the best estimation method (the most common is OLS),
- selecting the best set of explanatory variables (substantive criteria and formal statistics' criteria are used to identify variables that improve the explanation of the predicted variable),
- the validation of the model (logical verification, the significance of variables, the model fit to the data, residual properties, etc.),
- the determination of the explanatory variable values for the period for which the forecast is constructed (this is a crucial step),
- the choice of principle, according to which the forecast is constructed (the most common is a rule of unbiased prediction).

Explanatory variables

Key steps are selecting the best set of explanatory variables and the determination of their values in the forecasting period. The set of potential independent variables is defined through a subject-matter analysis of the phenomenon, where the chosen variables explain changes of the forecasted variable. The basis for that is the economic theory and the knowledge of the analysed phenomena. At first we need to take into account the variables that have a causal relationship[38] with the variable forecasted. The list of potential explanatory variables may be extended by the symptomatic variables in case of the absence of a sufficient number of causal explanatory variables. The independent variables should be good representatives of various aspects of the examined fragment of the economic reality. The existence of many potential variables causes colinearity and so-called spurious regression problems[39] which have to be solved.

If insignificant variables are in a set of exogenous variables of econometric model, then the way they are removed from the model is called into question. Leaving insignificant variables does not cause negative impact on the quality of the estimated parameters. It can, however, lead to the false conclusion that the insignificant variables affect the endogenous variable, and also insignificant exogenous variables unnecessarily complicate the model. Therefore, there is a need to establish special procedures of the variables selection for the model. Comprehensive description of these procedures can be found in Draper and Smith [1973]. These procedures can be broadly divided into two groups: the so-called "all regressions"[40] and the sequential procedures.

The sequential procedures rely on sequential introduction of exogenous variables into the model, as is the case in the selection procedure *a priori*, or sequential removal of variables from a set of exogenous variables, as in the case of the *a posteriori* elimination procedure. Among the sequential procedures, the so-called stepwise regression method can be distinguished. One can begin with no variables in the model and proceed forward (adding one variable at a time), or start with all potential variables in the model and proceed backward (removing one variable at a time).

---

[38] The causality issue will be discussed in section 4.2.2.

[39] The issue of spurious regression will be discussed in detail in section 2.2.3.

[40] Procedures of "all regressions" involve creating the set of "candidates for explanatory variables" all possible subsets and evaluation of regression models built on the basis of these subsets. The differences between the procedures depend on the criteria used for the model assessment.

Determination of explanatory variables for the future

The use of casual models in the forecasting seems very interesting. However, that idea is used less frequently than the forecasting based on time series models. One of the main reasons, along with complications with the selection of explanatory variables, is a need to know the values of explanatory variables in the period of the forecast. The incorrect values of explanatory variables for the forecasting period result in the unacceptable forecast although the econometric model is correct. The best explanatory variables of causal models are lagged variables, so-called leading variables. In that case the maximum horizon of the forecast without construction of additional forecasts of the dependent variable equals to time-lag between the forecasted and the explanatory variable. This reduces an important part of the forecast error that results from wrong predictions of the values of explanatory variables.

The values of explanatory variable for the forecast period can be established according to one of the following ways:

- on the planned level (only microeconomic variables),
- using the available forecasts made by various institutions,
- on the basis of the opinion of experts,
- using forecast obtained from time series models,
- using several methods at the same time, and by creating different variants (combination) of explanatory variables.

4.1.3. Model validation

Fitting statistics

To obtain a good model is not a simple task. It requires many steps and a lot of testing (verification of econometric model). The first step of the verification procedure is to check how well the regression model describes the data. The determination coefficient (R^2) may be used for this purpose.

Decomposition of variability of the endogenous variable for part explained by the model and the random variation allows us to construct a measure of model fit to empirical data. It indicates how much of the total variability of the endogenous variable is explained by the model. Coefficient R^2 takes values from the interval [0,1]. The closer the R^2 value is to 1, the better the model fits the empirical data and the better it describes the reality. However, R^2 is a tricky measure. It is relatively easy to obtain high values of it for non-stationary series, whereas R^2 is usually quite low in the case of stationary series. So it should be complemented with other statistics, standard errors among them. The standard error is also a good measure for comparing competitive models. When judge different models also information criteria measures are applicable.

Parameters significance

When estimating structural parameter of the model (β_j) one can test whether exogenous variables have significant influence on the endogenous variables. This test is based on performing *k* statistic tests [see *e.g.* Kennedy 2003, Chap. 4.2; Goldberger 1998, Chap. 20.2]. The null hypothesis of this tests assumes that $\beta_j = 0$, tested against the alternative hypothesis that the *j*-th structural parameter is different from zero. The test statistic is:

$$t_j = \frac{\left|\beta_j - b_j\right|}{S(b_j)}, \tag{4.7}$$

where:

b_j – the estimated value of the *j*-th structural parameter,

$S(b_j) = \sqrt{d_{jj}s^2}$ – the standard deviation of the estimated value of the *j*-th structural parameter, where s^2 is the variance of the residuals, d_{jj} is a diagonal element of matrix $(\mathbf{X}^T\mathbf{X})^{-1}$.

The test statistic given by the formula (4.7) has a *t*-Student distribution with *n*-*k*-1 degrees of freedom, assuming that the residuals have a normal distribution. The null hypothesis is rejected if the test statistic value is higher than the critical value for a given degree of freedom, and assumed significance level (usually $p = 0{,}05$). If the null hypothesis for the structural parameter cannot be rejected, it means that the exogenous variable does not significantly influence the endogenous variable.

Residual testing

Estimates of the structural parameters of the model (4.1-4.3) obtained with ordinary least squares (OLS) method have the desired properties (unbiasedness, consistency and efficiency), if some strong assumptions are met [see for example Kennedy 2003, Chap. 4.3]. These assumptions primarily concern the residuals of the model:

- Expected value of residuals is zero: $E(\varepsilon_i) = 0$ for each *I*,
- The residuals are spherical: $D^2(\varepsilon_i) = \sigma^2$ for every *i* (the variance of the random parameter is constant) and $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$ (residuals of the model are not correlated),
- Residuals of the model have a multi-dimensional normal distribution.

Furthermore, it is assumed that the model is invariant due to the observations (the relationship between the phenomena is stable in time), and the exogenous variables are non-random. In order to obtain estimates of structural parameters of the model, it is also necessary to include two formal assumptions. The first and rather obvious says that the model is linear or reducible to linear ($\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$)[41]. The second assumption is that the matrix **X** – observations on the exogenous variables – is a full column rank[42]. Assumptions, known as the assumptions of Gauss-Markov[43], form the so-called classical (standard) linear regression model. Classical linear regression model, apart from its simplicity, both in terms of its parameters estimation and interpretation of results, has numerous problems. They arise mainly from the fact that the assumptions of the model – that make it so attractive – are very rarely met. Failure of the model's formal assumptions causes that the approximation to the structural parameters of the model is not possible. The others cause failure of assumptions describing the properties of model residuals. This leads to a situation where the estimates obtained by the classical least squares model lose their desirable properties.

Estimators of the model structural parameters are biased, if the expected value of residuals is not equal to zero, when there are such *i*, that $E(\varepsilon_i) \neq 0$. This happens when at least one exogenous variable significantly correlated with the endogenous variable is omitted, or if the nonlinear function is estimated with a linear function. Estimates of the model structural parameters are then useless.

The estimators of parameters lose their effectiveness when the model residuals are not spherical. The standard errors of estimators are larger. This causes problems with testing the significance of the model structural parameters. Statistics of these tests have low values (see formula 4.7), which in turn may result in removal from the model variables significantly influencing the exogenous variable, which makes the evaluation of the remaining structural parameters worthless.

The failure to meet the assumption of sphericity of distribution consisting in a correlation between the residues of the model, called autocorrelation, results

---

41 In literature on econometrics there are many examples of non-linear functions, which by simple conversion can be reduced to the linear function [see for example Stock and Watson 2007, Chap. 8.1].

42 This assumption means that none of the matrix **X** columns is a linear combination of other columns, and the number of observations is larger than the number of estimated parameters. Failure of this assumption means that there is no inverse $\mathbf{X}^T\mathbf{X}$ matrix and therefore the $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ cannot be estimated.

43 Gauss-Markov theorem states that estimates obtained with the OLS method have desired properties (unbiasedness, consistency and efficiency) [see for example Stock and Watson 2007, Chap. 5.5].

in augmentation of the determination coefficient and therefore too affirmative assessment of the model quality. The failure to meet the assumption of normal distribution of the random parameter also causes problems with testing the structural parameters of significance, as the test statistic used for this purpose (see formula 4.7) has a t-Student distribution only if the residuals have normal distribution.

The remedy for the failure to meet assumptions of the random parameter sphericity is the use of estimation methods other than OLS. Very frequently generalized least squares (GLS) method are applicable in case of autocorrelation, non-normal distribution or heteroscedasticity in residuals.[44]

The rejection of the assumption of stability over time causes that it is necessary to use more complex models to obtain valuable estimates of structural parameters of the model (4.2). It may be achieved also by including additional parameters for structural changes or by the introduction of the so-called switching regime models.[45]

4.2. Stochastic processes and their properties

The starting point for determining the time series is the definition of a stochastic process $\{X_t\}$, which is defined as a sequence of random variables ordered by time index – t time series. The time series, in turn, is understood as the realisation of a stochastic process in the sample, as a single implementation of this random process. In other words, stochastic process is a description of the statistical variation of a phenomenon over time. The time series, however, which is the realisation of the stochastic process, is a time-ordered sequence of observed variable values [Lütkepohl 2007, pp. 3-4].

4.2.1. Stationarity

Stochastic processes

A key feature of both the stochastic process and a time series is its stationarity. A stochastic process is strictly stationary (in a broader sense) if for each subset of time indices (r, s, …, w) and for each integral number k overall distribution of random variables $\{x_r, x_s, \ldots, x_w\}$ is the same as the overall distribution of variables $\{x_{r+k}, x_{s+k}, \ldots, x_{w+k}\}$. The stochastic process is stationary in

[44] Details of generalized least squares method can be found in the literature [see for example Goldberger, Chap. 27].
[45] More on switching models see [Kennedy 2003, Chap. 14].

a broader sense if its multi-dimensional distribution of probability does not change over time. For practical reasons (the possibility of testing stationarity) it is convenient to use the notion of a weak stationarity of stochastic process. The process is weakly stationary (stationary in the narrower sense) if it satisfies three conditions [Box and Jenkins 1970; Charemza and Deadman 1997]:

- finite and constant in time, the expected value – $E(X_t) = const$,
- finite and constant over time variance – $V(X_t) = const$,
- the covariance between the observations from two periods depends only on the distance (gap) between them – $Cov(X_t, X_{t-p}) = const$, for all t.

The stochastic process that does not meet these conditions is a non-stationary process. A stationary stochastic processes generate a stationary time series and a non-stationary stochastic processes generate a non-stationary time series. Stationarity of the time series is defined analogously to stationarity of stochastic processes. In economics a weaker form of stationary is commonly employed. The time series is weakly stationarity when its expected value, variance and covariance do not change over time.

If the non-stationarity is only an effect of a non-constant average (expected value) then such a series is called a trend stationary. The way of transforming such series to stationary one is to remove the trend by fitting a trend line. If a series is non-stationary in variance a proper procedure is d-times data differencing $\Delta X_t = X_t - X_{t-1}$. Economic series can be also non-stationary in the mean and variance at the same time. Non-stationarity can be resulted also from deterministic and stochastic seasonality (Chap. 2.4).

Augmented Dickey-Fuller test

The time series stationarity might be tested by numerous statistical tests. The most widespread stationarity tests and implemented in most statistical packages are: Augmented Dickey-Fuller test (ADF) and the KPSS test (test of Kwiatkowski, Phillips, Schmidt and Shin).

In ADF test, the null hypothesis says that the time series is non-stationary, and alternative hypothesis is a contradiction of null hypothesis. The starting point in the construction of the statistic of ADF test is the regression equation of one of the following forms [Enders 2010]:

$$\Delta Y_t = \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \ldots + \gamma_k \Delta Y_{t-k} + \varepsilon_t \tag{4.8}$$

$$\Delta Y_t = \alpha + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \ldots + \gamma_k \Delta Y_{t-k} + \varepsilon_t \tag{4.9}$$

$$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \ldots + \gamma_k \Delta Y_{t-k} + \varepsilon_t , \tag{4.10}$$

where: Y_t – time series data, $\Delta Y_t = Y_t - Y_{t-1}$, α, β, δ, γ – model parameters, t – time variable.

The three subsequent cases of the ADF test are models: without constant, with constant and with constant and linear trend. Model 4.10 can be extended by a second order of the trend polynomial. In the case of deterministic seasonality all models can be extended with seasonal dummies variables. The structural parameters of these models are estimated using the OLS method. The purpose of the lagged components is to remove the autocorrelation of the random parameter. The number of lags is called augmentation. The statistic of this test is given by the formula 4.11:

$$\tau = \frac{\delta}{S(\delta)}, \tag{4.11}$$

where $S(\delta)$ is a standard deviation of the structural parameter δ.

The statistic τ less than the critical value results in rejection of the null hypothesis. The distribution of statistic τ is different from typical t distribution, despite the similarity of the statistics used in testing the significance of the structural parameters of the classical linear regression model (see equation 4.7). It follows the need to use specially prepared tables of critical values.[46]

KPSS test

The KPSS test null hypothesis and alternative hypotheses are reversed as compared to the Dickey-Fuller test. The null hypothesis states that the time series is stationary, while an alternative that it is non-stationary. Statistics of KPSS test have a complex structure and a very complicated probability distribution. Calculation of the value of KPSS test statistics can be summarized as follows [Maddala 2001, p. 552]:

- the *OLS* is used to estimate structural parameters and the residuals of the equation:

$$y_t = \delta t + \xi_t + \varepsilon_t, \tag{4.12}$$

  where ε_t is a stationary process and ξ_t is a random walk given by $\xi_t = \xi_{t-1} + u_t$ (u_t is a Gaussian white noise with zero average),
- the value of the test statistic (Lagrangian multiplier) is determined by the formula:

$$LM = \frac{\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_e^2}, \tag{4.13}$$

[46] More on ADF test see for example [Maddala 2001, p. 548].

where e_t are the residuals of a regression of y_t on a constant and a time trend, $\hat{\sigma}_e^2$ is the residual variance for this regression and S_t is the partial sum of e_t defined by equation:

$$S_t = \sum_{i=t}^{T} e_i , \quad t = 1, 2, .., T. \tag{4.14}$$

The null hypothesis is rejected if the value of the test statistic is larger than the critical value.

ADF and KPSS tests, with the null hypothesis formulated on the contrary, are used in a so-called confirmatory analysis [Maddala 2001, pp. 553-554]. According to this analysis, the rejection of the null hypothesis in the ADF test and finding no reason to reject the null hypothesis in the KPSS test "strongly suggests" the stationarity of tested time series, and the reverse situation "strongly suggests" the non-stationarity.

4.2.2. Causality

Causality is defined as the relation between one event (the cause) and the other event (the effect), where the second event is understood as a consequence of the first. The idea of causality is widespread in statistics and economics where the key question that can be addressed is how useful are some variables for explaining and forecasting others. One of the statistical concept of causality that is based on prediction is the Granger-causality definition. A time series X is said to Granger-cause Y if the current values of Y are better explained with lagged and current values of X than without it. The reason for proposing such a definition was that if an event X is a cause for another event Y, then the event X should precede the event Y [Hamilton 1994].

Causal relationships between exogenous variables and endogenous variable are tested in the *ADL* model by Granger causality test based on the results of the estimation of the model (4.18) and the model (4.19). The idea of this test is to verify whether the introduction of additional variable into the model with all the lags significantly reduce the variance of residuals[47].

The most common variant of the Granger test is the Wald variant [Lütkepohl 2007, p. 102]. The Wald variant of the F-test can also be used in the causality analysis[48]. This test is used originally to find whether the inclusion of a variable or a set of variables in the model significantly reduce the model vari-

[47] More on Granger causality test see: [Charemza, Deadman 1997; Lütkepohl 2007].
[48] The details of the F test applied in the regression analysis can be found in the handbooks by Kennedy [Kennedy 2003, Chap. 4.3] and Goldberger [Goldberger 1998, Chap. 20.3].

ance of residual, which in fact answers the same question as the classic Granger causality test. The use of the F-test is much simpler. Most of statistical packages supporting the regression analysis (e.g. the GRETL package) provide routinely the F-test statistics values and the type I error probability.

The causality analysis in the sense described above can be implemented only in case of stationary time series. In case of non-stationary time series it is possible to remove the non-stationarity by introducing deterministic variables for trend or seasonality, or transforming time series by the computing its first differences or logarithms [Enders 2010]. The Toda and Yamamoto (T-Y) research [1995] indicates however, that the first differences should not be used in the analysis. Instead the number of lags should be increased in the VAR[49] model (consisting of two ADL models) by the rank of integration (m) of the time series. That should be followed by the examination of the impact of the imposed restrictions (the additional m lags are not the subject of this restrictions) using the Wald version of the F test. That procedure (T-Y) can be used both in the case of cointegrated series, as well as stationary or non-cointegrated time series.

4.2.3. Long-term relationships

Spurious regression

The problem of spurious regression was identified for the first time by Granger and Newbold [1974] [50]. They concluded that even if non-stationary time series are randomly generated and "… it will be the rule rather than the exception" [Granger, Newbold 1974, p. 117] that econometric models estimated on the basis of this time series will make the appearance of a statistically significant relationship. The reason is that the distribution of the correlation coefficient between the non-stationary random variables is not unimodal as is the case when the variables are stationary (then probability density function of the correlation of coefficient distribution is centered at zero) but bimodal with the local maximums of the density functions shifted toward -1 and 1. It results in a greater probability of nonzero values of the correlation coefficient and related measurements, such as the determination coefficient. It also means that the distributions of the significance test statistics are different than assumed. It gives false results of these tests.

[49] The VAR (Vector Autoregressive Model) model will be discussed in section 4.4.
[50] The problem of the so called nonsense correlation has been present in the econometric literature at least since 1926. Granger and Newbold gave the precise description and explanation of the spurious regression problem [Phillips 1986].

The occurrence of spurious regression reveals in a quite characteristic manner [see for example Charemza and Deadman 1997, Chap. 5.2]. Estimation of regression models with a spurious regression gives good results – the values of determination coefficients are high, and the structural parameters of the model are significant, as proven by the statistics of *t*-test. It can be concluded that the model is correct, if the verification of the model is completed at that stage. The value of statistic of the Durbin-Watson test is however low in that case. The popular rule-of-thumb adopted for the identification of spurious regression says that the spurious regression occurs when the value of the Durbin-Watson statistic is less than R^2.

The main problem with the specification of econometric models for agricultural commodities prices results from the fact that the time series of that prices are generally non-stationary. Non-stationary can also be a time series of exogenous variables of these models. The values of R^2 can be high, and the values of statistics of *t* test may indicate the significance of the structural parameters of the model.

Cointegration

The methodology typical for a non-stationary time series analysis consists of two steps:

- the cointegration analysis of the levels of variables, and when time series are not cointegrated,
- the regression analysis of the first differences of variables.[51]

Co-integration analysis gives the possibility to identify the long-run equilibrium (relationships). The analysis of the correlation between the first differences is used to examine the short-run dynamics.

The detection and description of long-run relationships may be done through the so-called cointegration analysis [Enders 2010, Chap. 6; Johansen 2009; Lütkepohl 2007, Chap. 6.3]. Cointegration of the time series of two variables (X_T, Y_t) occurs when these variables are integrated of order d and their linear combination – $\beta_1 X_t + \beta_2 Y_t$, is integrated of order d - b ($d \geq b \geq 0$). Vector [β_1 β_2] is called a cointegrating vector. The cointegration vector components determines the long-run relationships between variables. The most common is the situation where the time series of observations on the variables is integrated of order one and their linear combination is stationary.

[51] A synthetic description of the methodology for examining the interaction in the case of non-stationary time series can be found in [Kennedy 2003, Chap. 19], and more detailed in works by Charemza and Deadman [1997] and Lütkepohl and Krättzig [2004], entirely dedicated to analysis of this methodology.

The Engle-Granger test [Engle and Granger 1987] investigating cointegration of two variables, requires running an *OLS* regression of the model:

$$Y_t = \beta_0 + \beta_2 X_t + \varepsilon_t, \tag{4.15}$$

where:

X_t, Y_t – variables, with which cointegration is tested,

β_0, β_2 – structural parameters,

ε_t – residuals.

The variables X_t and Y_t are cointegrated if the residuals ε_t are stationary. Introduction of the time-variable to the model (4.15) allows to identify the linear trend in the model residuals. The model can be extended also by seasonal component.

There may be up to m–1 linearly independent cointegrating vectors (where m is the number of variables). In case of many possible cointegration vectors this would be at least difficult to use the Engle-Granger test. Defects of the Engle-Granger test have not been observed in the Johansen procedure used for testing *VAR* models cointegration [Hamilton 1994, Chap. 20; Lütkepohl 2007, Chap. 8.2]. This procedure is described in detail in Section 4.4.1 where the *VECM* models are discussed.

4.3. One equation dynamic models

There is a variety of dynamic models classified according to the property of time series used for analysis, the number of equations in the model and the composition of the variable set. Models used for stationary time series are different from those used for the non-stationary data. Some models need to be supplemented with deterministic factor (as trend or seasonality for example) and some not. In some cases, there is a long-run relationship in other not. Some of the problems outlined above will be discussed in this subchapter.

In the current subchapter we assume that the following equation of causal relationship will be considered:

$$Y_t = \sum_{j=0}^{k} \beta_j X_{it} + \varepsilon_t \tag{4.16}$$

where endogenous variable Y_t is regressed on i-*th* exogenious variables X_{it}. Such a "static" specification is not suitable to fully express agricultural price dynamics. So other extensions of the model connected with properties of particular time series will be discussed.

4.3.1. Models for stationary series – the ADL model

For analysis and forecasting of a stationary time series the autoregressive distributed lag models can be applied. These models are basis for dynamic models specification. The model with distributed lags of the order q – $DL(q)$ (***Dis****tributed* ***L****ag*), where the time-distributed effect of the vector of independent variables (exogenous) on the dependent variable Y_t is modelled, is given by equation [Stock, Watson 2007, p. 180]:

$$Y_t = \beta_0 + \sum_{j=0}^{k}\sum_{i=0}^{q}\beta_{ij}X_{j\,t-i} + \omega_t \tag{4.17}$$

where:

β_0 – constant of the equation,

β_{ij} – vector of structural parameters of the model for the i-th lag of exogenous variables,

$X_{j,t-i}$ – matrix of j-th exogenious variables for i-th lag,

ω_t – t-th model residual.

A model describing the endogenous variable by its lagged values is called an autoregressive model of the order p – $AR(p)$ – where the number of lags is equal to p. The model is given by the formula [Box, Jenkins 1970]:

$$Y_t = \alpha_0 + \sum_{i=1}^{p} a_i Y_{t-i} + \varphi_t\ , \tag{4.18}$$

where:

α_0 – constant of an equation,

α_i – i-th structural parameter,

φ_t – t-th model residual.

Composition of these two models – $DL(q)$ and $AR(p)$ – provides the $ADL(p,q)$ model given by the formula:

$$Y_t = \mu + \sum_{i=1}^{p} a_i Y_{t-i} + \sum_{j=0}^{k}\sum_{i=0}^{q}\beta_{ij}X_{j\,t-i} + \varepsilon_t \tag{4.19}$$

where: μ – constant of the equation, ε_i – t-th model residual.

Quite frequently, for simplification issues, ADL models are constructed with the same order of autoregressive processes p=q. The modified ADL model can possess different order of particular endogenous and exogenous processes.

The use of ADL models in the description of economic relationships and in forecasting raises a number of problems which do not occur in case of models estimated on the basis of cross-sectional data. These problems are related to the model identification, the estimation of its parameters, and the interpretation of results and their use in forecasting.

The problems associated with the specification of dynamic models concern two issues: the existence of the spurious regression and determination of the number of lags for the exogenous variables and for the endogenous variable as well. Spurious regression occurs when an attempt is made to construct an econometric model based on non-stationary time series. Then, "… the fitted coefficient are statistically significant when there is no 'true relationship' between the dependent variables and the regressors" [Phillips 1998, p. 1300]. This problem results from the fact that dynamic models are estimated on the basis of time series data.

Another major problem associated with the ADL models specification, and dynamic models in general, raises the issue of lag rank, both exogenous variables and the endogenous variable. The lag order may depend on the characteristics of the time series used in the construction of the model. In the case of quarterly data, the lag order may be 4 or a multiple of that number, for monthly data – 12 or a multiple, etc. The introduction of too many lags reduces the number of degrees of freedom, which in turn results in lower precision in model estimation. This issue becomes particularly important when the number of exogenous variables of the model is large, and a time series is not too long.

The choice of the lag order may be accomplished using two solutions. The first involves the use of the information criteria that determines the degree of loss of information associated with the adoption of the specific lag order. Regardless which of the information criteria is accepted, this lag order is accepted, for which the value of the information criterion is the least[52]. The second solution is less formal and more intuitive – the accepted lag order should give accepted value of the autocorrelation coefficient of residuals along with statistical significance of model parameters.

4.3.2. Models for non-stationary data

ADL models in a form of equation 4.19 are relatively rarely used for analysis and forecasting real economic processes. Most of economic processes, agricultural processes among them, are non-stationary which is a key problem in forecasting. Economic processes can be non-stationary in the mean or in the variance. Non-stationarity in the mean assumes that the fluctuations around deterministic trend is transitory whereas non-stationarity in the variance assumes that

[52] More on information criteria and their application can be found in Lütkepohl [2007, pp. 146-157].

a random shock has a permanent effect on the system. Different specification of model can be assumed then.

Models for trend stationary series

Let's assume that endogenous and exogenous processes can be described by the deterministic and autoregressive components [Zieliński 1995, Kufel 2002]:

$$Y_t = T_{yt} + S_{yt} + u_{yt}\,, \tag{4.20}$$

$$X_{it} = T_{x_i t} + S_{x_i t} + u_{x_i t}\,, \tag{4.21}$$

where:

T_{yt} and T_{xit} are deterministic trends in endogenous Y and exogenous X processes (variables) explained with the use of polynomial trends,

S_{yt} and S_{xit} are deterministic seasonal component in endogenous and exogenous processes (variables) captured with the use of seasonal dummies,

u_{yt} and u_{xit} are stationary processes (deviations) around deterministic components in endogenous and exogenous processes which can be described by the following equations:

$$u_{yt} = \phi_1 u_{yt-1} + \phi_2 u_{yt-2} + \ldots + \phi_p u_{yt-p} + \varepsilon_{yt}\,, \tag{4.22}$$

$$u_{xit} = \phi_1 u_{xit-1} + \phi_2 u_{xit-2} + \ldots + \phi_p u_{xit-q} + \varepsilon_{xit}\,. \tag{4.23}$$

After substituting 4.22 and 4.23 into 4.20 and 4.21 and then merging it into equation 4.16 we obtain the following model [Kufel 2002]:

$$Y_t = T_t + S_t + \sum_{i=1}^{p} a_i Y_{t-i} + \sum_{j=0}^{k}\sum_{i=0}^{q} \beta_{ij} X_{j\,t-i} + \varepsilon_t\,. \tag{4.24}$$

We can notice that the model in question is an extension of the ADL model according to equation 4.19. This model can be treated also as an extension of the descriptive time series model 2.17 with current and lagged causal explanatory variables.

The model includes deterministic (seasonal component S_t and trend component T_t) and stochastic segment (current and lagged explanatory variables as well as lagged endogenous variables). Seasonal component is captured with the use of seasonal dummies. The deterministic trend component is included in the form of polynomial trends of the order equal to the highest order of the polynomial trend among all analysed processes (dependent and independent). The inclusion of the trend and seasonal components to the model means that from each process (series) the non-stationarity in the mean was eliminated and parameters refer to the dependency on stationary level [Piłatowska 2004].

Models for variance non-stationary series

In the case of a variance non-stationary series inclusion of deterministic trend into equation (4.24) does not remove problem of spurious regression. So another solution have to be considered. They depend however on existence of the long-run relationship among variables and simultaneous existence of non-stationarity in the mean.

In the case of lack of long-run relationship (no cointegration) between non-stationary variables the most widely used procedure assumes taking differences of data. Then ADL model based on differenced series has the following form:

$$\Delta Y_t = \mu + S_t + \sum_{i=1}^{p} \alpha_i \Delta Y_{t-i} + \sum_{j=0}^{k} \sum_{i=0}^{q} \beta_{ij} \Delta X_{j\ t-i} + \varepsilon_t, \tag{4.25}$$

where ΔY_t, ΔX_{tj} are the first differences of endogenous variable and exogenous variables, rest designations as in eq. 4.19.

The differencing of data removes not only a stochastic trend but also a deterministic trend. More precisely, one order of differencing reduces the order of polynomial trend by one. If the maximum order of polynomial of trend line would be one, then only constant and seasonality (which is not removed by non-seasonal differencing) is left in the equation after differentiation. Such a model takes into account only short-run dynamics neglecting long-run trends.

The model 4.25 can be transformed into the model for levels and presented in a form similar to that given by the formula 4.24. However, it would include lower order of polynomial trend in comparison to the model 4.24. Moreover, its autoregressive part, in comparison to the model 4.24, would be extended by additional number of lags equal to the order of integration (differencing) of particular dependent and independent processes [Zieliński 1995, Kufel 2002].

Error Correction Model

The cointegration idea assumes that a combination of processes non-stationary in variance is stationary (section 4.2.3). The relationship for cointegrated processes can be expressed in the form of error correction model [Engle and Granger 1987]:

$$\Delta Y_t = p \sum_{i=1}^{r} \alpha_i \Delta Y_{t-i} + \sum_{j=0}^{k} q \sum_{i=0}^{r} \beta_{ij} \Delta X_{j\ t-i} + \gamma EC_{t-1} + \varepsilon_t, \tag{4.26}$$

where: γ – error correction coefficient, measures the speed of convergence to long-run equilibrium, EC_{t-1} is stationary error correction term obtained from the cointegrating equation $EC_{t-1} = Y_{t-1} - \sum_{j=1}^{k} \theta_j X_j$ where θ_j are long-run coefficients (cointegration vector), rest designations as above.

The error correction coefficient γ informs about speed of convergence to the long-run equilibrium path. It shows how much of the deviation from the long-term path is corrected in a subsequent period. The system will be restored to equilibrium, if the value of the γ coefficient belongs to the interval (0; -1). If $\gamma > 0$ there is no error correction mechanism, the variables are not cointegrated and when $\gamma < -1$ there are oscillations around the long-term trajectory of the increasing amplitude. Parameters α_i and β_{ij} of the model relate to short-run dynamics to equilibrium [Lütkepohl 2007, p. 247].

The error correction model can be extended also by deterministic components. Among them are he constant, the trend, the seasonal component or the variables for structural change.

4.3.3. Concepts of dynamic model specification

The specification of dynamic causal models can be done according to various concepts. The key issues are non-stationarity of data, the causality relationship between variables and specification of deterministic part of the modal as well as number of lags in dependent and in independent variables.

Most of economic series are non-stationary so the choice of the type of stationarising of series seems to be crucial. However, Piłatowska [2005] proves that there is no substantial difference in forecast errors between forecast obtained according to the strategy *always take levels* (include deterministic trend) and *always difference* even though the identification of the type of non-stationary was wrong. This statement is true provided that these models fulfill postulate of congruence. In other words, when the whole information about internal structure of analysed processes is taken into account. In such a case there is also no problem of spurious regression. In the next two sections we will describe two concepts of dynamic model specification.

The concept of congruent modeling according to Zieliński

One of the most interesting forecasting concept is the congruent modeling. The model congruence in Zieliński sense is understood as the congruence of harmonic structure of an endogenous process with joint harmonic structure of endogenous processes and residual process [Piłatowska 2008]. The dynamic congruent model is such a casual model that considers information about the internal structure of examined processes (trend, seasonality, autoregression) in that way that residuals have the white noise properties. The idea of congruence might be applicable for models which are given by equations: 4.19, 4.24 and 4.25. It means that it can be appropriate for stationary, trend stationary, integrated or seasonal processes.

The explanatory variables are chosen on the base of the economic theory. The process of selecting the model components and lags, called congruent dynamic model specification, involves the following steps [Zieliński 1995, Kufel 2002, Piłatowska 2008]:

1. The examination of the internal structure of the endogenous process and all exogenous processes (eq. 4.20-4.21). It includes:
   – the identification of the trend and its separation,
   – separation of the seasonal component,
   – assuming of the order of integration,
   – specification of maximum autoregressive order of individual processes (without trend, seasonality and differenced if needed).
2. The specification of initial model that consists of the maximum order of a polynomial trend line, seasonality and the maximum autoregression order for each process (eq. 4.19, 4.24-4.25).
3. The estimation of the initial congruent model containing all the specified components. Usually OLS method is applied.
4. The elimination of non-significant variables with the use of a posteriori method is applied after estimation. Elimination of variables is done according to *t*-statistic (see eq. 4.7) and the examination of the residuals properties.
5. The interpretation of the estimated values of the structural parameters and the evaluation of model fitting to the data. Lagged dependent variable are interpreted as substitute elements which appears in the model when important explanatory variables are omitted or dependence of forecasted variable on explanatory variable for different frequency components is not the same. Current and lagged explanatory processes have causal interpretation.

The procedure outlined above is a manual one. However, Błażejowski *et al.* [2009] have developed automatic procedure in GRETL program similar to Autometrics, described in next section.

The determination of the forecast requires establishing the values of explanatory processes (variables) for the future. So, even though possessing a well estimated model, the forecast might be inaccurate when forecasts of explanatory variables are incorrect.

General to specific modeling according to Hendry

General-to-specific (Gets) model selection procedure, often referred to as the LSE methodological approach, is one of the few coherent and relatively comprehensive methodological bases for applied econometric modeling. It is the automatic model selection algorithm which is provided by Autometrics in the PcGive program. *Gets* involves the formulation of a "general" unrestricted model that is congruent with the data and then eliminating non-significant variables. This leads to a simpler "specific" congruent model that encompasses rival models. "Congruency" relates to the matching of the model with the evidence in the data with respect to the criteria such as homoscedastic, normal, innovation errors; parameter constancy; and weak exogeneity of the conditioning variables for the parameters of interest. Encompassing is concerned with avoiding loss of information in the reduction process [Hendry 1995, p. 365].

The Gets algorithm implemented in Autometrics is based on five main components (steps) [see: Hendry and Krolzig 2001, Hendry and Krolzig 2005; Castle et. al. 2013; Hendry and Pretis 2012]. They are described below.

The first step is a formulation of the general unrestricted model (GUM) based on theory, previous evidence and existing data. The basis for GUM is data generation process (DGP) which reveals an economic mechanism that operates in the real world. In practice there is impossible to model precisely all variables in economy so DGP needs to be reduced to manageable size in the so-called "local DGP" (LDGP). Therefore, the GUM should nest the LGDP. This step is similar to the Zieliński approach, however, there is some difference in internal structure specification (trend, seasonality, autoregression). In Gets the lag length for particular processes is usually higher than in the Zieliński approach because an analysis of internal structure is omitted [Kufel 2004].

The next step involves an estimation and a pre-search lag reduction. It requires the selection of the set of mis-specification tests, their forms, and significance levels. There is the possibility to use Liberal or Conservative strategies during specifying model. In this step insignificant lags are removed, speeding up selection procedures and reducing the fraction of irrelevant variables selected. Autometrics removes the least significant variable as determined by the lowest absolute *t*-ratio.

The third step includes a multiple-path reduction (tree search) that explores all feasible reduction paths to avoid path dependence. Each removal constitutes one branch of the tree. For every reduction, there is an unique sub-tree which is then followed – each removal is back-tested against the initial GUM using an F-test. Branches are followed until no further variable can be removed at the pre-specified level of significance. They examine multiple search paths, thus avoiding “path dependency”, which can seriously affect the properties of a simplification algorithm based on a single search path; e.g. a simple decision rule, such as successively removing the variable with the lowest absolute t-value, can easily result in being stuck in a search path that has deleted relevant variables. They also emphasize the importance of considering only model reductions that do not fail diagnostic tests in order to retain congruence, and the use of overlapping sub-sample testing to aid in the overall assessment of the ‘reliability’ of the significance of the coefficients.

In the fourth step, the validity of each reduction is checked. A wide range of diagnostic tests is applied and they include: tests for normality, heteroscedasticity, test for parameter constancy, test for residual autocorrelation and autoregressive conditional heteroscedasticity. Both congruence and encompassing are checked by Autometrics when each terminal model is reached after path searches. If all reductions and diagnostic tests are acceptable, and all remaining variables are statistically significant that model becomes a terminal selection.

Selection of the final, unique, model by comparing all terminal models which passed diagnostic tests is the final step. When all paths have been explored and all separate terminal models have been found, they are tested against their union to find an undominated encompassing candidate. To select a unique model (in case of a few final candidates), the likelihood Schwarz information criterion (SIC) is used. In the end the significance of every variable in the final model is assessed in two over-lapping sub-samples to check the reliability of the selection.

After selection of the final model the forecast might be calculated. Also the competing terminal models might be used for preparing forecast so the multi-path forecasts can be compared.

4.4. VAR and VECM models

This section presents a basis of the dynamic VAR and VECM models. They can be regarded as an alternative to: multi-equation models of interdependent equations, single-equation causal dynamic models presented in previous section and time series models presented in Chapters 2 and 3.

4.4.1. VAR and VECM

Vector autoregressive model

The analysis of the relations between variables may indicate that they influence each other. This requires the use of a VAR methodology (Vector Auto-Regressive) developed by Sims [Sims 1980], as an alternative to the classical multi-equations model of interdependent equation[53]. The basic form of the VAR model [Charemza, Deadman 1997, Chap. 6; Enders 2010 Chap. 5.5; Mills 2002] is as follows:

$$\mathbf{x}_t = \mathbf{A}_0\mathbf{d}_t + \sum_{i=1}^{r}\mathbf{A}_i\mathbf{x}_{t-i} + \mathbf{e}_t\,, \tag{4.27}$$

where:

$\mathbf{x}_t = [x_{1t}, ..., x_{mt}]^{\mathrm{T}}$ is a vector of observation on the current values of the variables,

$\mathbf{d}_t = [d_{0\mathrm{t}}, ..., d_{kt}]^{\mathrm{T}}$ is a vector k+1 of deterministic components of equation (intercept, temporal variable, binary variables, etc.),

$\mathbf{A}_0$ – is a matrix of parameters in the $\mathbf{d}_t$, vector variables,

$\mathbf{A}_i$ – is a matrix of parameters in the delayed variables of vector $\mathbf{x}_t$, where maximum delay row is equal to r,

$\mathbf{e}_t = [\mathbf{e}_{1t}, ..., \mathbf{e}_{mt}]^{\mathrm{T}}$ – vectors of the model equations residuals.

Vectors of the model equation residuals should satisfy the classical assumptions (zero mean, constant variance, absence of autocorrelation), while covariance between residuals of the individual equations can be different from zero. The lag order (r) should be selected to reflect the natural interactions (for example, for quarterly data the lag order should not be less than 4) and to remove autocorrelation.[54]

The *OLS* VAR model estimators have the desired properties only when the time series of observations on the variables are stationary. In the case of non-stationary (integrated) time series, the VAR model can be applied to the first differences or, when the variables are cointegrated, the VECM model (Vector Error Correction Model) should be implemented.

Johansen procedure

The Johansen procedure ought to be applied when the model contains many equations and when endogenous variables in a one equation are exoge-

---

[53] This is a result of the so called "Sims critics".

[54] The procedures use for the selection of lag rank used for the *ADL* models and described in subsection 4.1.2 also apply here.

nous variables in other equations. The VAR model must be converted under this procedure to VECM model (Vector Error Correction Model):

$$\Delta \mathbf{x}_t = \mathbf{\Psi}_0 \mathbf{d}_t + \mathbf{\Pi} \mathbf{x}_{t-1} + \sum_{i=1}^{r-1} \mathbf{\Pi}_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t , \tag{4.28}$$

where: $\mathbf{\Psi}_0$ – the matrix of parameters of the $\mathbf{d}_t$ vector variables, $\mathbf{\Pi} = \sum_{i=1}^{k} \mathbf{A}_i - \mathbf{I}$;

$\mathbf{\Pi}_i = \sum_{j=i+1}^{k} \mathbf{A}_j$, $\boldsymbol{\varepsilon}_t$ – model residuals.

The rank of $\mathbf{\Pi}$ matrix is used for cointegration testing in the Johansen procedure. This rank is equal to the number of independent cointegration vectors [Johansen 1988]. The two characteristic of the matrix $\mathbf{\Pi}$ estimator:

$$\lambda_{trace}(R) = -N \sum_{i=R+1}^{m} \ln(1-\lambda_i); \qquad \lambda_{\max}(R) = -N \ln(1-\lambda_{R+1}), \tag{4.29}$$

where: λ_i – estimated eigenvalues, N – number of observations, λ_{trace} and λ_{max} are the test statistics in the Johansen procedure.

The first statistic is used to test the null hypothesis that the number of independent cointegrating vectors is less than or equal to R, against the alternative hypothesis that the number of cointegration vectors is greater than R. The second statistic is used to test the null hypothesis that the number of different cointegration vectors is equal to R, against the alternative hypothesis that there are R+1 of them. Both tests are right-hand side tests.

The test used in the Johansen procedure is an iterative test. The strategy for finding the cointegration rank of a given set of k variables is to test a sequence of null hypothesis:

$$H_0: \text{rank}(\mathbf{\Pi})=0,\ H_0: \text{rank}(\mathbf{\Pi})=1,\ \ldots,\ H_0: \text{rank}(\mathbf{\Pi})=k\text{-}1, \tag{4.30}$$

and cease the test when H_0 cannot be rejected. The cointegrating rank is equal to the rank for which the sequence of tests is terminated. The Johansen procedure might produce three results [Johansen 1988]:

- rank($\mathbf{\Pi}$) = 0, the model (4.28) reduces to a VAR model in first differences,
- rank($\mathbf{\Pi}$) is greater than 0 and less than k, then the number of cointegrating vectors is equal to that rank;
- rank($\mathbf{\Pi}$) = k – none of the null hypothesis of the sequence (4.30) can be rejected – then the time series of variables are stationary and the model (4.28) is a VAR model of variables levels.

The VAR model can be applied in the case of non-stationary and non-cointegrated series for the first differences of variables. In the case of such a modification of the model (4.27), when vectors $\mathbf{x}_t$ and $\mathbf{x}_{t-i}$ are the first differences only the information about short-term impact of each variable on the endogenous variables is obtained. The VAR model will have in that case the following form:

$$\Delta\mathbf{x}_t = \mathbf{A}_0\mathbf{d}_t + \sum_{i=1}^{r}\mathbf{A}_i\Delta\mathbf{x}_{t-i} + \mathbf{e}_t\,, \tag{4.31}$$

where:

$\mathbf{\Delta x}_t = [\Delta x_{t1}\ \ldots\ \Delta x_{tk}]^T$ – vector of observations on the current values of the first differences in dependent variables,

$\mathbf{d}_t = [d_0\ \ \Delta d_t\ \ \Delta d_{1\,t\text{-}1,}\ \ldots\ \Delta d_{1\,t\text{-}r}\ \ldots\ \Delta d_{l\,t\text{-}r}]^T$ – vector of exogenous components of equation, which components are respectively: constant equation and current and delayed values of the first differences of the exogenous variables,

$\mathbf{A}_0$ – the vector $\mathbf{d}_t$ variables,

$\mathbf{A}_i$ – matrix of parameters of lagged values of vector $\mathbf{x}_t$,

$\mathbf{e}_t = [\mathbf{e}_{1t}\ \ldots\ \mathbf{e}_{kt}]^T$ – vectors of the model residuals.

4.4.2. Structural model and forecasting

Exogeneity

In the case of multi-equation models, it is important to identify exogenous ones (serving only as explanatory variables) from the whole set of the variables taken into account in the analysis. Exogenous variables should be excluded from the set of explanatory variables by the model, or to put it more technically, should be excluded from the vector **x** of models (4.27), (4.28) and (4.31). The arbitrary distinction between exogenous and endogenous variables (explained by the model) was one of the elements of the so-called "Sims criticism".

The methodology using time series in the modelling of the problem of dependences, exogeneity of variables can be weak or strong [Lütkepohl 2007, pp. 387-390; Maddala 2001, Chap. 9.10][55]. A variable, for example x_t is weakly exogenous with respect to estimated vector of model parameters, such as $\boldsymbol{\psi}$, if inference of conditional $\boldsymbol{\psi}$ with respect to x_t is not associated with loss of information. This means that one can effectively perform inference about the components of vector $\boldsymbol{\psi}$ conditionally only with respect to x_t.

The process X_t is weakly exogenous for the parameter vector $\boldsymbol{\psi}$ involved in the study, if the marginal probability density does not contain significant information to estimate vector $\boldsymbol{\psi}$. In order to establish a strong exogeneity it is also

[55] There is also so-called superexogeneity [see Maddala 2001, p. 378].

necessary that the process Y_t is not the Granger-cause of process X_t. The procedure for testing the weak exogeneity is given by Engle *et al.* [1983]. Strong exogeneity test procedure follows directly from its definition and involves the use of weak exogeneity test, and then the test of causality [Engle *et al.* 1983].

Impulse response function

The correlation of the residuals of the VAR model equations allows the construction of the so-called structural models [Hamilton 1994, pp. 324-325; Lütkepohl 2007, pp. 358-368]. The structural VAR models makes it possible to construct the impulse response function (IRF), determining the distribution in time values of the *j*-th variable in response to changes (innovations) of the *k*-th variable[56]. Most often IRF is presented as a graph showing the change in the reaction of *j*-th variable to the change (shock) of the *k*-th variable equal to one standard deviation of the of *k*-th variable residuals. Analysis of the impulse response function comprises three elements: the direction of the impulse, the strength of impulse, and the distribution in time.

The correlation of the residuals of the VAR model equations gives the ability to build structural models. The structural VAR model is [Hamilton 1994, pp. 324-325; Lütkepohl 2007, pp. 358-368] given by the formula:

$$\mathbf{B}\mathbf{x}_t = \boldsymbol{\Gamma}_0\mathbf{d}_t + \sum_{i=1}^{r}\boldsymbol{\Gamma}_i\mathbf{x}_{t-i} + \boldsymbol{\xi}_t \tag{4.32}$$

where:

$\mathbf{B}$ – matrix of parameters with non-lagged variables of vector $\mathbf{x}_t$,

$\boldsymbol{\Gamma}_0$ – matrix of parameters with variables of vector $\mathbf{d}_t$,

$\boldsymbol{\Gamma}_i$ – matrix of parameters with lagged variables of vector $\mathbf{x}_t$, (i = 1, ...,r),

$\boldsymbol{\xi}_t$ – vector of model residuals.

Multiplying equation (4.24) by $\mathbf{B}^{-1}$ allows us to show the relationship between the model (4.27) and its structural form (4.32):

$$\mathbf{x}_t = \mathbf{B}^{-1}\boldsymbol{\Gamma}_0\mathbf{d}_t + \sum_{i=1}^{r}\mathbf{B}^{-1}\boldsymbol{\Gamma}_i\mathbf{x}_{t-i} + \mathbf{B}^{-1}\boldsymbol{\xi}_t\,. \tag{4.33}$$

If we assume: $\mathbf{B}^{-1}\,\boldsymbol{\Gamma}_0 = \mathbf{A}_0$, $\mathbf{B}^{-1}\,\boldsymbol{\Gamma}_{\mathrm{i}} = \mathbf{A}_{\mathrm{i}}$, $\mathbf{B}^{-1}\,\boldsymbol{\xi}_t = \mathbf{e}_t$, we will obtain the model given by equation 4.27.

[56] Description and interpretation of the impulse response function can be found in Hamilton [1994, pp. 318-323] and Lütkepohl [2007, pp. 51-63].

The relationship between the model (4.27) and (4.32) comes from the fact that $\mathbf{e}_t=\mathbf{B}^{-1}\xi_t$ .This allows us to analyse the interactions between variables of the vector $\mathbf{x}_t$ because model (4.27) can be reduced to VMA model[57], when $r = 1$:

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}_1^i \mathbf{e}_{t-i} , \tag{4.34}$$

where $\boldsymbol{\mu}$ is a vector of the $\mathbf{x}$ vector variables mean values. As $\mathbf{e}_t=\mathbf{B}^{-1}\xi_t$ we obtain:

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}_1^i \mathbf{B}^{-1} \xi_{t-i} , \tag{4.35}$$

and when we assume that $\boldsymbol{\Phi}_i = \mathbf{A}_1^i \mathbf{B}^{-1}$, we get:

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \xi_{t-i} . \tag{4.36}$$

The elements of $\boldsymbol{\Phi}_i$ matrices are used to measure the effect of residuals ξ_t for variables $\mathbf{x}_t$. The elements of $\boldsymbol{\Phi}_i$ matrix – $\phi_{jk}(i)$ – measure the effect of change in $\xi_{k\, t-i}$ on the j-th variable. Subsequent values $\phi_{jk}(i)$ for $i = 0, 1, ..., T$ give the impulse response function, which defines the behavior of j-th variable in response to changes in the residuals of k-th variable. The most common way of presenting the IRF is a graph showing the change over time as a reaction of j-th variable to the change in the residuals of k-th variable equal to one standard deviation.

The IRF values are determined by the order of variables in the $\mathbf{x}$ vector. This order is all the more important, when correlation coefficients between residuals of the VAR model equations are higher [Lütkepohl 2007, pp. 358-368]. When this correlation coefficients are so high that the ordering of the variables affects the value of the *IRF*, the way of ordering is determined based on the decomposition of the forecast error [Lütkepohl 2007, pp. 358-368].

Forecasting

The primary problem in the case of causal models is the need to make a prediction of the exogenous variables values to obtain a forecast. This is one of the main sources of forecast errors. Even the best model will give poor predictions, if these values are incorrect. Forecasting based on the VAR models is free from this problem if the vector **d** elements (constant, time variable and binary variables) are the only exogenous variables of the VAR model (see model 4.28). This makes the VAR model a very convenient tool for forecasting. If the exoge-

[57] *Vector Moving Average*. Similar to single-equation autoregressive models (AR), the VAR model can be represented as a moving average – VMA [Lütkepohl 2007, pp. 423-426].

nous variables are present in the model then there is the need for their forecast in the period of forecasting.

Generating forecasts based on the VAR model generally takes place with multi-step iterating procedure (see for example Stock and Watson 2001). Forecast for period $T+1$ for model (4.28) estimated for $t = 1, \ldots, T$ is generated based on the values of model variables for time T. Forecasts for the later periods are based on the model variables forecasts previously formulated. In the case of the structural VAR models the impulse response functions (IRF) can be used for forecasting. The VAR models are used generally to create short-term predictions. The forecast of a longer time horizon will be less accurate due to the transfer of forecast errors.

Forecasts based on econometric models cannot be treated in automatic way. Their quality depends on many factors, especially the quality of statistical data. It should be noted, however, that even the best data will not provide a correct estimate if the model is inconsistent.

4.5. Application of selected causal models to agricultural price forecasting

4.5.1. Forecasting pig prices with the use of congruent model

We try to construct a one-equation congruent model for monthly live pig price in Poland in a period January 2000–June 2013 in the first step. There are a lot of alternatives for a model specification. The most obvious model might be that which assume that the pig price in Poland is a function of the world pig price and exchange rate (Chap. 1.4).

We try to build another model which assumes that the pig prices in Poland depend on domestic factors, which are obviously influenced by global conditions. The set of explanatory variables would consist in this case of variable cost of production (feeders) and variable representing prices of substitute goods. As a proxy for feeders will be farm wheat prices (other grains are highly correlated with wheat). Poultry meat is the main substitute for pork meat in Poland (the consumption of beef and veal is below 3 kg per capita). Therefore, poultry farm prices will serve as another explanatory variable. The main equation is in this case: $P.pig_t = f(P.wheat_t, P.poultry_t, \varepsilon_i)$. The causality test confirms that wheat prices series is a Granger cause for pig prices. The causality between pig and poultry prices seems to be mutual.

Internal structure of processes

In the first step we will examine the internal structure of the endogenous process and all the exogenous processes. The time series structure was analysed on the basis of logs data (X-12-ARIMA method shows that for wheat and poultry prices multiplicative model is preferable, some information criteria indicate the additive model but others the multiplicative model in the case of pig prices).

The pig and poultry prices are highly correlated so they have similar internal structure. A quadratic trend is present in both series (according to F test, R^2 is significantly higher in the model with quadratic trend as compared to the model with linear trend). In both series seasonality is present and there are no structural breaks which might be detected by automatic X-12-ARIMA. Both series are trend stationary (as proven by ADF test with second order polynomial trend). A null hypothesis cannot be rejected, however, in the model with constant (eq. 4.8-4.10).

In the case of wheat prices some structural breaks might be present. Applying automatic X-12-ARIMA procedure allows us to detect five level shift structural breaks (in August-November of the years: 2003, 2004, 2007, and 2010). When dummy variables for structural breaks are included, a linear trend is indicated. The second order polynomial trend outperforms a linear one if the model is without these variables. According to X-12-ARIMA procedure the series contains a statistically significant seasonality. Application of ADF test without structural breaks indicates that wheat price series is non-stationary. After inclusion of level shift dummies to the ADF test equation (with liner trend), we can reject a null hypothesis for levels of price data.

In the next step we specified the maximum autoregressive order of individual processes (for data without trend and seasonality) with the use of PACF (the Quenouille test). A summary of this analysis is given in the Table 4.1.

Table 4.1. Internal structure of analysed processes (log of data)

Processes	Structural breaks	Polynomial order of trend	Seasonality	Autoregression order
Pig prices (Y_t)	No	2	Yes	AR(6)
Wheat prices (X_{1t})	No	2	Yes	AR(2)
	Yes	1	Yes	AR(1)
Poultry prices (X_{2t})	No	2	Yes	AR(2)

Source: own calculations.

Model estimation

The information about internal structure of analysed processes (Table 4.1) allows us to specify an initial model that consists of the maximum order of a polynomial trend line, seasonality and the maximum autoregression order for each process. Another possibility (when internal structure analysis is omitted) is to include linear or quadratic trend, seasonality and 12-month lags for each process. This significantly increases the number of explanatory variables.

Additional cross-correlation analysis shows that the wheat prices (X_1) may serve as the leading variable for the pig price series (Y_1). The estimated led-length is around 6-8 months. Therefore lags for the wheat prices (X_1) were extended up to 8 months. The initial congruent model is as follows:

$$Y_t = \mu + \sum_{i=1}^{2}\delta_i t^i + \sum_{i=1}^{5} d_i LS_i + \sum_{i=1}^{11} D_i S_i + \sum_{i=1}^{6} a_i Y_{t-i} + \sum_{i=0}^{8}\beta_{1i} X_{1,t-i} + \sum_{i=0}^{2}\beta_{2i} X_{2,t-i} + \varepsilon_t, \quad (4.37)$$

where:

t – time variable,

LS – level shift dummies,

S_i – seasonal dummy variables that take the value of 1 for i-th season and 0 otherwise, rest designations as above in the section.

Table 4.2. Estimated congruent model (log of data) with selected statistics

Parameter	Estimate	P-Value	Parameter	Estimate	P-Value
Const	0.44267	0.1761	LS1	0.04917	0.0074
Y_{t-1}	0.89991	0.0000	S1	-0.04553	0.0004
Y_{t-3}	-0.11384	0.0174	S3	0.04181	0.0009
X_{1t-1}	-0.04289	0.0448	S6	0.04790	0.0002
X_{1t-7}	0.05971	0.0051	S7	0.03966	0.0025
X_{2t}	0.13284	0.0212	S10	-0.05162	0.0000
t	-0.00200	0.0068	S11	-0.03662	0.0037
t^2	0.00001	0.0103	-	-	-
R-squared adjusted 94.61; Standard error of estimation 0.040; Durbin-Watson statistic 1.965; Durbin *h* statistic 0.451; Doornik-Hansen normality χ^2 test statistic 1.591 (p= 0.451).					

Source: own calculations.

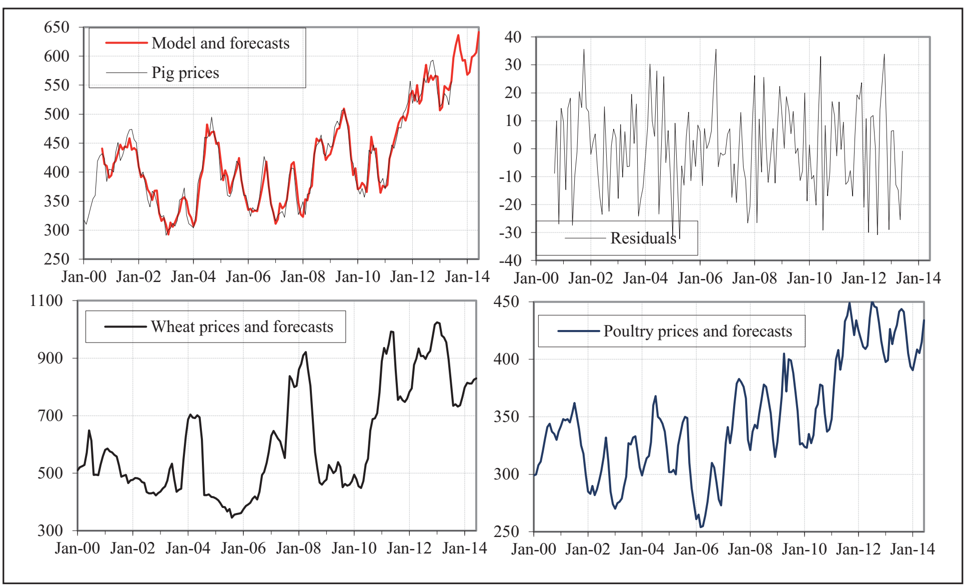
The initial model was estimated (OLS) and all the insignificant variables were removed with the use of *a posteriori* method in the next step. The estimation results of the final model, along with selected statistics, are given in Table 4.2. Lagged dependent variables are interpreted as substitute elements which appear in the model when important explanatory variables are omitted or there is some kind of persistency in the forecasted variable. The positive impact of the poultry prices on pig price is in the line with the theory of economy. The wheat prices in the long-run (lagged by 7 months) have positive impact on the pig prices, however, in the short term (lagged by 1 month) the increase in the wheat prices may cause the decrease of the pig sale prices in reaction for the increase of fodders.

The model seems to be acceptable in terms of fitting. All the variables are statistically significant and residuals possess desired properties. There is no autocorrelation in residuals. The residuals have a normal distribution and there are no structural changes during the period of research.

Forecasting

The values of all the explanatory variables have to be known or have to be predicted (or assumed) to obtain a forecast. Deterministic variables (time, dummies) are known. To calculate forecasts for causal explanatory variables we will use regARIMA model which is nested in X-12-ARIMA procedure (Chap. 2).

Figure 4.1. Data and forecasts for live pig prices calculated with the use of congruent modelling (Table 4.2) and explanatory causal variables with their forecasts



Source: authors' own compilation based on the data of the Central Statistical Office.

The forecasts of explanatory variables as well as forecasts of pig prices are given in Figure 4.1. The reliability of forecasts will not be discussed. The main goal was to demonstrate the procedure application[58]. However, we have to bear in mind that the forecasts of the pig prices highly depend on forecasts of explanatory variables.

4.5.2. Commodity price forecasting with VAR/VECM models

Another approach of forecasting agricultural prices is VAR/VECM methodology. There is no need for forecasting values of explanatory variables in this case because all of them are endogenous and their values are derived from their past values and past forecasts. Model is also restricted to the 2-5 endogenous variables so there is a need for estimation of several models and linking them in the integrated system to calculate forecast for more variables. In such a system, forecasted (dependent) variables from one models may serve as an explanatory variable in others.

Model specification and estimation

Our task was to estimate VAR or VECM model for three variables from chapter 4.5.1 (Fig. 4.1). It should be underlined that all the variables are non-stationary according to ADF test with constant. However, inclusion of the quadratic trend (the pig and poultry price series), the liner trend and LS structural dummies (the wheat price series) allows us to reject null hypothesis stating that variables are non-stationary.

There are a few possibilities for forecasting of those prices. The first one is to use VAR model for levels and include all deterministic variables (for trends, seasonality and structural changes) as exogenous variables. Another possibility is to test long-run relationship (cointegration) for levels of price series with the use of the Johansen procedure. If there is such a relationship, the VECM framework can be applied. If not – VAR model for first differences can be estimated. The $\mathbf{\Pi}$ matrix is a full rank matrix. We decided therefore to estimate unrestricted VAR model for levels to be consistent with the analysis performed in chapter 4.5.1. All the information criteria indicate that the model with two lags should be preferred. The Table 4.3 contains estimated coefficients for

[58] In authors opinion the use of trend polynomials out of sample may lead to biased forecasts. Another possibility would be the application of segmented linear trends instead polynomial ones. The application of linear trends with two segments (2000-2005 and since 2006) does not change significantly the set of explanatory variables of the final model (table 4.2) but the forecasts obtained for pig prices are 5-15% lower than those in Figure 4.1.

endogenous lagged variables (exogenous deterministic coefficients were not included in the table due to clarity reasons).

Table 4.3. Estimated VAR model (log of data)

Parameter	Pig prices model		Wheat prices model		Poultry prices model	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
const	0.4734	0.1675	1.4917	0.0003	0.7075	0.0055
Pig_{t-1}	0.9901	0.0000	-0.0095	0.9248	0.1222	0.0537
Pig_{t-2}	-0.1460	0.0902	-0.0891	0.3793	-0.1209	0.0564
$Wheat_{t-1}$	-0.1155	0.0586	1.2899	0.0000	0.0496	0.2664
$Wheat_{t-2}$	0.1192	0.0433	-0.4024	0.0000	-0.0221	0.6076
$Poultry_{t-1}$	0.0115	0.9232	0.0574	0.6822	0.9789	0.0000
$Poultry_{t-2}$	0.0686	0.5720	-0.0840	0.5572	-0.1357	0.1290
Statistics	R-squared adj. 93.90; Standard error 0.0421; Durbin-Watson 2.095;		R-squared adj. 97.25; Standard error 0.0497; Durbin-Watson 1.973;		R-squared adj. 94.86; Standard error 0.0309; Durbin-Watson 2.022;	

Source: own calculations.

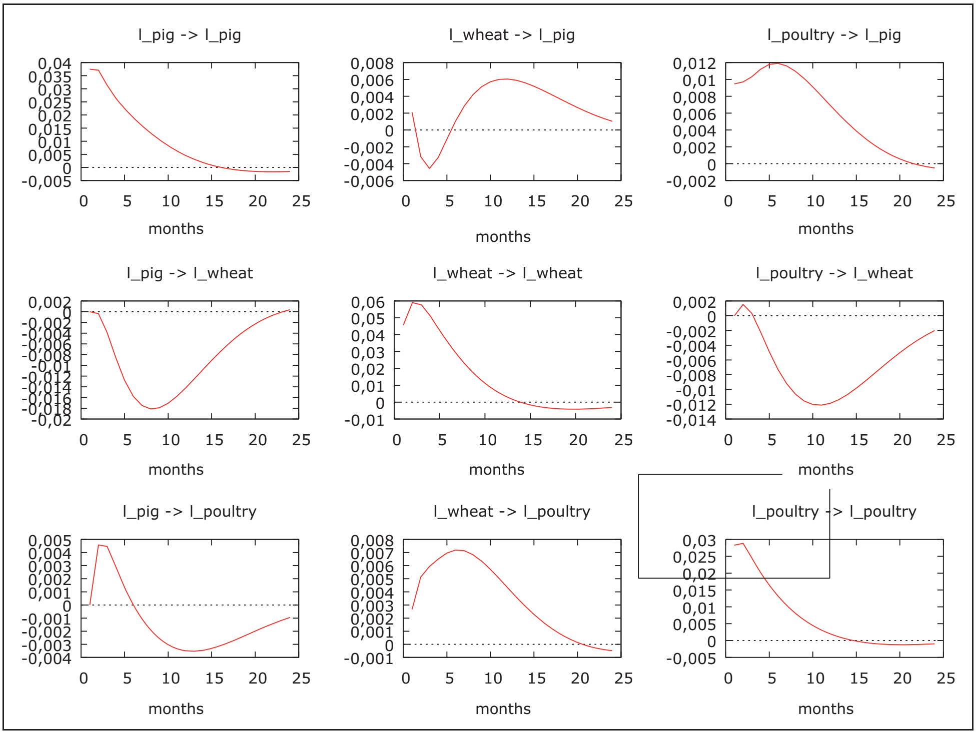
All the models fit well the data what is confirmed by the coefficients of determination and the standard errors of estimation. There is no significant autocorrelations in residuals of all the equations (see D-W statistics and Ljung-Box test). The ARCH effect for the first 12 months is not present (Engle's LM-ARCH test) however some statistically significant ARCH effects for first two lags are visible in case of the wheat and the poultry prices. The Doornik-Hansen multivariate normality test indicates lack of normal distribution of residuals, despite the inclusion of LS variables. Only residuals of the pig prices model have a normal distribution. Lack of normality in residuals of agricultural commodity prices models is rather the rule than the exception. This effect can be partially mitigated with some structural breaks dummies but in some cases there is a need for models' extension and the use of non-liner dependences (ARCH, GARCH models).

Price analysis and forecasting

The causal relationships between variables can be seen via impulse response functions (IRF, see Figure 4.2). The impulse response function analysis refers to four elements: the direction of the impulse impact, the impulse strength, the impulse distribution in time and the impulse expiration rate. All these components of IRF analysis can be used in forecasting of agricultural commodity prices. The Figure 4.2 gives some good examples.

A positive impact of wheat prices on pig (10-13 months lag) and poultry prices (6-7 months lag) is plainly visible on the graph depicted in Figure 4.2. Estimated lags reveal the reaction of pig prices on changes in wheat prices and it differs slightly from those assumed on the base of cross-correlation analysis in section 4.5.1. According to the IRF the poultry prices lead the pig prices by 5-7 months. However, on the basis of causality test it was assumed (section 4.5.1) that there is no significant lag between pig and poultry prices.

Figure 4.2. Impulse response function for variables included in VAR model

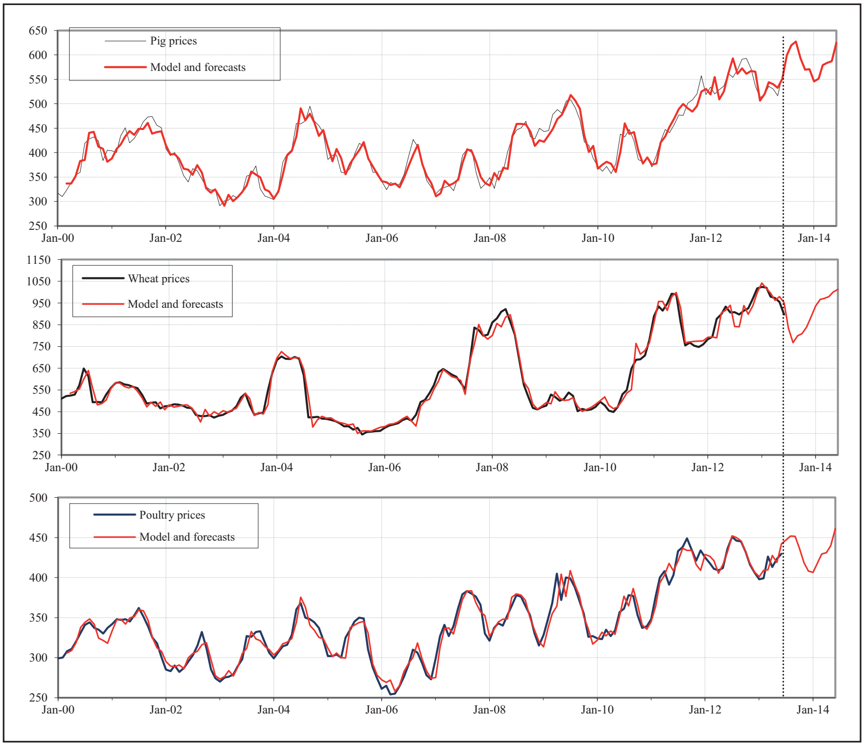


Source: own calculations.

Forecast of all prices from VAR model are depicted in Figure 4.3. They indicate on high wheat and meat prices in the 12-month period. Forecasts for the pig prices are similar to the forecasts obtained from the congruent model (Fig. 4.1). However, higher forecasted poultry and wheat prices in VAR model can be noticed.

The comparison of Figures 4.1 and 4.3 shows more similarities. Approximation of price changes made by both models, congruent and VAR model, are very similar and very close in both cases. The fact that the predictions generated by both models reflect the pattern of the agricultural commodity price changes is also visible.

Figure 4.3. Price forecast calculated with the use of VAR model (Table 4.3)



Source: authors' own compilation based on the data of the Central Statistical Office.

5. Partial equilibrium models of the agricultural sector

There are various methods to analyse the links between commodity markets and assess the impact of different policy measures. The most commonly applied methods include general and partial equilibrium models, which can also be used to predict the future state of the market [Baumel 2001; Wisner *et al.* 2002; Hamulczuk 2011]. The most well-known partial equilibrium models of the agricultural sector are AGLINK-COSIMO and FAPRI models. Every year they are used to generate long-term projections for the most important agricultural markets worldwide. These models do not refer particularly to Poland or other individual Member States, but to the European Union as a whole. The projections for individual EU Member States may be done with the use of the AGMEMOD model. The following sections of the paper provide information on the fundamentals of partial equilibrium models, characterise the above three models and show their possible applications.

5.1. Foundation of partial equilibrium models

Market equilibrium

The market equilibrium is defined as the state of the market in which the quantity demanded and the quantity supplied are equal and thus the manufacturers do not have the motivation to change the production structure, the consumers are not motivated to change their consumption patterns, the production volume is equal to its consumption, and the analysed economy reaches the highest level on the indifference curve. This is the Walrasian understanding of equilibrium which characterises the economy at rest. The premise behind that equilibrium model is the competitiveness of all markets and subordination to the market mechanism. In the Neumann terms, the economy is in equilibrium if it can evenly increase the production without alterations to its structure, while, at the same time, maintaining full compliance between technological and economic growth. The neoclassical approach to equilibrium is an intermediate form between that of Walras and Neumann ones, according to which the economy is in equilibrium if it allows a steady growth of all basic economic values: the factors of production, production and consumption [Dąbrowski 2009].

The equilibrium models, as given by Walras and Marshall, were and still are subject to criticism, which highlights the inadequacy of the neoclassical assumptions for general equilibrium models when faced with the observed reality. Critics indicate primarily the imperfect mobility and divisibility of factors of production, imperfect information (regarding both natural conditions and the behaviour of market agents) or limited rationality of market agents. This means that the market does not reach equilibrium in classical terms (supra-optimal), but it may drive at the optimal equilibrium. Optimal equilibrium is, theoretically, achievable provided that one can overcome defects in coordination on respective markets. At the same time, it is essential to assume rationality of agents' actions and the associated risk aversion.

The market equilibrium in the whole economy, i.e. all markets at the same time, is called the general equilibrium. If we consider the equilibrium of individual markets or sectors, it is termed the partial equilibrium. The general equilibrium models have a wider formula that assumes interactions between all sectors of the economy, including the agriculture which is significant advantage of this type of models. On the other hand, due to a highly aggregated structure of the general equilibrium models, they poorly reflect interactions inside the sectors. The partial equilibrium models allow for analysis of the sector at a much higher level of disaggregation (detail) than the general equilibrium models. Due to the detailed definition of the relationships between individual markets and the instruments of economic policy, the partial equilibrium models are often applied to evaluate changes in the agricultural markets, where the government policy is quite important. They also allow for detailed analysis of the relationships and links within and between sectors. Also the relationships between domestic agricultural markets and the situation on the foreign and global markets is of great importance [Banse, Tangermann 1996; Tongeren *et al.* 2001].

Partial equilibrium model

The partial equilibrium concept was derived from the theory of supply and demand by Marshall and concerns only one market or sector. Partial equilibrium occurs when the volume of realised demand is equal to the volume of supply on the market of the *i*-th commodity. Static equilibrium, which allows for simultaneous determination of the price and quantity of product, is defined algebraically by two equations (5.1, 5.2). The third equation expresses balance between supply and demand [Tomek, Robinson 2003]:

$$Q^d_t = \alpha - \beta P_t \tag{5.1}$$

$$Q^s_t = -\delta + \gamma P_t \tag{5.2}$$

$$Q^d_t = Q^s_t \tag{5.3}$$

where: Q^d_t – demand, Q^s_t – supply, P_t – prices, t – time, α , β, δ, γ – structural parameters.

The main drawback of the static analysis is the inability to assess how the equilibrium is achieved, if the initial state of the system was in disequilibrium (e.g. as a result of shock). Thus, the dynamic approach to analysis includes the time factor which allows for considering interim relationships. The dynamic approach also enables us to include into the model the expectations of market agents and to take into account the producers' responses to prices from previous periods. The concept of dynamic inclusion of market adjustments seems to be more in line with the reality, where a significant period of time elapses between the decision on production and the appearance of the product on the market.

In the simplest terms, the expectations can be naive, described in the form of the so-called cobweb model of supply and demand. In this model, the demand is determined by the equation (5.1), while the volume of supply (production) is a function of past, but not current prices. Consequently the equation (5.2) is modified to $Q^s_t = -\delta + \gamma P_{t-1}$. Thus, the equilibrium price in the current period depends not only on the elasticity of supply and demand curves in the short--term, but it also reflects the situation in the previous periods and is equal to $P_t = (\alpha + \delta)/(\beta + \gamma)P_{t-1}$.

This system of equations, when solved, allows us to determine the future market prices. The idea of partial equilibrium model presented above is very simplistic, because it takes into account only the price of one product while assuming fixed prices of all other goods. Including additional products from the sector with the analysis allows for cross-market interactions. On the demand side, this means that the demand for a given product (*i*) will depend not only on its current price, but also on the prices of other products which are substitutes or complementary ones. The supply side is also subject to modifications. Producers can allocate their resources (factors of production) to various alternative uses. Thus, the volume of production will depend on the relationship between the expected prices of a given product (*i*), and the prices of other goods. If we include the costs (valuation of factors of production), the partial equilibrium models become more and more complex.

The partial equilibrium model maintains its character despite including additional details, since a number of interactions are unilateral. Many variables influencing the state of the sector come from the external sources, which are mainly macro-economic conditions such as the following variables: population figures, technology, GDP, interest rates and the exchange rate. The model does not consider equilibriums in other sectors or the equilibrium in global terms. If we considered the equilibrium of the total (n) related markets (sectors), we would have to deal with the state of the general equilibrium.

The partial equilibrium models of the agricultural sector cover a number of other interactions in order to approximate the reality as closely as possible. These interactions may be unilateral or multilateral and they include, for instance, the instruments of trade policy, sectoral policies and links between the market and global conditions. Hence the model is approximated to the reality, where producers and consumers make their decisions in the world of imperfect competition. For example, the decision-making process of producers is guided not only by the price-cost relationship, but also the potential non-market benefits that they can achieve on account of government intervention.

Partial equilibrium models for agricultural sector

There are two basic reasons for using partial equilibrium models. First, they allow for market forecasting. Secondly, they allow for carrying out simulations. Forecasts derived from equilibrium models are often called projections or baseline scenarios. Projection is the most probable picture of the reality in the light of current knowledge and assumptions regarding exogenous variables. The term *projection* seems to be safer than *forecasting* for institution preparing predictions. When there are some deviation from reality one can suggest that this is due to incorrect assumptions (based on other sources). The term projection is especially used in the case of public institutions, e.g. the National Bank of Poland (NBP), which points out that, in contrast to commercial banks, their actions have a significant impact on the market. To avoid predicting their own actions, the NBP publishes projections rather than forecasts. Another reason for using term projection is the microeconomic nature of forecasts.

The key difference between a forecast and a projection is the nature of the assumptions. If these assumptions are the most probable and not only hypothetical, then the projections may be called forecasts. The long-term projections for the agricultural sector in the case of a number of partial equilibrium models are regularly published, together with a comprehensive rationale and interpretation of the results, in the form of reports called the Agricultural Outlook.

Practically, only three institutions: the USDA (United States Department of Agriculture), FAPRI (Food and Agricultural Policy Research Institute) and the OECD and FAO jointly (Organisation for Economic Co-operation and Development, the Food and Agriculture Organisation) prepare annual reports (Agricultural Outlooks) covering the characteristics of the world agricultural markets and their projections for the period of 8-10 years. Such projections are based on certain assumptions regarding the formation of exogenous variables, such as weather, macroeconomic conditions and assumptions about the development of the agricultural and trade policies.

The baseline scenario may be used as a point of reference for alternative scenarios which assume different path of exogenous variables. Such comparisons are called simulations. For example, the impact assessment of the reform regarding abolition of milk quotas in the agricultural sector is a simulation.

5.2. AGLINK-COSIMO model

5.2.1. Background and model coverage

Model foundation

The most popular partial equilibrium model is the AGLINK-COSIMO model, which is a combination of the AGLINK model, developed since 1992 by the OECD, and the COSIMO model – by the FAO. These two models were combined in 2004, and the 10-year projections published as OECD-FAO Agricultural Outlook provide a visible effect of co-operation between them. This model focuses not only on identifying future trends, but also allows a medium-term assessment of the potential impact of the changes in agricultural and trade policies on agriculture. Analyses of sensitivity to exogenous shocks constitute an important element of the publications.

The AGLINK model has its roots in the MTM static model (Ministerial Trade Mandate). The original premise behind the MTM model was analysis of the effects of reducing protection and trade barriers. The first projection and simulation results were published in 1987. However, due to its static nature, the MTM model did not meet the expectations, so in 1989 the work began on the AGLINK model that would have greater ability to assess a shock response, such as a policy change. The first report (the OECD Agricultural Outlook) prepared on the basis of the new model was released in 1995. Since then, these publications were issued regularly in the first half of the year [Conforti and Londero 2001; Uebayashi 2008].

AGLINK model is a dynamic, recursive partial equilibrium model for agriculture in OECD and selected non-OECD countries and regions. It includes supply, demand and prices of major agricultural commodities produced, consumed and exchanged in countries (regions) represented in the model. The overall structure of the model takes into account the economic character of individual countries, particularly with regard to agricultural policy.

In 2004, COSIMO component (COmmodity SImulation MOdel) was added to the AGLINK model. The general structure of the COSIMO model, developed by FAO experts, was tailored to the structure of the AGLINK model. As regards the behavioural parameters, COSIMO model is a continuation of the work performed under the WFM (World Food Model). Inclusion of the COSIMO module allowed for extending the AGLINK model with more detailed analyses for the non-OECD countries, mainly the developing countries. However, it needs to be kept in mind that not all modules (commodity markets) of the AGLINK and COSIMO models are fully integrated with each other.

In general, the properties of the AGLINK-COSIMO model can be summarised in three points [OECD 2007]:

1. AGLINK-COSIMO is a recursive, dynamic, partial equilibrium model for the most important agricultural commodity markets in the world. Non--agricultural markets are not modelled, and their impact on agriculture is captured in exogenous terms.
2. It is assumed that respective agricultural markets are competitive, which means that the sellers and the buyers do not have the competitive advantages resulting from their monopoly position, and the market prices result from market equilibrium of supply and demand at a global or regional level.
3. It is assumed that agricultural commodities produced and traded in different countries are considered by buyers to be perfect substitutes. This means that importers and consumers do not differentiate products in respect of their countries of origin.

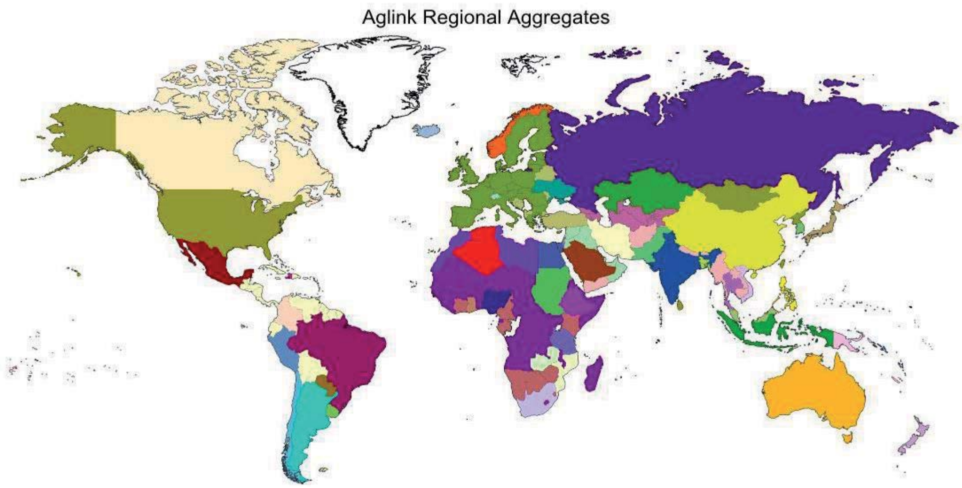
Spatial and commodity coverage

In terms of territory, the AGLINK-COSIMO model covers virtually the entire world (Figure 5.1), broken down by countries and regions. In 2012, it was divided into 42 countries and regions [OECD/FAO 2012]. At the same time, the status of some countries (regions) in the model is exogenous, meaning that their commodity markets are not modelled, but taken as assumptions. Moreover, it should be noted that the European Union is considered as one region (EU-27 aggregate).

Modelling covers most of the major agricultural commodity markets, including: cereals (wheat, feed grains, and rice), oilseeds (broken down into a variety of plants and the production utilization – seeds, fodder or oil), sugar, milk and its products, meat (beef and veal, pork, poultry, lamb), eggs, fish and seafood. Recently, the model for the biofuel component has also been added. One can say that there is an individual commodity market model for each region (or country), but the level of detail in modelling individual markets varies depending on the country or region.

Individual commodity markets are interrelated by substitution and complementary relationships. The demand for food is a function of prices, income and population figures. The demand for fodder is a function of livestock number and livestock prices. The demand for biofuels is, however, conditioned by institutional requirements. Production of various agricultural products is usually a function of productivity, past prices (including relative prices), and the potential benefits resulting from agricultural policy [Uebayashi 2008].

Figure 5.1. Example of spatial aggregation of the AGLINK-COSIMO model



Source: Pérez Domínguez et al. 2012.

It can be assumed that the models for different countries (regions) are, in a sense, independent. However, there is a procedure for aggregation of individual modules of the AGLINK-COSIMO model when creating the baseline projection. In the end, the whole model is optimised to produce the projections. One can also prepare the model for each country individually, treating other variables as exogenous. The very structure of the aggregated model and the method of attaining the equilibrium are not clearly defined, and the lack of current literature does not help in understanding all the relationships.

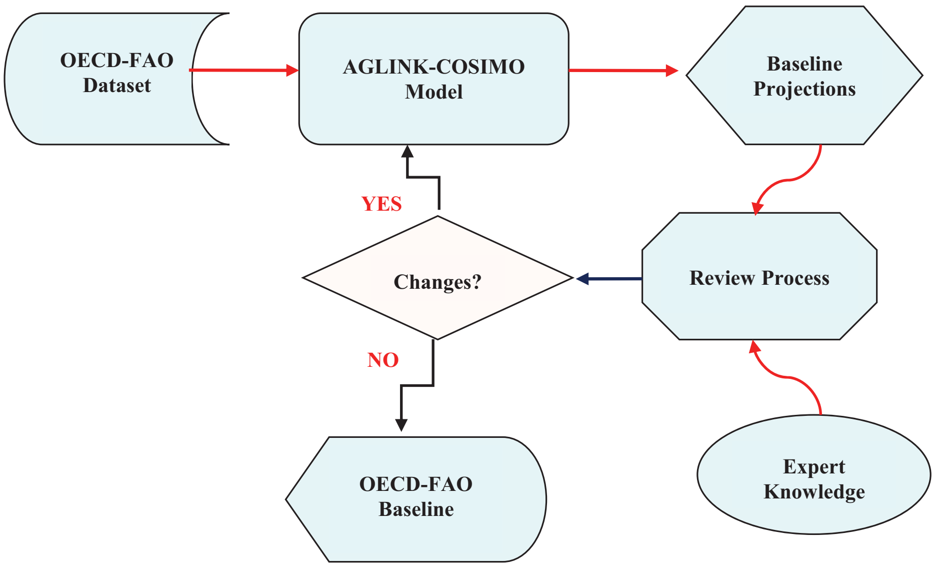
5.2.2. Modeling markets

Baseline process

The AGLINK-COSIMO model is the most well-known tool for producing baseline and alternative scenarios in a 10-year horizon. The baseline process is not automatic, as it is often in case of econometric models, but it contains a great deal of expertise and it is a type of algorithm (Figure 5.2).

Given that AGLINK-COSIMO is a partial equilibrium model, it is necessary to adopt a number of assumptions about exogenous factors, population growth, production technology or economic policy (macroeconomic, trade and agriculture). The detailed presentation of agricultural policies is the strength of the AGLINK-COSIMO model and distinguishes it from other models [Burrell (ed.) 2010]. These issues will be discussed later.

Figure 5.2. The OECD-FAO baseline process



Source: Blanco-Fonseca 2010.

It is necessary to feed into the model a set of assumptions for exogenous macroeconomic variables. There are four core indicators among the macroeconomic variables [OECD 2007]:

1. GDP expressed as an index and constituting a proxy variable informing about changes in consumers' income,
2. consumer expenditure deflators and GDP deflator (to capture real price changes and production costs),
3. exchange rates,
4. crude oil prices.

Projections of these variables are assumed on the basis of the analysis of OECD Economics Department (Economic Outlook), the publication of the World Bank, as well as other available sources. Moreover, average weather conditions and productivity changes, in accordance with the trends observed in the historical period, are assumed for the period covered by projection horizon. Production costs are approximated as cost index which aggregates e.g. the GDP deflator, oil prices, fertilizer prices, exchange rates. Cost indexes vary depending on the type of production [OECD/FAO 2012].

Some analyses are conducted with the use of the stochastic model rather than the deterministic model in recent years. The deterministic model presupposes one set of exogenous variables (one scenario), whereas multiple sets of exogenous variables generated by random samplings, are fed into the model for stochastic experiments. The model is simulated for each set of assumptions and, thus, multiple sets of solutions are obtained. This is then used as a basis for optimization of the formal AGLINK-COSIMO model which gives various results in the form of the distribution function. These results form, in turn, the basis for the assessment of uncertainty of the baseline projections for each commodity market [OECD/FAO 2012].

Relevant projections are based on the balance sheet data for respective commodity markets, which come, mainly, from domestic sources or OECD and FAOSTAT databases. In the case of the OECD module (AGLINK model), some part of the data is obtained via questionnaires distributed at the beginning of the year and concerning the development prospects of individual markets and the evolution of agricultural and trade policies in the period covered by the projections. These data (supplemented by other sources) constitute the starting point for projections of national and regional modules [Conforti, Londero 2001; Blanco-Fonseca 2010].

As for the non-OECD countries, the preliminary projections of the COSIMO model are a combination of results from a formal model and the opinions of FAO experts. Various external sources are taken into account in both cases to supplement the knowledge about the main factors determining the agricultural markets perspective (information about exogenous variables. These assumptions are the basis for the initial calibration of models for each country [Blanco-Fonseca 2010, OECD/FAO 2012].

The next step involves the combination of national and regional models and their optimization, which produces preliminary global baseline projections. These are then compared with assessments obtained from the questionnaires and the initial projections of the COSIMO module. Verification of the results is carried out first by FAO and OECD experts, followed by national (regional) experts of the OECD Working Commodity Groups (made at the beginning of the year). Numerous experts from the field of agricultural policy and national agricultural markets are involved in the preparation of the baseline projection and its evaluation, which has its advantages and disadvantages. On the one hand, the participation of experts from different countries makes it more probable that the distinguishing features of individual countries will be taken into account, on the other, it makes it difficult to consider all, often extreme opinions [Uebayashi 2008, OECD/FAO 2012].

The assumptions are modified in the areas of major differences based on the above opinions. Thus the system for modelling agricultural markets under the AGLINK-COSIMO model allows for cross-compliance of projections that follows from reaching a consensus as regards formation of factors influencing the agricultural market [OECD 2007, OECD/FAO 2012]. A new baseline scenario is generated on the grounds of the formal AGLINK-COSIMO model when changes are introduced. The updated baseline scenario is again subjected to validation and in the absence of any major objections; it constitutes the basis for preparing a preliminary outline of the publication (Agricultural Outlook). It is then discussed by the Senior Management Committee in FAO and by the Working Party on Agricultural Policies and Markets of the Committee for Agriculture in OECD. The report is published after considering the comments.

Equilibrium price formation

The equilibrium market price is determined at the level at which there is a market clearing, which means that global demand is equal to global supply. Reference prices recorded in countries, regions or ports are considered as benchmarks for the world prices. Behavioral equations linking supply and demand with prices are mostly log-linear. The coefficients of these equations reflect partial elasticity and come from different sources. Some coefficients are obtained by estimating equations with econometric methods, others come from the other models (WFM or database of elasticity coefficients in FAPRI model), or are adopted on the basis of economic literature. Due to the dynamic nature of the model, some equations (mainly related to production) take into account the time delays of up to several years, which have an obvious effect on the path to attaining market equilibrium, which resembles a cobweb model of supply and demand [OECD 2007, Uebayashi 2008].

It is assumed, that most countries (regions) are small open economies, and thus that the level of prices quoted there do not differ from the world prices. Domestic prices are then a function of the world prices (converted to the currency of the given country). The differences between the domestic price and the world price result from transport costs, product quality and impact of trade policies (tariffs, taxes, subsidies, etc.) [OECD 2007].

There are different methods to establish the balance of trade (net exports) depending on the status of the country, which, in turn, is determined by agricultural and trade policies. In the extreme case, for the countries with restrictive trade policies, it is possible to make an assumption about exports or imports (e.g. based on import quotas). However, in the absence of trade barriers, the net export is a residual variable of domestic production and consumption.

5.3. FAPRI model

5.3.1. Basic information

Model foundation and dissemination

The FAPRI Institute (Food and Agricultural Policy Research Institute) is a joint venture of the Iowa State University and the University of Missouri, Columbia, formed in 1989. FAPRI-Iowa State University maintains the international modeling structure for grains, oilseeds, livestock, dairy, sugar and the U.S. crop insurance model. FAPRI-Missouri maintains the U.S. modeling structure for grains, oilseeds, livestock, and dairy, along with models for the international cotton sector and the European Union. Over time, the consortium was enlarged with other entities (such as the University of Arkansas and the University of Wisconsin) that are engaged in modelling and preparation of the long-term projections. However, because of budget constraints, FAPRI did not develop a joint report in 2012. Instead a separate Outlook available as the FAPRI-ISU 2012 World Agricultural Outlook was developed by the Iowa State University.

The main beneficiaries of FAPRI studies (and the founders of research) are the Senate Committee on Agriculture, Nutrition and Forestry and the Committee on Agriculture of the U.S. House of Representatives. Apart from them, the studies are used by USDA, other government agencies, farmer organizations etc., as the projections and simulations generated on the basis of the model are publicized free of charge on the FAPRI website. Expandability of the model is, undoubtedly, related to the evaluation of the impact of changes in U.S. agricultural policy (the Farm Bill) on individual markets, as well as the analysis of the impact of changes in the EU agricultural policy (CAP) and the international agreements under GATT/WTO [Meyers *et al.* 2010].

FAPRI approach

The annual reports (FAPRI U.S. and the World Agricultural Outlook) are prepared on the basis of a vast database, the results of modelling and the substantive review process carried out by experts [FAPRI 2011]. The entire report is prepared using an iterative approach which covers modelling elements with a high degree of expertise. Although this process has evolved over the past 25 years, it is generally referred to as the FAPRI approach.

Modelling has its origins in preparing the reports (Outlook) by USDA--ERS. This is due to the fact that at the time when the institute was founded, many FAPRI employees had experience in working at USDA-ERS. The process of generating a report in the case of FAPRI consists more in the use of modelling than in the case of USDA, where the Outlook is based on expert knowledge. However, the role of experts and analysts, whose expertise allows for reflecting the specificity of individual markets as closely as possible, is still very important [Meyers *et al.* 2010].

The process of preparing the report lasts two baseline weeks. It is an iterative process similar to that presented in chapter 5.2. The first step covers individual modelling of U.S. and international commodity markets and exogenous variables. Next, the researchers involved in the modelling of individual markets participate in discussion panels and present their results. The key variables in this case are the prices and the volume of net exports. The panel meetings allow to obtain additional information on related markets, and to appropriately correct the models. The iterative procedure (correction of the model equations and discussions) is continued until all markets attain equilibrium. This produces the baseline scenario, whose initial projections are evaluated by a panel of FAPRI internal experts, representatives of different USDA departments, international organizations, consulting or industry. Substantive comments regarding the feasibility of projections are taken into consideration before the final publication of the report [FAPRI 2011].

Deterministic and stochastic model

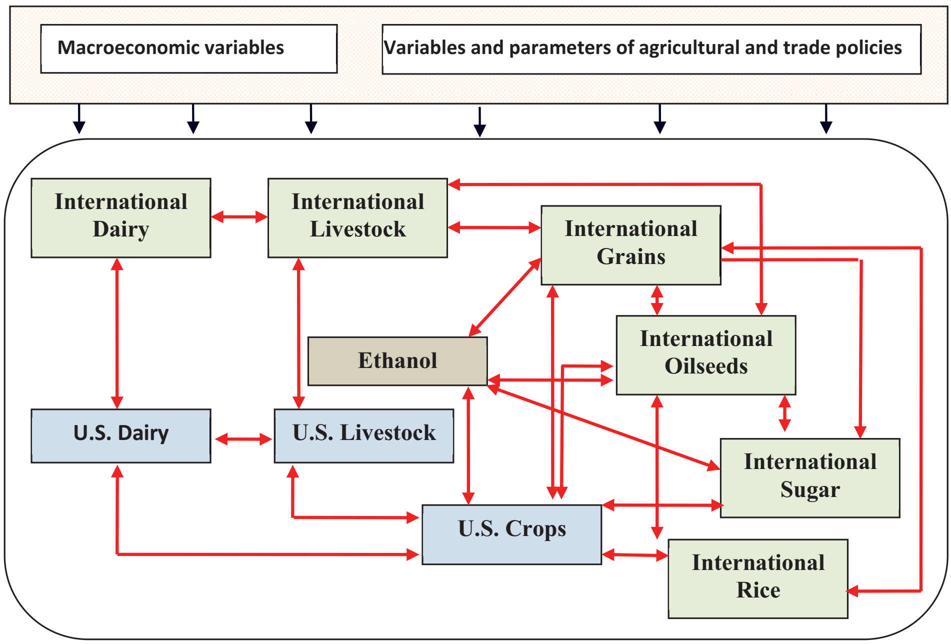
The FAPRI generally specializes in responses to the "what if" question. For many years, the deterministic model, which allowed the analysts obtaining point estimates of individual variables describing different markets, was the main tool used in the analyses. But due to some limitations of this approach, the FAPRI implemented the stochastic model and improves it gradually. This stochastic model is used for analysis of the U.S. agricultural markets and biofuels market. However, projections for the global markets still come from the deterministic model. The stochastic model allows for identifying sources of variation in agricultural markets. More on the advantages of the stochastic modelling and the FAPRI stochastic equations structure can be found in the following studies: [Westhoff *et al.* 2006, Westhoff *et al.* 2008, FAPRI 2011a].

5.3.2. The model structure

Model coverage

The FAPRI model is a set of partial equilibrium models, including models for the U.S. market and international markets in cereals, oilseeds, cotton, rice, sugar, milk and animal products (Figure 5.3). Recently, due to growing significance of demand for agricultural production from fuel industry the biofuel component (market) has been developed. The FAPRI model includes more than three thousand equations explaining the behavior of variables that determine the agricultural market (U.S. and global) and its individual components.

Figure 5.3. Interactions in FAPRI model



Source: based on [Meyers et al. 2010].

Models of individual markets are dynamic models representing at present 26 most important, from an economic viewpoint, countries (regions) and the remainder is represented as an aggregate – Rest of the World. It may be noted that similar models for U.S. and other global markets are characterized by a relatively high diversity, given the detailed modelling of different aspects of reality. The main focus is on the U.S. markets, which significantly outweigh the other markets in terms of the details regarding agricultural policy. The European Union is included in the model as the aggregate EU-27. By 2004 Poland was taken into account in the projections individually, and since then it has been considered as part of the EU-27 [Hamulczuk 2011].

Model links

The structure of the model consists of three components: the exogenous component, the component covering the U.S. markets and the international component (Figure 5.3). On the one hand, models for each commodity market are partially independent of each other, because they are managed by different people and institutions within the consortium. On the other, individual markets are correlated via demand or supply side, as well as prices. The degree of correlation increases along with the number of consultation meetings within the consortium. The directions of these relations are indicated by arrows in Figure 5.3. For example, models (markets) of milk and animal products allow for specifying the demand for feed grain, and the macroeconomic variables determine consumer demand. Other vegetable markets (cereals, oilseeds, rice and sugar) provide information determining the relative profitability indicators of individual plants, and thus are the basis for land allocation for crops.

Data for individual countries, regions or the world are in the form of balance sheets: Initial stocks + Production + Imports = Ending stocks + Domestic use + Exports and they come from the USDA-FAS databases and other sources. Production is derived from crop area or livestock population and from production yields (e.g. yield per hectare, yield per cow or slaughter weight). Domestic utilization is divided into food consumption, feed use and industrial use. In order to attain equilibrium on a given market, one of the variables (usually export or import) has to act as a residual variable. The domestic (regional) prices usually are modelled as a function of the world prices (market clearing prices) using price transmission equations [FAPRI 2011].

Exogenous variables

Models of individual markets are also linked to variables of macroeconomic, sectoral and trade policy acting as exogenous variables. Agricultural policy instruments include: export subsidies, tariffs and export quotas, intervention prices and rates of compulsory set-aside. Recently, the mandatory biofuels blending indicators have become an important element of the policy influencing the demand for some crops. The agricultural and trade policies in the model affect the decisions of market agents on the demand or supply side. The baseline projections assume that there is no change in policies or that the occurring changes are in accordance with the accepted and known agreements. The FAPRI model captures the policies with different degree of detail. The policies for the United States are covered in more detail than in the case of other countries, including the EU [Blanco-Fonseca 2010].

Macroeconomic variables, such as GDP growth, population figures, exchange rates or crude oil prices are among the exogenous variables. The International Monetary Fund (IMF) is the source of such macroeconomic data. These variables are important when it comes to long-term projections, since the GDP and population growth determine the changes in consumer demand in the world and in respective countries. The exchange rates are also important, as they determine the relative profitability of production in spatial (international) terms and the directions of foreign trade. In the case of the FAPRI model, the projections of macroeconomic variables are taken mostly from the IHS Global Insight or the US Bureau of the Census [FAPRI 2011].

5.4. AGMEMOD model

5.4.1. The model origin and structure

Model background

The AGMEMOD[59] is an econometric, dynamic, multi-product, partial-equilibrium modelling system which was built in the aim to undertake a model-based economic analysis of the potential impact of policy or other changes in the agri-food sector of each EU Member State and the EU as a whole. The AGMEMOD model was developed under the 5th and 6th EU Framework Project (FP5 and FP7) and constituted a more elaborate version of the GOLD model[60] which was equivalent of FAPRI model for several major EU countries. The FAPRI-GOLD model was tailored, under the FP5 and FP6, to the specificity of other European countries. The countries' models are based on a structure common for all the EU countries and with common procedures for data collection, estimation and validation, but take into account local conditions and rely on the local experts' knowledge. Consequently, the AGMEMOD model works as a system of aggregated local models and is able to produce forecasts and scenario analyses of various policy and external conditions' changes for the Member States separately as well as for the entire EU [Westhoff 2001, Hanrahan 2001, Donnellan *et al.* 2002; Chantreuil, Hanrahan 2007]. The extending of the model to the following countries: Turkey, Ukraine, Russia, Kazakhstan or the Balkan countries is an ongoing process.

Model structure

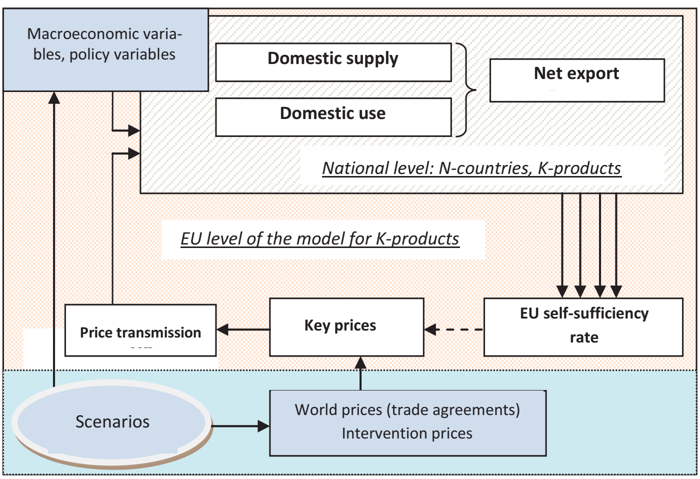
Each model of a country consists of a set of sub-models of the main agricultural products: grains, oilseeds and the derived products, industrial plants, milk and dairy products, livestock and meat as well as some other, of lesser importance and more locally grown products. It should be noted that not all commodity markets are included in the national models. For each product in each

[59] The acronym is derived from the words: AGriculture, Macro, Economic, MODelling.
[60] GOLD – Grains, Oilseeds, Livestock, Dairy.

country the respective domestic prices (market-clearing prices) are modelled. The general structure of the model is presented in Figure 5.4.

The variables entering in each sub-model represent consecutive positions in the balance sheet of each market. The statistical data used come from Eurostat, national sources or are the estimates by the country experts. On the supply side the beginning stocks, production and imports are being considered and on the demand side the domestic use, exports and ending stock are modelled. Each market is modelled by a set of behavioral and identity equations. The most important variable among the above is the production, whose level is a function of price expectations (based on past prices), non-market benefits (agricultural policy) and the costs of production. Price relations are the most important variables determining the demand of a product in question and substitute goods for the current year.

Figure 5.4. Structure of the AGMEMOD model



Source: Chantreuil et al. (ed.) [2012].

Exogenous variables and policy implementation

The AGMEMOD model contains endogenous and exogenous variables. The endogenous variables are mostly prices and the variables determining the supply and demand in the market of each product in every EU country. The exogenous variables include a set of variables describing the general macroeconomic conditions influencing the agricultural market (GDP, inflation, exchange rates, population), world agricultural prices as well as CAP and trade regulations.

The manner of agricultural policy implementation is an important feature of the model, which includes the typical CAP instruments, such as quotas, direct payments and intervention prices. In addition, the model covers variables of agricultural policy conditioned by international agreements under GATT/WTO. The assumptions about the impact of direct payments on the volume of agricultural production are different under each model. More about the problems associated with the implementation of agricultural policies can be found in the following papers: Binfield *et al.* [2005], Conforti [2001] or Donnellan *et al.* [2002].

It is necessary to make certain assumptions about exogenous variables to determine the future direction of development on the markets (baseline or other scenario). Macroeconomic variables are adopted on the basis of projections of the European Commission, the OECD and the national governments. It is also assumed that the weather conditions remain on average level. World prices (if they are recognized as exogenous) are linked to projections of OECD-FAO, FAPRI and USDA [Chantreuil, Hanrahan 2007].

5.4.2. The modeling process

Market equilibrium

The equilibrium is reached in the AGMEMOD model in each market of each country independently. The characteristic feature of the model is that the price does not serve as a variable which would lead to the equality between the supply and demand in the separate market at a given moment of time, but is exogenous for the supply and demand variables at a given moment of time. Therefore one of the positions of the trade balance sheet, in most cases imports or exports, is treated as a closing variable.

The equilibrium in each market is reached also at the level of the whole EU. This implies that the EU net export variable is used as the closing variable at the EU level. The necessary condition for the model to be solved is that the equality between supply and demand in each domestic market has to be maintained.

Interactions in the AGMEMOD model occur at two levels: spatial (between countries) and product (between the markets of each product). Interactions between respective product markets in one country are attained through substitution or complementarity of production or consumption. Such relationships between the products can be exemplified by land allocation between different types of plant production or the use of plants for feed and industrial purposes, which is determined by the price level. The plant and animal markets are linked through variables representing the demand for feed for livestock [Chantreuil *et al.* (eds.) 2012].

Price formation

The key element in equilibrium models is the selection of a method for modelling and forecasting the prices of individual products. Price equation is the most important way of linking the domestic market to markets of other countries and the global market. Equilibrium models employ two ways of modelling world prices: the world prices of products are an exogenous variable (the small open economy assumption) or the world prices of products are modelled endogenously, i.e. they are the endogenous variable in the model [Tongeren *et al.* 2001]. In the AGMEMOD model the world prices are the external variable and are not subject to modelling (exogenous). Hence, the projections should make certain assumptions about their development.

The price transmission between the global market and the individual Member States used in the AGMEMOD model is a two-step process. This means that the world prices influence the EU representative prices (key prices), while the key prices drives the domestic prices. The key price is the price of the product in the country which is its most important producer in the EU and provides a benchmark for other countries [Chantreuil *et al.* 2008].

The key prices depend on the respective world prices (and their forecasts), the variables expressing the CAP, WTO agreements and other variables, including self-sufficiency in the EU. By adopting this method of modelling, it is expected that in the absence of trade restrictions, the key price projections will not depart substantially from the world price projections. Greater differences may exist in the case of commodities (markets) subject to considerable regulations under the CAP or the trade policy.

The second type of model equations are the ones used for modelling domestic prices. In most cases, the domestic price of a given commodity depends on a simultaneous development of the key price, lagged domestic and EU (or key country) self-sufficiency rates and other variables. The inclusion of self-sufficiency ratio enables the path of domestic price formation to deviate from the path of key or international price formation. Depending on whether a country is a net exporter or net importer, the domestic prices may be lower or higher than the key prices.

Recently, the work on endogenisation of the world prices have been ongoing. The endogenisation of the rest of the world could be done by the comparison of the EU net-export supply with the potential rest of the world demand. As a result, the world market price can be specified as a function of EU net-exports and could also be influenced by demand shifters such as the world GDP and population as well as by policy variables. With endogenised world market prices, the small country assumption is no longer valid. The world prices are estimated using seemingly unrelated regression models (SUR) [Chantreuil *et al.* 2008; Listorti, Esposti 2008; Banse *et al.* 2012].

Estimation, validation and utilization of the model

The behavioral equations of the model are mostly individually estimated with econometric techniques (generalized least squares). However, in the situations when relatively short time series were accessible, the quality of data was unsatisfactory, important structural breaks were observed or values of estimated parameters were inconsistent with the economic theory (such as the positive price elasticity of demand or the negative reaction of a domestic price to the key price), the calibration techniques were used [AGMEMOD 2005]. This technique was applied especially in the case of New Member States allowing for reliable parameters estimation and consequently for a long-run forecasting even though only relatively short time series were available.

The validation played an important role in the construction of the model. Apart from general econometric tests on parameters and residuals, the baseline results were analysed by national experts from the point of view of their feasibility. The experts assess the results generated by the model. A negative assessment leads to re-estimation or re-calibration of models (equations) with a view to obtaining more rational results. An important part of the model validation is the correct response to external shocks that guarantees the simulation abilities of a model [Chantreuil *et al.* (ed.) 2012].

The projection on agricultural markets covers medium- and long-term periods. Potential scenarios express the possible range of changes in various variants of the CAP and other exogenous variables. So far, the forecasts generated by the AGMEMOD model have not been published regularly in a form similar to that of the FAPRI or AGLINK-COSIMO models. However, there are discussions on preparing Outlooks for the individual countries and the European Union as a whole in the future.

5.5. Ex-post projections of the world agricultural commodity prices – the Aglink-Cosimo model

The use of projections made on the basis of partial equilibrium models for economic decision-making requires that these projections do not deviate from reality. Therefore, the objective of this section is to present the ex-post world price forecasts against the actual data. We will concentrate on price projections however projections of production, consumption and trade are available. In particular, the focus is on assessing to what degree long-term projections formulated using partial equilibrium models, supported by substantive knowledge of experts, can be a reliable source of market information.

The predictions obtained on the basis of the Aglink-Cosimo model and published in annual Agricultural Outlooks issued by the OECD-FAO will be presented. Choosing Aglink-Cosimo model stems from two factors. Firstly, the OECD-FAO publications are the most recognizable projections in the environment of agricultural economists. Secondly, a high correlation between the projections (forecasts errors as well) of different partial equilibrium models is noticeable, hence we will focus our attention on that model.

5.5.1. Price projections of plant origin commodities

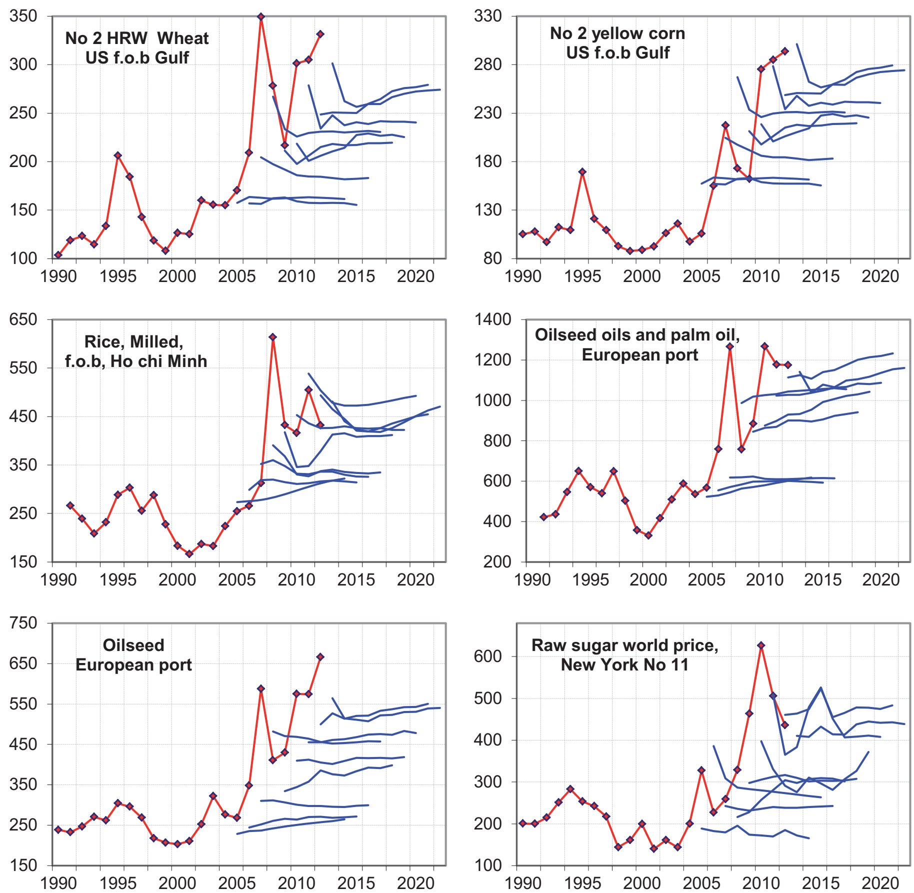
The medium- and long-term world price projections for the selected commodities of plant origin are presented in Figure 5.5. We present only one world price for each product even though for some products two or three world price projections are published. The actual prices in the subsequent years are indicated by a red line whereas price projections are marked by a solid blue line. The evolution of the prices of these commodities in the past – regardless of geographical location and price quotations – shows a lot of similarities. The price rises can be seen in 1994-1996 and since 2006-2007 to the end of time series. All these price increases were due to a low level of stocks in the world.

The increase of commodity prices in the first period was supply driven whereas a rise of prices in recent years is mostly demand driven. The most important factors affecting growth in demand in recent years, according to the literature include: economic growth in developing countries, changes in food consumption patterns towards products of animal origin in developing countries (demand for feeds), depreciation of the U.S. dollar, speculative influences related to the interests of financial investors, an increase in the consumption of agricultural commodities for biofuel production, etc. [Abbot *et al.* 2011; IFPRI 2011; Nazlioglu and Soytas 2012].

The comparison of ex-post projections with prices quoted makes us look critically at the possibility of agricultural prices forecasting with the use of large partial equilibrium models. The projection of world commodity prices in the analysed period cannot be considered accurate. The price projections published assumed no sudden increase in prices after 2006. Most of projections underestimate the future level of commodity prices of plant origin. In many cases forecast errors exceed 50%.

The low accuracy of the projection is caused by the wrong assumption about the factors mentioned earlier. Among them, in the authors' opinion, an underestimation of the impact of biofuels on the increase in demand for grains, rapeseed, or sugar cane is the most important. In recent years, due to the statutorily defined minimum thresholds for the use of biofuels in the industry, their production increased dramatically. An increase in the production of biofuels is the consequence of policies of developed countries supporting the development of fuels from renewable sources [Gilbert and Morgan 2010]. Presently, almost 90% of global biofuel production comes from the USA, the EU and Brazil so biofuel policies in these countries have the highest impact on world prices.

Figure 5.5. Baseline world price projection of the selected plant origin products from AGLINK-COSIMO model [USD per ton]



Source: author's own elaboration based on OECD-FAO Agricultural Outlooks.

In the biofuel era more direct interdependence between energy and agricultural and food prices is observed. According to Tyner [2010] the correlation between corn and oil prices in the USA in years 1988-2005 was -0.26. The situation changed together with biofuel market development. As agricultural commodities are increasingly being used as a raw material for biofuels, the linkage between energy and agricultural markets begins to intensify, although the nature and strength of this relation is not clear.

Biofuel policies shifted the world agricultural prices upward. It has to be emphasized that under the higher price regime the stocks are lower and prices become more sensitive to shocks in supply. This is where the additional problem arises, namely the increase in volatility of world agricultural prices. Tyner *et al.* [2012] stress that under the non-flexible biofuel policy (demand) the drought may highly impact the corn and ethanol prices. The strong drought under high blending levels might result even in 60% increase of the US corn prices. The projections presented in Figure 5.5 are computed for the typical, average weather conditions.

The estimates of the impact of biofuel production vary among the researchers. Davies [2012] estimates using Aglink-Cosimo that the removal of the biofuel supports by the EU and the US would lead in medium term to respectively 80% and 90% reduction of bioethanol production. This might result in significant decrease of the world prices of such commodities as coarse grains, corn, oilseeds, vegetable oils and wheat. The removal of the EU biofuel policies would lead to the decrease of world prices by 2-5% whereas the abolition the US policies supporting biofuel production might result in the decrease of world prices by 5-14% [Davies 2012].

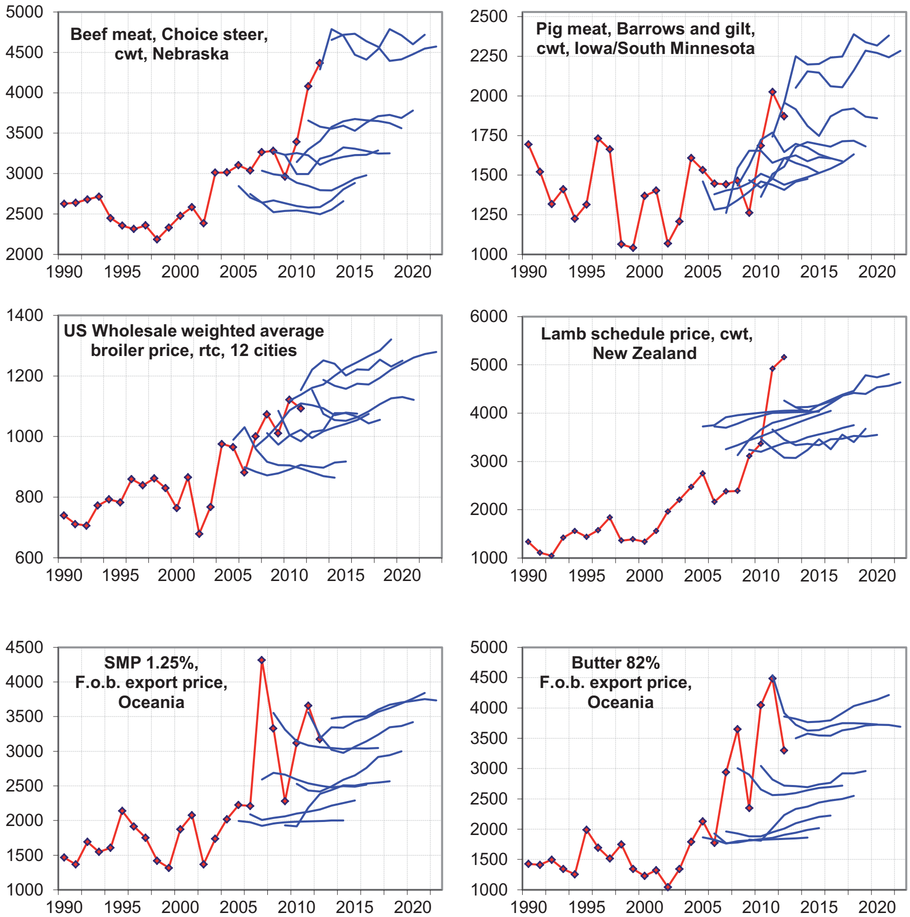
5.5.2. Price projections of animal origin commodities

In this section the prices and the ex-post projection of prices of animal products are presented. Figure 5.6 shows the above characteristics for four kinds of meat and two major dairy products.

The price variability of animal origin products is significantly lower than the variability of the prices of cereals, rice and oilseeds. For this reason, the accuracy of the meat price projections is higher than the price projections of plant products. Most importantly, the path of price changes enclosed in the projections is in most cases consistent with the direction of actual price.

Very frequently the expected price increase was justified by the demand growth in developing countries. It was an assumption, which contributed to obtain correct direction of projections. However, more often the projections were under- than overestimated. The reason for this was the underestimation of price projections for crops underlying the production of feeds.

Figure 5.6. Baseline world price projection of the selected meat and dairy products from AGLINK-COSIMO model [USD per ton]



The projections of milk products in 2005-2007 were done for European prices: SMP - F.o.b. export price, non-fat dry milk, extra grade, Northern Europe; Butter - F.o.b. export price, 82% butterfat, northern Europe. However, the level of prices in EU and Oceania is comparable.

Source: author's own elaboration based on OECD-FAO Agricultural Outlooks.

Meat prices are to some extent related to the prices of plant products, particularly cereals and oilseeds. In the periods of high prices of grains and oilseeds, the profitability of livestock production decreases, so producers reduce the number of animals kept or decide to withdraw from the market. The consequence of reduced supply is the increase in the price of meat. For this reason, almost all factors responsible for the increase in prices of grains and oilseed products have their impact on the price of meat. Consequently, we can conclude that biofuel policy is responsible for underestimation of meat price projections.

It should be underlined that the reactions of producers to changes in the profitability are much faster in the production of poultry and pork meat than beef in particular, due to differences in the length of the production cycle. There were also some supply shocks. Since 2004, prices have increased by reducing the supply of beef arising from the BSE epidemic and reduction of the poultry trade, which was a consequence of the emergence of cases of avian influenza in 2003 and 2007. Consumer reaction on above turmoil and significant increase in production in developing countries was assumed when baseline scenario was developed.

It should also be noted that the various types of meat can be regarded to some extent as a good substitute. Religious taboos forbidding the consumption of certain types of meat, on a global scale do not significantly affect the substitutability.

In the case of milk production and trade we have to bear in mind that the raw milk is not the subject of international trade and price quotation. The main components of raw milk are fats and proteins. The most important products subject to trade is skim milk powder (SMP) and butter (Figure 5.6). The SMP is a protein based product whereas for the production of butter fat is used. The price paid to a farmer is a function of milk product prices.

The largest exporters of skimmed milk powder are New Zealand, the USA, Australia and the EU countries. Usually prices quoted in the region of Oceania and the EU are regarded as world prices. The direction of price changes in both markets is similar, but the price of milk recorded in Europe is slightly higher.

According to the OECD-FAO reports the increase in the milk product prices was expected, which was justified by the increase in demand for dairy products and the increase in the cost of production in recent years. The increase in prices is associated with the steady growth of production in developing countries, particularly in China and India. However, Figure 5.6 indicates that most projections were underestimated. Like in the case of the projection of other agricultural commodities prices, it was not possible to predict the sudden increase in the level of prices in 2007-2008, and equally rapid decline in prices in the following year. It seems that some kind of surprise was the scale of growth in demand for dairy products in recent years in developing countries.

In can be concluded that the accuracy of projections of world prices based on partial equilibrium models is not high. These models work much better in predicting the global production and consumption of particular goods. But even here, along with a decrease in aggregation of variables (to national level) errors increase.

Medium- and long-term world price projections should have to be evaluated not only in terms of their accuracy. Forecasts (projections) can warn against unfavorable events so price projections might be regarded as warn-signal tool for policy makers. Thus, the projections probably led to the activation of agricultural policy.

A strong pressure on state policy appeared to overcome the effects of low prices or take action to reverse this trend when price projections indicated downtrend. For such a measure should be considered biofuel policies which led to alternative use of crop production for energy and fuel production. On the other hand, recently available projections assuming sustained high prices of agricultural products may cause a reduction in pressure through a policy on price increases. Thus, expectations about the high level of prices do not necessarily have to be fulfilled.

5.6. Domestic prices modelling and forecasting on the basis of world price projections

One may wonder if the medium- and long-range projections of world prices might be the basis of the decision-making process for market agents. The accuracy of such projections is not high as it was described in section 5.5. However, the number of sources of price projections is limited. Therefore, these projections are worth considering in the decision making process. One can take into account also own expectations by adjusting the projections if the assumptions underlying price projections are not in line with own expectations.

The reports (Agricultural Outlooks) include usually projected world prices of agricultural commodities. The projection of market prices is what is lacking for most countries. In this case, one try to estimate domestic price development with the use of world price projections. The simple conversion of the world prices for the domestic price is not sufficient in many cases, due to deviation of domestic prices from world prices as a consequence of various factors. This section presents also the empirical issues dealing with medium- and long-term forecasting of agricultural commodity prices for countries that are not present in the partial equilibrium model.

5.6.1. Price formation

The relationships between domestic and world prices are indisputable, both in short and long term. Thus it has to be assumed that domestic prices are largely a function of prices on the international markets. Additionally, the level of domestic prices is influenced by exchange rates, agricultural policy and local supply and demand conditions.

Basic form

It is necessary to assess the impact of individual factors influencing domestic prices when forecasting is based on quantitative models (also in informal models). The baseline model may be the form of the model where price changes in Poland are the function of world prices, exchange rates and deviations from the law of one price[61]. This equation for the good *j* can be given in the log-linear form:

$$\ln P_t^K = \beta_0 + \beta_1 \ln E_t + \beta_2 \ln P_t^W + \varepsilon_t, \qquad (5.4)$$

where: P_t^K – domestic prices of the commodity denominated in national currency in time *t*, P_t^W – world prices of the commodity denominated in another currency (herein USD), E_t – exchange rate (e.g. PLN/USD), $\beta_0, \beta_1, \beta_2$ – model parameters, ε_t – random component.

At this point the basic questions are as follows: what should be the magnitude of the estimated parameters, to which extent the estimates are consistent with the economic theory and market conditions, and whether the form of the model presented above is sufficient, or it should rather be supplemented with other variables. The equation (5.4) can be estimated using the classical (or generalized) method of least squares, or it can be calibrated based on expert estimates. In both cases, the values of coefficients β_1, β_2 should be positive and express a percentage response of domestic prices to changes in the world prices (in USD) and changes in the exchange rate. Thus, if the markets are highly integrated, these values should be close to one, if not – they should be close to zero.

The future values of explanatory variables have to be assumed. The world prices can be assumed on the basis of the available projections of FAPRI, AGLINK-COSIMO and USDA models. Perhaps it would be a good idea to average the projections derived from different equilibrium models. Another solution would be the adjustment of the world prices by experts at their own prefer-

[61] Alternative specification is that a domestic price is a function of world price expressed in domestic currency. Thus we have assumed that the elasticity of domestic prices in respect of changes in currency is 1.

ence. It is also crucial to adopt relevant level of the exchange rate. One can use the forecasts of changes in exchange rates according to the assumption included in Agricultural Outlooks or adopt the average value resulting from various sources. Due to the great difficulty in forecasting exchange rates, the practitioners often resort to assuming the exchange rate at the last quoted level. Another possibility is to assume exchange rate at the average level in the last few years.

Additional deterministic variables

The relationship of the domestic agricultural prices to the world prices changes quite often. The changes in the conditions of trade exchange highly influence the estimates. In certain periods the responses of domestic prices to changes in world prices might be low (under more restrictive trade policy), and at present this influence seems to be greater. It can be stressed that the estimated parameters (eq. 5.4) show always the average effect. However, from the prognostic perspective, it is better if the estimated parameters correspond to the current and forthcoming situation. It is better to calibrate them or modify the equation 5.4 by including additional deterministic variables in the case of considerable differences between the econometric estimates and the expected values of parameters.

The model should therefore include in some way the convergence of domestic prices to global market prices, or the EU prices. The effect of changes in market conditions, especially trade policy, can be captured using artificial variables. If the change was sudden (e.g. beef prices at the moment of Poland's accession to the EU), it would be simply necessary to include a binary variable that contained zeroes until the moment of accession to the EU, and equaled 1 from that moment onwards. By this method, we can get a better estimation of parameters and improve the quality of the model.

The model 5.4 can be extended by variables *ln(t)* or *1/t* if the process of converging is slow. In both cases, it should be rather assumed that the time variable *t* does not start with value 1, but with the value of e.g. 3, 5 or 8. This can be done with the use of trial and error method by analyzing the obtained estimates and statistics showing the adjustment of the model. It might be also possible that the model will contain both types of convergence on the market.

It is possible to include a larger number of variables. Other extensions of the model (eq. 5.4) could be linked to a different structure of delays. For example, it can be assumed that the domestic prices depend on the world prices of the current and previous year. It should be, however, remembered that excessive expansion of the model can take the form of data-mining. In the case of time series containing several observations it is difficult to estimate complex models.

Calibration and additional variables

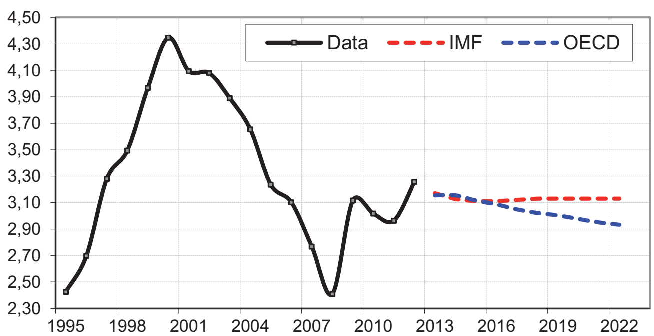
One can perform the calibration process if the estimates are unsatisfactory. For example, let us assume that the elasticity of Polish prices to changes in world prices is 0.9 and the elasticity of Polish prices to the change in the PLN/USD rate amounts to 0.8. This means that the estimates of parameters amount to $\beta_1 = 0.8$ and $\beta_2 = 0.9$. Then the equation (5.4) is transformed into $(\ln P_t^K - 0.8 \ln E_t - 0.9 \ln P_t^W) = \beta_0 + \varepsilon_t$. At this point, it is possible to make an econometric estimation for the adjusted time series. Depending on what regularities are represented by the transformed time series: constant, trend or structural changes, different set of explanatory variables can be considered for modelling the calculated differences.

Models estimated or derived as a result of calibration can be assessed using standard measures (fitting data, significance of parameters, distribution of residuals). However, the reality of the estimated parameters showing the strength of world price transmission and foreign exchange impulses has to be the basic criterion for the acceptability of the model.

5.6.2. Forecasts of selected agricultural commodity prices in Poland

We tried to construct the econometric model which allowed for medium- and long-term prediction of Polish agricultural prices of selected commodities with the use of projection from Aglink-Cosimo model. A few commodities prices were chosen to show empirical solution for forecasting problems. The models were constructed according to procedure explained in the section 5.6.1 for period of 1996-2012.

Figure 5.7. PLN/USD exchange rate assumptions



Source: author's own elaboration based on CSO data and IMF and OECD projections.

To calculate forecasts till 2022 the future world agricultural prices were assumed according to the projections from OECD-FAO Agricultural Outlook 2013 (Fig. 5.5-5.6). The PLN/USD exchange rates were taken from the International Monetary Fund database (IMF, World Economic Outlook Database, April 2013). The value of PLN/USD exchange rates according to IMF and the alternative according to OECD (Economic Outlook No 93, June 2013) is shown in the Figure 5.7. As it can be seen the volatility of exchange rates is extremely high so it is hard to predict their future changes. Therefore, the projections are often close to the last observed level.

Wheat prices

Wheat is the most important cereal produced and consumed in Poland. The analysis of long-term trends in the market of cereals indicates that although there are a number of barriers in international trade, especially in the early years of the analysed period, the prices on the Polish market are highly dependent on the situation of supply and demand on the world market. The wheat price level in Poland is close to the level of world prices.

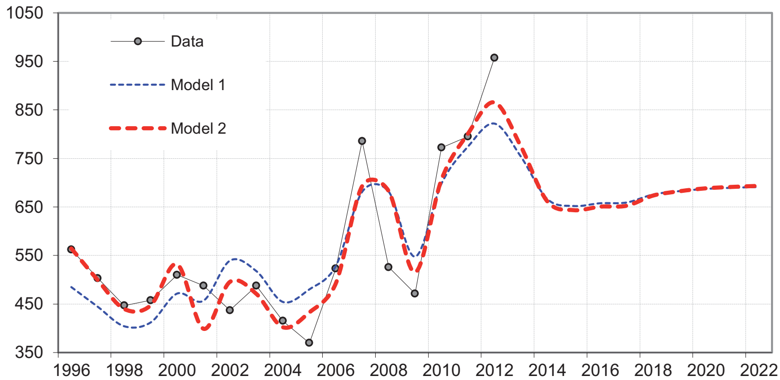
Table 5.1. Estimated models for Polish wheat prices (logs of data)

Parameter	Model 1		Model 2	
	Coefficient	P-value	Coefficient	P-value
constant	1.104	0.427	-0.143	0.903
log_PLN/USD	0.806	0.000	0.960	0.014
log_US_Wheat	0.797	0.079	1.040	0.000
LS2001	-	-	-0.249	0.010
Statistics	R-sq. adj. 65.48; St. error 0.155; Durbin-Watson 1.407		R-sq. adj. 78.04; St. error 0.124; Durbin-Watson 2.583	

Source: author's own elaboration based on CSO and OECD-FAO data.

Two models (based on logs data in season) were estimated (table 5.1) with the use of OLS method. In the first one, the elasticity of Polish wheat prices in respect of exchange rate and world prices is around 0.80 (coefficient is treated as elasticity in log based model). This value is acceptable, however the second model was estimated due to moderate properties of residuals (D-W statistic). In the second model the additional variable was included (LS2001) to catch an increase of wheat prices in Poland after signing of the pre-accession agreement. In the second model the elasticity coefficients are close to 1. This additional variable has improved model properties.

Figure 5.8. Forecast for Polish wheat prices (PLN/ton) – models from Table 5.1



Source: author's own elaboration based on CSO and OECD-FAO data and projections from IMF and OECD-FAO 2013.

Theoretical values and forecasts for wheat prices in Poland derived from both models are shown in Figure 5.8. One can notice that although there are significant differences between models in the estimation period, the forecasts are very close to each other. In the following years it is anticipated that the price of wheat on the domestic market will decrease to the level of around 650-700 PLN per ton.

Live pig prices

Polish pig and pork prices are linked with world prices as strong as it is in the case of grain market (table 5.2). The relatively low elasticity of domestic prices to changes in world prices and changes in the exchange rate may result from higher trade barriers and the fact that intra-EU trade prevails. Since the trend in residuals was observed, we use the model with additional variables. The first model was extended by trend variable (1/t, where *t* starts from 6) to capture convergence or divergence of domestic prices to world prices. The justification for this procedure is an observed tendency in the price ratio between prices in Poland and in other countries.

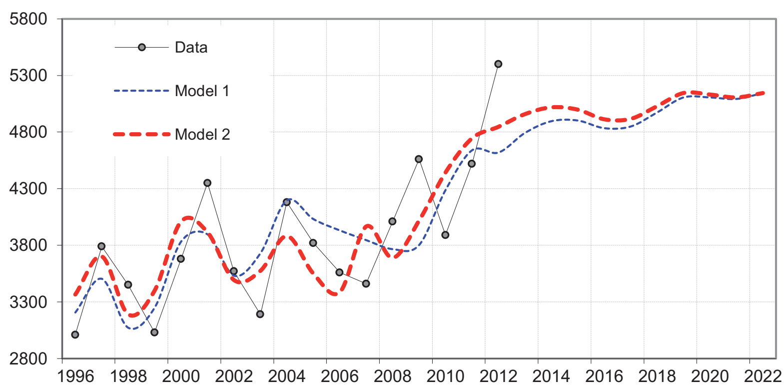
The level shift dummy starting in 2007 was added instead time variable in the second model. It might work even better because since 2007 we observe a structural change in the Polish pig meat market. The livestock number has dropped over 30% and the prices have increased as self-sufficiency in meat production has deteriorated. It resulted from low concentration and the efficiency of production. The estimated increase in prices was 22.5% (Table 5.2) and we got more reasonable values of other elasticity coefficients. The second model performs better in comparison to the first one in terms of fitting and parameter significance[62].

Table 5.2. Estimated models for Polish live pig prices (logs of data)

Parameter	Model 1		Model 2	
	Coefficient	P-value	Coefficient	P-value
constant	5.052	0.003	4.479	0.004
log_PLN/USD (t)	0.255	0.221	0.570	0.024
log_US_Pig (t)	0.424	0.025	0.413	0.024
1/t (5)	-2.343	0.012	-	-
LS2007	-	-	0.225	0.007
Statistics	R-sq. adj. 47.58; St. error 0.112; Durbin-Watson 1.87		R-sq. adj. 51.23; St. error 0.108; Durbin-Watson 2.261	

Source: author's own elaboration based on CSO and OECD-FAO data.

Figure 5.9. Forecast for Polish live pig prices (PLN/ton) - model from Table 5.2



Source: author's own elaboration based on CSO and OECD-FAO data and projections from IMF and OECD-FAO 2013.

The forecasts of live pig prices in Poland under the assumption of world pig prices (Fig. 5.6) and exchange rates (Fig. 5.7) are in Figure 5.9. The forecasts from first the model are slightly higher than from the second one. However, all assume that the prices will converge to the long-term trend. This is an analogous path similarly to the case of wheat prices forecasts (Fig. 5.8).

[62] As the pig prices in Poland are time lagged as compared with the US prices (turning point of estimates are slightly before real data, Fig. 5.9) another concept may assume inclusion of current and past independent variables. However, in the case of short series it may lead to data mining.

SMP wholesale price

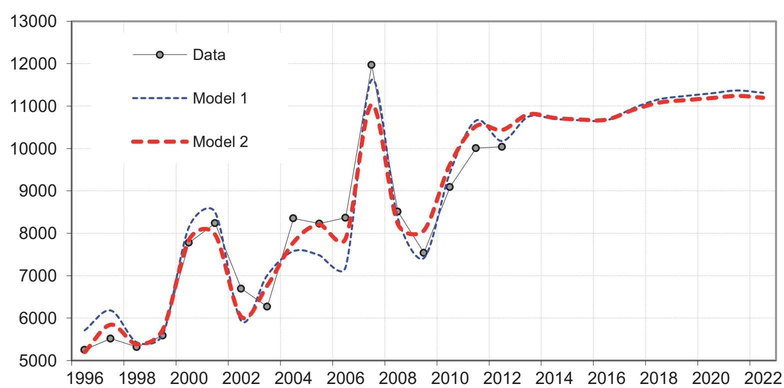
Polish dairy market is strongly linked to world market due to the fact that high share of production is exported. One of the products being exported is a skimmed milk powder (SMP). Table 5.3 and Figure 5.10 contain models for Polish SMP prices as well as their forecasts.

Table 5.3. Estimated models for Polish SMP wholesale prices (log of data)

Parameter	Model 1		Model 2	
	Coefficient	P-value	Coefficient	P-value
constant	1.456	0.085	2.703	0.002
log_PLN/USD	0.784	0.000	0.894	0.000
log_Ocenia_SMP	0.849	0.000	0.657	0.000
UE2004	-	-	0.196	0.004
Statistics	R-sq. adj. 88.17; St. error 0.084; Durbin-Watson 1.580		R-sq. adj. 93.48; St. error 0.062; Durbin-Watson 1.774	

Source: author's own elaboration based on CSO and OECD-FAO data.

Figure 5.10. Forecast for Polish SMP wholesale prices (PLN/ton) - model from Table 5.3



Source: author's own elaboration based on CSO and OECD-FAO data and projections from IMF and OECD-FAO 2013.

Estimated coefficients for world prices as well as for exchange rate (Table 5.3) confirm high integration of Polish market with the world one. Amendment of the first model by inclusion of level shift dummy reflecting joining the EU market improved the model in term of fitting and residual distribution. The Polish SMP prices have increased significantly after the accession to the European Union (19.6%).

There are no significant differences in forecast produced from both models (Fig. 5.10). Calculated forecasts indicate high SMP prices in Poland in the future. The rationale for that path of forecasted prices is rising demand in developing countries.

Such forecasts (concerning all prices) do not take into account short term variability caused by weather conditions, diseases, speculations or other random shocks. It seems that model specification plays less crucial role (forecasts from competitive models are close to each other) than assumptions about explanatory variables. As it was said before, the market agents can utilize such predictions in a decision making process, however they must be supplemented by additional market information, from which exchange rates forecasts can be crucial.

Summary

The income maximization is the main purpose of every entrepreneur from the microeconomic theory perspective. It is widely observed in practice that price risk is one of the main sources of income drops in agriculture, due to the significant time lapse between the purchase of the inputs and the sales of the output. We can fully appreciate the real value of the agricultural commodity prices forecasting when we realize that agricultural economists define price risk as the difference between expected and actual prices.

Appropriate forecasting procedures can reduce uncertainty about the formation of prices in the future. However, agricultural producers are not the only ones who deal with price risk. Food processors, retailers or other market players are exposed to unexpected price changes as well. So there is a need for market agents to prepare individual market forecasts. Forecasting is frequently regarded as the microeconomic process of discovering future prices. Therefore, the agricultural price forecasting is also a way of gaining a competitive advantage over other market players. The available projections made by public institutions, as a part of market information system, cannot assure an extra advantage. Market agents can use analyses and projections prepared by institutions as the information for establishing the basis of their own forecasts.

The primary aim of this volume is to support practitioners and analysts operating on the agricultural markets by presenting the variety of statistical methods helpful in forecasting commodity prices. It provides the comprehensive review of the statistical and econometric methods from the relatively simple to the more complex ones. The presented methods allow for obtaining the information about the pattern and relationships that form price changes in the past and extrapolate them into the future. Availability of data, the forecasting horizon, the patterns existing in time series or the nature of relationships between exogenous and endogenous variables are the main factors influencing the choice of appropriate model for forecasting the agricultural commodity prices. Also the knowledge of market economy and statistics is of great importance for forecasters. The familiarity with market forces together with the knowledge of econometric rules is highly desirable for any forecaster.

However, we assume that knowledge about advanced econometric methods is not widespread. In this case, the optimal strategy is to use simple statistical models to explore the patterns and relationships in prices, such as the time series decomposition model, the regression and correlation analysis, which may be a good starting point. One can therefore combine facts from numerical models with market domain knowledge. Simple models can be also employed in the case of short time series or missing data. However, as shown by some studies, the use of more complex procedures does not always lead to improvement in the accuracy of forecasts. Using very complex model can also reduce understanding of the mechanisms on which the model is based, since the forecaster perceives it as a black box. Hence, there may be an obstacle to the recognition of such forecasts as reliable.

Time series models are one of the econometric methods groups frequently used in commodity prices forecasting. Time series forecasting makes no attempt to discover the factors influencing prices behavior. The only information needed is a past data of predicted prices. One of the most important reasons to use time series models is that there is no need to make assumptions about the values of explanatory variables in the forecasted period. According to the studies, the time series of agricultural commodity prices are characterized by a complex structure. Analyses show that cyclical variation is the most important in forecasting agricultural prices. The specificity of agricultural production causes frequent and unexpected disturbances in the time series structure, which are reflected in structural breaks. Among simple time series model recommended for forecasting agricultural prices there are: time series decomposition model for series with cyclical fluctuations, econometric models with trend, seasonal dummies and autoregressive component as well as Holt-Winters exponential smoothing models with a damping factor. Among more complex time series models X-12--ARIMA models and TRAMO/SEATS models are recommended. Some of them are available in a free statistical package like Gretl, R or DEMETRA+. Using these programs reduces the cost of such models application by a wider group of market participants.

Simplicity and low data requirements are the most important advantages of time series methods; nevertheless the studies confirm that there are causal interdependencies between agricultural commodity prices. Thus, short term forecasts of agricultural commodity prices can be calculated also as a function of other variables. The one equation congruent model and the multiple equations VAR/VECM models are among models that utilize information about other factors. These models are quite flexible and allow for capturing both the regularities existing in time series, as well as the interactions between domestic prices, world prices and external factors. The lag-length analysis is especially important when specifying a model. However, the construction of a proper empirical model describing the behavior of prices is a difficult task, often doomed to failure. Even if one manages to estimate a good model for the past, there is no guarantee that the assumptions made regarding extrapolation to the future will be met.

The short-term forecasts of agricultural prices are important for operational decision making: when to buy, when to sell or what to produce. However, when it comes to production adjustment or investment perspective medium- and long-run forecasts are the crucial ones. Governments are interested in the future state of the market in the long perspective as well. The medium and long-run predictions of prices can be made by extrapolating trends or with the use of large-scale partial equilibrium models. Partial equilibrium models are usually used to evaluate changes in the government policy or to prepare market projections.

The most well-known partial equilibrium models of the agricultural sector include AGLINK-COSIMO and FAPRI models. Long-term projections for the agricultural sector prepared with the use of above two models, among them for world commodity prices, are regularly published together with a comprehensive rationale and interpretation of the results. However, the comparison of the quoted prices with the ex-post projections indicates low accuracy of such projections. The incorrect assumptions about the future changes of the macroeconomic environment, random supply shocks or underestimation of policy effects are regarded as the main factors of price projection errors.

Bearing in mind that there is no feasible alternative to projections of the partial equilibrium models, they can be considered as one of the tools useful in construction medium- and long-term projections of the domestic agricultural prices. It is possible to estimate the models of price transmission from global to domestic markets by using the available statistical data and projections of world prices. The domestic price is a function of the world price and the exchange rate in its simplest form.

It is emphasized in literature that currently there is no general agreement among economists which methods generate highest accuracy forecasts. There are usually a few competitive methods that can be employed depending on the available data and market characteristics. However, forecasters should be aware that there can exist significant differences between forecast obtained on the basis these models. Predictions based on formal models do not have to be acceptable in each case, as they represents only one of the sources of information on the future course of events over time. So the forecaster has to decide what forecast will be the final one. In practice, one can build forecasts in a less formal way by combining market related patterns, relationships and econometric forecasts with intuition and non-statistical knowledge, including information that is not reflected in the historical data. The better knowledge about market mechanism the higher probability to obtain accurate forecasts. Therefore, the identification of the main drivers of agricultural commodity price is crucial in forecasting process.

Literature

1. Abbott P.C., Hurt C., Tyner W.E., *What's driving food prices?* Oak Brook, IL: Farm Foundation Issue Report, 2008.
2. Abbott P.C., Hurt C., Tyner W.E., *What's Driving Food Prices in 2011*? Oak Brook, IL: Farm Foundation Issue Report, 2011.
3. Adya M., Armstrong J.S., Collopy F., Kennedy M., *An application of rule-based forecasting to a situation lacking domain knowledge*, International Journal of Forecasting Volume 16, Issue 4, October 2001, p. 477-484.
4. AGMEMOD Partnership 2005*, Impact analysis of the CAP reform on main agricultural commodities. Final Report*, European Commission Directorate General JRC – IPTS, 2005.
5. Ardeni P.G., *Does the Law of One Price Really Hold for Commodity Prices?* American Journal of Agricultural Economics, Vol. 71, No. 3, 1989.
6. Armstrong J.S., Adya M., Collopy F., *Rule-Based Forecasting: Using Judgment in Time-Series Extrapolation*. [in:] Principles of Forecasting: A Handbook for Researchers and Practitioners (ed. J. Scott Armstrong). Kluwer, 2001.
7. Banse M., Ledebur O., Salamon P., van Leeuwen M., Bouma F., Salputra G., Fellmann T., Nekhay O., *Impact of Russia and Ukraine on the international price formation and the EU markets - A Model based analysis*. 123rd EAAE Seminar, February 23-24, Dublin Ireland, 2012.
8. Banse M., Tangermann S., *Agricultural implications of Hungary's Accession to the EU – Partial versus general equilibrium effects.* 50th EAAE Seminar "Economic Transition and the Greening of Policies: New Challenges for Agriculture and Agribusiness in Europe", 1996.
9. Baulch B., *Transfer costs, spatial arbitrage, and testing for food market integration,* American Journal of Agricultural Economics 79, 1997.
10. Baumel C.P., *How U.S. Grain Export Projections from Large Scale Agricultural Sector Models Compare with Reality*, 2001. http://www.iatp.org/iatp/publications.cfm?accountID=258&refID=72945
11. Bell W.R., Sotiris E., *Seasonal Adjustment to Facilitate Forecasting: Empirical Results*, JSM Proceedings Paper, October 2010.
12. Binfield J., Meyers W., Westhoff P., *Challenges of incorporating EU enlargement and CAP reform in the GOLD model framework*, 89th EAAE Seminar paper, FAPRI, University of Missouri, Columbia, USA 2005.
13. Bisgaard S., Kulahci M., *Time Series Analysis and Forecasting by Example*. John Willey&Sons, New Jersey, 2011.

14. Blanco-Fonseca M., *WP4 Baseline, Deliverable: D4.1. Common Agricultural Policy Regional Impact – The Rural Development Dimension. Literature Review of Methodologies to Generate Baselines for Agriculture and Land Use*, JRC, IPTS, 2010.
15. Błażejowski M., Kufel P., Kufel T., *Automatic procedure of building congruent dynamic model in Gretl*, [in:] I. Díaz-Emparanza, P. Mariel, M.V. Esteban (eds.) *Econometrics with Gretl: Proceedings of the Gretl Conference 2009*, Bilbao, Spain, May 28-29, 2009.
16. Box G.E.P., Jenkins G.M., *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day, 1970.
17. Brockwell P.J., Davis R.A., *Introduction to Time Series and Forecasting*. Second Edition, Springer 2002.
18. Burrell A. (ed.), *Impacts of the EU Biofuel Target on Agricultural Markets and Land Use. A Comparative Modelling Assessment*. JRC, European Commission, 2008.
19. Caporello G., Maravall A., *Program TSW. Revised Reference Manual*, Banco de España 2004, http://www.bde.es/servicio/software/tramo/tswrm.pdf).
20. Castle J.L., Hendry D.F., Kitov O.I., *Forecasting and Nowcasting Macroeconomic Variables: A Methodological Overview*, University of Oxford, Economics Series Working Papers no 674, 2013.
21. Chantreuil F., Hanrahan K., *AGMEMOD EU Agricultural Market Outlook*. Drustvo Agrarnih Ekonomistov Slovenije, DAES, 2007.
22. Chantreuil F., Hanrahan K.F., van Leeuwen M. (eds.), *The Future of EU Agricultural Markets by AGMEMOD*, Springer, 2012.
23. Chantreuil F., Tabeau A., van Leeuwen M., *Estimation of impact of EU agricultural policies on the world market prices.* Paper for presentation at the 107th EAAE Seminar "Modelling of Agricultural and Rural Development Policies". Sevilla, Spain, January 29th - February 1st, 2008.
24. Charemza W., Deadman D., *New Directions in Econometric Practice: General to Specific Modelling, Cointegration, and Vector Autoregression*, Edward Elgar Publishers, Lyme, N.H. 1997.
25. Cieślak M. (ed.), *Prognozowanie gospodarcze: metody i zastosowania*, PWN, Warszawa 2005.
26. Collopy F., Armstrong J.S., *Rule-based forecasting: development and validation of an expert systems approach to combining time series extrapolations.* Management Science 38, 1992, p. 1392–1414.
27. Conforti P., Londero P., A*GLINK: The OECD partial equilibrium model*, Working Paper n.8, INEA, The National Institute of Agricultural Economics, Roma, Italy, 2001.

28. Conforti P., *The Common Agricultural Policy in main Partial Equilibrium models*. Osservatorio sulle Politiche Agricole dell'UE, INEA, Working Paper No. 8, 2001.
29. Davies G., *Removing Biofuel Support Policies: An Assessment of Projected Impacts on Global Agricultural Markets using the AGLINK-COSIMO model*, Defra, 2012.
30. Dąbrowski I., *Teoria równowagi ogólnej. Rys historyczny i obecny status w ekonomii*. SGH, Warszawa 2009.
31. Donnellan T., Hanrahan K., McQuinn K., Riordan B., *Policy analysis with AG-MEMOD Model: Dealing with diversity in the EU Agri-Food Sector*. RERC, EAAE Congress, 2002.
32. Doornik J.A., *Autometrics*, [in:] J.L. Castle and N. Shephard (eds.), *The Methodology and Practice of Econometrics: Festschrift in Honour of David F. Hendry*, Oxford University Press, Oxford 2009.
33. Draper N.R., Smith H., *Analiza regresji stosowana*, PWN, Warszawa 1973.
34. Drummond H.E., Goodwin J.W., *Agricultural Economics*, Prentice-Hall International, 2004.
35. Enders W., *Applied Econometric Time Series*, Willey, New York 2010.
36. Engle R.F., Granger C.W.J., *Co-integration and Error Correction*, Econometrica, vol. 55, 1987.
37. Engle R.F., Hendry D.F., Richard J.-F., *Exogeneity*, Econometrica, vol. 51, 1983.
38. Enke S., *Equilibrium among spatially separated markets: solution by electrical analogue,* Econometrica 19, 40–7, 1951.
39. Esposti R., Camaioni B., *Technical Report on the Modeling Structure*, AGMEMOD WP2 D2, Project no. SSPE-CT-2005-021543, 2007.
40. *ESS Guidelines on Seasonal Adjustment*, 2009, http://epp.eurostat.ec. europa.eu/cache/ITY_OFFPUB/KS-RA-09-006/EN/KS-RA-09-006-EN.PDF
41. Evans M.K., *Practical business forecasting*, John Wiley & Sons Inc., Oxford, 2003.
42. Falk B., Roy A., *Forecasting using the trend model with autoregressive errors*. International Journal of Forecasting 21, 2005, s. 291– 302.
43. FAPRI 2011, *FAPRI U.S. and World Agricultural Outlook. FAPRI Staff Reports 1999-2010*, FSR 1, Food and Agricultural Policy Research Institute. Iowa State University and the University of Missouri-Columbia, 2011, http://www.fapri.iastate.edu/outlook/
44. FAPRI 2011a, *MU Stochastic U.S. Crop Model Documentation*, FAPRI-MU Report #09-11, 2011.

45. Ferris J.N., *Agricultural Prices and Commodity Market Analysis*, Michigan State University Press, Copyright 2005 by John N. Ferris.
46. Figiel S., *Cenowa efektywność rynku towarowego na przykładzie zbóż w Polsce*, Wydawnictwo UW-M, Olsztyn 2002.
47. Findley D.F., Monsell B.C., Bell W.R., Otto M.C., Chen B-C., *New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program*, Journal of Business and Economic Statistics 16, 1998, pp. 127-152, http://www.census.gov/ts/papers/jbes98.pdf.
48. Gardner B.L., *The Economic of Agricultural Policies*, McGraw-Hill Publishing Company, New York 1990.
49. Gardner B.L., *The Farm-Retail Price Spread in a Competitive Food Industry,* American Journal of Agricultural Economics, Vol. 57, No. 3, pp. 399-409, 1975.
50. Gardner E.S., Jr., McKenzie E., *Forecasting trends in time series*, Management Science, 31, 1985.
51. Gilbert C.L., Morgan C.W., *Food price volatility*. Philosophical Transactions of The Royal Society 365, 2010.
52. Goldberger A.S., *Introductory Econometrics*, Harvard University Press, Cambridge Mass 1998.
53. Gómez V., Maravall A., *Seasonal Adjustment and Signal Extraction in Economic Time Series*, [in:] Tsay, R.S. (eds.), *A Course in Advanced Time Series Analysis*, Wiley and Sons, New York, pp. 202-246, 2001.
54. Goodwin J.W., *Agricultural Price Analysis and Forecasting*, John Willey & Sons Inc. 1994.
55. Granger C.J.W., Newbold P., *Spurious Regression in Econometrics*, Journal of Econometrics, vol. 2, s. 111-120, 1974.
56. Grudkowska S., Buono D., Palate J., Ciebiera W., *Advanced Tools for Time Series Analysis and Seasonal Adjustment in the New JDemetra+*, JSM Proceedings Paper 2013.
57. Grudkowska S., *Demetra+ User Manual*, 2013.
58. Grudkowska S., Paśnicka E., *X-12-ARIMA I TRAMO/SEATS – empiryczne porównanie metod wyrównania sezonowego w kontekście długości próby*, Narodowy Bank Polski, „Materiały i Studia" no. 220, 2007.
59. Gruszczyński M., Kuszewski T., Podgórska M. (eds.), *Ekonometria i badania operacyjne*. PWN, Warszawa 2009.
60. *Guide to Seasonal Adjustment with X-12-ARIMA*, Office for National Statistics (ONS), 2007, http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/time-series-analysis/guide-to-seasonal-adjustment.pdf.

61. Hamilton J.D., *Time series analysis*, Princeton University Press 1994.
62. Hamulczuk M. (ed.), *Prognozowanie cen surowców rolnych z wykorzystanie, modeli szeregów czasowych*, Raport PW nr 10, IERIGŻ-PIB, Warszawa 2011.
63. Hamulczuk M., *Dokładność długookresowych projekcji na rynku rolnym - przykład modelu FAPRI i rynku pszenicy*, Roczniki Nauk Rolniczych. Seria G - Ekonomika Rolnictwa, T. 98, z. 1, 2011.
64. Hamulczuk M., Klimkowski C., *Response of the Polish Wheat Prices to the World's Crude Oil Prices*, Acta Oeconomica et Informatica,Volume 15, Number 2, 2012.
65. Hamulczuk M., Stańko S., *Ekonomiczne skutki likwidacji kwot mlecznych w Unii Europejskiej – wyniki symulacji z wykorzystaniem modelu AGMEMOD*, Zagadnienia Ekonomiki Rolnej, nr 4, 2009.
66. Hanrahan K., *The GOLD Model 2.1: An introductory manual*. Mimeo, Rural EconomyResearch Centre, Dublin, 2001.
67. Heijman W., Krzyżanowska Z., Gędek S., Kowalski Z., *Ekonomika rolnictwa. Zarys teorii*, Fundacja Rozwój SGGW, Warszawa 1997.
68. Hendry D.F., *Dynamic Econometrics,* Oxford University Press 1995.
69. Hendry D.F., Krolzig H.-M., *Automatic Econometric Model Selection*, Timberlake Consultants Press, London 2001.
70. Hendry D.F., Krolzig H.-M., *The properties of automatic Gets modelling. Economic Journal*, 115, 2005.
71. Hendry D.F., Pretis F., *Anthropogenic influences on atmospheric CO2*, [in:] Fouquet R. (ed.), *Handbook on Energy and Climate Change*, Cheltenham: Edward Elgar 2012.
72. Hill B., *An Introduction to Economics for Students of Agriculture*, Pergamon Press 1990.
73. Holt C.C., *Forecasting seasonal and trends by exponentially weighted moving averages*, Office of Naval Research, Research memorandum 52, 1957.
74. Hyndman R., Koehler A.B., Ord J.K., Snyder R.D., *Forecasting with Exponential Smoothing: The State Space Approach,* Springer, 2008.
75. IFPRI 2011, *Food Price Volatility: Implications for ACP countries*. Brussels, 2011.
76. *Introduction to Seasonal Adjustment*, The Central Bureau of Statistics, http://www.cbs.gov.il/www/publications/tseries/seasonal12/intro.pdf
77. Johansen S., *Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models*. Econometrica vol. 59 pp. 1551–1580, 1991.

78. Johansen S., *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford 2009.
79. Kennedy P. A., *Guide to Econometrics,* MIT Press, Cambridge Mass, 2003.
80. Kowalski A., Rembisz W., *Rynek rolny i interwencjonizm a efektywność i sprawiedliwość społeczna*, IERiGŻ-PIB, Warszawa 2005.
81. Kufel T., *Postulat zgodności w dynamicznych modelach ekonometrycznych*, Wydawnictwo UMK, Toruń 2002.
82. Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, PWN, Warszawa, 2007.
83. Kufel T., *General-to-Specific Modelling vs. Congruent Modelling in PcGets*, [in:]: Zieliński Z. (ed.) *Dynamic Econometric Models*, vol. 6, pp. 83–92. Nicolaus Copernicus University Press, Toruń 2004.
84. Kusideł E., *Modelowanie wektorowo-autoregresyjne VAR. Metodologia i zastosowanie w badaniach ekonomicznych*, Absolwent, Łódź 2000.
85. Ladiray D., Quenneville B., *Seasonal Adjustment with the X-11 Method*, Lecture Notes in Statistics, New York: Springer-Verlag 2001.
86. Listorti G., Esposti R., *Making the world market price endogenous within the AGMEMOD modelling framework: an econometric solution*. Paper prepared for presentation at the 107th EAAE Seminar "Modelling Agricultural and Rural Development Policies". Sevilla, Spain, January 29th - February 1st, 2008.
87. Lothian J., Morry M., *A Test for Presence of Identifiable Seasonality when Using the X-11 Program*, Working Paper, Time Series Research Division, Statistics Canada, Ottawa, ON, Canada, 1978 http://www.census.gov/ts/papers/LothianMorry1978.pdf.
88. Lütkepohl H., Krättzig M., (eds.) *Applied Time Series Econometrics*, Cambridge Univeristy Press, 2007.
89. Lütkepohl H., *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin 2007.
90. Maddala G.S., Kim I.-M., *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, 1998.
91. Maddala G.S., *Ekonometria*, PWN, Warszawa 2006.
92. Makridakis S., Wheelwright S.C., Hyndman R.J., *Forecasting Method and Applications*, John Wiley & Sons, Nowy Jork, 1998.
93. Meyers W.H., Westhoff P., Fabiosa J., Hayes D., *The FAPRI Global Modeling System and Outlook Process*, Journal of International Agricultural Trade and Development Volume 6, Issues 1, 2010.
94. Mills T.C., *The Econometric Modeling of Financial Time Series*, Cambridge University Press, Cambridge 2002.

95. Monsell B.C., *The X-13A-S Seasonal Adjustment Program*, http://www.fcsm.gov/07papers/Monsell.II-B.pdf.
96. Moschini G., Hennessy D.A., *Uncertainty, Risk Aversion and Risk Management for Agricultural Producers*, Handbook of Agricultural Economics, Elsevier Science Publishers, Amsterdam 2000.
97. Nazlioglu S., Soytas U., *Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis*. Energy Economics 34, 2012.
98. OECD 2007, *Documentation of the Aglink-Cosimo Model*, Working Party on Agricultural Policies and Markets. OECD, Paris, 2007, http://www.oecd.org/site/oecd-faoagriculturaloutlook/45172745.pdf
99. OECD/FAO 2012, *OECD-FAO Agricultural Outlook 2012-2021*, OECD Publishing and FAO, http://dx.doi.org/10.1787/agr_outlook-2012-en
100. Ord K., Fildes R., *Principles of Business Forecasting*, South-Western Cengage Learning 2013.
101. Osińska M. (ed.), *Ekonometria współczesna*. Dom Organizatora, Toruń, 2007.
102. Pedregal D.J., Young P.C., *Statistical Approaches to Modelling and Forecasting Time Series* [in:] *A Companion to Economic Forecasting*, ed. By Clements M.P., Hendry D.F., Blackwell Publishing, Oxford 2005.
103. Pegels, C.C., *Exponential forecasting: some new variations.* Management Science, 12, 1969, p. 311–315.
104. Pérez Domínguez I., Britz W., Gopalakrishnan B.N., *Post-model Analysis in large-scale models: the examples of Aglink-Cosimo, CAPRI and GTAP2012*, Paper prepared for 15th Annual Conference on Global Economic Analysis "New Challenges for Global Trade and Sustainable Development" Geneva, Switzerland, June 27-29, 2012.
105. Perron P., *Dealing with Structural Breaks*, Handbook of Econometrics Vol. 1: Econometric Theory, 2005.
106. Phillips P.C.B., *New Tools for Understanding Spurious Regression*, Econometrica, vol. 66, 1998.
107. Piłatowska M., *Efekty błędnej identyfikacji niestacjonarności procesów ekonomicznych dla wariancji błędu prognozy*, [in:] T. Kufel, M. Piłatowska (eds.), *Dynamiczne modele ekonometryczne: materiały na IX Ogólnopolskie Seminarium Naukowe*, 6-8 września 2005UMK Toruń 2005.
108. Piłatowska M., *Realization of the Congruence Postulate as a Method of Avoiding the Effect of a Spurious Relationship*, [in:] Z. Zieliński (ed.) *Dynamic Econometric Models* vol. 6, Nicolaus Copernicus University Press, Toruń 2004.

109. Piłatowska M., *The econometric models satisfying the congruence postulate: an overview*, Dynamic Econometric Models, Vol. 8, 2008.
110. Planas C., *The analysis of seasonality in economic statistics: a survey of recent developments*, Questiio, vol, 22, Eurostat 1998.
111. Radetzki M.A., *Handbook of Primary Commodities in Global Economy*, Cambridge University Press, 2010.
112. Rapsomanikis G., Hallam D., Conforti P., *Market integration and price transmission in selected food and cash crop markets of developing countries: review and applications,* FAO, Commodity Market Review, FAO Commodities and Trade Division, Rome, 2003.
113. Rathmann R., Szklo A., Schaeffer R., *Land use competition for production of food and liquid biofuels: An analysis of the arguments in the current debate,* Renewable Energy 35 (2010) 14–22, 2010.
114. Rembeza J., *Transmisja cen w gospodarce polskiej*, Wydawnictwo Uczelnianie PK, Koszalin 2010.
115. Rembisz W., *Krytyczna analiza podstaw i ewolucji interwencji w rolnictwie*, Współczesna Ekonomia, nr 4, WSFiZ, Warszawa 2010.
116. Réquillart V., Bouamra-Mechemache Z., Jongeneel R., Penel C., *Economic analysis of the effects of the expiry of the EU milk quota system,* Report prepared for the European Commission, IDEI, Toulouse 2008.
117. Ritson Ch., *Agricultural Economics. Principles and Policy*, Collins, London 1977.
118. Samuelson P., *Spatial price equilibrium and linear programming,* American Economic Review 42, 283–303, 1952.
119. Serra T., Zilberman D., *Price volatility in ethanol markets,* European Review of Agricultural Economics, Vol. 38(2), 2011.
120. Shiskin J., *Electronic Computers and Business Indicators*, Journal of Business, 30, 1957, pp. 219-267, http://www.nber.org/books/juli57-1.
121. Sims C.A., *Macroeconomics and Reality,* Econometrica, vol. 49, 1980.
122. Stańko S. (ed.), *Prognozowanie w agrobiznesie. Przykłady zastosowania*. Wyd. SGGW, 2013.
123. Stock J.H., Watson M.W., *Introduction to Econometrics*, second edition, Pearson 2007.
124. Stock J.H., Watson M.W., *Vector Autoregressions*. Journal of Economic Perspectives, vol.15, pp. 101-115, 2001.
125. Takayama T., Judge G., *Spatial and Temporal Price Allocation Models*, Amsterdam: North-Holland, 1971.
126. Taylor J.W., *Exponential Smoothing with a Damped Multiplicative Trend*, International Journal of Forecasting, 2003, Vol. 19, pp. 715-725.

127. Toda, H.Y., Yamamoto T., *Statistical inferences in vector autoregressions with possibly integrated processes*. Journal of Econometrics, 66, 1995.
128. Tomek W.G., Robinson K.L., *Agricultural product prices*, Cornell University Press, 2003.
129. Tongeren F., van Meijl H., van Surry Y., *Global models applied to agricultural and trade policies: A review and assessment*. Agricultural Economics, no 26, 2001.
130. Tsay R.S., Analysis of Financial time series, John Wiley & Sons, 2010.
131. Tyner W.E., Taheripour F., Hurt Ch., *Potential Impacts of a Partial Waiver of the Ethanol Blending Rules*, Farm Foundation, Purdue University, 2012.
132. Tyner W.E., Taheripour F., *Policy Options for Integrated Energy and Agricultural Markets,* Review of Agricultural Economics, Vol. 30, Nr 3, 2008.
133. Tyner W.E., *The Integration of Energy and Agricultural Markets*, Agricultural Economics, Vol. 41 Issue Supplement s1, November 2010.
134. Uebayashi A., *Agricultural Outlook and Its Baseline Process using AGLINK Model*, OECD, 2008, http://www.unece.org/fileadmin/DAM/stats/documents/ces/sem.44/wp.7.e.pdf
135. Varian H.R., *Intermediate Microeconomics: A Modern Approach* (Fifth Edition), W.W. Norton & Company, New York, London, 1999.
136. Wang D., Tomek W.G., *Commodity Prices and Unit Root Tests*. American Journal of Agricultural Economics, 89, 2007.
137. Waught F.V., *Demand and price analysis. Some examples from agriculture*, Technical biulletin 1316, USDA-ERS, 1964.
138. Westhoff P., Brown S., Hart C., *When Point Estimates Miss the Point: Stochastic Modeling of WTO Restrictions*. Journal of International Agricultural Trade and Development 2, 2006.
139. Westhoff P., Brown S.; Binfield J., *Why Stochastics Matter: Analyzing Farm and Biofuel Policies*. Paper to the 107th EAAE Seminar “Modeling of Agricultural and Rural Development Policies”, Sevilla, Spain, 29 January - 1 February 2008.
140. Westhoff P., *The European Union Grain, Oilseed, Livestock and Dairy (EU GOLD) Model*, FAPRI at the University of Missouri, 2001.
141. Wisner R.N., McVey M., Baumel C.P., Curtiss C.F., *Are Large-Scale Agricultural-Sector Economic Models Suitable for Forecasting*? 2002, http://www2.econ.iastate.edu/faculty/wisner/largescalemodels.pdf
142. Wright B., Cafiero C., *Grain reserves and food security in the Middle East and North Africa*, Food Sec. 3 (Suppl 1), S61–S76, 2011.
143. Wu L.S.-Y., Hosking J., Ravishanker N., *Reallocation Outliers in Time Series*, Journal of the Royal Statistical Society, Series C, no.2, 1993.

144. *X-12-ARIMA Reference Manual*, Time Series Research Staff, Centre for Statistical Research and Methodology, US Census Bureau, 2007, http://www.census.gov/ts/x12a/v03/x12adocV03.pdf.
145. *X-13ARIMA-SEATS Reference Manual*, Time Series Research Staff, Centre for Statistical Research and Methodology, US Census Bureau, 2013, http://www.census.gov/ts/x13as/docX13ASHTML.pdf.
146. Zieliński Z., *Liniowe modele zgodne opisujące zależności sumacyjnych (zintegrowanych) procesów ekonomicznych*, [in:] Zeliaś A. (ed.) Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych, AE Kraków 1995.

COPY FREE

Print run: 280 copies, 11,23 publisher`s sheets
Printing and binding: EXPOL Włocławek