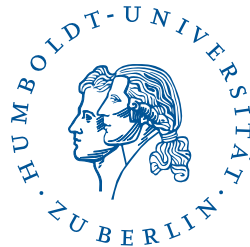


HUMBOLDT UNIVERSITÄT ZU BERLIN
SCHOOL OF BUSINESS AND ECONOMICS
LADISLAUS VON BORTKIEWICZ CHAIR OF STATISTICS



Time Series Forecasting of the Development of the Insurance Industry in Poland



BACHELOR THESIS
OF
Wiktor Olszowy
(537124)

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
BACHELOR OF SCIENCE IN ECONOMICS

SUBMITTED TO:

Prof. Dr. Wolfgang Härdle
Dr. Sigbert Klinke

BERLIN, 17 JUNE 2013

Abstract

This work investigates the development of the insurance industry in Poland over the last twelve years (2001–2012) and makes forecasts of this development for all quarters of the year 2013. We consider Gross Written Premiums (GWP) as the best indicator showing the size of the insurance industry. Our aim is to discover relations between GWP and other time series regarding Polish economy: profitability ratio of technical activity for the entire insurance industry, Gross Domestic Product, inflation and consumer confidence indicators. Firstly, we conduct univariate analysis of all the six time series, find trends and seasonal effects, model the residuals, as well as apply SARIMA models. For each series the corresponding forecasts are presented. In the second part we conduct multivariate time series analysis, in particular we model our data with VAR and look for Granger causalities.

Keywords: Polish Insurance Industry, SARIMA Models, Seasonal Effects, VAR, Granger Causality

Acknowledgements

First of all, I would like to thank Dr. Sigbert Klinke for his enthusiastic supervision throughout the work and technical support regarding the programming language R and the document markup language \LaTeX . Without his guidance this Bachelor thesis would not be possible. I want to acknowledge the help of other members of the Statistics and Econometrics Chairs of the Humboldt University too, particularly PD. Bernd Droge, whom I have consulted several times in matters related to this work. My sincere thanks are extended to Mateusz Jakitowicz, who was my supervisor during an internship at ERGO Hestia in 2012 and who has in fact chosen the topic for the thesis. Throughout the work I have regularly discussed the results with him.

I cannot find words to express my sincere gratitude to my family and to my friends, Awdesch, Emil, Karolina and Piotr. I would like to thank them all for the encouragement, motivation and help they have given me. Last but not least I would like to thank my hometown Gdańsk for awarding me the G.D. Fahrenheit Scholarship. The scholarship has enabled me to concentrate on learning and has made the studies abroad considerably easier.

List of abbreviations

ACF	Autocorrelation function
ADF	Augmented Dickey–Fuller
AIC	Akaike information criterion
AR	Autoregressive
ARCH	Autoregressive conditional heteroscedasticity
CCCI	Current consumer confidence indicator
DWN	Discrete white noise
EFF	Profitability/Efficiency ratio of technical activity
FGLS	Feasible generalized least squares
GDP	Gross Domestic Product
GLS	Generalized least squares
GUS	Polish Central Statistical Office
GWP	Gross Written Premiums
i.i.d.	Independently and identically distributed
INF	Inflation
JB	Jarque–Bera
KS	Kolmogorov–Smirnov
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
NBP	National Bank of Poland
LCCI	Leading consumer confidence indicator
MA	Moving average
ML	Maximum likelihood
OLS	Ordinary least squares
PACF	Partial autocorrelation function
PLN	Polish new zloty
Q.	Quarter
SARIMA	Seasonal autoregressive integrated moving average
VAR	Vector autoregressive

Contents

1. Introduction	1
2. Data	2
3. Methodology	4
3.1. Seasonality	4
3.2. Correlation and autocorrelation	5
3.3. White noise and random walk	6
3.4. Stationarity	7
3.5. AR(p) and MA(q)	8
3.6. SARIMA(p,d,q)(P,D,Q)	8
3.7. Regression	9
3.8. Structural breaks	11
3.9. VAR(p)	11
3.10. Granger causality	13
4. Univariate Time Series Analysis	14
4.1. Gross Written Premiums	14
4.2. Profitability ratio of technical activity	18
4.3. Gross Domestic Product	19
4.4. Inflation	22
4.5. Consumer confidence indicators	23
5. Multivariate Time Series Analysis	26
5.1. Correlations	26
5.2. VAR and Granger causality	27
6. Conclusions	31
A. Appendix	32

List of Figures

4.1.	Gross Written Premiums [million PLN], quarter data	14
4.2.	Forecasted values of Polish GWP till the end of 2016 [million PLN]	15
4.3.	Forecasts of log(GWP) based on the SARIMA model. The blue area corresponds to the confidence interval of the 80% significance level	16
4.4.	Application of the Chow forecast test to detect structural breaks within GWP (harmonic model)	17
4.5.	Forecasts of EFF based on the SARIMA model. The blue area corresponds to the confidence interval of the 80% significance level	18
4.6.	GDP time series [million PLN], quarter data	19
4.7.	Correlograms of the residuals from the AR process which has been computed for the residuals from the GDP OLS regression	20
4.8.	Forecasted values of Polish GDP till the end of 2016 [million PLN]	21
4.9.	Inflation time series [%], quarter data	22
4.10.	Correlograms of the inflation series after subtracting seasonal effects	23
4.11.	Consumer confidence indicators	24
4.12.	Correlograms of the differenced CCCI and LCCI time series	25
5.1.	Correlations between each two time series	26
5.2.	p-values of the Granger causality test	28
5.3.	Forecasted residuals and corresponding confidence intervals for the analysed VAR(1) model	30
A.1.	Correlograms of the residuals from the regression model with dummy variables, GWP	32
A.2.	Correlograms of the residuals from the harmonic seasonal model regression, GWP	32
A.3.	Correlograms of the residuals from the AR process which has been computed for the residuals from the GWP OLS regression (dummies model)	33
A.4.	Correlograms of the residuals from the AR process which has been computed for the residuals from the GWP OLS regression (harmonic seasonal model)	33
A.5.	Correlograms of the residuals from the SARIMA model for GWP	34
A.6.	Application of the Chow forecast test to detect structural breaks within GWP (dummies model)	34
A.7.	Correlograms of the residuals from the regression model for EFF	35
A.8.	Correlograms of the residuals from the SARIMA model for EFF	35
A.9.	Scatter plots of all the pairs of considered series	36
A.10.	Forecasted values of Polish GWP till the end of 2016 [million PLN], including the VAR(1) model of the residuals	37

List of Tables

4.1.	OLS and FGLS regressions for the GWP trend	15
4.2.	Forecasts of the development of the Polish GWP in 2013 and 2014 [million PLN], based on the dummies model, the harmonic model and SARIMA, respectively. For '13 Q. 1 the values are 18388.0, 16924.0 and 17837.1 . . .	17
4.3.	OLS and FGLS regressions for the GDP trend	20
4.4.	Forecasted values of Polish GDP till the end of 2014 [million PLN], not knowing and knowing the value for '13 Q. 1 (394739.6)	21
4.5.	Additive seasonal effects + mean of the series	22
4.6.	t–statistics of the OLS and FGLS estimates of the trend	23
5.1.	Stationarity and normality tests for the series which are used in the VAR modelling	27
5.2.	Coefficients of the VAR(1) model involving GWP, EFF, GDP and LCCI	29
6.1.	Forecasts for the year 2013, based on univariate time series analysis (except for the second row); green colour corresponds to the already known values	31
A.1.	Forecasts of GWP, EFF, GDP and LCCI in 2013 and 2014 [GWP and GDP in million PLN], based on univariate trends and the VAR(1) model of the residuals	37

1. Introduction

After the political and economic transition in 1989–1990 Poland witnessed several years of economic stagnation, after which its economy started to boost. On May 1, 2004 Poland joined the European Union and benefiting from the EU subsidies it became one of the fastest growing economies in Eastern Europe. During the recent financial crisis Poland managed to maintain GDP increase as one of few countries in Europe.

Along with GDP, the insurance industry has developed, as measured by Gross Written Premiums – the total sum of insurance premiums. According to my own calculations, which can be found in the attached `multi.R` file, for the time period 2001-2012 the insurance industry in Poland grew on average 9.8% per annum, whereas Polish GDP grew by 6.7% (in both cases current prices, no inflation considered). In 2001 Gross Written Premiums constituted 2.87% of GDP, in 2012 it was already 3.93%. My own interest in the subject of Polish insurance industry arose last year when I completed an internship at ERGO Hestia, where I was mainly conducting statistical analyses regarding the insurance market.

In the first part of the work I will consider the univariate analysis of GWP. To detect a trend, regression models will be used and the autocorrelation of the error terms will be accounted for. As an alternative to this approach SARIMA modelling will be applied. The same methodology will be then used for other considered series which can have an influence on GWP. Profitability ratio of technical activity shows how effective the insurance companies are, in terms of investing money from insurance premiums and setting prices of their products. Like Gross Domestic Product, which describes the general situation of the economy, there may be a relationship between this index and Gross Written Premiums. Inflation and consumer confidence indicators are other macroeconomic indices for which the influence upon GWP should be investigated. For all these time series we make forecasts.

In the end we are going to detect causal relationships of the five macroeconomic indices with GWP (considering residuals from the univariate analysis) and make forecasts based on a vector autoregressive model, for which eventually only the indices with proved causality upon GWP will be considered. All the analysed models can be treated as a framework for further work and the forecasting models can be used in future when new values of the indices will be known. For that one can use the codes, which have been attached to the thesis on a compact disk. All the graphs, statistical tests and other calculations including forecasts have been made in the programming language R, under version 2.15.2. I have used the `fBasics`, `fields`, `forecast`, `fUnitRoots`, `nlme` and `vars` packages.

2. Data

Altogether we have data from twelve years: 2001–2012, for each year for all four quarters. Values of Gross Written Premiums, as well as of the efficiency/profitability ratio of technical activity were downloaded on the 9th of May from the website of Polish Financial Supervision Authority (KNF). KNF has published quarter analysis of the insurance industry since 2002, but because next to every entry there is the corresponding value from the previous year (for comparison), there are values from 2001 as well. Before 2001 quarter data regarding the entire Polish insurance industry had not been collected. This is why for other time series we take 2001 as the starting point too. With the exception of 2012 I have always considered corrected values of GWP and EFF published one year later. For the year 2012 I have taken the current values, not the corrected ones, which will appear in 2014. KNF reports referring to the first quarter of 2013 will appear on the 20th of June.

Values of GWP can be found in the attached financial reports of the Polish insurance industry, in worksheet 'Tabl.B.3.' in cell 8B. However, for 2002 and 2003 the reports were chaotic and the structure differed. The units are thousands Polish Zloty, for further work we will consider millions. The reports include only aggregated values (from the first quarter of the year till the analysed quarter). For a meaningful analysis we subtract two sequent values, only for the first quarter we copy the value, just as it is in the report.

As for the efficiency ratio of technical activity, the exact way of its measure is described in the legislation of the Polish finance minister regarding the insurance sector from the 28th of December 2009 (see the bibliography). It is computed using numerous values from the financial reports and its values are displayed in the efficiency ratios reports. Since again the values are aggregated and we are interested in the efficiency of the sector in every quarter, we have to compute the index by ourselves. Approximately it is the quotient of the balance on technical account and earned premiums (for easiness we omit other indices, which have relatively small values). Since earned premiums change in a very similar way as GWP, we consider in the quotient instead of them the Gross Written Premiums. Values of the balance on technical account are in the same worksheet as GWP, in cell 26B.

The third time series to be considered is the Polish Gross Domestic Product. The data is compiled by Polish Central Statistical Office (GUS) in line with the principles of the European System of National and Regional Accounts (ESA 1995). GDP of a country represents the final result of the activity of all entities of the economy. It is the sum of consumption, investment and government expenditures within a time period, in our case a quarter. I downloaded the data from the GUS website on the 27th of May. The reader can find the Excel dataset `GUS_quarterly_indicators.xls` attached to the thesis. In the worksheet 'NATIONAL ACCOUNTS NACE Rev.2' in row 7 the values of GDP are listed. The units are millions Polish Zloty.

Polish inflation is being recorded by GUS as well, but usually publications of the National Bank of Poland (NBP) are used as a source of data regarding changing price levels. I use the data provided by NBP as well. The inflation rate is published every month, but since for other series we only have quarter data, we consider multiplied price levels within each quarter. We divide this product through $100^3 = 1000000$, subtract from it 1 and treat the result as a percentage. NBP publishes estimates of many different kinds of inflation. We consider CPI (consumer price index) for which the value of 100 corresponds to the price level in the previous month (see the attached file `NBP_inflation.xls`, worksheet 'data', column G). The last time I downloaded the inflation rates from the NBP website was on the 27th of May.

The data for the last two time series, namely current consumer confidence indicator (CCCI) and leading consumer confidence indicator (LCCI), comes from the website of GUS and was downloaded for the last time on the 28th of April. The values can be found in the attached `GUS_monthly_indicators.xls` file (rows 167 and 168 of worksheet 'Selected_indicators_p_I', respectively). Both these indices are calculated on the basis of household surveys which contain a set of questions directly addressing the consumers. The aim of the survey is to measure the consumers' assessment of the national economy's condition, as well as their assessment of their own financial situation. The Central Statistical Office and the National Bank of Poland are responsible for the organisation of the survey. Till the end of 2003 the survey had been conducted every quarter. Afterwards both indicators were measured on a monthly basis. As already discussed, we are mainly interested in multivariate analysis and forecasts for GWP, which is measured only every three months. That is why I take for CCCI and LCCI for every quarter the value from the corresponding first month.

Generally, the indices represent the differences between positive and negative answers. According to the 'Methodological Notes' of GUS (p. 17, see the bibliography), CCCI is the arithmetic mean of the evaluations of the previous and predicted (for the next twelve months) changes concerning the household's financial condition, as well as the general economic situation and important purchases currently made. Again referring to the already mentioned GUS document, LCCI is the arithmetic mean of the evaluations of changes in the household's financial condition, the economic situation of the country, trends in unemployment and saving propensity, all over the next twelve months. Both CCCI and LCCI may range from -100 to $+100$, where a positive value means that the majority of the consumers have a positive attitude. In fact, in the analysed time period it has never occurred for CCCI or LCCI to have a positive value. The negative values mean in this context the prevalence of pessimistic attitudes.

3. Methodology

Our analysis of data will refer to both uni- and multivariate time series. Before analysing the data I would like to recall some definitions, as well as list and discuss all the statistical methods which will be used at a later stage. The notation used in the analytical part of my thesis will be the one used in this 'Methodology' part. Generally I follow the notation of Cowpertwait et al. (2009) and Hamilton (1994).

3.1. Seasonality

Many time series consist of both a trend and some seasonal effects. There are two main categories of seasonal effects: additive and multiplicative ones. The additive decomposition model is described as follows:

$$x_t = m_t + s_t + z_t \quad (3.1)$$

x_t is the analysed series at time t , m_t is the trend, s_t is the seasonal effect and z_t is an error term. Using the same notation we can analyse in the following way a model with multiplicative effects:

$$x_t = m_t \cdot s_t \cdot z_t \quad (3.2)$$

To estimate the seasonal effects we need first to calculate a trend with some filter. A popular way of doing this is by applying moving averages, it means averaging the values around a value, so that the entire period (usually the year) is covered and the seasonal effects are omitted. Since in case of my study the data is for quarters, as an example consider:

$$\hat{m}_t = \frac{\frac{1}{2} \cdot x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2} \cdot x_{t+2}}{4} \quad (3.3)$$

This approach does not allow us to calculate the trend for the first and last three terms. As for the seasonal effects' terms, we can now consider a series constructed by subtracting the trend term from x_t . The additive seasonal effect of quarter i corresponds to the average of the differences $x_t - \hat{m}_t$ for all quarters i within the time interval. If we have an integer number of cycles within the data, the average of the additive effects equals zero. For multiplicative effects the same approach is followed, but instead of using the difference $x_t - \hat{m}_t$, we use the quotient x_t/\hat{m}_t . If the number of cycles (say, years) is an integer number, the terms average unity.

Because the thesis is about making forecasts, moving averages are not very helpful – they give a trend one should not use for extrapolation. Throughout the thesis the moving averages trend will only be used in order to estimate the seasonal effects when a deterministic trend cannot be used. For the above procedure the function `decompose` from the R package `stats` will be used.

3.2. Correlation and autocorrelation

We will often need a way to measure the linear association between two variables (or just of a variable with itself after a shift). For this we need the concept of covariance. When x and y are the variables to be considered, their covariance function is (μ_x denotes the expected value of x and μ_y is the expected value of y):

$$\gamma(x, y) = \mathbb{E}[(x - \mu_x)(y - \mu_y)] \quad (3.4)$$

In order to get a dimensionless measure, we can divide the covariance function through the multiplied standard deviations of x and y (see Equation 3.5) and so we obtain the correlation. Because of the construction, it can only assume values between -1 and 1, where the extremes mean, respectively, a perfect negative and a perfect positive linear association.

$$\rho(x, y) = \frac{\gamma(x, y)}{\sigma_x \sigma_y} \quad (3.5)$$

However, sometimes we are interested in calculating the correlation when one of the variables is lagged. This leads to the concept of cross-correlation $\rho_k(x, y)$, for which the cross-autocovariance function is used:

$$\gamma_k(x, y) = \mathbb{E}[(x_{t+k} - \mu_x)(y_t - \mu_y)] \quad (3.6)$$

In the analysis of time series the correlation of a variable (a time series) with itself, the so-called autocorrelation or serial correlation, plays an important role. It can be seen as cross-correlation where the second variable is just the lagged first one. The value of autocorrelation depends on the autocovariance function, which is defined for lag k in an analogous way to Equation 3.6 (y substituted by x):

$$\gamma_k(x) = \mathbb{E}[(x_{t+k} - \mu_x)(x_t - \mu_x)] \quad (3.7)$$

The autocorrelation function (in R the function `acf`) $\rho_k(x)$ is the autocovariance function divided by the variance of the variable. Partial autocorrelation (`pacf`) corresponds to the autocorrelation function with the difference that here we remove the effect of any correlations due to the terms at shorter lags. The graphical visualisations of the `acf` and `pacf` functions are called correlograms and will be often used throughout the thesis. On the R generated correlograms the dotted lines indicate confidence intervals for $\rho_k = 0$. The formula used to produce these confidence intervals is (Cowpertwait et al., p. 36; n is the length of the time series):

$$-\frac{1}{n} \pm \frac{2}{\sqrt{n}} \quad (3.8)$$

The considered significance level is 5%, thus even in the case of a real autocorrelation function equal to zero at all lags, we would expect 5% of all of the estimates to lie beyond the confidence interval. Since the above formula (Equation 3.8) refers to testing the significance of separate lags, we will also use the Ljung-Box test. It tests whether there is a group of autocorrelations which is different from zero. Because it tests the overall randomness at each distinct lag, the Ljung-Box test is a portmanteau test (a test with well defined null hypothesis and more loosely defined alternative hypothesis). The R implementation of this test is the `Box.test` function with the parameter `type="Ljung-Box"`. The function returns a p-value, which is the probability to obtain a value of the test

statistic at least as extreme as the one which has been observed, assuming the null hypothesis holds. If the p-value is below the standard significance level of 0.05, the null hypothesis can be rejected. The p-values are used in most statistical tests throughout the thesis. Please note that in case of normal distribution no serial correlation is equivalent to independent distribution.

3.3. White noise and random walk

One of the most important concepts of time series is the so called white noise. It is a series notated $\{w_t\}$, whereas t is the time index: $t = 1, 2, \dots, n$. The series $\{w_t\}$ is called discrete white noise (DWN), when w_1, w_2, \dots, w_n are both identically and independently distributed (i.i.d.) with the mean 0. It means that:

- (i) $\mathbb{E}[w_t] = 0$
- (ii) all the w_t have the same variance, equal to some σ^2
- (iii) $cov(w_i, w_j) = 0$ for all $i \neq j$

When w_t additionally comes from a normal distribution, it is called Gaussian white noise. If $\{x_t\}$ is a time series and $x_t = x_{t-1} + w_t$, where $\{w_t\}$ is discrete white noise, then $\{x_t\}$ is called the random walk. An important characteristic of random walk is that after differencing it, one obtains a white noise time series.

To test whether a sample comes from a normal distribution, we will follow a common practice and use the Jarque–Bera test. The null hypothesis of this test is that the skewness of the analysed distribution (measure of asymmetry, third moment of the distribution function) equals zero and the kurtosis (measure of the peakedness, the fourth moment) equals three, just like in the case of a normal distribution. However, if we cannot reject the null hypothesis, it is possible that the sample comes from a non-normal distribution for which only higher moments differ (or that we make Type II error). In the calculations the `jarqueberaTest` from package `fBasics` will be used. In Equation 3.9 the corresponding test statistic has been displayed (Lütkepohl 2007, p. 45). n stands for the sample size and \hat{u}_i^s are the standardized values from the sample (for example residuals from a model). Under the null hypothesis the test statistic follows the $\chi^2(2)$ distribution. We assume the data to be normal if JB is smaller than the corresponding critical value.

$$JB = \frac{n}{6} \left[n^{-1} \sum_{i=1}^n (\hat{u}_i^s)^3 \right]^2 + \frac{n}{24} \left[n^{-1} \sum_{i=1}^n (\hat{u}_i^s)^4 - 3 \right]^2 \quad (3.9)$$

An alternative to the Jarque–Bera test is the Kolmogorov–Smirnov test. It can be used to compare data from a sample (empirical distribution F_n) with a theoretical continuous distribution (F_0), for instance the normal distribution. Equation 3.10 presents the test statistic. Again, we assume the data to follow a normal distribution if the corresponding p-value lies above 0.05, the standard significance level. To conduct the test we will use like previously the `fBasics` package and its `ksnormTest` function (which is a wrapper function regarding `ks.test` from the `stats` package). The function works only for data with a mean of zero and a standard deviation of one (which, however, is not specified in

the documentation, but follows from an analysis of the R source code). To achieve this, we can standardize the values, for example using the basic `scale` function. Since we are going to operate on small samples, the probability of making a Type II error is high and thus both normality tests should be applied.

$$D_n = \sup|F_n(X) - F_0(X)| \tag{3.10}$$

3.4. Stationarity

A series $\{x_t\}$ is strictly stationary when the joint statistical distribution of $(x_{t_1}, \dots, x_{t_n})$ does not differ from the joint distribution of $(x_{t_1+\tau}, \dots, x_{t_n+\tau})$ for all t_1, \dots, t_n and for all τ (Chan 2002, p. 16). However, this is a very restrictive condition. For us of more interest is the concept of second-order stationarity, the so-called weak stationarity. The conditions for a time series to be weakly stationary are:

- (i) $\mathbb{E}[x_t] = \mu$ for all t
- (ii) $cov(x_t, x_{t+\tau}) = \gamma(\tau)$ for all t and for all τ

Whilst strict stationarity implies weak stationarity, the converse relation does not hold in general except for the normal distribution (Chan 2002, p. 17). Weakly stationary time series are thus characterized by mean and variance which are constant over time, whereas the covariance depends only on the lag. In the following, whenever we use the term stationarity we will refer to weak stationarity. In order to be able to apply some time series models, like ARMA, stationarity is a requirement.

There are many statistical approaches regarding testing stationarity of a time series. We will apply both ADF and KPSS tests. Augmented Dickey Fuller (ADF) is based on the simple Ordinary Least Squares regression of the analysed time series on its own lagged values (for which higher-order autoregressive terms are used). In the regression both a constant and a time trend are considered. Calculating higher-order AR terms in the ADF test controls for serial correlation and distinguishes the procedure from the older simple Dickey–Fuller test (Hamilton 1994, p. 516). The null hypothesis of the ADF test is the non-stationarity of the time series. In the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test the null hypothesis of no unit roots corresponds to the stationarity of the data. The series is expressed as the sum of deterministic trend, random walk and stationary error. The test is the Lagrange multiplier test for which the null hypothesis says that the corresponding random walk has a variance of zero.

For the tests we will use package `fUnitRoots` and its `adfTest` and `urkpssTest` functions. In case of `adfTest` we have to specify the argument `lags`: the maximum number of lags used for error term correction (by default it is one). In order to analyse possible seasonal effects we consider `lags` to be five. When applying the `adfTest` function we obtain a p-value, whereas for `urkpssTest` R gives the value of the test statistic and four critical values: 0.347, 0.463, 0.574 and 0.739 for the 10%, 5%, 2.5% and 1% significance levels respectively. We reject the null hypothesis when the value of the test statistic is large.

3.5. AR(p) and MA(q)

A time series $\{x_t\}$ which follows an autoregressive process of order p can be defined as:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + w_t \quad (3.11)$$

where $\{w_t\}$ is discrete white noise, it means it has mean equal to zero and variance equal to σ^2 . The AR process can be explained using the backward shift operator \mathbf{B} (θ_p is a polynomial of order p):

$$\theta_p(\mathbf{B})x_t = (1 - \alpha_1 \mathbf{B} - \alpha_2 \mathbf{B}^2 - \dots - \alpha_p \mathbf{B}^p)x_t = w_t \quad (3.12)$$

We call the $\theta_p(\mathbf{B}) = 0$ a characteristic equation, for which all roots (both real and complex) must exceed unity in absolute value for the series to be stationary. For the analysis of the order of an AR process the `pacf` graphs can be used, since the `pacf` at lag k is the k th coefficient of a fitted $AR(k)$ model (Cowpertwait et al., p. 81). Apart from it, we can use the `ar` function (package `stats`). The calculated parameters refer to a model in which we subtract the mean from every element, so that the new series has a mean of zero. This has to be considered, although in case of modelling the residuals from a regression model we already have a mean of zero. Assuming normality of the data and using the asymptotic variance of the parameter estimates, we can calculate with the following formula the 95% confidence intervals of the AR parameters (Cowpertwait et al., p. 84; in the book 2 has been used as an approximation of 1.96, i.e. the 97.5 percentile point of the normal distribution; \mathbb{AV} corresponds to the asymptotic variance):

$$\left[\hat{\alpha}_i - 1.96 * \sqrt{\mathbb{AV}[\hat{\alpha}_i]}; \quad \hat{\alpha}_i + 1.96 * \sqrt{\mathbb{AV}[\hat{\alpha}_i]} \right] \quad (3.13)$$

A time series $\{x_t\}$ follows a moving average process of order q when:

$$x_t = \beta_1 w_{t-1} + \beta_2 w_{t-2} + \dots + \beta_q w_{t-q} + w_t \quad (3.14)$$

where $\{w_t\}$ is again DWN. Here we can use the backward shift operator \mathbf{B} too (ϕ_q is a polynomial of order q):

$$\phi_q(\mathbf{B})w_t = (1 + \beta_1 \mathbf{B} + \beta_2 \mathbf{B}^2 + \dots + \beta_q \mathbf{B}^q)w_t = x_t \quad (3.15)$$

A moving average model is said to be invertible if it can be expressed as a stationary autoregressive process of infinite order without an error term (Cowpertwait et al., p. 123). This can be understood as though the roots of $\phi_q(\mathbf{B})$ all exceed unity in absolute value.

3.6. SARIMA(p,d,q)(P,D,Q)

Autoregressive moving average (ARMA) is the combination of AR and MA processes. It is defined in the following way:

$$x_t = \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + \beta_1 w_{t-1} + \dots + \beta_q w_{t-q} + w_t \quad (3.16)$$

It is stationary and invertible in the same sense as AR and MA processes, but it can often be the case that the best fitted ARMA model requires fewer parameters than a single AR or MA process (Cowpertwait et al., p. 127).

Usually the time series we want to analyse is not stationary, e.g. it can include trends. Trends can be either deterministic, like in the case of a linear trend, or stochastic, like in the case of a random walk. Differencing the series can remove the trends. We call a time series an autoregressive integrated moving average (ARIMA) process of order (p,d,q) , when after differencing the series d -times, it follows an ARMA (p,q) process.

If there are seasonal effects in the data, the ARIMA model can be extended to include them. A seasonal ARIMA model (SARIMA) is defined using six parameters $(p,d,q)(P,D,Q)$, where (P,D,Q) refer to the previous cycle (last year for example). It is a non-stationary time series model (Cowpertwait et al., p. 137).

Function `auto.arima` from package `forecast` returns the best SARIMA model according to an information criterion. We will choose the commonly used Akaike information criterion (see Equation 3.23; in the VAR section there is an explanation of the notation, although the formula refers to the normally distributed data), so that the parameter `ic="aic"`. The parameter `stepwise` is by default `TRUE`, because the non-stepwise selection can be very slow, particularly in case of seasonal models (reference manual of the 'forecast' package; see the bibliography). We will specify, however, `stepwise="FALSE"`, so that the search for the best model will consider all the models within the order constraints. As we analyse quarter data, the default values for the order constraints are too high. We will use `max.p=3`, `max.q=3`, `max.P=1`, `max.Q=1`. The order of first differencing (`d`) will be chosen automatically by the function using the KPSS test. The order of seasonal differencing (`D`) will be assumed to be zero.

The same package includes function `forecast`, which enables for the model generated by `auto.arima` the calculation of forecasted values for the next `h` periods. The parameter `level` refers to the significance level, in case of time series forecasting the standard values are 80 and 95.

3.7. Regression

Centred moving average and some other popular smoothing procedures (as `loess` for example) do not produce a formula which we could extrapolate in order to make a forecast (Cowpertwait et al., p. 22). Another approach is to find a deterministic trend, for instance a regression line. If the forecast is to be made for the next few periods, the trend is unlikely to change in a dramatic way. However, when the autocorrelation of the error terms is positive, the standard errors of the estimated OLS regression parameters tend to be underestimated, thus the calculated significance of statistical tests can be false.

An alternative to OLS could be to apply the generalized least squares (GLS) regression. Since we know neither the order of the autocorrelation nor the values of its parameters, we have to estimate those and thus we shall use feasible generalized least squares (FGLS). Since in reality one hardly ever knows the real autocorrelation structure, estimates are

used and in the literature GLS can be sometimes found as a synonym for FGLS. To estimate the autocorrelation of the error terms the `ar` function (or alternatively the `pacf` function) will be applied on the residuals from the OLS regression. It is very often the case that autocorrelation is only of the first order and thus we analyse only ρ (corresponds to α_1 in Equation 3.11), for which the estimate is $\hat{\rho}$. In the formula for the estimated regression coefficients the $\hat{\Omega}$ matrix has to be used, see Equation 3.17 (X stands for the matrix which includes the regressors, and y is the dependent variable). The notation follows Hamilton (1994, p. 220 – 222).

$$\beta_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}y), \text{ where } \hat{\Omega} = \frac{1}{1 - \rho^2} \begin{bmatrix} 1 & \hat{\rho} & \hat{\rho}^2 & \dots & \hat{\rho}^{T-1} \\ \hat{\rho} & 1 & \hat{\rho} & \dots & \hat{\rho}^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}^{T-1} & \hat{\rho}^{T-2} & \hat{\rho}^{T-3} & \dots & 1 \end{bmatrix} \quad (3.17)$$

Because of the estimated Ω within the formula, FGLS is not a linear estimator. The estimation of Ω is unbiased, but the inverse of $\hat{\Omega}$ which is used in the FGLS estimation is biased, thus FGLS is biased in finite samples. OLS is more straightforward and that is why we will use the OLS estimates of the parameters and FGLS will only be used in order to check the significance of the parameters. This approach is often followed in the literature (e.g. Cowpertwait et al., p. 109–115). To conduct the OLS regression we will use the `lm` function from the standard `stats` package. For the FGLS regression function `gls` from package `nlme` will be applied. Its parameter `method` is by default "REML" and gives a fit by maximizing the restricted log-likelihood function. However, the formula explained above refers to the "ML" method. We will use throughout the work the default.

The regression model (Equation 3.18) itself is based on Equation 3.1, where for m_t we impute $\alpha \cdot t$ to account for linear increase or decrease throughout our time period. As s_t we consider a factorial variable of 4 different levels referring to 4 quarters.

$$x_t = \alpha t + s_t + z_t = \begin{cases} \alpha t + \beta_1 + z_t & t = 1, 5, \dots \\ \alpha t + \beta_2 + z_t & t = 2, 6, \dots \\ \alpha t + \beta_3 + z_t & t = 3, 7, \dots \\ \alpha t + \beta_4 + z_t & t = 4, 8, \dots \end{cases} \quad (3.18)$$

We often have data with an exponentially increasing trend, which can be described as in Equation 3.2: $x_t = m'_t \cdot s'_t \cdot z'_t$. When we take logarithms of both sides of it we get: $y_t = \log x_t = \log m'_t + \log s'_t + \log z'_t = m_t + s_t + z_t$ (Cowpertwait et al., p. 109). The right-hand-side corresponds to Equation 3.1, and also represents a linear model which we can estimate using OLS and FGLS.

Instead of using separate indices for the seasonal effects, we could use smooth functions, for example sine and cosine functions, which are part of the so-called harmonic seasonal model. The advantage of using that model is the parameter-efficiency in comparison to separate indices (Cowpertwait et al., p. 101). The model has been derived from the deconstruction of the sine function:

$$A \sin(2\pi ft + \phi) = \alpha_s \sin(2\pi ft) + \alpha_c \cos(2\pi ft) \quad (3.19)$$

where f is the frequency of the wave (number of cycles within a sampling interval), A is amplitude, ϕ phase shift, $\alpha_s = A \cos \phi$ and $\alpha_c = A \sin \phi$. The sum of the sine and cosine functions is a linear expression, where α_s and α_c are the parameters to be estimated, again either with OLS or FGLS. Referring to Equation 3.1 and the notation used in Cowpertwait et al.:

$$x_t = m_t + \sum_{i=1}^{\lfloor s/2 \rfloor} \{s_i \sin(2\pi it/s) + c_i \cos(2\pi it/s)\} + z_t \quad (3.20)$$

where m_t is the trend including intercept, s_i and c_i are parameters to be estimated and s is the number of seasons. Since in our case s is an even number, the sine term at the frequency of $1/2$ assumes the value of zero and can thus be omitted.

3.8. Structural breaks

An essential part of the validity inspection of a model is the search for possible structural breaks within the considered time period. It can be seen as well as a way to detect outliers in the data. For this purpose the Chow test can be applied. In the literature three common variants of this test are reported: sample-split, break-point and forecast tests. We will use the forecast test, where the approach is to compare residual variance from the full sample (i.e. for the entire time period) with the residual variance from the first subperiod. The corresponding test statistic is (Lütkepohl 2007, p. 49):

$$\lambda_{CF} = \frac{T\hat{\sigma}_u^2 - T_1\hat{\sigma}_{(1)}^2}{T_1\hat{\sigma}_{(1)}^2} \cdot \frac{T_1 - K}{T - T_1} \quad (3.21)$$

where K denotes the number of regressors in the restricted and stable model, T stands for the full sample size and T_1 for the size of the first subperiod. Residual variance from the full sample is $\hat{\sigma}_u^2$ and for the first subperiod $\hat{\sigma}_{(1)}^2$. Under the null hypothesis of the parameter constancy the test statistic of this Chow test follows an approximate $F(T - T_1, T_1 - K)$ distribution. When the value of the test statistic lies above the corresponding critical value from the F distribution (we will use a significance level of 5%), the null hypothesis should be rejected. Estimation of the p-values via bootstrapping (resampling) is possible, but will not be applied to our data.

3.9. VAR(p)

An extension of the concept of AR processes are the vector autoregressive processes (VAR), where each time series from a given set of series depends on the lags up to lag p of all the series in the set. The mathematical notation is presented in Equation 3.22. Following the notation of Lütkepohl 2005 (p. 69) $y_t = (y_{1t}, \dots, y_{Kt})'$, where K is the dimension of the multiple time series, it means the number of time series we model. t is a time index, assuming values from 1 to T , thus T is the sample size (the number of the time periods). $\nu = (\nu_1, \dots, \nu_K)'$ is the intercept vector, A_i corresponds to the values within the $(K \times K)$ coefficient matrix and u_t is the white noise term with a nonsingular covariance matrix Σ_u . y_t is known to be generated by a both stationary and stable VAR(p) process.

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \quad (3.22)$$

To analyse VAR processes we will use the R package `vars`. Function `VAR` estimates via OLS per equation the values of the VAR parameters (assuming the order is equal to a given p). `VARselect` will be used to choose the order of the VAR process. Under the assumption of normally distributed data it considers four different criteria: Akaike information criterion (AIC), Hannan–Quinn criterion (HQ), Schwarz criterion (SC) and final prediction error (FPE), which one can see in Equations 3.23– 3.26. The order of the VAR process fitted to the data is denoted by m and $\tilde{\Sigma}_u(m)$ is the maximum likelihood estimation of the white noise covariance matrix (notation corresponds to Lütkepohl 2005, p. 146–150).

`VARselect` sequentially increases the analysed order m up to the specified value of the parameter `lag.max`, which has by default the value 10, and for each m calculates the values of the four criteria (reference manual of the 'vars' package, see the bibliography). After considering all values for m it returns for each criterion the optimal order (i.e. for which the criterion has the smallest value). Because for the given `lag.max` for every m the same sample for y is taken (without the first `lag.max` elements), considering higher values of `lag.max` can return a smaller optimal order than the optimal order for a smaller `lag.max`.

$$\text{AIC}(m) = \log |\tilde{\Sigma}_u(m)| + \frac{2mK^2}{T} \quad (3.23)$$

$$\text{HQ}(m) = \log |\tilde{\Sigma}_u(m)| + \frac{2 \log \log T}{T} mK^2 \quad (3.24)$$

$$\text{SC}(m) = \log |\tilde{\Sigma}_u(m)| + \frac{\log T}{T} mK^2 \quad (3.25)$$

$$\text{FPE}(m) = \left[\frac{T + Km + 1}{T - Km - 1} \right]^K \det \tilde{\Sigma}_u(m) \quad (3.26)$$

Within the `vars` package there are more functions which we will use. To make forecasts based on the model generated by the `VAR` function the `predict` function can be applied, where we have to specify the number of time periods ahead (the parameter `n.ahead`), as well as the significance level `ci`, which by default is 0.95. For a visual representation of the forecasts the `fanchart` function is helpful. It plots the mean predictions and the corresponding confidence intervals. As the forecasts are made on the assumption of normal distribution of the error terms, we have to check it in a multivariate way, for example with the `normality.test`, which is a generalization of the already discussed Jarque–Bera test. Even if the error terms are normally distributed, it is not clear whether they are independent. For that we will apply the `serial.test` function with the parameter `type="PT.asymptotic"`, which corresponds to the multivariate portmanteau test.

Some time series exhibit changing variance over time: the so-called volatility clustering, where periods of lower volatility can be followed by periods of higher volatility or the other way round. In such cases apart from modelling the mean, we can also be interested in forecasting changes in the variance. For this purpose autoregressive conditional heteroscedasticity (ARCH) models are used. In Equation 3.27 an univariate ARCH(m) process of an error term w_t has been shown (Hamilton 1994, p. 659 and 665). v_t is i.i.d. with zero mean and variance equal to one.

$$w_t = \sqrt{h_t} \cdot v_t, \quad h_t = \zeta + \delta_1 w_{t-1}^2 + \delta_2 w_{t-2}^2 + \dots + \delta_m w_{t-m}^2 \quad (3.27)$$

For univariate analysis a visual check suffices, for a VAR model we will use a numerical test: multivariate `arch.test` (again package `vars`). The null hypothesis of the corresponding univariate test is that w_t is i.i.d. $N(0, \sigma^2)$ (Hamilton 1994, p. 664 and 665). This test looks at the R^2 of the OLS regression of \hat{w}_t on its past m values, which multiplied by the sample size T converges in distribution to a χ_m^2 variable.

3.10. Granger causality

Within a VAR(p) framework we can check for causalities between the variables. A popular concept is here the Granger causality, which exists between two variables y_{1t} and y_{2t} if one of the variables helps to improve the forecasts of the other one (in a statistically significant way). If we denote the optimal h -step prediction of y_{1t} at time t based on the set of all relevant information in the universe Ω_t as $y_{1,t+h|\Omega_t}$, then y_{2t} does not Granger cause y_{1t} if and only if (Lütkepohl 2007, p. 144):

$$y_{1,t+h|\Omega_t} = y_{1,t+h|\Omega_t \setminus \{y_{2,s} | s \leq t\}}, \quad h = 1, 2, \dots \quad (3.28)$$

This causality concept considers the correlation between a value of one variable with the lagged values of another. To test the significance of these correlations conventional F-tests are carried out to verify if the null hypothesis of the corresponding coefficients in the VAR framework (for the bivariate case see Equation 3.29, sometimes additionally constants are considered) being 0 can be rejected. If the null hypothesis is rejected, it speaks for a Granger causality. In fact a proved Granger causality does not necessarily mean that such a causality exists. It is always possible that the two considered series are causally affected by another, which we have not considered. The Granger causality is mainly applied for differenced series (where the stationarity should be tested). To work with the level variables we could use the cointegration techniques, which are not part of this work.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \sum_{i=1}^p \begin{bmatrix} \alpha_{11,i}, \alpha_{12,i} \\ \alpha_{21,i}, \alpha_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix} + u_t \quad (3.29)$$

To test for causality we can either use the `grangertest` from package `lmtest` or the function `causality` from the already discussed package `vars`. In the first case we only consider bivariate series and the test is a Wald test, in which the unrestricted model where y is explained by its own lags and the lags of x (up to the specified `order`) is compared with the restricted model where y is explained only by its own lags. The second function, `causality`, is based on the F-test and is generally more sophisticated. First of all, it allows the analysis of multivariate series. As an argument it gets an object generated by the `VAR` function. Through the `cause` parameter the user specifies the variables whose influence on the rest should be tested. For both the functions the null hypothesis is that there is no causal relation between the variables. The function `causality` conducts additionally the instantaneous causality test, but we will not make use of it.

4. Univariate Time Series Analysis

4.1. Gross Written Premiums

In Figure 4.1 the red line shows the values of Gross Written Premiums. After general visual inspection an increasing trend in the data is obvious. Moreover, we see seasonal variation is not strong (but exists) and the year 2008 seems to be an outlier. The increasing trend looks like an exponential one and we will deal with it as such. The orange line represents the OLS regression where seasons have been treated as dummy variables. The harmonic seasonal model corresponds to the green line. Because $SIN[1]$ and $COS[2]$ are insignificant (t -values of the FGLS regression 0.620 and 0.763, respectively), we consider only three variables: $time(GWP)$, intercept and $COS[1]$. Values of the OLS regressions' parameters and the t -values from the FGLS regressions are displayed in Table 4.1. The residuals from both the models follow an $AR(1)$ process (see the corresponding correlograms in the Appendix, Figures A.1 and A.2). The α_1 parameters are 0.658 and 0.641.

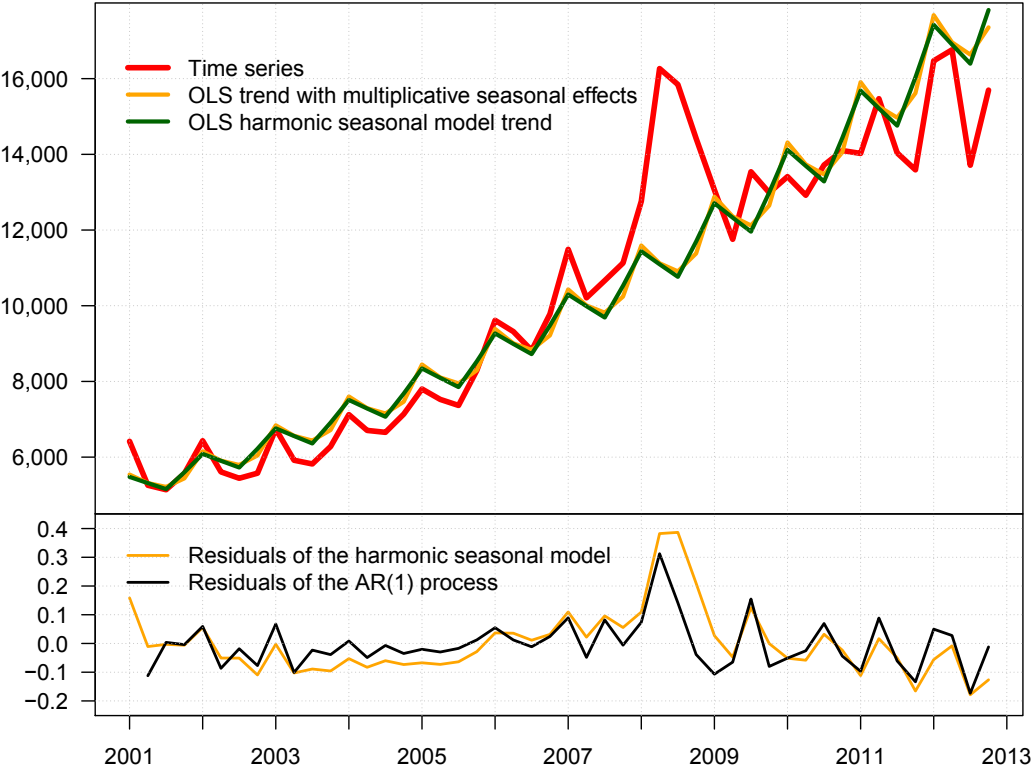


Figure 4.1.: Gross Written Premiums [million PLN], quarter data

Model with dummies	OLS parameter	FGLS t-value	Harmonic model	OLS parameter	FGLS t-value
time(GWP)	0.10549	8.077	time(GWP)	0.10520	7.829
quarter 1	-202.46362	-7.701	COS[, 1]	0.05671	3.608
quarter 2	-202.53142	-7.703	intercept	-201.94997	-7.462
quarter 3	-202.57718	-7.705			
quarter 4	-202.56117	-7.704			

Table 4.1.: OLS and FGLS regressions for the GWP trend

When we compare the two OLS regression models using the Akaike information criterion, the superiority of the harmonic seasonal model over the one with dummy variables can be confirmed (AIC of -65.594 as compared to -62.659 for the model with dummy variables). The variance of the AR residuals is, in turn, smaller for the model with dummies (0.00672 instead of 0.00706 for the harmonic seasonal model). Since it is difficult to say which model is better, we proceed to the validation tests of both of them.

As can be seen in the correlograms (Figures A.3 and A.4 in the Appendix) there are no significant lags, so we may assume the AR error terms from both models are independently distributed. The next issue we should consider is the possible normal distribution of the AR error terms. The normality of the error terms would make the forecasting easier, although within the methodological framework of the work no appropriate methods have been introduced to make forecasts based on combined models. For the Jarque–Bera test we obtain p-values below 0.01 for both the models and for the Kolmogorov–Smirnov test the p-values are 0.604 and 0.475 for the model with dummy variables and for the harmonic one, respectively. All in all, it is uncertain which model is the better one, but because of the parameter efficiency we will use in the multivariate analysis the harmonic model.

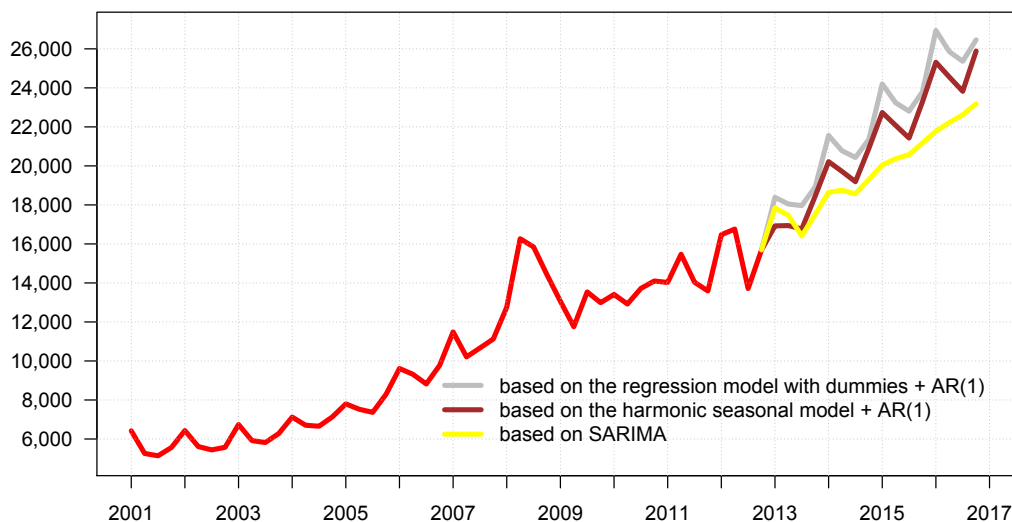


Figure 4.2.: Forecasted values of Polish GWP till the end of 2016 [million PLN]

According to the derived values of the model parameters, Equation 4.1 for the harmonic seasonal model has been constructed (the intercepts from the trend model inside and outside the AR process have been considered together as one general intercept). Using it I have calculated forecasts of GWP till the end of 2016. The numerical results till the end of 2014 have been presented in Table 4.2 and the graphical ones till the end of 2016 can be seen in Figure 4.2. The table and the plot consider additionally the model with dummy variables. In all the cases the predicted values based on the harmonic seasonal model are smaller.

$$\begin{aligned} \text{GWP}_{t+1} = & \exp \{ 0.105198 \cdot \text{time}(\text{GWP}_{t+1}) + 0.05671 \cdot \cos(2 \cdot \pi \cdot \text{time}(\text{GWP}_{t+1})) \\ & + 0.641 \cdot [\log(\text{GWP}_t) - (0.105198 \cdot \text{time}(\text{GWP}_t) + 0.05671 \cdot \cos(2 \cdot \pi \cdot \text{time}(\text{GWP}_t)))] \\ & - 72.50004 \} \end{aligned} \tag{4.1}$$

Apart from models based on a deterministic trend we can consider the seasonal autoregressive integrated moving average processes (SARIMA). Function `auto.arima` with the specified parameters following the assumptions formulated in the 'Methodology' part of the work suggests to choose SARIMA(0,1,2)(1,0,0)_[4]. The corresponding forecasts can be found in the same graph and table as for the regression models (Figure 4.2 and Table 4.2). Because in this case we consider a single model (not a combination of two, like previously), it is straightforward to calculate confidence intervals of the predicted values. One can see those for the time period till the end of 2016 in Figure 4.3 (the considered significance level is 80%). According to the Jarque–Bera test, the model's error terms are normally distributed (p-value of 0.153). The Kolmogorov–Smirnov test confirms the hypothesis of normality, giving a p-value of 0.905. The confidence intervals are thus reliable. When it comes to the comparison of the SARIMA model with the regression ones, no explicit answer can be given, due to different models we cannot apply Akaike information criterion.

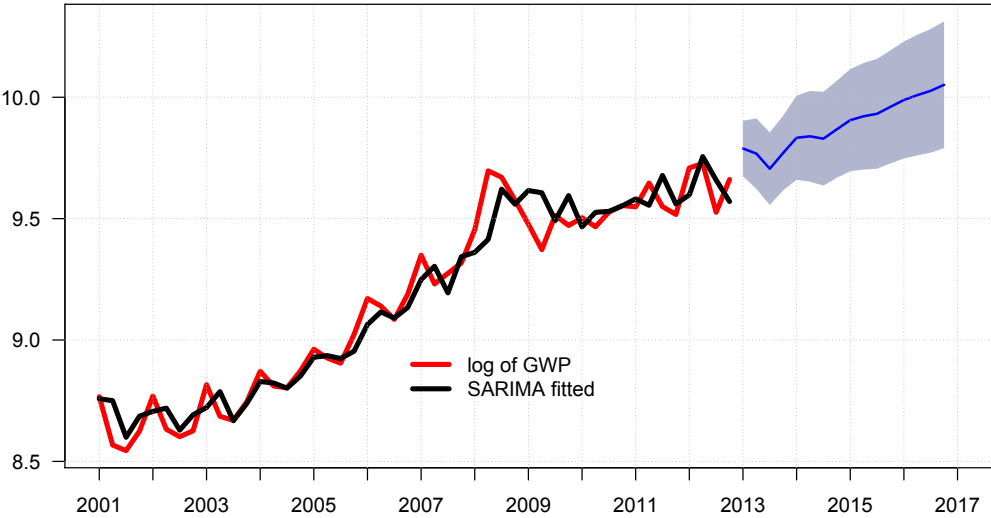


Figure 4.3.: Forecasts of log(GWP) based on the SARIMA model. The blue area corresponds to the confidence interval of the 80% significance level

	'13 Q. 2	'13 Q. 3	'13 Q. 4	'14 Q. 1	'14 Q. 2	'14 Q. 3	'14 Q. 4
dummies:	18045.0	17963.6	18925.7	21561.0	20773.9	20432.0	21355.8
harmonics:	16940.5	16767.0	18454.4	20217.8	19716.3	19190.8	20897.1
SARIMA:	17463.4	16407.6	17513.7	18642.6	18752.2	18569.6	19302.1

Table 4.2.: Forecasts of the development of the Polish GWP in 2013 and 2014 [million PLN], based on the dummies model, the harmonic model and SARIMA, respectively. For '13 Q. 1 the values are 18388.0, 16924.0 and 17837.1

However, the year 2008 seems to be very different from the rest. GWP was in 2008 much above the trend and the previously stable trend with stable seasonal pattern changed then. The extraordinarily high values in 2008 were an effect of fast developing Polish economy and relatively cheap insurance products on the market. We would like to test whether it corresponds to a structural break within the analysed harmonic seasonal model. In Figure 4.4 one can see the test statistics and the critical values of the Chow forecast test for all the possible structural breaks from the beginning of 2003 till the end of 2010. Till the second quarter of 2008 the test statistics lie above the critical values from the F distribution, thus allowing the rejection of the null hypothesis of the parameter constancy. The later values lie below the critical values of the test, what sounds reasonable, since then the considered subperiod (which we compare with the entire model) already includes year 2008. The exact quarter of the structural break is still not clear, but year 2008 changed the trend indeed. In the Appendix a corresponding plot with the results of the Chow forecast test for the OLS regression model with dummy variables has been shown (Figure A.6). The two plots look much the same. Since we are interested in estimating a trend, as well as in making forecasts and in analysis of the data in a multivariate way, we will not consider it.

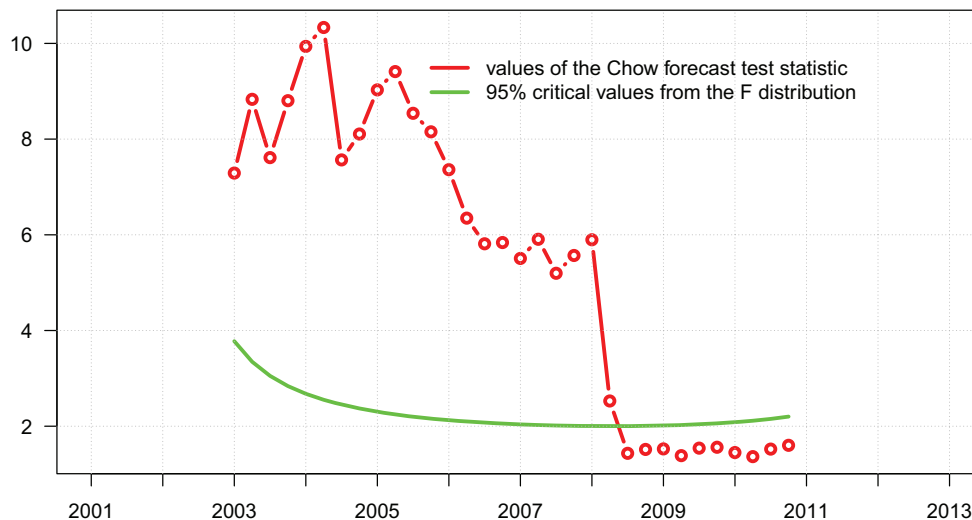


Figure 4.4.: Application of the Chow forecast test to detect structural breaks within GWP (harmonic model)

4.2. Profitability ratio of technical activity

In Figure 4.5 the red line shows the values of the profitability ratio within the last twelve years. The p-values of the Jarque–Bera and Kolmogorov–Smirnov tests are 0.498 and 0.447, so the data is normally distributed. For an OLS regression on time where the seasonal effects are represented as dummy variables all the parameters are insignificant (the absolute values of all the five t-values are below 1.96). The regression only on time leads to insignificant parameters too, where again the absolute values of the t-values are below 1.96. For a regression considering only the four dummy variables we get significant parameters, but this regression explains only 6.6% of the variance within the data. As can be seen in the correlograms (Figure A.7) the residuals from the regression model involving only dummy variables for the quarters are not independently distributed and so we continue the search for an appropriate model.

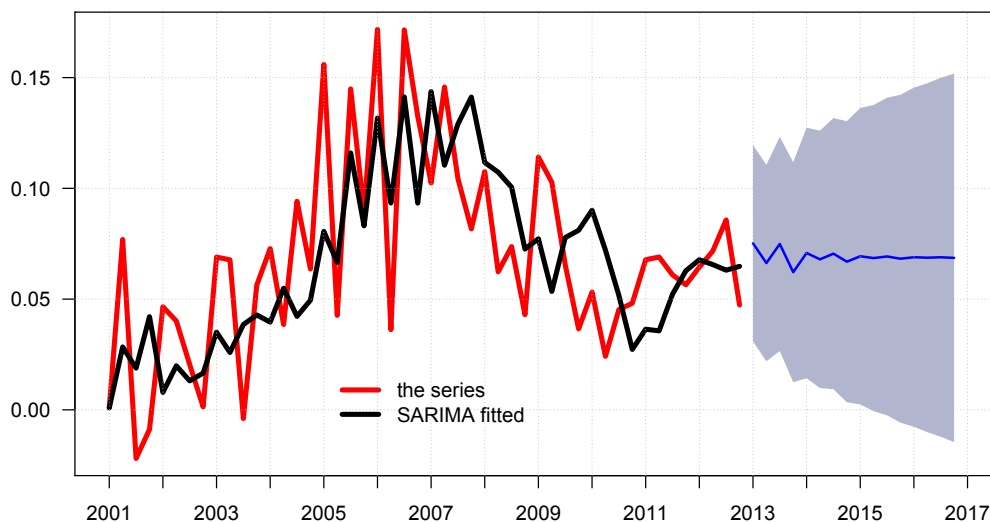


Figure 4.5.: Forecasts of EFF based on the SARIMA model. The blue area corresponds to the confidence interval of the 80% significance level

The function `auto.arima` with values of the parameters following the specification in the 'Methodology part' of the work suggests to apply $SARIMA(1,1,1)(1,0,0)_{[4]}$. Based on this model we can calculate the forecasts till the end of 2016. These are shown in Figure 4.5. The correlograms for the residuals of this SARIMA model can be seen in the Appendix (Figure A.8). Because there is no significant lag, the independence of the error terms can be confirmed. As for the normality tests regarding the error terms, we get a p-value of 0.323 for the Jarque–Bera test and a p-value of 0.546 for the Kolmogorov–Smirnov test. Based on the proved normality of the error terms the confidence intervals for the forecasts have been constructed (see Figure 4.5). The forecasted mean values for EFF are approximately 0.070. For the four quarters of 2013 the exact values are 0.075, 0.066, 0.075 and 0.062.

4.3. Gross Domestic Product

As can be seen from Figure 4.6, Polish Gross Domestic Product varies strongly between quarters. For the last twelve years Polish GDP always has had its highest value in the fourth quarter and the lowest value for the first one. Apart from an increasing trend, which explains this feature in part, the fourth quarter includes Christmas holidays when people tend to spend much more money than usual. Moreover, at the end of the year companies try to improve their financial results, having in mind the performance reviews at the beginning of the following year.

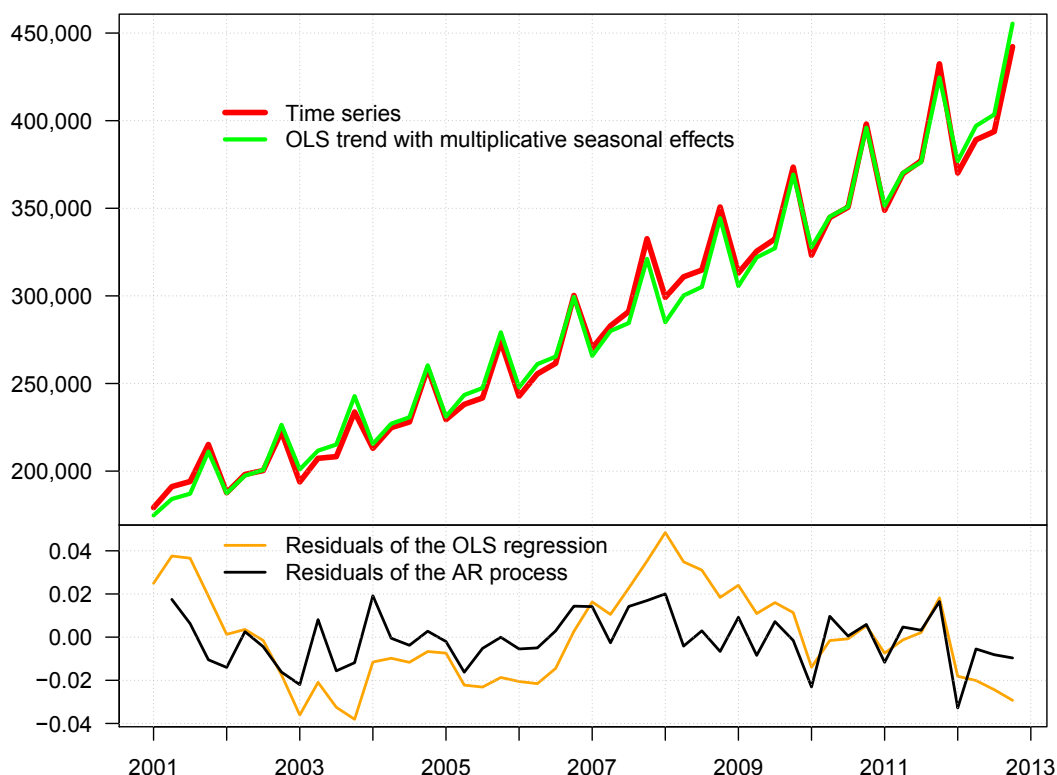


Figure 4.6.: GDP time series [million PLN], quarter data

In the graph one can observe a slightly exponentially increasing trend. After general graphical inspection we conclude that seasonal effects obviously exist. This allows us to estimate the trend using linear regression with dummy variables representing different quarters. The explanatory variable is time, whereas the dependent variable is the logarithm of GDP. We compute the OLS regression, which gives us significant parameters, but for which residuals are serially correlated (specifically, they follow an AR(1) process with $\alpha_1 = 0.806$). Thus we conduct a FGLS regression, for which we obtain slightly different and still significant terms. In Table 4.3 there are the values of the parameters of the OLS regression for the model with dummy variables, as well as values of the parameters of the OLS regression for the harmonic seasonal model. Next to every estimated parameter the t-value for a corresponding FGLS parameter is shown. Because the absolute value of each of these t-values is above 1.96, all the parameters (of both models) can be considered to be significant. Using trigonometric functions often allows the reduction of the number of

parameters, which is not the case for our GDP series. That is why we choose to apply the model with dummy variables, which is generally easier in use.

Model with dummies	OLS parameter	FGLS t-value	Harmonic model	OLS parameter	FGLS t-value
time(GDP)	0.06987	14.684	time(GDP)	0.06987	14.684
quarter 1	-127.74097	-13.299	intercept	-127.68981	-13.293
quarter 2	-127.70640	-13.294	SIN[, 1]	-0.05109	-28.255
quarter 3	-127.70768	-13.294	COS[, 1]	-0.01664	-9.650
quarter 4	-127.60422	-13.283	COS[, 2]	-0.03451	-37.921

Table 4.3.: OLS and FGLS regressions for the GDP trend

The orange line in Figure 4.6 represents the residuals from the OLS regression (based on the model with dummy variables). Even visually their autoregressive behaviour is clear. The residuals of the AR process are depicted with the black line. They are normally distributed: p-values of the Jarque–Bera and Kolmogorov–Smirnov tests are 0.765 and 0.968, respectively. In Figure 4.7 one can see the correlograms from the AR(1) process. The only significant lag is at the eleventh quarter, which is probably random. To be sure we conduct the Ljung–Box test, which gives a p-value of 0.805, so we assume there is no serial correlation. Since no multiples of four are significant, our model has fully accounted for seasonal variation.

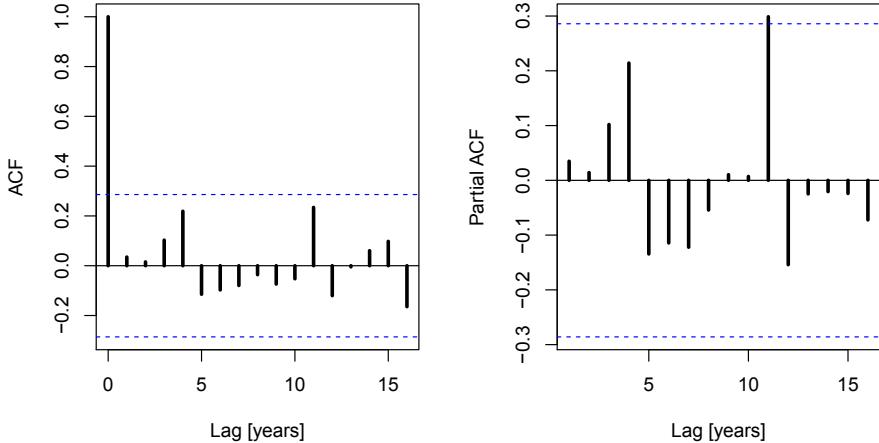


Figure 4.7.: Correlograms of the residuals from the AR process which has been computed for the residuals from the GDP OLS regression

The following equation summarizes the model to be used while forecasting. It considers the parameters of the OLS regression for the model with dummies, as well as the autocorrelation of the error terms. Variable $\text{time}(\text{GDP}_t)$ corresponds to the year and quarter, e.g. for the fourth quarter of 2012 it would assume the value of 2012.75 and for the first quarter of 2013 it would be 2013.00.

$$\begin{aligned}
\text{GDP}_{t+1} = \exp & \left(0.06987 \cdot \text{time}(\text{GDP}_{t+1}) - \begin{cases} 127.74097 & \text{if } t+1 \cong \text{Q. 1} \\ 127.70640 & \text{if } t+1 \cong \text{Q. 2} \\ 127.70768 & \text{if } t+1 \cong \text{Q. 3} \\ 127.60422 & \text{if } t+1 \cong \text{Q. 4} \end{cases} \right) \\
& + 0.806 \left[\log(\text{GDP}_t) - 0.06987 \cdot \text{time}(\text{GDP}_t) + \begin{cases} 127.74097 & \text{if } t \cong \text{Q. 1} \\ 127.70640 & \text{if } t \cong \text{Q. 2} \\ 127.70768 & \text{if } t \cong \text{Q. 3} \\ 127.60422 & \text{if } t \cong \text{Q. 4} \end{cases} \right]
\end{aligned} \tag{4.2}$$

Based on the above formula, I have calculated forecasts for all quarters from 2013 to 2016. They are included in Figure 4.8. Numerical values of these forecasts till the end of 2014 are shown in Table 4.4. First line refers to forecasts made using the knowledge available prior to '13 Q. 1, while second line refers to the forecasts made when already knowing the value for that time period. Forecasts regarding time after 2013 should be treated only as a scenario.

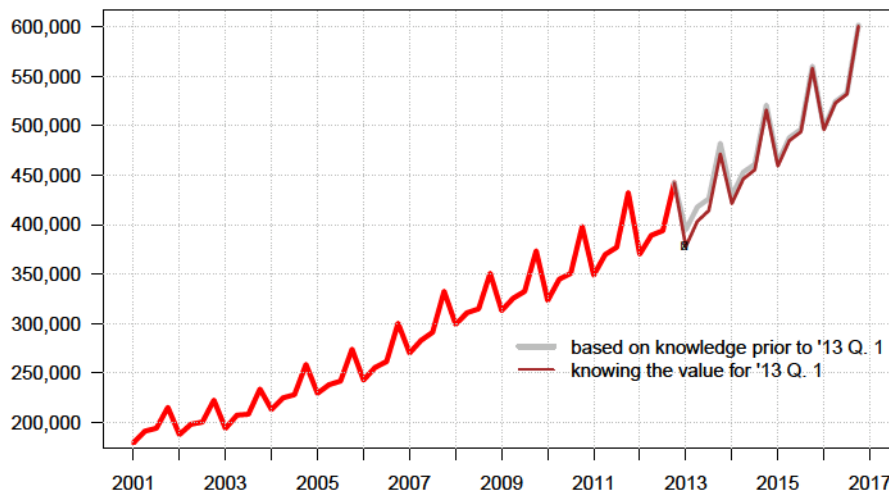


Figure 4.8.: Forecasted values of Polish GDP till the end of 2016 [million PLN]

	'13 Q. 2	'13 Q. 3	'13 Q. 4	'14 Q. 1	'14 Q. 2	'14 Q. 3	'14 Q. 4
<'13 Q. 1	417732.9	426117.6	482326.6	429119.0	452915.7	461022.9	520940.4
>'13 Q. 1	403234.3	414155.0	471381.4	421251.9	446210.4	455513.0	515915.8

Table 4.4.: Forecasted values of Polish GDP till the end of 2014 [million PLN], not knowing and knowing the value for '13 Q. 1 (394739.6)

4.4. Inflation

In Figure 4.9 the red line indicates quarter inflation rates in Poland. While trying to conduct a linear regression on time, we get insignificant coefficients (already in case of OLS). The hypothesis of there being no trend is not supported by the ADF test (p-value of 0.286, however, for `lags=1` it lies below 0.01), but it is confirmed by the KPSS test (value of the test statistic 0.183). This is why we will not use the `decompose` function and the reference point to calculate the seasonal effects will be the mean, which has the value of 0.691 [%].

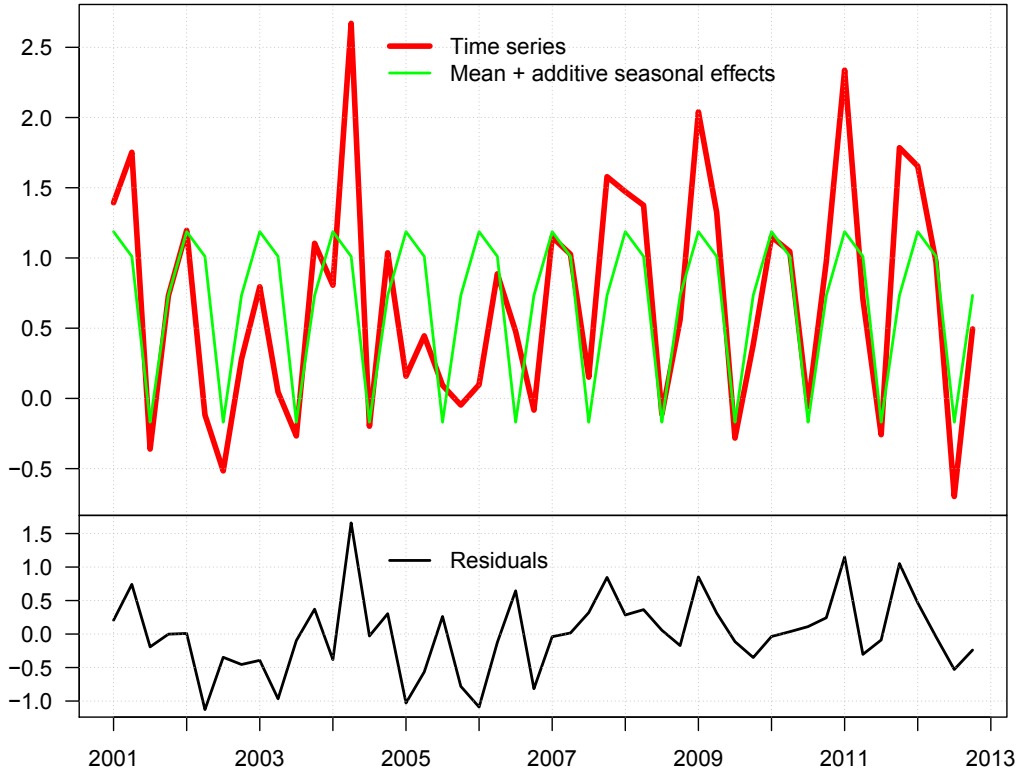


Figure 4.9.: Inflation time series [%], quarter data

On the basis of Figure 4.9 it can be easily concluded that inflation was usually highest in the first quarter and lowest in the third one. Since there are no unambiguously increasing or decreasing trends throughout the time period and the changes over quarters do not seem to depend on the previous levels, we will use the additive seasonal model. The estimated parameters of the seasonal effects are reported in Table 4.5.

Quarter 1	Quarter 2	Quarter 3	Quarter 4
1.187	1.011	-0.169	0.733

Table 4.5.: Additive seasonal effects + mean of the series

Because of including seasonal effects, the variance of the data decreases by 45% from 0.62 to 0.34. The residuals are shown in the lower part of Figure 4.9. They are normally distributed (p-values of the Jarque–Bera test 0.523 and of the Kolmogorov–Smirnov 0.737) and as can be seen in the correlograms in Figure 4.10 there is only one significant value at the lag of the eleventh quarter (however, for GDP we already have had a significant eleventh lag), which we cannot support with any theory, thus we treat it as random. Additionally, the p-value of the Ljung–Box test of 0.385 confirms the hypothesis of independently distributed data. For forecasts we can simply use Table 4.5.

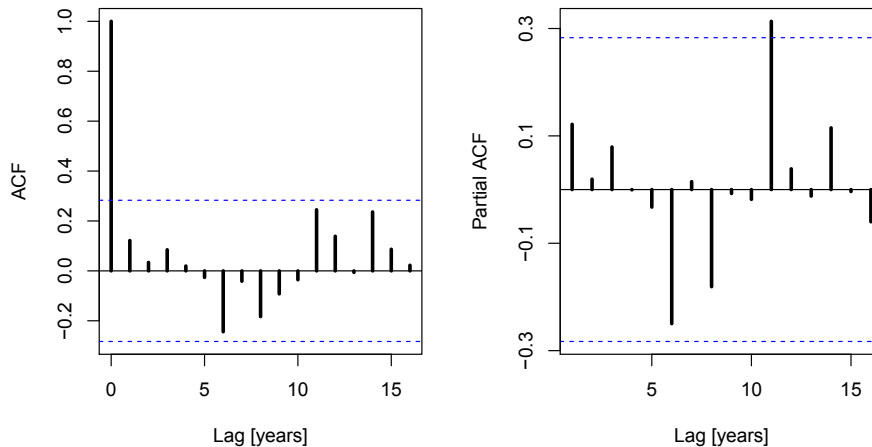


Figure 4.10.: Correlograms of the inflation series after subtracting seasonal effects

4.5. Consumer confidence indicators

Finally we want to analyse consumer confidence indicators. As can be seen in Figure 4.11 the two time series, namely current consumer confidence indicator and leading consumer confidence indicator, do not largely differ from each other. With the exception of the last two quarters of 2007, CCCI has always been higher than LCCI, suggesting that generally Polish consumers have rather bad expectations about the future. What is more, although some positive trend is visible, the FGLS coefficients are not significant (estimated AR(1) correlation of the error terms: 0.878 and 0.877). This can be seen in Table 4.6.

		OLS	FGLS
Current consumer confidence indicator	Intercept	-4.049	-0.084
	time	4.018	0.078
Leading consumer confidence indicator	Intercept	-3.878	-0.194
	time	3.848	0.190

Table 4.6.: t-statistics of the OLS and FGLS estimates of the trend

Graphically we can reject the hypothesis of significant seasonal variation. Although the `decompose` function estimates the additive seasonal effects to be non-zero, the difference between the largest and the smallest effect is only 1.37 for CCCI and 1.30 for LCCI, respectively, which is an extremely small value as compared to the values of the analysed time series. The non-zero values are thus insignificant.

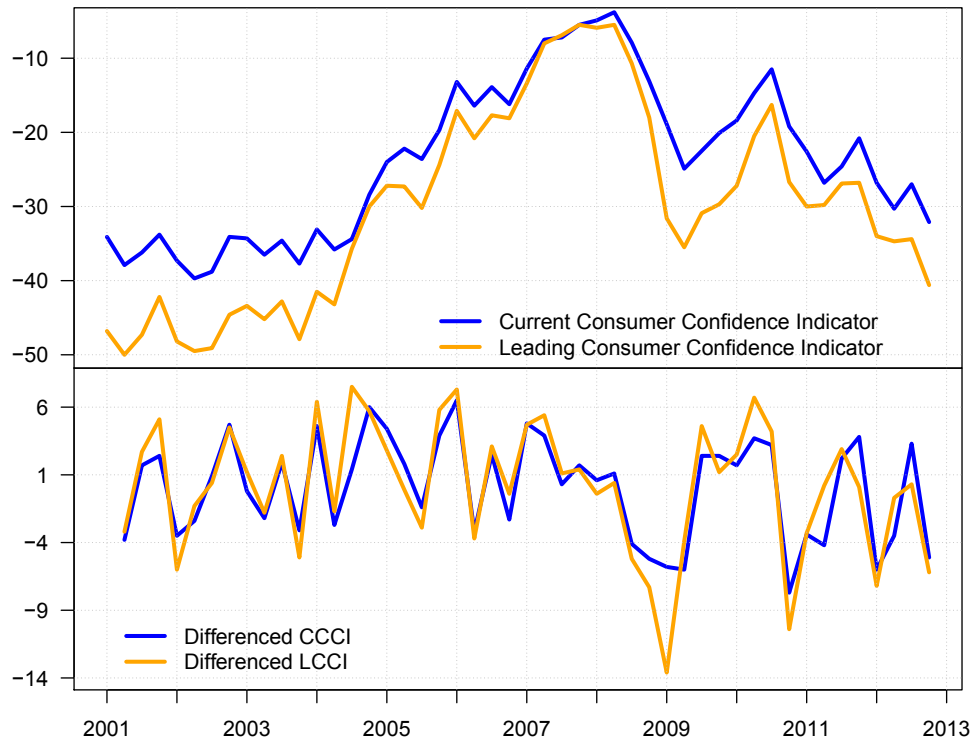


Figure 4.11.: Consumer confidence indicators

The R function `auto.arima` suggests the first series follows an ARIMA(0,1,0) process and the second one an ARIMA(0,1,1) process, for which AIC is 281.002. Assuming LCCI has the same order as CCCI (0,1,0), we get a very similar value for AIC of 281.036. Thus the decision based on `auto.arima` is not unambiguous and the random walk may be a good choice even for LCCI. Function `ar` estimates for CCCI the series to be an AR(1) process with $\alpha = 0.922$ and treating LCCI as an AR(1) process with $\alpha = 0.914$. Both series are normally distributed, the p-values of the Jarque-Bera and Kolmogorov-Smirnov tests are (0.242, 0.438) for CCCI and (0.324, 0.731) for LCCI. Thus we can calculate the confidence intervals of the AR parameter. On the 5% significance level the confidence interval for CCCI is [0.808; 1.036] and for LCCI [0.794; 1.034]. Since both these intervals include 1, which corresponds to the random walk, which is simple to use, we treat the two series as realizations of random walk processes. The residuals (in case of a random walk the same as the differences) are normally distributed (p-values of the JB test 0.236 and 0.180 for CCCI and LCCI, respectively, and for the KS test 0.318 and 0.768). The residuals are stationary: p-value of the ADF test for both series is below 0.01 and the test statistics of the KPSS test are 0.266 and 0.267. The following correlograms (Figure 4.12) show there is no serial correlation in the differenced data.

Since the random walk fully accounts for autocorrelation of the analysed time series, our forecasts are only based on the last known value (the means of the residuals were in our case approximately zero). This is why for the first quarter of 2013 we would expect just the values from the last quarter of 2012, for CCCI -32.1 and for LCCI -40.6. The real values were very similar: -30.8 and -40.9, respectively. For the other quarters of 2013 and for 2014 the forecasts of CCCI and LCCI are again the last known values (as for the time of writing: from the first quarter of 2013): -30.8 for CCCI and -40.9 for LCCI.

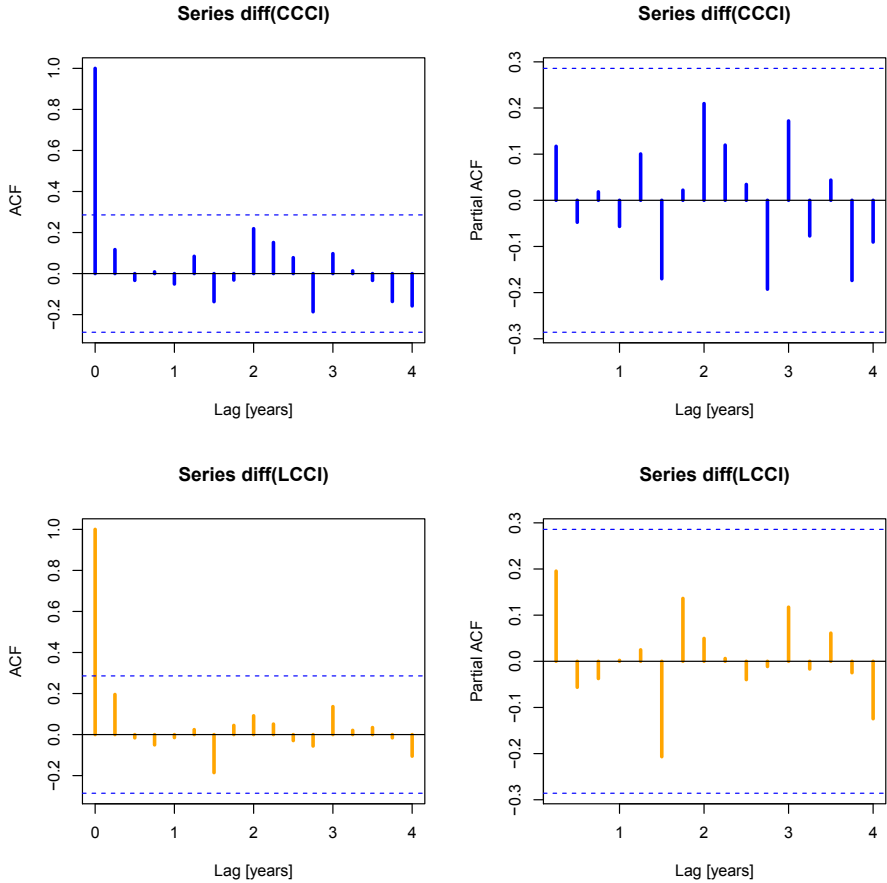


Figure 4.12.: Correlograms of the differenced CCCI and LCCI time series

5. Multivariate Time Series Analysis

5.1. Correlations

Before analysing VAR and Granger causalities it is worth having a look at the following correlations between the considered six time series (Figure 5.1). In Figure A.9 in the Appendix one can see the corresponding scatter plots (for each pair of the series). The highest correlation, of 0.980, is between the two consumer confidence indicators, the current and the leading ones. It should not be surprising taking into account the already discussed (in the 'Data' part) similarities of the questions in the corresponding surveys. The second highest correlation is between Gross Domestic Product and Gross Written Premiums (0.910), which is for us of particular interest, since the aim of the work is to make forecasts regarding the development of the insurance industry in Poland and GWP is a good indicator of that. The high correlation between GWP and GDP does not necessarily mean a causal relationship. This will be investigated in the next section of the work.

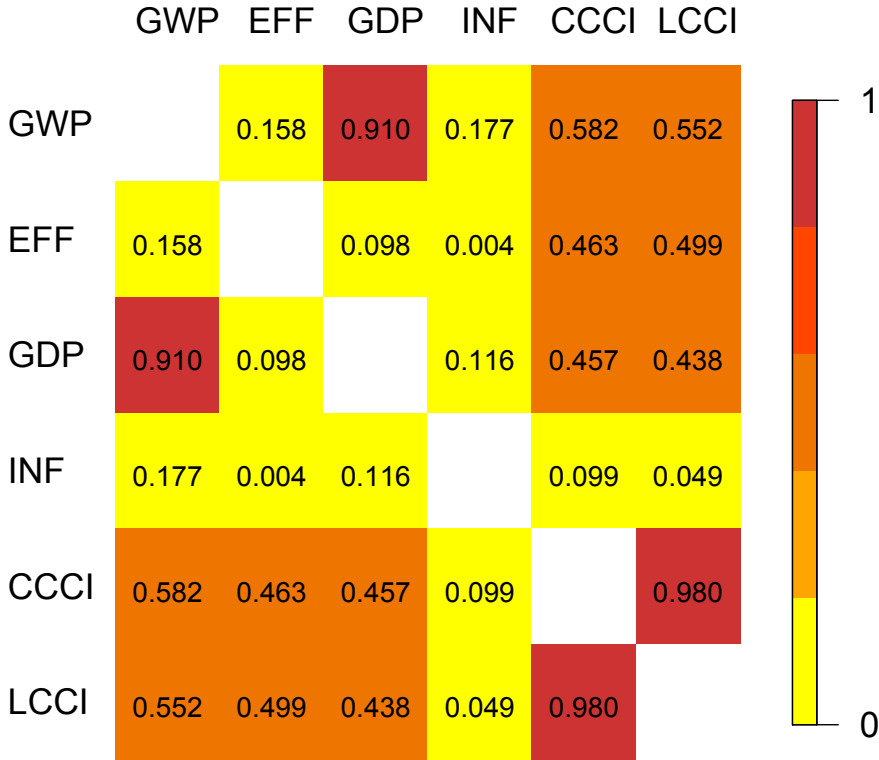


Figure 5.1.: Correlations between each two time series

The relationship of inflation with the other series is weak. Between inflation and GWP the correlation is 0.177, for GDP it is 0.116. The difference is difficult to be interpreted, but maybe it is not significant. The positive values seem, however, to be natural, since we consider nominal values of GWP and GDP, not real ones. But the fact that the correlations between inflation and (CCCI, LCCI) are positive is strange. Inflation lowers one's real income and makes a person feel financially less confident. For the short run and thus for the CCCI it should be a negative relationship. The calculated values are small and because of the small sample size (48) may be insignificant. The correlations between (CCCI, LCCI) and (GWP, EFF, GDP), for which the values are displayed in the last two rows and the first three columns, are all around 0.500, although the correlation with GDP is slightly higher for CCCI than for LCCI. This corresponds to the logic that CCCI refers to the current economic situation, for which GDP is an exact measure. Other correlations involving EFF are small, but positive.

5.2. VAR and Granger causality

Now we can proceed to the modelling of the residuals we have obtained from the univariate analysis. Because of seasonal deviations we do not use only differenced series. For GWP and GDP these are residuals from the regression models (harmonic model and regression with dummy variables, respectively), the discussed univariate autocorrelation of the error terms will not be considered. For EFF we cannot take residuals from the SARIMA model, since our SARIMA model already accounts for a MA process. Thus we take the differenced series. For INF we consider residuals from the simple regression model where we have calculated only the seasonal effects; for both CCCI and LCCI we take the differenced series. In case of GDP and EFF we still do not know the values from the first quarter of 2013, so for other series we do not consider that quarter either. In Table 5.1 one can see results of the stationarity and normality tests applied on the series of those residuals. Except for GWP all the series are both stationary and normally distributed. For GWP the p-value of the ADF test is 0.114, which corresponds approximately to the 10% significance level, so the result of the test should not be treated as a problem. On the other hand, the p-value of the Jarque-Bera test lies much below the critical value of 0.05. Because we conduct so many tests, an individual p-value below standard significance levels is not so relevant.

Considered series	p-value of the ADF test	test statistic of the KPSS test	p-value of the JB test	p-value of the KS test
resid. of GWP	0.114	0.173	< 0.01	0.131
resid. of EFF	< 0.01	0.175	0.925	0.679
resid. of GDP	0.031	0.136	0.375	0.886
resid. of INF	< 0.01	0.234	0.523	0.737
resid. of CCCI	< 0.01	0.266	0.236	0.318
resid. of LCCI	< 0.01	0.267	0.180	0.768

Table 5.1.: Stationarity and normality tests for the series which are used in the VAR modelling

Within those six series we would like to find possible causal relationships. For that we have to choose a suitable VAR model, for example using `VARselect`. When we specify the argument `lag.max=5`, so that it accounts for seasonal variation, three of the four considered criteria choose $p=1$ and only Akaike information criterion speaks for $p=5$. All the information criteria, as well as the final prediction error, consider the number of parameters within the model. However, since we want to obtain a model which can be applied in a not very complex way, the relation: the smaller the order, the better, is of even more importance. So we follow $p=1$ and compute the p -values of the bivariate Granger causality tests. The null hypothesis of this test is the non-causality, thus a small p -value corresponds to a probable causal relation.

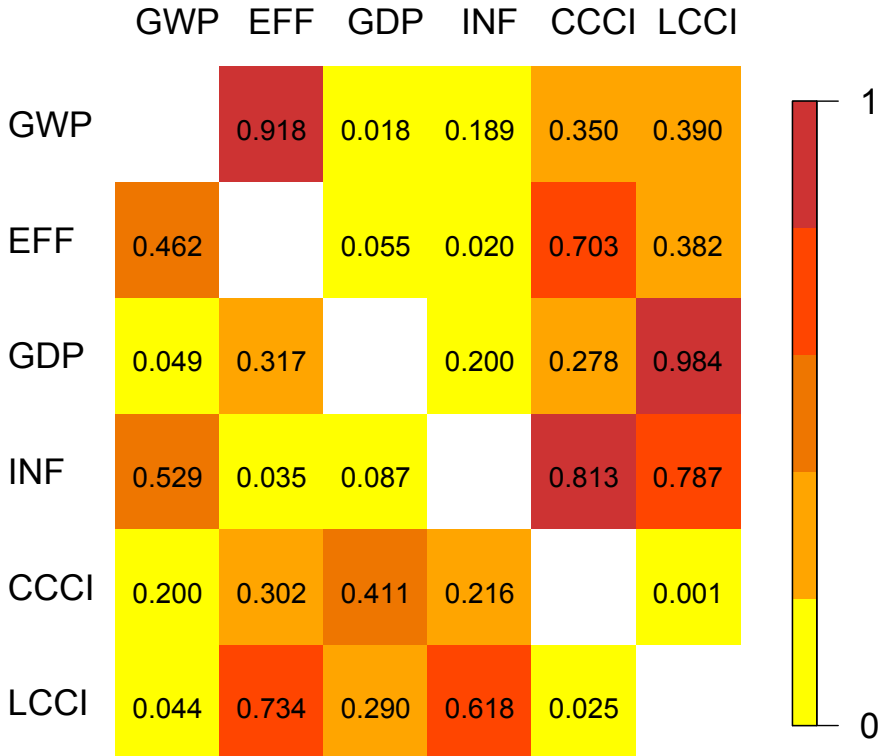


Figure 5.2.: p -values of the Granger causality test

In Figure 5.2 all pairs of the six series have been considered. The value in row i and column j is the p -value of the test that the variable indicated in row i does not Granger cause the variable in column j . We are primarily interested in the values within the first column, since those correspond to the causalities on GWP. Below 0.05 there are the p -values of both GDP and LCCI. For EFF the p -value is 0.462, but nevertheless we would like to investigate this relation, since apart from GWP this is the only data describing the insurance industry. So in the final VAR model we consider GWP, EFF, GDP and LCCI. The joint causal relation of (EFF, GDP, LCCI) on GWP is significant as computed by the `causality` function, which gives for this test a p -value of 0.074. Normally we consider the 5% significance level, but 7.4% does not differ so largely and in fact the 10% significance level is as well quite popular in applied statistics. As a remark, although CCCI and LCCI

are so strongly correlated (0.980), LCCI has a significant influence on GWP and CCCI has not. This refers to the logic that a person purchasing assets like a house or a car considers his future economic situation rather than the current one, and when purchasing new assets one often buys new insurance products.

For the new VAR model, only with GWP, EFF, GDP and LCCI, we again consider `VARselect` with `lag.max=5`. All four criteria speak for $p=1$. Table 5.2 shows the coefficients of the following VAR(1) model ('r' stands for residuals). Note that the values do not correspond to the strength of the causal relation. In the VAR framework we consider as well autocorrelation terms. For GWP we get approximately 0.520, in the univariate analysis it was 0.641, and for GDP the autocorrelation within the VAR framework is 0.715, which as well does not differ so much from the univariate estimate of 0.806. The p-value of the multivariate Jarque–Bera test for the residuals of this VAR(1) model is 0.362, so we assume the final residuals are normally distributed. For the multivariate Portmanteau test we get a p-value of 0.978, so we can treat the final residuals as independently distributed. The `arch.test` returns a p-value of 0.999, thus we are not able to reject the null hypothesis of constant variance of the final residuals. The model seems to be valid and so we compute the forecasts. The visualisation of these forecasts together with confidence intervals at the significance level of 80% is shown in Figure 5.3

	$r(\text{GWP})_{t-1}$	$r(\text{EFF})_{t-1}$	$r(\text{GDP})_{t-1}$	$r(\text{LCCI})_{t-1}$	constant
$r(\text{GWP})_t$	0.52007	0.03537	1.71320	0.00277	-0.00382
$r(\text{EFF})_t$	0.02070	-0.71661	-0.63292	-0.00143	0.00114
$r(\text{GDP})_t$	0.03885	0.03586	0.71547	0.00010	-0.00149
$r(\text{LCCI})_t$	-13.27024	-7.18356	1.33966	0.15491	0.16649

Table 5.2.: Coefficients of the VAR(1) model involving GWP, EFF, GDP and LCCI

Based on the forecasted residuals of the four considered series, we can now make new forecasts for GWP, EFF, GDP and LCCI. These are presented in the Appendix (Table A.1). For the first quarter of 2013 the forecasts are 16855.6 for GWP, 1.11 for EFF, 392463.4 for GDP and -39.48 for LCCI. In Figure A.10 (see the Appendix) all the discussed four prediction models for GWP have been displayed together, including the harmonic model trend with the VAR(1) model of the residuals. One can see that the new forecasts are very similar to the harmonic model with an AR(1) model describing the residuals, although now the values are slightly lower. In the long run the forecasts converge to the underlying harmonic seasonal trend, thus the causality has an impact on forecasts only over the next few quarters.

If we analyse just the insurance industry, we could consider a VAR framework consisting only of residuals from GWP and EFF. `VARselect` again suggests to choose VAR(1). The influence of EFF on GWP within the VAR model becomes negative (in the vector autoregressive model considering the four series this influence was positive). In the second direction (influence of GWP upon the profitability ratio EFF) the influence is negative too. The corresponding Granger causalities are, however, statistically insignificant (p-values of 0.462 and 0.918, respectively).

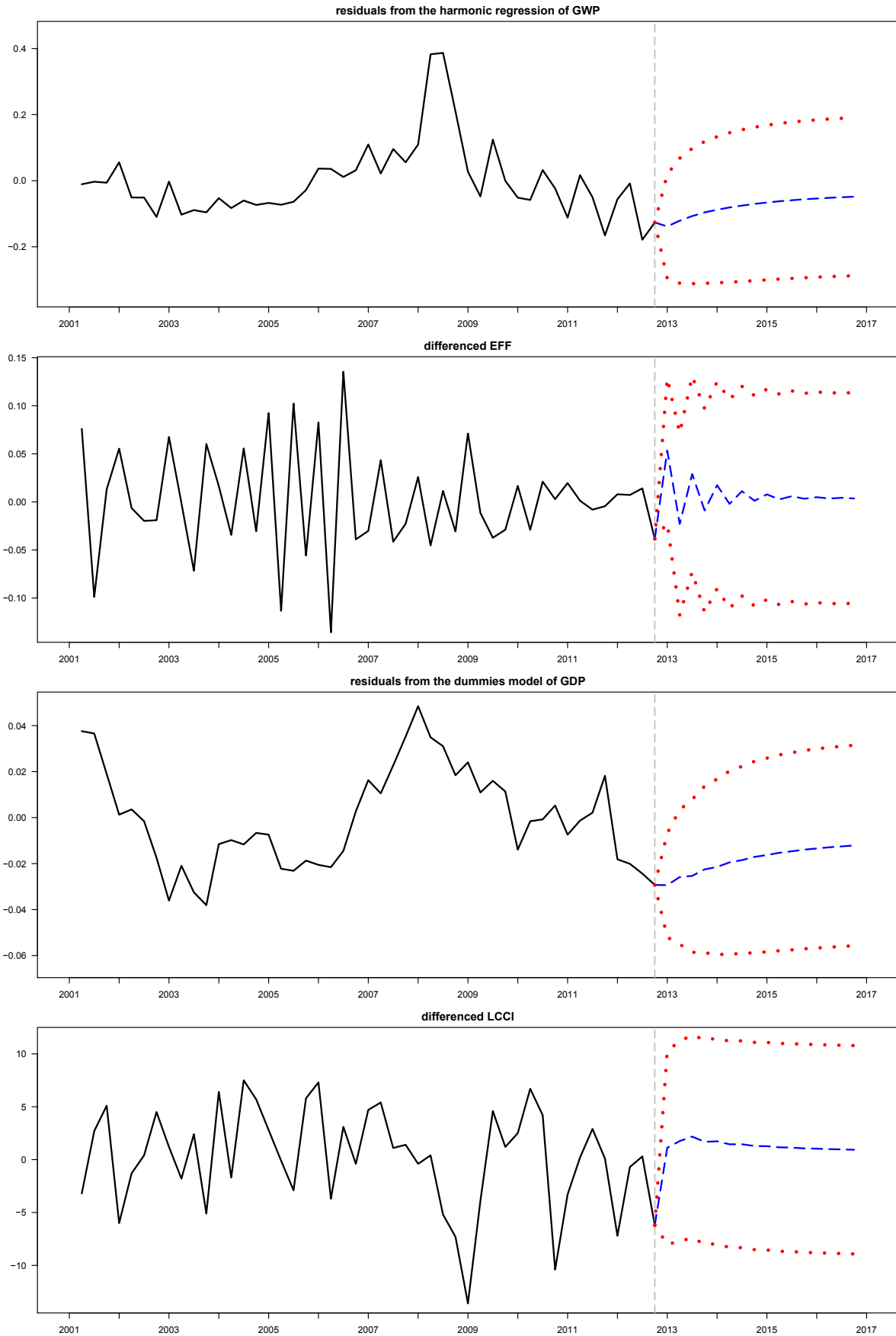


Figure 5.3.: Forecasted residuals and corresponding confidence intervals for the analysed VAR(1) model

6. Conclusions

The aim of the work has been to analyse the last twelve years of the development of the Polish insurance industry and to make forecasts, based on both univariate and multivariate methods. Some of the corresponding results are displayed in Table 6.1. Within the work additionally numerical values of the forecasts for 2014 can be found, as well as the graphical visualisation of the predicted values till the end of 2016. However, since the variance of estimated forecasts increases the longer the time period between, it makes little sense to consider values after 2013. For every forecast the formula or approach is given, so that after some time the procedure can be easily repeated.

	Q. 1	Q. 2	Q. 3	Q. 4
GWP [million PLN], harmonic model	16924.0	16940.5	16767.0	18454.4
GWP, harmonic model + VAR(1)	16855.6	16634.6	16364.5	17981.0
GDP [million PLN], dummies regression	377.815	403.234	414.155	471.381
EFF, SARIMA	0.075	0.066	0.075	0.062
LCCI, random walk	-40.9	-40.9	-40.9	-40.9

Table 6.1.: Forecasts for the year 2013, based on univariate time series analysis (except for the second row); green colour corresponds to the already known values

The above forecasts rely on the assumption that the previous trends will continue, which is not so certain. The last twelve years have been very prosperous for the Polish economy and it will be difficult to maintain the same increase ratio of GDP and GWP. In this context it seems to be more appropriate to consider for GWP in both the univariate and multivariate case the harmonic seasonal model, for which the predicted values are lower, rather than the regression model with dummy variables or SARIMA. The parameter efficiency of the model is another advantage.

Regarding the VAR framework, a joint Granger causal relation of the profitability ratio EFF, GDP and LCCI on GWP has been proved, the corresponding positive coefficients speak for positive causal effects. The forecasts of GWP following from the VAR(1) analysis are displayed in the second row of Table 6.1. This model should be treated as the best one. Within the work its validity has been proved.

As an important remark I have to underline that the proved normality of some samples, which can be used to make confidence intervals of the forecasts, can be spurious. Normality tests like the Jarque–Bera and Kolmogorov–Smirnov tests consider normal distribution of the elements as the null hypothesis, and since the number of time periods in my work has been small, the probability of making the Type II error has been relatively high, but we have not been able to measure it.

A. Appendix

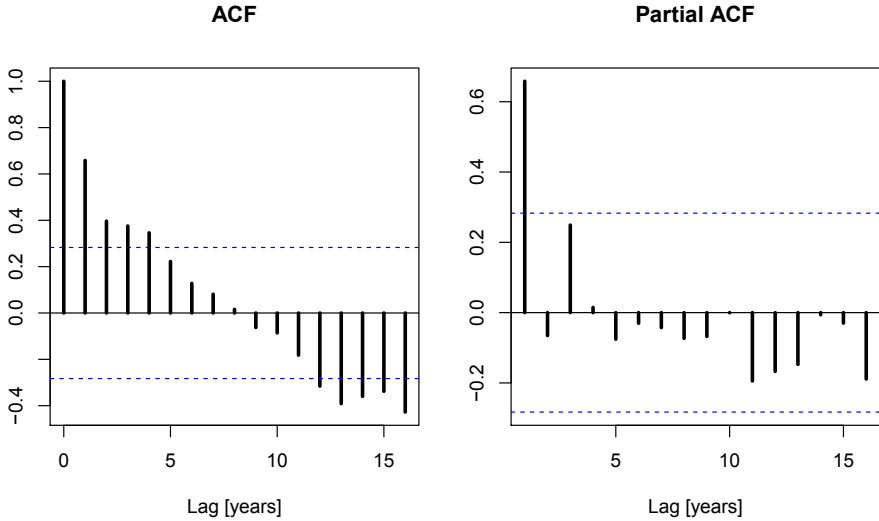


Figure A.1.: Correlograms of the residuals from the regression model with dummy variables, GWP

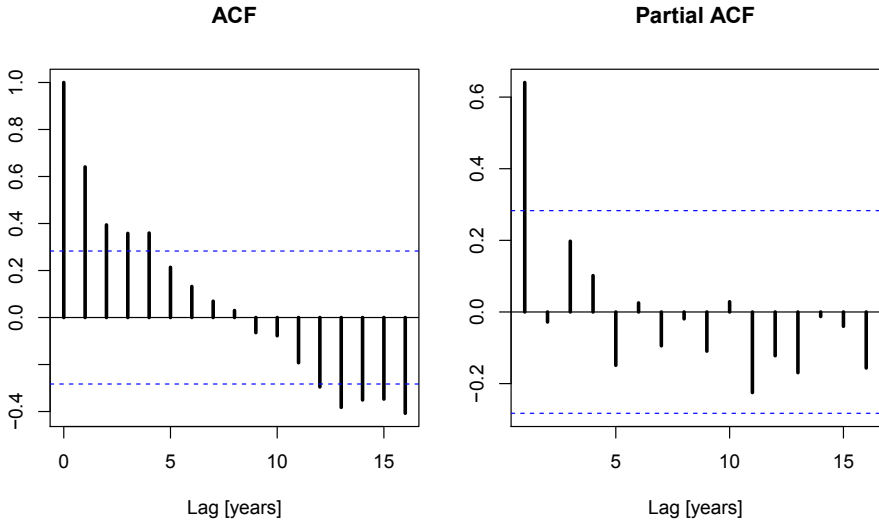


Figure A.2.: Correlograms of the residuals from the harmonic seasonal model regression, GWP

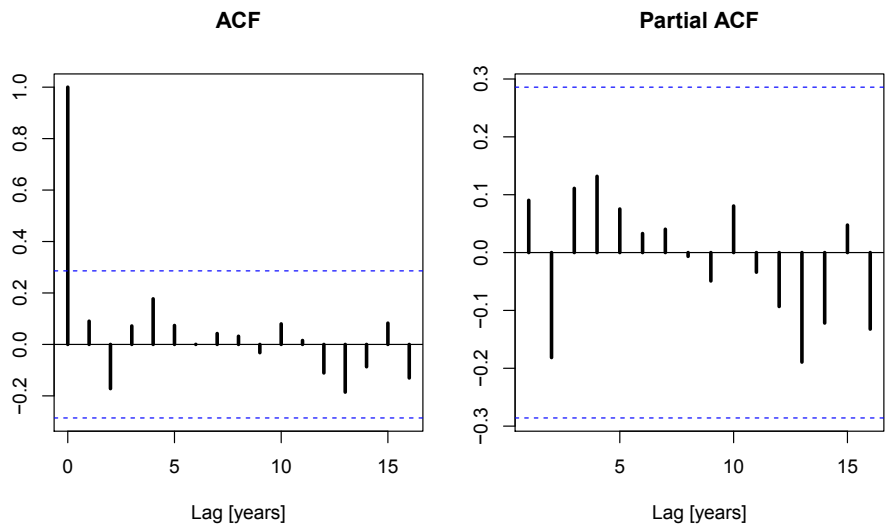


Figure A.3.: Correlograms of the residuals from the AR process which has been computed for the residuals from the GWP OLS regression (dummies model)

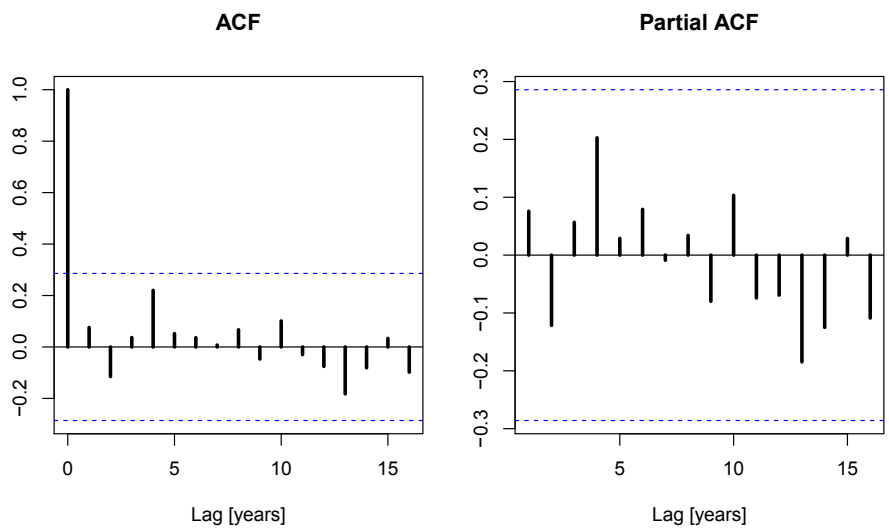


Figure A.4.: Correlograms of the residuals from the AR process which has been computed for the residuals from the GWP OLS regression (harmonic seasonal model)

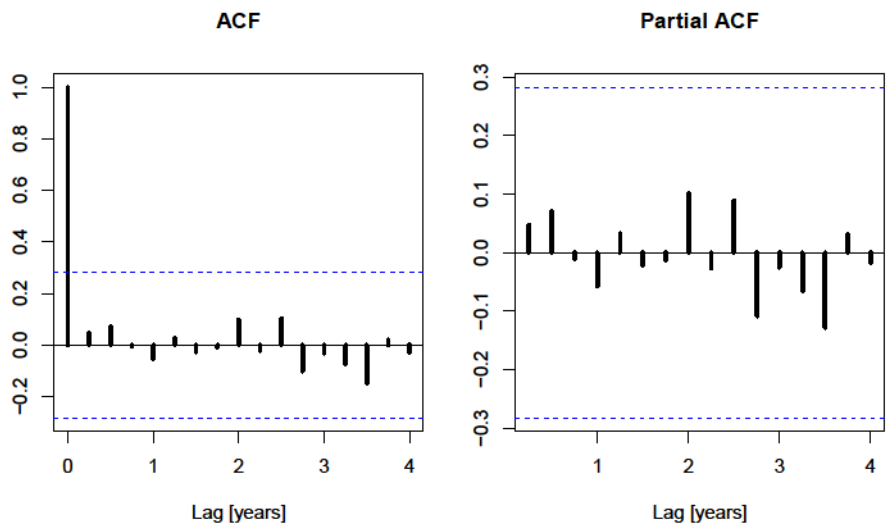


Figure A.5.: Correlograms of the residuals from the SARIMA model for GWP

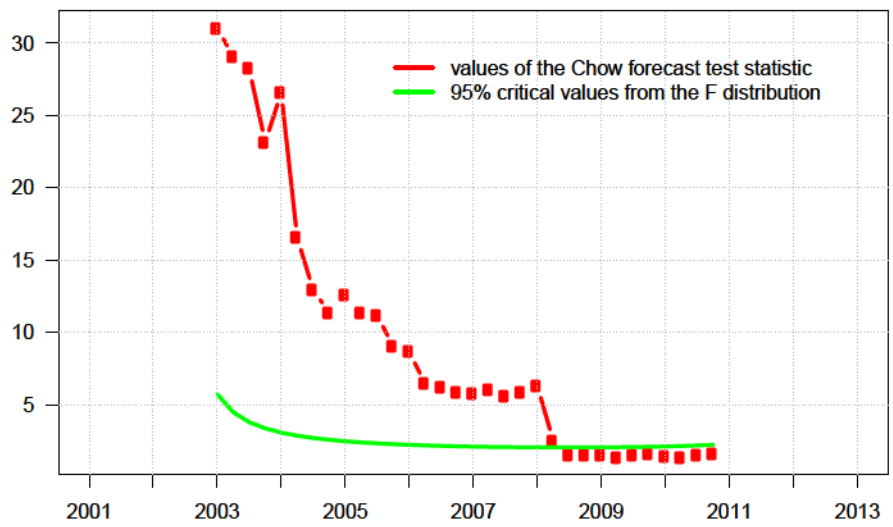


Figure A.6.: Application of the Chow forecast test to detect structural breaks within GWP (dummies model)

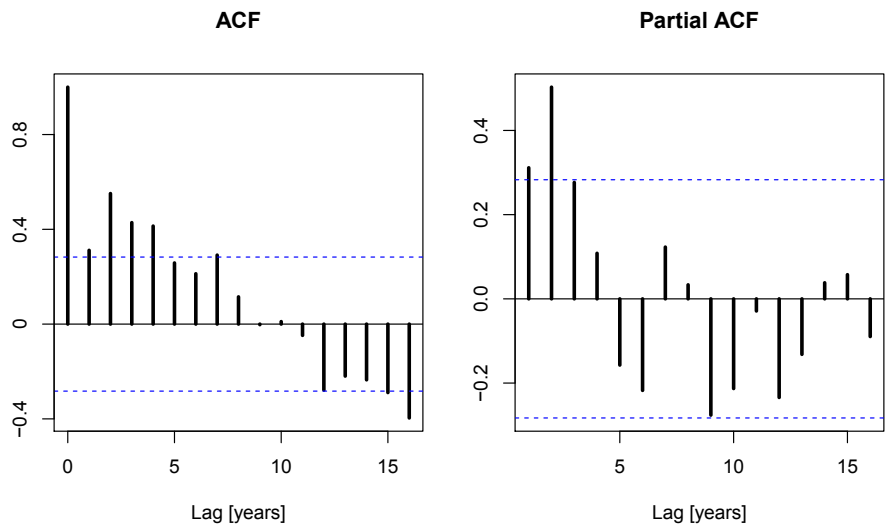


Figure A.7.: Correlograms of the residuals from the regression model for EFF

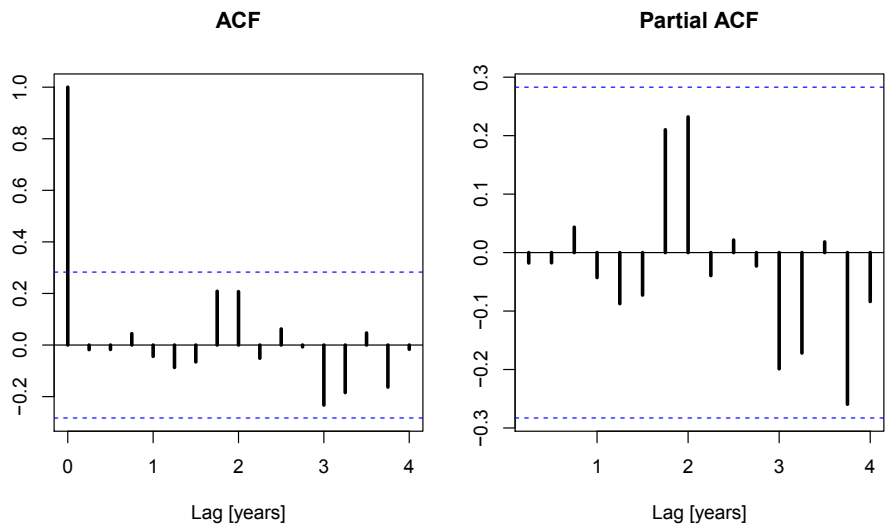


Figure A.8.: Correlograms of the residuals from the SARIMA model for EFF

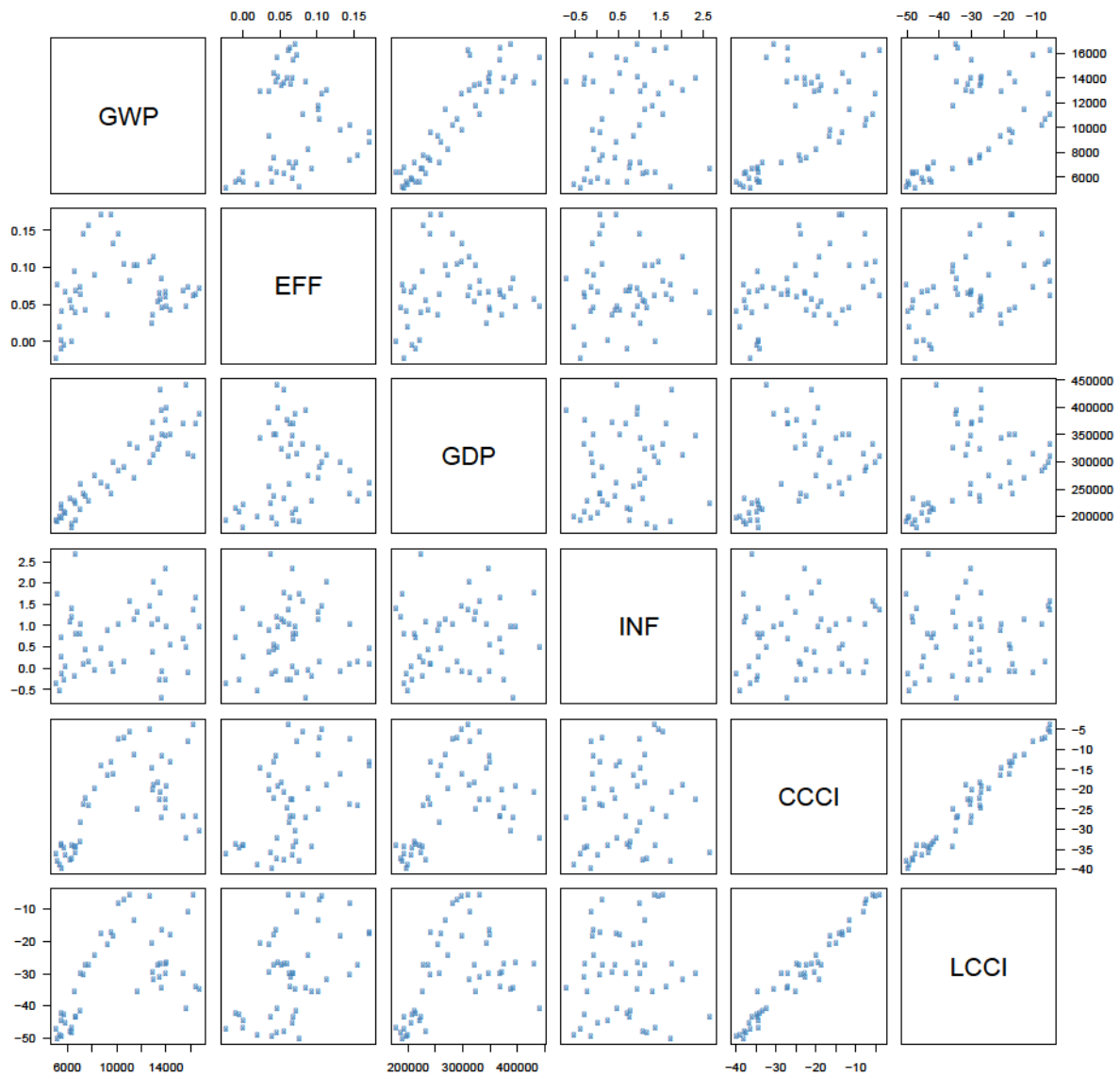


Figure A.9.: Scatter plots of all the pairs of considered series

	'13 Q. 2	'13 Q. 3	'13 Q. 4	'14 Q. 1	'14 Q. 2	'14 Q. 3	'14 Q. 4
GWP:	16634.6	16364.5	17981.0	19693.2	19237.5	18766.1	20494.0
EFF:	1.02	1.08	1.04	1.07	1.05	1.06	1.05
GDP:	414888.2	421875.1	477452.9	424208.6	447765.6	455517.3	514789.7
LCCI:	-38.85	-38.42	-38.92	-38.87	-39.15	-39.15	-39.31

Table A.1.: Forecasts of GWP, EFF, GDP and LCCI in 2013 and 2014 [GWP and GDP in million PLN], based on univariate trends and the VAR(1) model of the residuals

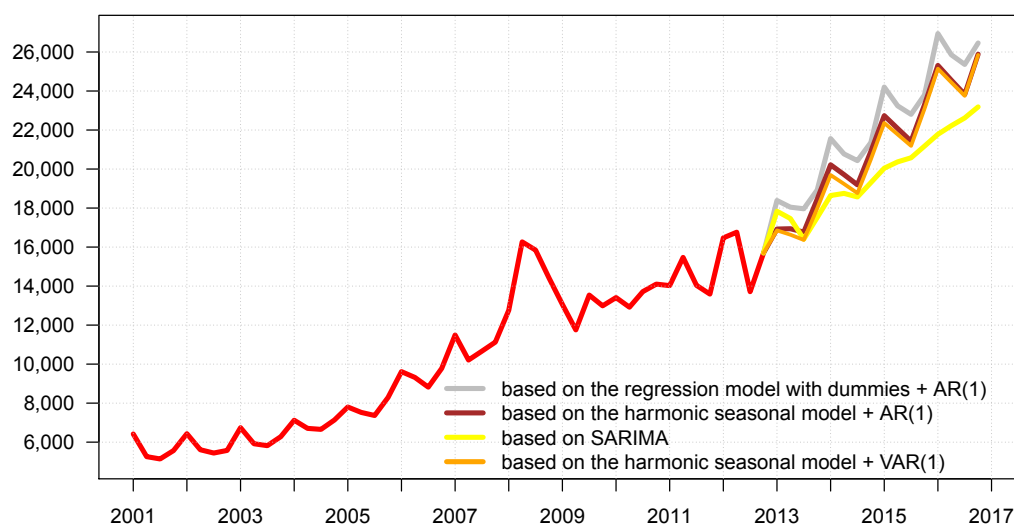


Figure A.10.: Forecasted values of Polish GWP till the end of 2016 [million PLN], including the VAR(1) model of the residuals

Bibliography

- [1] Biecek, P.: *Przewodnik po pakiecie R*. Oficyna Wydawnicza GIS 2008 [Polish]
- [2] Chan, N.H.: *Time Series. Applications to Finance*. Wiley 2002
- [3] Cowpertwait, P.S.P.; Metcalfe, A.V.: *Introductory Time Series with R*. Springer 2009
- [4] GUS: http://www.stat.gov.pl/cps/rde/xbcr/gus/2013_wyjasnienia_metodyczne.pdf
Methodological notes on 'Poland – macroeconomic indices'.
- [5] Hamilton, J.D.: *Time Series Analysis*. Princeton University Press 1994
- [6] Hyndman, R.J.: <http://cran.r-project.org/web/packages/forecast/forecast.pdf>
Reference manual of the R package 'forecast'.
- [7] Lütkepohl, H.: *New Introduction to Multiple Time Series Analysis*. Springer 2005
- [8] Lütkepohl, H.; Krätzig M.: *Applied Time Series Econometrics*. Cambridge University Press 2007
- [9] Pfaff, B.; Stigler, M.: <http://cran.r-project.org/web/packages/vars/vars.pdf>
Reference manual of the R package 'vars'.
- [10] Polish Finance Ministry: [legislation of the insurance sector from the 28th of December 2009](#)

Declaration of Authorship

I hereby confirm that I have authored this Bachelor thesis independently and without use of other than the indicated sources. Where I have consulted the published work of others, in any form (e.g. ideas, equations, figures, text, tables), this is always explicitly attributed.

Berlin, 17 June 2013

Wiktor Olszowy