Lappeenranta University of Technology

Course: 3219 Business Forecasting

SPSS Project: Consumption Forecast

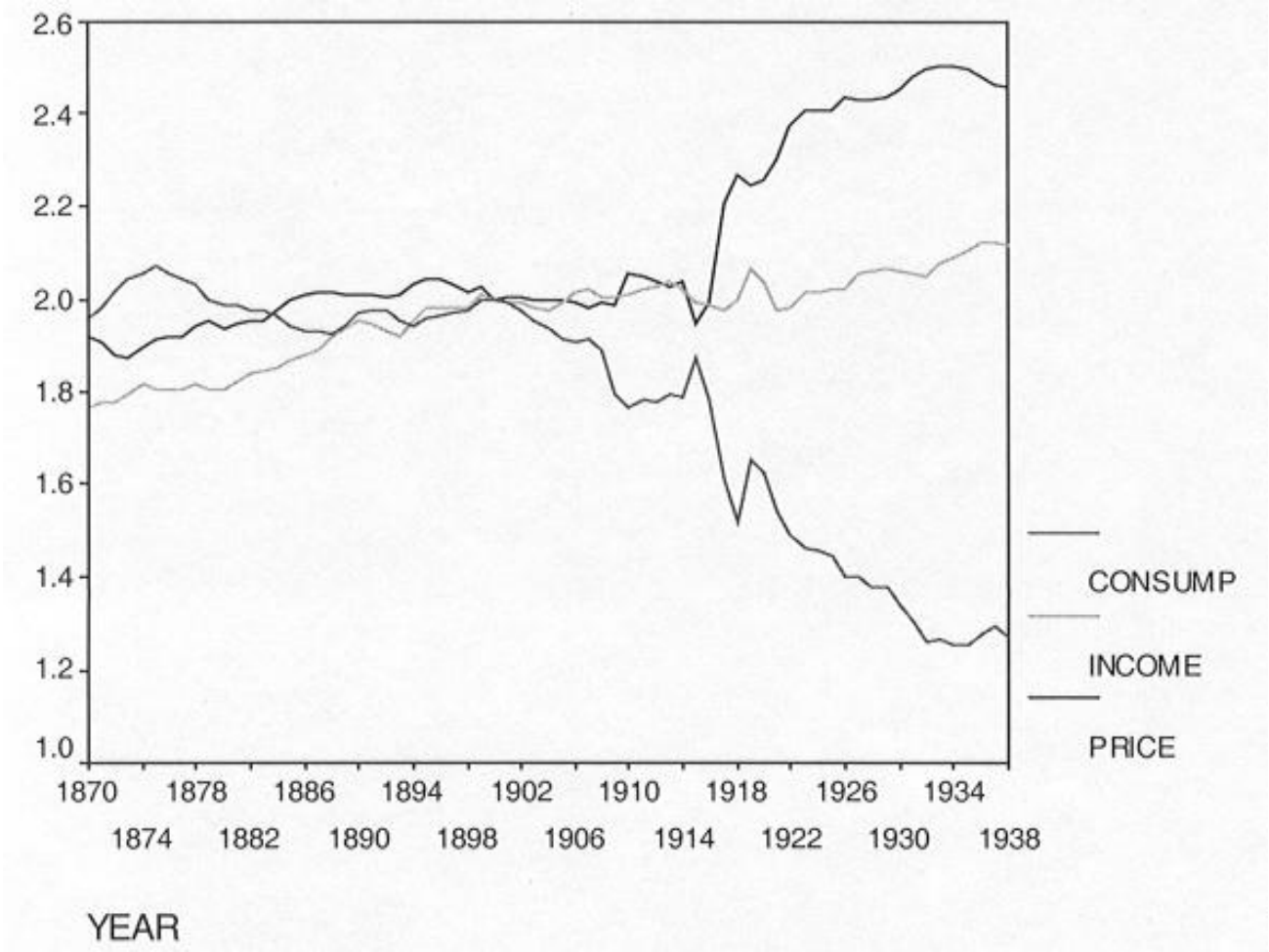
Student names: Thoralf Czichy, Bart Timmermans

Student email: czichy@rcs.urz.tu-dresden.de
Bart.Timmermans@adalvas.be

Table of Contents

1 Description of the Data

The data is provided in form of annual figures from 1870 to 1938. Based on the data we will develop a forecast for the consumption of the following 10 years from 1939 to 1948. The first step we did was to make a sequence graph of the available data to get a visualized idea of what the data is representing and how the data is interrelated.



The available data consists of three time series for the consumption (C), the price level (P) and income level (I) throughout the years from 1870 to 1938. These variables are expected to be related. So, e.g. ceteris paribus consumption will go up if the price level decreases and vice versa. Consumption is expected to increase if the income level is also increasing and vice versa. Of course there are more factors then just these two determining the economic consumption level -e.g. the individual savings,

investment behavior and governmental actions. Throughout our analysis we will use only these three available time series (C, P,I) to make a forecast.

First we take a look at the price level time series. The years 1870 to 1914 are characterized by slow growth of the prices. This trend is only slightly influenced by the usually cyclical and random variations. In 1915 there is a small but sharp decline which is succeeded by a very high increase in the price level through the years 1916 to 1923. The following years from 1924 to 1938 are again characterized by a more moderate and slow growth in the price level - similar to the situation before 1915. On the other hand a small decrease during the last four years (1935 to 1938) suggest that the curve from 1915 to 1938 is shaped like an upside down "U" (quadratic model). This small decrease could indicate the beginning of a decreasing price level in the following years nullifying the previous increase.

The income level time series seems to follow a linear trend, which is only influenced by random noise. Later on we will forecast this variable using a linear regression.

The consumption time series seems to remain constant during the years 1870 to 1914. In 1915 there is a small increase in the consumption, which corresponds to the decrease of the price level in the same year. This increase is followed by a sharp decline in the consumption level throughout the years 1916 to 1923 again corresponding to a similar increase in the price level in these years. After 1923 these struggles seem to quiet down and a situation of a slow decrease seems to emerge. A slow increase in the consumption curve at the end of the time series suggests, similar to the decrease in the price level, the beginning of a re-covering of the consumption level. This would result in an increase of the consumption level in the following years.

The struggles of consumption and price level through the years 1914 to 1923 are probably caused by the world war one from 1914 to 1918. We can clearly see this in the graphs. It is difficult to predict the data since we were not entirely sure in what matter our data is influenced by this fact. It is known that the world war one hit the world growth and had serious economic consequences.

2 Selection of the Techniques Used for the Forecasting

For forecasting the consumption of the following 10 years we will use a simple linear model with two predictor variables.

$$C_t = a + b_1 * I_{t-s_1} + b_2 * P_{t-s_2}$$

C_t :Consumption in year t

I_{t-s_1} :Income in year $t - s_1$

P_{t-s_2} :Price in year $t - s_2$

The consumption is modeled as a function of income and price. The constants a, b_1 and b_2 have to be estimated by multiple linear regression. We also consider the possibility, that the consumption is better described by the income or prices of preceding years by introducing the time lag constants s_1 and s_2. We chose this model because it is generally accepted, that the consumption depends on price and income level. A linear model is chosen because the data suggests that (observation of the graph). On the other hand it is simple to understand and easy to use and we do not expect that more difficult models will add significant accuracy.

To forecast values for the consumption on the basis of that model we need the values for price and income for the following years. We will forecast these values simply on the basis of the available time series.

For the forecast of the income we will use a simple linear regression with the income depending on the year.

$$I_t = a + b_1 * t$$

Again we chose a linear model, because visual observation of the graph suggests a linear trend. Seasonal decomposition is not useful because there are no seasons in the available data and a cyclical influence is not visible.

For the forecast of the price level we will use the Autoregressive Integrated Moving Average Model (ARIMA). This is a quite sophisticated approach. But the curve does not fit a linear trend or seasonal decomposition and a simple exponential smoothing is better applicable for short-term forecasting. We think that ARIMA as a combination of an autoregression approach and moving averages approach will give a good prediction for this time series.

3 Making the Forecast and the Results

3.1 The Forecast of the Income

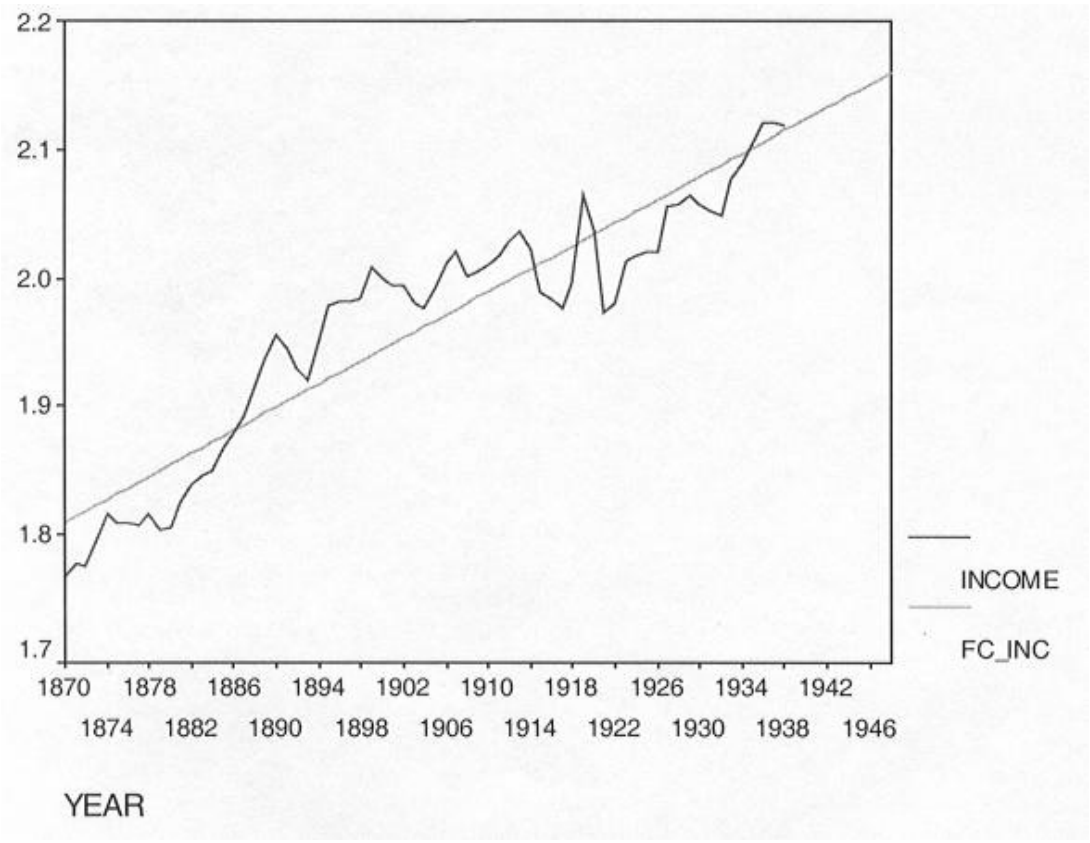
For the income forecast we assume a linear trend. As a result we get a formula, which we can use to make the forecast. We used the regression algorithm of SPSS. The regression results in following values.

$$a = -6.587$$

$$b_1 = 0.00449$$

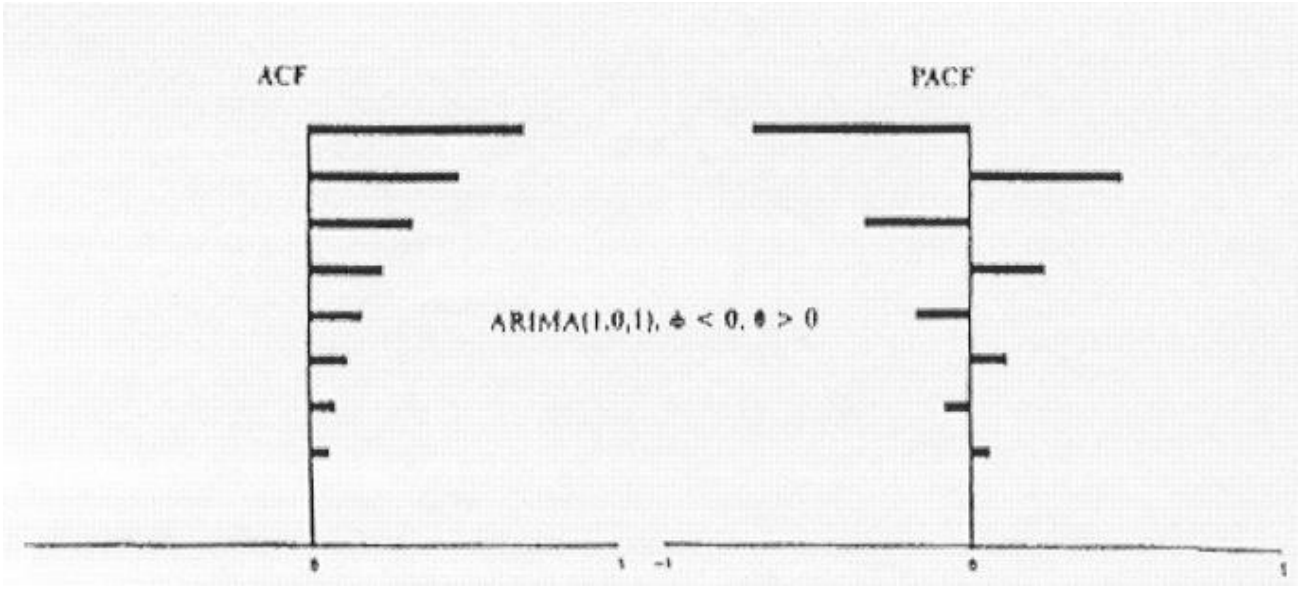
This results into following formula, which can be used to make predictions. The variable t is the year to be predicted.

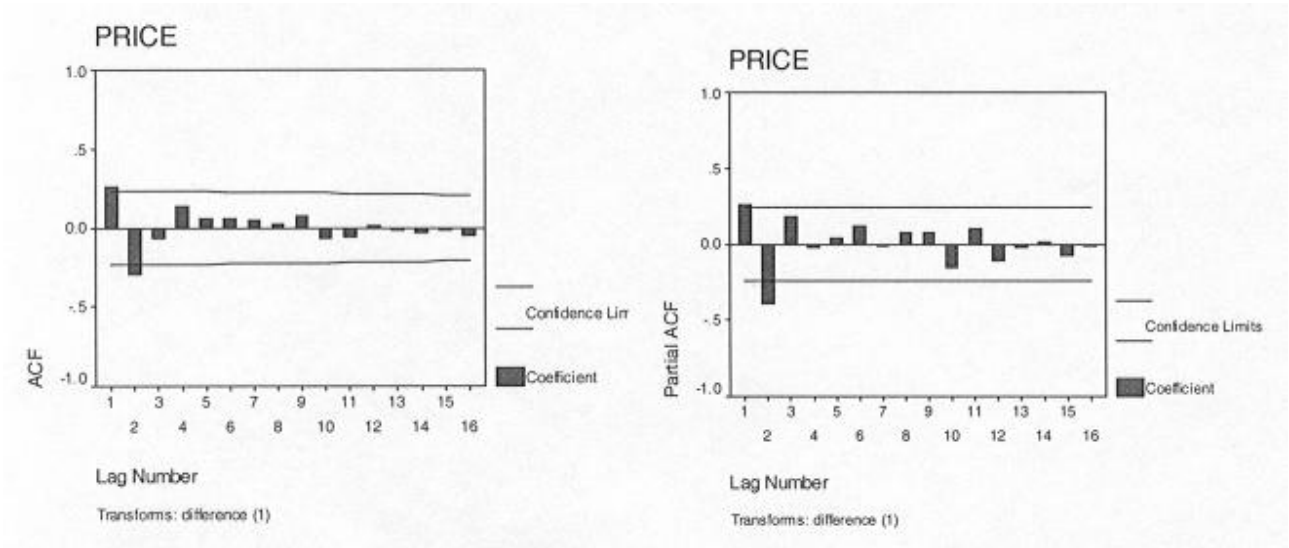
$$I_t = -6.587 + 0.00449 * t$$



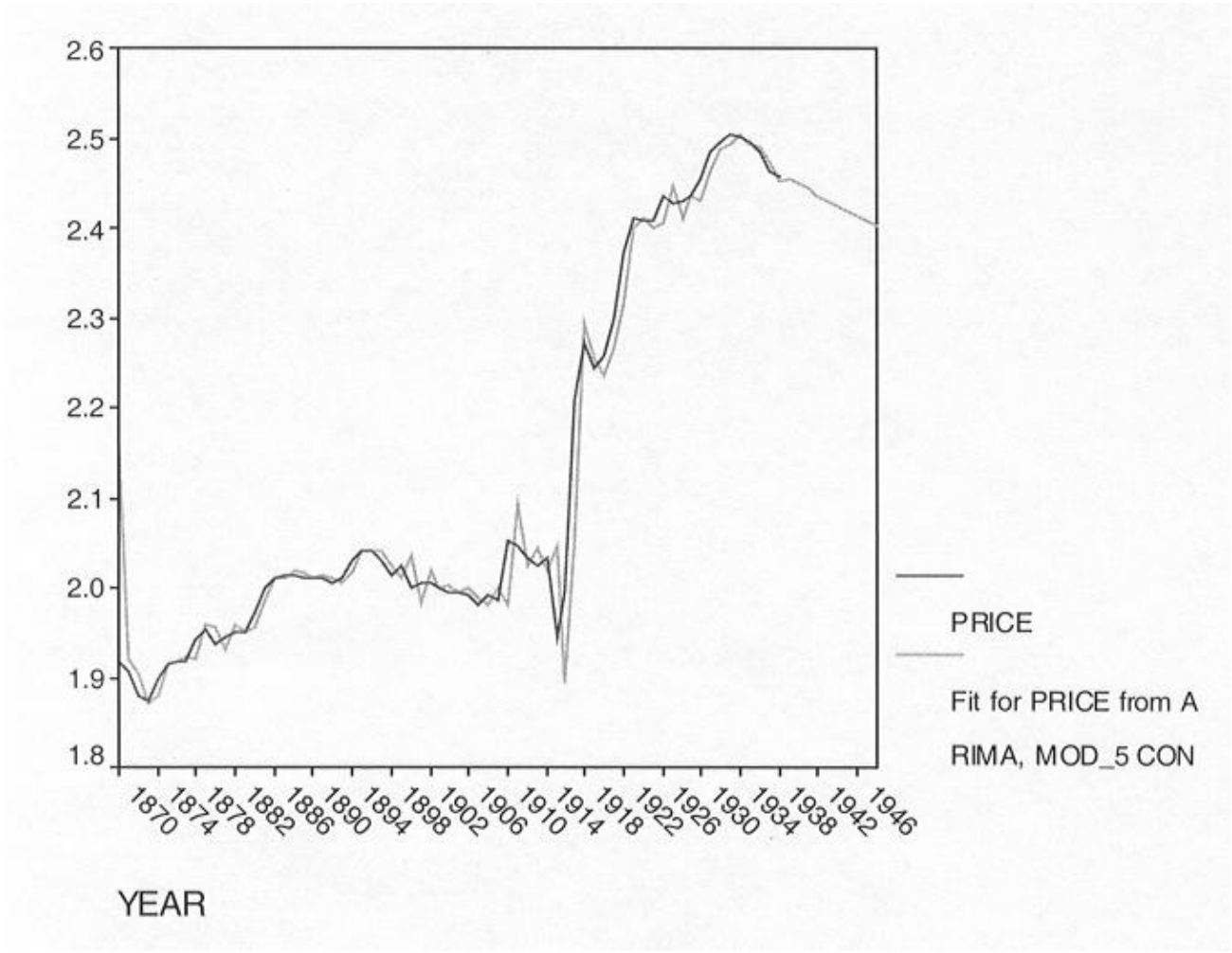
3.2 The Forecast of the Price

The forecast for the price will be made using ARIMA. First the series has to be made stationary .This is achieved by differencing it once. Furthermore we have to determine p and q for the ARIMA model. This is based on the shape of the autocorrelation and partial autocorrelation function.



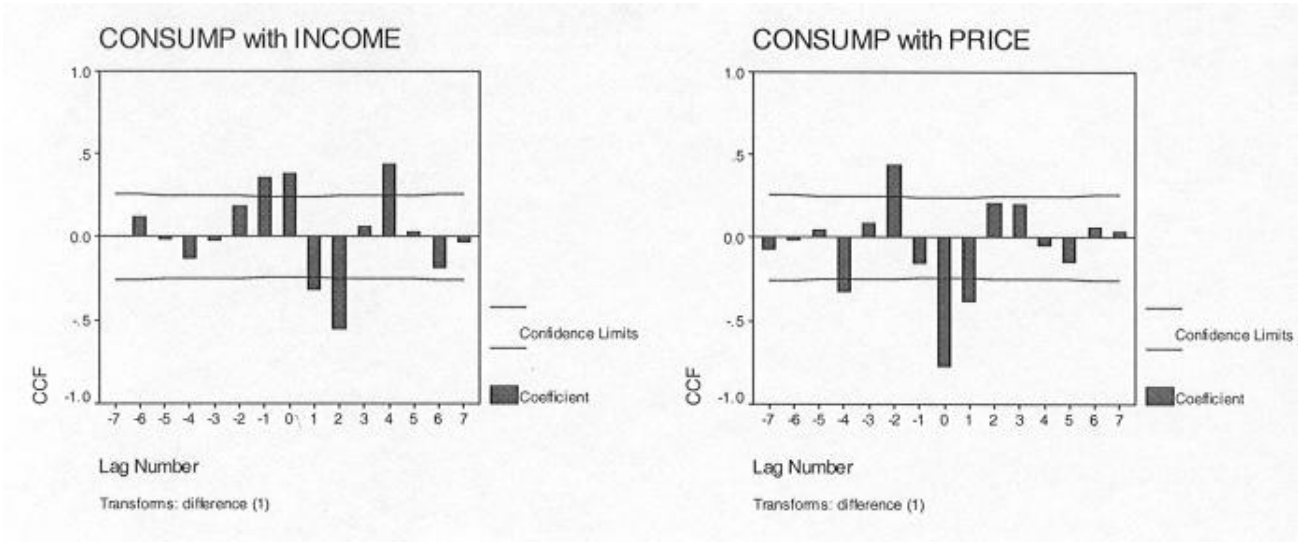


Although it is not exactly the shape of the ideal functions (upper picture), we think that the actual shape (lower picture) is closest to the ARIMA model where p equals one and q equals one. The forecast on the basis of that decision looks like that.



3.3 The Forecast of the Consumption

To get a forecast of the consumption first the values of the formula (see 2 Selection of the Techniques Used for the Forecasting) have to be estimated. To get the values for the time lags s_1 and s_2 we have to take a look at the cross correlation functions.



We can see that between price and consumption the highest peak is at zero -which results in a time lag of zero for the price (= s_2). The highest peak for the time lag between income and consumption is at two. This would mean, that there is a high correlation between the consumption of the current year and the income two years in advance. This assumption seems to be quite non-typical. There are also significant peaks at minus one, zero, one and four. Therefore we decided to set the value for the time lag of the income to zero (= s_1), which also delivers good results.

On the basis of the price series and the income series we calculated the values for a, b_1, b_2 by using the SPSS algorithm for multiple regression.

$$a = 4.607$$

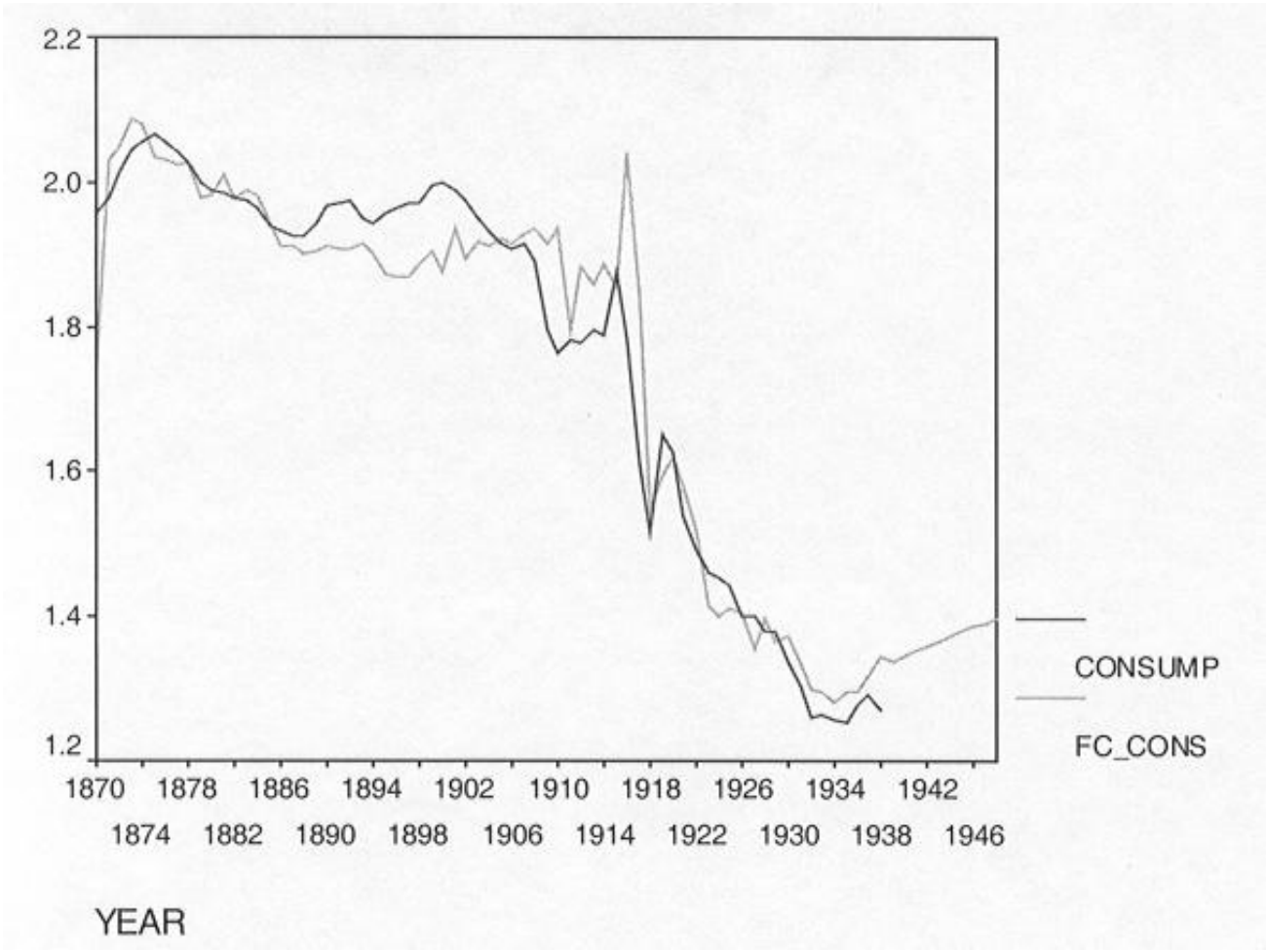
$$b_1 = -0.120$$

$$b_2 = -1.228$$

This results into following formula, which can be used to make forecasts.

$$C_t = 4.607 - 0.120 * I_t - 1.228 * P_t$$

Graphically the actual forecast looks like that.



4 Comparison of the Techniques Used and Goodness of the Forecasts

4.1 The Income Time Series and its Forecast

The graph of the income time series (see 3.1 The Forecast of the Income) is fluctuating around the regression line. Taking a look at the SPSS error statistic the linear regression approach seems to be good. The Mean Absolute Percentage Error (MAPE) of 1.5384% indicates a good forecast. Also the Sum of Squared Errors (SSE), which equals 0.0802, compared to income values in the range of 1.8 to 2.1 seems to be okay. We expect good income forecasts using this approach.

4.2 The Price Time Series and its Forecast

Although the autocorrelation and partial autocorrelation function did not perfectly fit the forecast seems to be good, using ARIMA. According to the SPSS error statistic the forecast for the past values

compared to the real values is good. The Mean Absolute Percentage Error (MAPE) of 1.01 % is quite good. The Sum of Squared Errors (SSE) equals 0.1248 and also indicates a good forecast. The Durbin-Watson value of 1.5335 is close to 2, which indicates that there is no or only a low degree of autocorrelation. Compared to the past we expect good forecasts for the price level in the following years.

4.3 The Consumption Time Series and its Forecast

The final forecast of the consumption can be evaluated using the SPSS error statistics. We can compare the estimated values of the past with the actual values of the past. If the forecast for the past fits the real past values we will expect the same for the future. The Mean Absolute Percentage Error (MAPE) of 2.9124% is not as good as for the forecasts for the price and income level, but still satisfying. The Sum of the Squared Errors (SSE) equals to 0.3642, which is not bad compared to the consumption values in a range between 1.2 and 2.0. This leads us to the conclusion, that the actual forecast of the 10 following years will be satisfying. Of course this is only applicable with the assumption, that there are no major changes in the assumptions under which this forecast has been made (see 5 Evaluating the Reliability of the Results).

5 Evaluating the Reliability of the Results

Although our forecast compared to the past delivers good results it is difficult to say, whether our model will give a good prediction or not. As already mentioned in the beginning of this paper (see 1 Description of the Data) we can consider the slight decrease in prices and the increase in consumption at the end of our data as the beginning of a recovering process from the recession caused by the first world war or just as random noise or a short recovering. We also can see this as a problem of different results of a long and short term forecast. But this is a general problem of a forecaster.

As we have seen from the decrease in consumption, caused by world war one the economy and therefore the consumption is heavily influenced by politic and other governmental actions. To add more reliability to this report more qualitative and intuitive methods, like brainstorming and scenario techniques, could be used. This could increase the understanding of the subject matter and its endogenous and exogenous variables influencing the development of the consumption.

So, as we know nowadays, the world war two or the oil crisis in 1973 had all a certain influence on the consumption of nearly all countries in the world. Of course it is difficult to forecast these events by just using statistical-empirical methods relying in "hard" data. Therefore we again suggest the use of further methods to build an awareness for the assumptions under which our forecast is made and also to build an understanding for the options in which the future might develop.