Time series analysis project

Forecasting the inflation rate in Switzerland using a Seasonal ARIMA model

May 19, 2019

Contents

1 Introduction 2

2 Description of the data 2

3 Data analysis 3

4 Discussion 8

5 Conculsion 8

1 Introduction

The aim of this project is to forecast the montly inflation rate in Switzerland during the year 2019. To achieve this goal a proper analysis of an univariate time series is needed. The first step is to clean the raw data set and to plot it in order to visually inspect the raw time series. The plot indicates a negative trend, a decreasing variance over time and no seasonality. Since there is evidence for a polynomial trend of degree three, the trend is removed by using lag-1 differentiation to get a stationary time series. Then, the empirical autocorrelation (ACF) and the partial autocorrelation function (PACF) are investigated. The PACF is rather complicated and clearly shows periodic spikes at lags 1, s, 2s, 3s and 4s where s = 12, since one has monthly data in the data set. Therefore, a seasonal ARIMA model is considered as a good model for the process. The best SARIMA model by considering the AICc and the significance of the estimated parameters is a SARIMA(5,0,5) x (1,0,2)[12]. Next, the previous model is fitted to the stationary time series to get the coefficients of the model and the validity of the model is investigated. Especially by looking at the ACF of the residuals of the model and by performing the Ljung-Box test, one concludes that the model is valid. That is, the residual time series is a white noise and therefore it is appropriate to make 12 months ahead predictions of the residual time series. The last step is to transform the predictions of the residual time series to get the predictions of the initial time series about the monthly inflation rate.

2 Description of the data

The data set is gathered form the website of the Organization for Economic Co-operation and Developement (OECD). The initial chart shown on the website was filtered for Switzerland and monthly intervals from January 1956 until December 2018. Then the raw data was downloaded in the CSV format. Since the downloaded data came together with some metadata, it was necessary to clean the data set by extracting the two columns "TIME" and "VALUE". The entries in the column "VALUE" correspond to Switzerland's inflation rate in percent for a specific month in a given year. In total, the data set consists of 756 data points.

The OECD measures inflation by consumer price index (CPI). The inflation is defined as follows: Inflation is the average change in the price level of a basket, which consists of goods and services often bought by households in an economy over a time interval. Specifically, the basket is broken down to the three categories energy, food and a total without energy and food. The calculation of the CPI is based on the weighted average of the mentioned categories in the basket, see OECD (2019). As a reference year for the CPI they use an index with 2015=100. Therefore, the entries of the time series in the column "VALUE" represent the change of the CPI in percent from the actual period to the prior period, which is the definition of the inflation rate.

3 Data analysis

A good starting point of the data analysis is to plot the raw time series data as indicated in Figure 1 to visually inspect trend and seasonality components of the process.

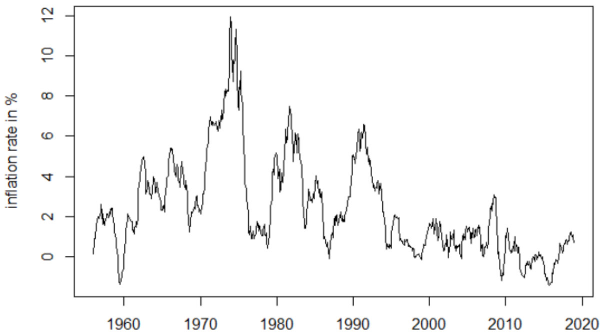

Figure 1: The raw time series

At first glance, one can see the the variance decreases over time and that there is an outlier in the 1970s, which was due to the huge increase in oil prices before the oil crises 1979. Overall there seems to be a negative trend and no seasonality is identifiable. The empirical analysis by using linear regression underlines that there indeed is a ploynomial trend of degree three and that there is no evidence for seasonality. Therefore, the trend has the form $m_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$. Afterwards, the trend is removed by using the lag-1 difference operator. Since the mean of the time series after differencing is slightly positive one has to subtract its mean to get the stationary and zero-mean time series plotted in Figure 2. Moreover, by applying the Ljung-Box test for 100 lags, see Ljung and Box (1978), one can clearly reject the null hypothesis that the process is an iid-noise and the application of the Augmented Dickey-Fuller test, see Said and Dickey (1984), underlines that the process is stationary.

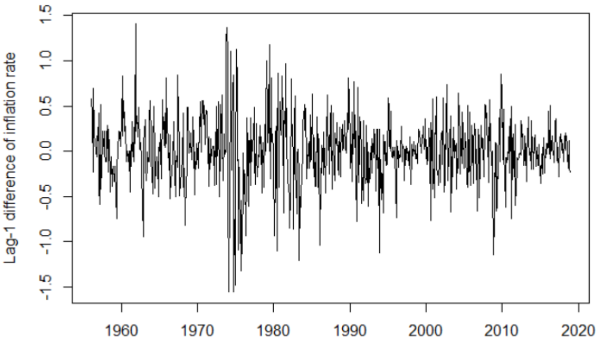

Figure 2: Lag-1 difference of inflation rate

Although the variance is not constant over time due to the volatile property of the inflation rate, the lag-1 difference method reduced it considerably compared to the raw time series.

Then next step in the analysis is to analyze the sample autocorrelation and the sample partial autocorrelation functions plotted in Figure 3 and Figure 4 respectively. In the sample ACF one can identify three significant spikes at lags 1, 2, 6, 12 and 13. In contrast to that one observes a very complicated sample PACF with lags at $1, s, 2s, 3s$ and $4s$ where $s = 12$ and therefore corresponds to the period of the seasonality function.

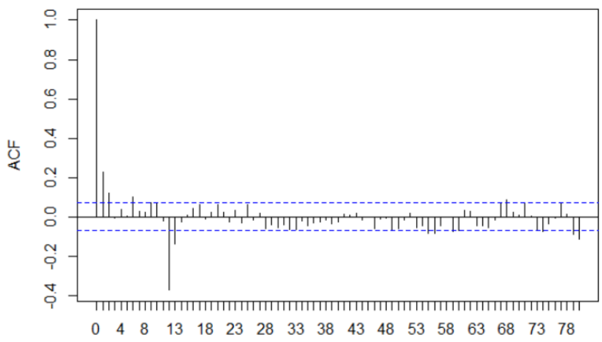

Figure 3: Sample ACF

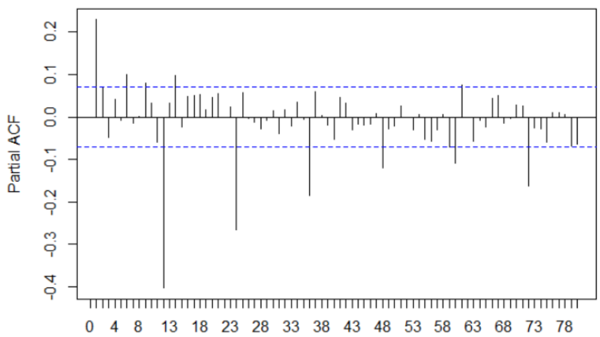

Figure 4: Sample PACF

This plot leads to the conclusion that a $SARIMA(p,0,q)$x$(P,0,Q)[12]$ is an appropriate model, since it is able to replicate ACFs and PACFs with periodic spikes. With this information one can guess the parameters of the model. If one considers only the values of the ACF and PACF where the lags are a multiple of s, one ends up with the values $P = 1$ and $Q = 4$ for highly significant lags since both, ACF and PACF, can be considered as roughly exponential decreasing. With respect to the parameters p and q one can see spikes at lags at 1, 2, 6 and 1, 6 and 9 and then nothing for the ACF and the PACF respectively and a roughly exponential decrease for both. Therefore, it could be some $ARMA(p,q)$ process with $p > 0$ and $q > 0$.

After having guessed a model it is time to choose the best model with respect to the Akaike Information Criterion by fitting the data to some SARIMA models with different parameters for p, P, q, and Q by using the ARIMA() function in R, see Hyndman and Khandakar (2008). The output suggests a $SARIMA(5,0,5)$ x $(1,0,4)[12]$ model, which accords approximately with the guess about the parameters given above. Nevertheless, the last two coefficients of Θ are not significant and the AICc decreases if one cancels these two parameters from the model. Therefore, the final model can be written in the following way: $\phi(B)\Phi(B^s)X_t = \theta(B)\Theta(B^s)Z_t$ where $Z_t \sim WN(0,\sigma^2)$ and $\phi(z) = 1-\phi_1 z - \ldots - \phi_5 z^5$, $\Phi(z) = 1-\Phi_1 z$, $\theta(z) = 1-\theta_1 z - \ldots - \theta_5 z^5$ and $\Theta(z) = 1-\Theta_1 z - \Theta_2 z^2$. In total one has to estimate $1+p+q+P+Q = 14$ parameters. The values of the estimated parameters and their standard errors (rounded up to two significant digits) are given in Table 1. One sees by using the test for significance of the coefficients at the 5% level, that the parameters ϕ_4, θ_1, θ_2, θ_4, θ_5, Φ_1, Θ_1 and Θ_2 are significant. The estimated value of σ^2 is 0.09056.

Once the final model is fitted to the data, one has to investigate the validity of the model. First, one checks the ACF of the residuals of the model which should behave like

ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	θ_1	θ_2	θ_3	θ_4	θ_5	Φ_1	Θ_1	Θ_2
-0.25	-0.01	0.12	0.55	0.32	0.48	0.28	-0.04	-0.54	-0.48	0.58	-1.19	0.26
0.21	0.10	0.11	0.12	0.20	0.21	0.11	0.1	0.12	0.18	0.20	0.23	0.18

Table 1: Coefficients and their standard errors of the $SARIMA(5,0,5)$x$(1,0,2)[12]$ model

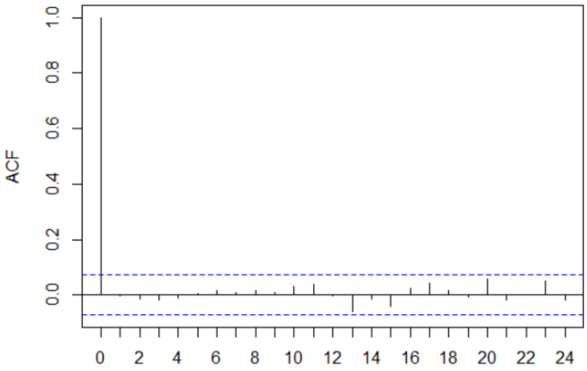

Figure 5: ACF of the model residuals

an iid-noise. This is visualized in Figure 5. The residuals clearly behave like a white noise since there are no significant spikes. Second, one applies the Ljung-Box test on the residuals, see Ljung and Box (1978), which tests the null hypothesis that a process is an iid-noise. In Figure 6 the p-values of the Ljung-Box test for the residuals of the model are plotted. For each of the 24 lags one can not reject the null hypothesis that the model residuals follow an iid-noise. That is, the residuals do not contain any time dependence. Therefore, the suggested $SARIMA(5,0,5)$x$(1,0,2)[12]$ model is valid.

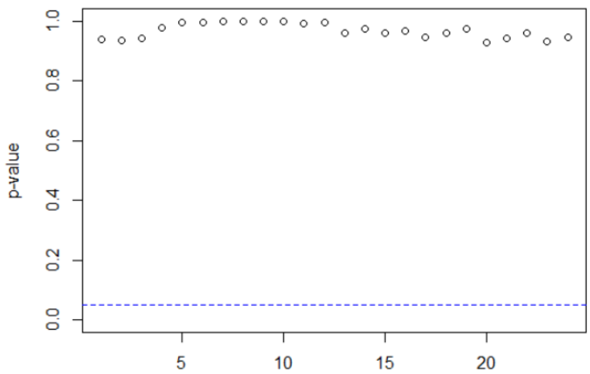

Figure 6: Ljung-Box test for the model residuals

The last step of the analysis is to forecast the values of the initial time series 12 months ahead. To achieve this goal, one first predicts the residual time series from the model. Then, one has to transform the residual predictions back to predictions in the original scale of the data set. Since the initial time series is differenciated once to remove the trend and is centered to get a zero-mean stationary time series, one applies the following equations to obtain the predictions in the original scale: $\hat{Y}_{t+1} = Y_t + z_{t+1} + \mu$, $\hat{Y}_{t+2} = \hat{Y}_{t+1} + z_{t+2} + \mu = Y_t + z_{t+1} + z_{t+2} + 2\mu$, ... and $\hat{Y}_{t+12} = \hat{Y}_{t+11} + z_{t+12} + \mu = Y_t + z_{t+1} + ... + z_{t+12} + 12\mu$ where Y_t is the last observed value of the initial time series, z_{t+1} is the first predicted value of the residual time series from the $SARIMA(5,0,5)$x$(1,0,2)[12]$

Year 2019											
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
0.61	0.52	0.42	0.51	0.34	0.31	0.19	0.22	0.27	0.26	0.39	0.49

Table 2: 12 months ahead prediction of the inflation rate in %

model and μ is the mean of the residual time series before it was centered. The predictions of the next 12 months of the initial time series, which is the inflation rate in Switzerland, are given in Table 2. The associated plot of the forecasts is given in Figure 7. In order to get a better visualization only the inflation rates from January 2010 until December 2019 are shown. In addition, confidence bands are included in the plot. These confidence bands are only valid 95% confidence bands, if the residuals of the model are normally distributed. This can be tested by using the Shapiro–Wilk test, see Shapiro and Wilk (1965). One tests the null hypothesis that the residuals follow a normal distribution. Unfortunately, the p-value at the 5% level is very low. So the confidence bands can not be interpreted as 95% confidence bands because we reject the null hypothesis, that the residuals of the model are normally distributed. Nevertheless, these confidence bands give a rough idea about the uncertainty of the forecast.

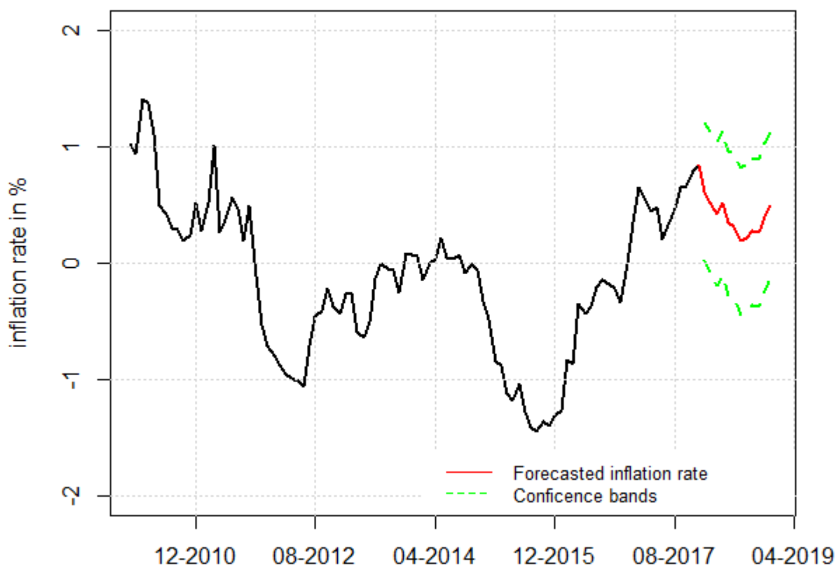

Figure 7: 12 month ahead forecasts of the inflation rate in % in the year 2019

An interesting method to test the prediction quality is to forecast the values of a year we already know. In the following, this method is applied to the reduced initial time series from January 1956 until December 2017. Therefore, one predicts the monthly inflation rate in the year 2018 with the $SARIMA(5, 0, 5)\text{x}(1, 0, 2)[12]$ model and one compares these values with the actually observed inflation rates. Exactly the same process as described in the beginning of the data analysis part is applied to the new time series. One obtains valid results, that is in short, the process is stationary and the residuals of the model are an iid-noise.

The difference of the observed and the forecasted values of the inflation rate are visualized in Figure 8 and the corresponding values are given in Table 3. Again, only the inflation rates from January 2010 until December 2018 are shown in Figure 8 to ensure

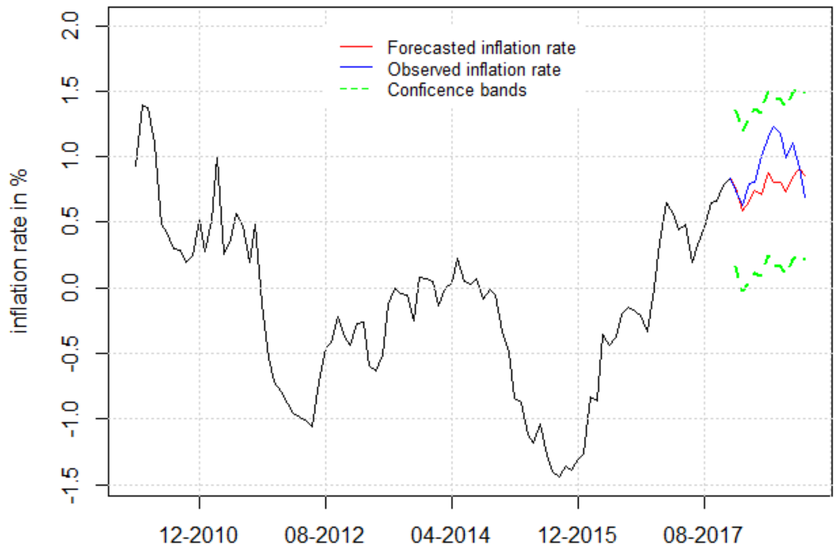

Figure 8: Difference of the inflation rate in % between forecasted and observed values in the year 2018

Year 2018												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
observed	0.74	0.63	0.80	0.80	1.00	1.15	1.23	1.18	0.99	1.11	0.92	0.69
forecasted	0.76	0.59	0.66	0.73	0.70	0.86	0.79	0.77	0.70	0.81	0.85	0.79
difference	0.02	-0.04	-0.14	-0.07	-0.30	-0.29	-0.44	-0.41	-0.29	-0.30	-0.07	0.10

Table 3: Differences between observed and forecasted inflation rates in % in the year 2018

a meaningful visualization. The real values of the inflation rates are rounded up to two significant digits. Over the horizon of 12 months, the difference in absolute values of the observed and the forecasted inflation rates is 0.2%.

4 Discussion

The analyzed time series about the inflation rate in Switzerland from January 1956 until December 2018 still indicates periodic spikes in the ACF and PACF although no seasonal component is detected visually and by using empirical methods to decompose the series. The trend is eliminated in a simple way by using the lag-1 operator. Nevertheless, the remaining process is stationary and therefore it is reasonable to fit some ARIMA model to it. Due to the periodic spikes in the ACF and PACF it seems reasonable to model the remaining series with a SARIMA model. The evaluation of the AICc for SARIMA models with $p = 1, 2, 3, 4, 5; P = 1, 2, 3; q = 1, 2, 3, 4, 5$ and $Q = 1, 2, 3, 4$ reveals that a $SARIMA(5, 0, 5)$x$(1, 0, 2)[12]$ model is appropriate to model the remaining time series. Accordingly, to fit the data to this model, in total 14 parameters have to be estimated, which is a rather complicated model with a lot of computations. The parameters of the autoregressive part show a poor significance because only ϕ_4 was significantly different from zero at the 5% level. Though, all the parameters are included into the model since the predictions were the most accurate by using all of them. The model is valid, because the ACF clearly shows the behaviour of an iid-noise and the null hypothesis of the Ljung-Box test that the residuals of the modeled process is a white noise could not be rejected.

In terms of prediction the model yields relatively precise forecasts for the first four months in the year 2018. Afterwards, the forecasts clearly underestimate the observed inflation rate. The fact that ARIMA models are able to forecast only the short-term inflation rate is a phenomenon widely discussed in the literature, for examples see Pufnik et al. (2006) and Zhang et al. (2013).

5 Conculsion

This small project focuses on forecasting the inflation rate in Switzerland by using ARIMA models, specifically SARIMA models. Only the aggregated CPI data is considered. An extension of this approach would be to apply ARIMA models on disaggregated CPI data. That is, by using ARIMA models for each of the CPI components (energy, food and a total without energy and food) and by aggregating them together. The literature shows that this method yields more accurate short-term predictions of the CPI in Switzerland than the one used in this project, see Huwiler and Kaufmann (2013). Moreover, there exist other famous, but more complicated econometric models to forecast the inflation rate, for example vector autoregressive models (VAR) or autoregressive conditional heteroskedasticity models (ARCH), see Cologni and Manera (2008) and Fountas et al. (2004) respecitvely. It would be interesting to apply these models in a further step to the discussed time series data in this project and to compare the results.

List of Figures

1 The raw time series . . . 3
2 Lag-1 difference of inflation rate . . . 3
3 Sample ACF . . . 4
4 Sample PACF . . . 4
5 ACF of the model residuals . . . 5
6 Ljung-Box test for the model residuals . . . 5
7 12 month ahead forecasts of the inflation rate in % in the year 2019 . . . 6
8 Difference of the inflation rate in % between forecasted and observed values in the year 2018 . . . 7

List of Tables

1 Coefficients and their standard errors of the $SARIMA(5,0,5)$x$(1,0,2)[12]$ model . . . 5
2 12 months ahead prediction of the inflation rate in % . . . 6
3 Differences between observed and forecasted inflation rates in % in the year 2018 . . . 7

References

Cologni, A. and Manera, M. (2008). Oil prices, inflation and interest rates in a structural cointegrated var model for the g-7 countries. *Energy economics*, 30(3):856–888.

Fountas, S., Ioannidis, A., and Karanasos, M. (2004). Inflation, inflation uncertainty and a common european monetary policy. *The Manchester School*, 72(2):221–242.

Huwiler, M. and Kaufmann, D. (2013). Combining disaggregate forecasts for inflation: The snb's arima model: the snb's arima model. *Swiss National Bank Economic Study*, 7.

Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22.

Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

OECD (2019). Inflation (cpi). `https://data.oecd.org/price/inflation-cpi.htm#indicator-chart`. Accessed on 2019-05-09.

Pufnik, A., Kunovac, D., et al. (2006). Short-term forecasting of inflation in croatia with seasonal arima processes. Technical report.

Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.

Zhang, F. W., Che, W. G., Xu, B. B., and Xu, J. Z. (2013). The research of arma model in cpi time series. In *Applied Mechanics and Materials*, volume 347, pages 3099–3103. Trans Tech Publ.