**Econometric analysis of the Oil Price and its influence on industrial production in Austria**

Macroeconomterics: Empirical Project

SS2020

Katrin Aug (12021412) & Leah Bryson (01609098)

1. Introduction

* Hypothesis
* Description of our (raw) data, our approach and methods

1. Univariate analysis
   1. Univariate analysis of Total production
      1. First differences
      2. Model selection
   2. Univariate analysis of the Oil Price
      1. Model selection
2. Multivariate analysis
   1. Model diagnostics

3.2 Impulse Response function

1. Appendix
2. References
3. Introduction

Oil is also known as black gold, as it is a valuable resource used primarily in the transportation and industrial sectors. The strong dependence of industrialized nations on oil makes the oil price an interesting indicator of economic activity. Austrian GDP, and thus the prosperity of Austrian society, is composed of many factors, with industry playing a crucial role.

In this short paper, we would like to analyze the hypothesis that the price of oil influences the total Austrian production.

For this purpose, we use the monthly oil price in USD, as well as the monthly total production in Austria from 2000 - 2019. We decided to not include data for 2020, as the Corona crisis would significantly distort the data.

Oil is traded in different forms and there are several oil prices which can be used as a benchmark oil price, like brent crude oil etc. For reasons of data availability, we use a dataset with WTI Crude Oil Future prices.

The second data set is the Production of Total Industry in Austria. It is seasonally adjusted and the Index is set to 100 in 2015 (baseline).

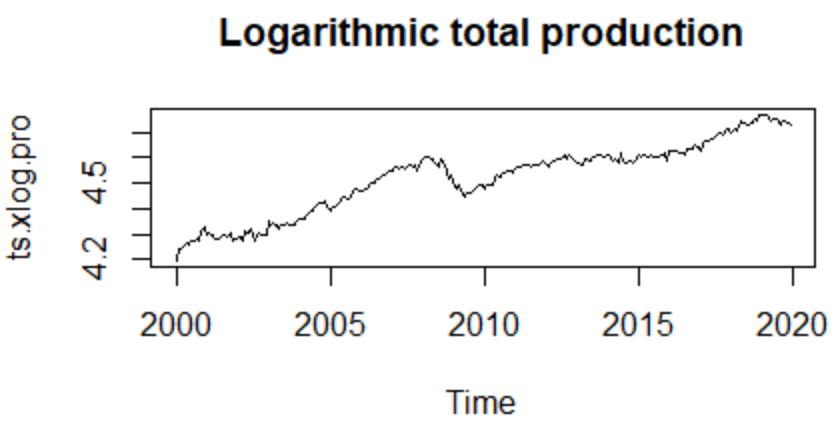
In the following, we will first perform a univariate analysis using an ARIMA model of the two datasets and then set up a multivariate model.

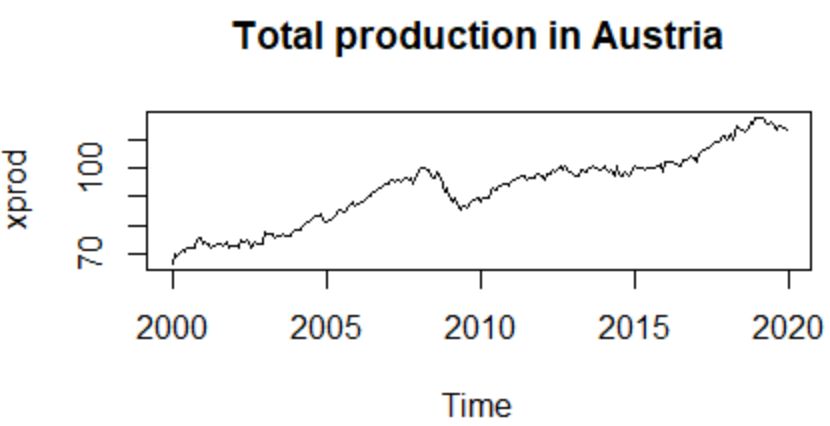
For this we use the program R and the packages timeseries, forecast, tidyverse, mFilter and TSstudio.

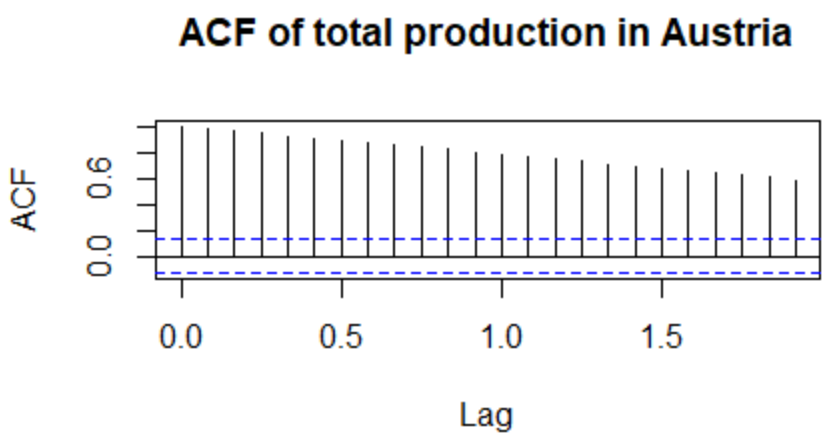
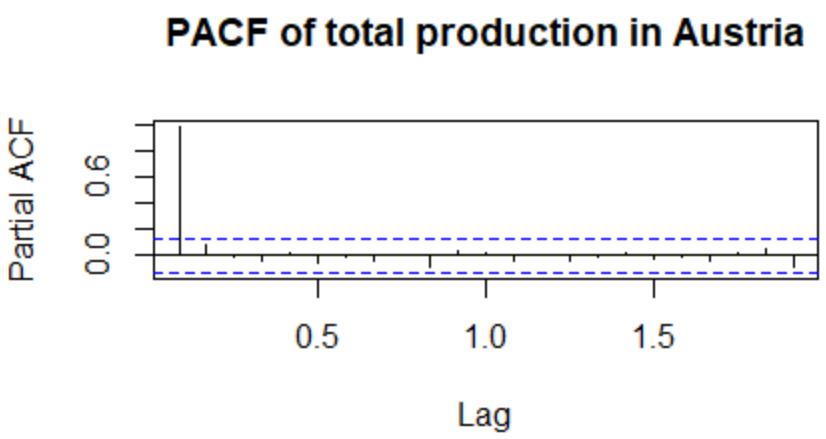
1. **Univariate analysis**
   1. **Total production**

First, we take a look on the absolute values of total production from 2000 to 2019:

As you can see, production has steadily increased, except for a decline after the 2008 financial crisis.

It is recommended to use the logarithm of national account data. However, as you can see, they look almost the same. As we have monthly data, we can argue that we do not have exponential growth leading to the logarithmic data looking almost as the original ones. This is why we decided to work with the non-logarithmic data for now.

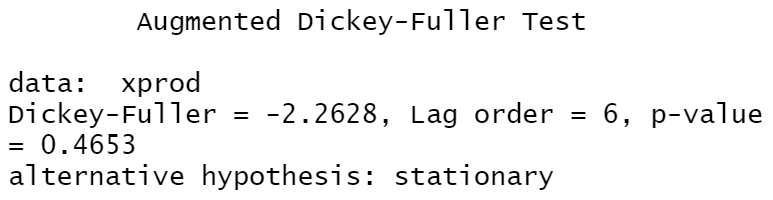


To better understand this time-series we examine the ACF and the PACF:

The ACF plot shows the correlation coefficients between the time series and its lagged values, in this case a geometric decay. The PACF plot shows a sharp drop after the first lag, indicating an AR(1) process as only the first lag crosses the 95% confidence interval (dashed blue line).

Even though the geometric decay in the ACF indicates that the time series is non-stationary, we also test for this. To do so we use the augmented Dickey-fuller test, which tests for a unit root under its null-hypothesis. It gives us a p-value of 0.4653, which indicates that we cannot reject the hypothesis at a significance level of 5%.

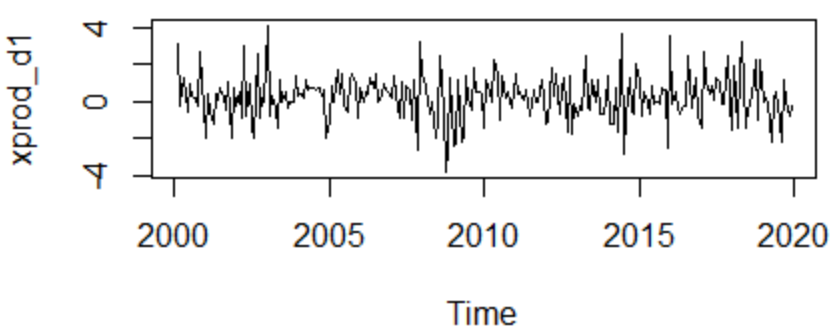
We therefore conclude that our time-series is a non-stationary process, which is not much of a surprise since we were working with absolute data up to now.



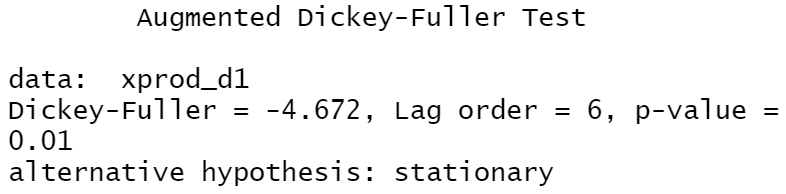
**2.1.1 First differences**

To tackle the non-stationarity problem we take first-differences, which gives us the following

plot:

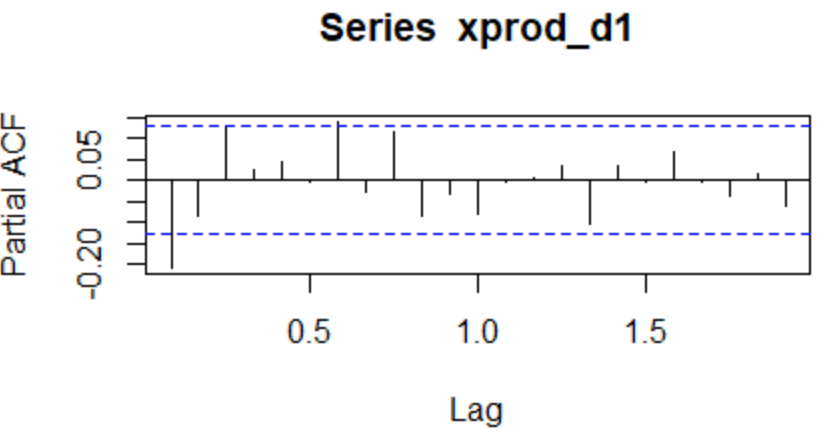
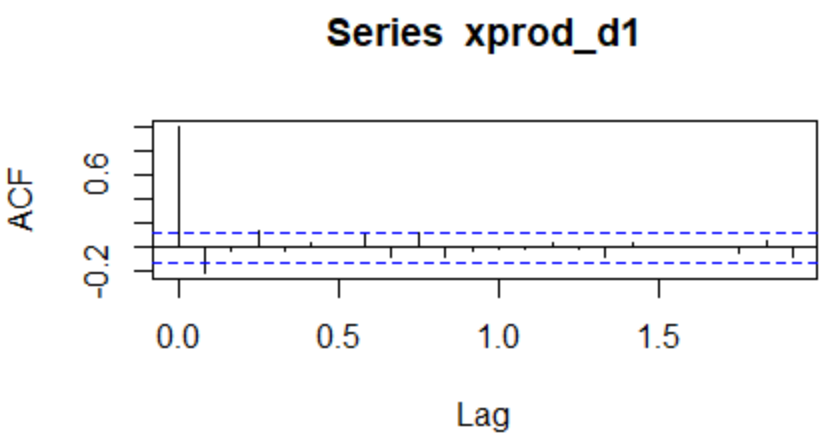


It is evident that the process, as well as the ACF below, look way more stationary than before. To make sure we are indeed dealing with a stationary process we once again conduct an augmented Dickey-Fuller test:



With a p-value of 0.01, we are now able to reject the null hypothesis, even on a 1% significance level, and conclude this is a stationary process.

Again, we plot the ACF and PACF to gain some more information:



The ACF cuts of after two lags, which might indicate an AR(2) process.

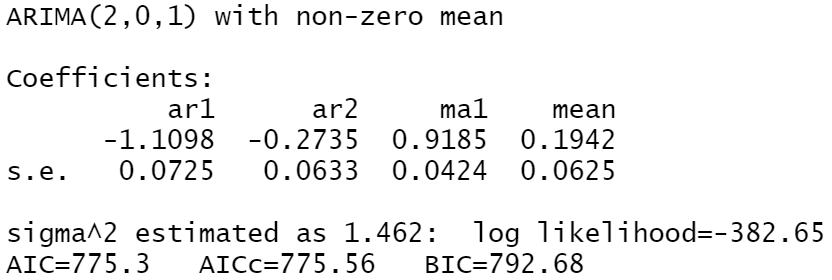
Both in the ACF & PACF we see significance not only on the first lags, but up to the sixth. However, as they barely even cross the blue line, we expect that they can be neglected. To further investigate this, we now proceed to the model selection.

**2.1.2 Model selection**

To find an appropriate ARMA model, we use the *auto.arima()* function in R which chooses the best ARIMA model based on some provided information criteria. We decide to use both the AIC and the BIC as information criteria and examine the results to then choose one of them. We got the following results:

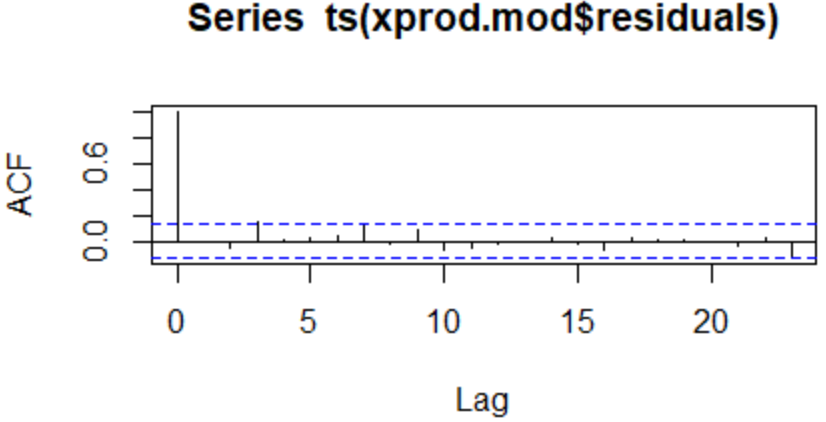
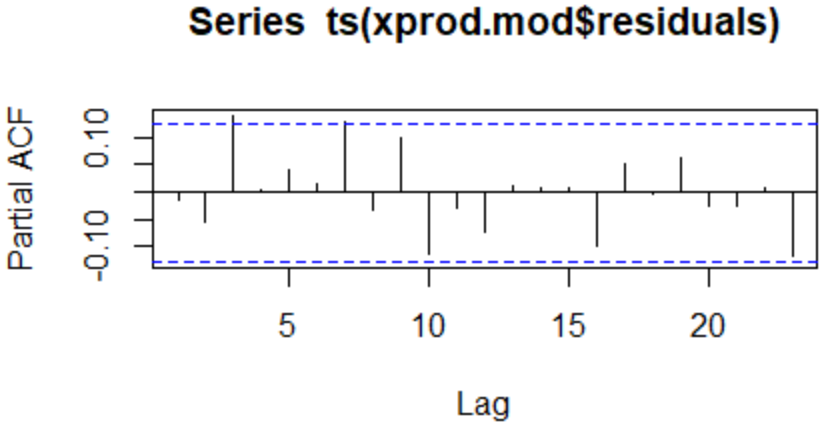
|  |  |  |  |
| --- | --- | --- | --- |
| Information criteria | Model obtained | AIC value | BIC value |
| AIC | ARIMA(2,0,1) | **775.3** | **792.68** |
| BIC | ARIMA(1,0,0) | 781.06 | 791.49 |

We see that both the AIC and the BIC value are slightly better when using AIC as information criteria. We see that it is indeed an AR(2) process, which is what we already suspected from looking at the ACF.

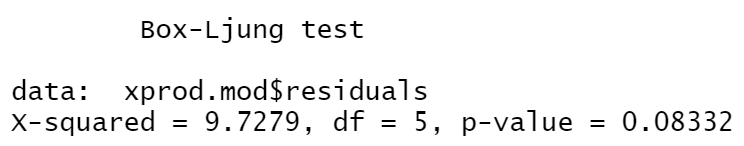
It has to be noted that the model has a non-zero mean. The mean obtained by the *auto.arima* function is 0.1942, indicating that the production increases on average (constant in the model).

We therefore obtain the following model:

Lastly, we check if there actually exists correlation between the residuals. For this, we use the Box-Ljung test (also called portmanteau or Q test which uses the Ljung-Box statistic). To determine the number of lags for the test we look at the (P)ACF functions:

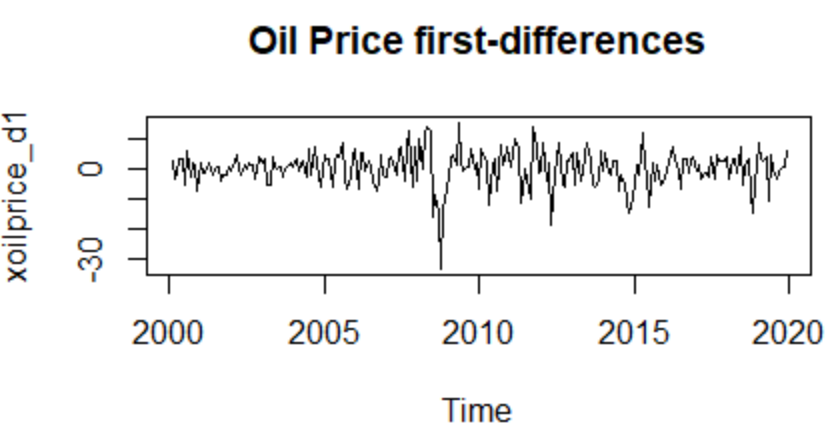


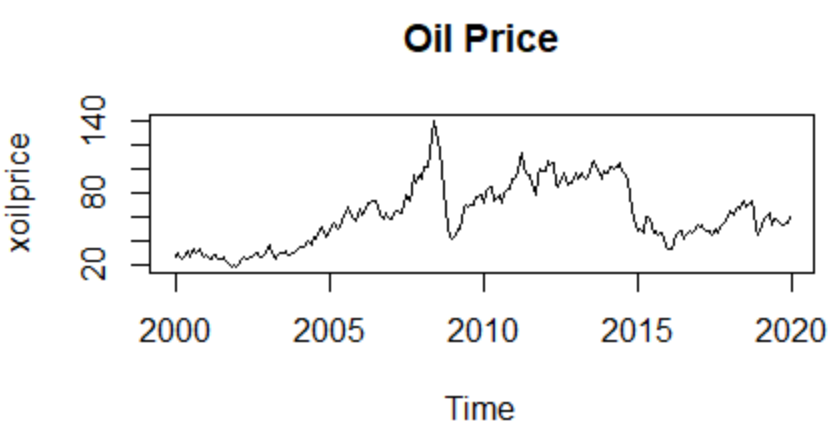
We can see that the last lag where the correlation exceeds the 95% confidence interval is the 7th. Having a p-value of 0.08332 we can clearly not reject the null-hypothesis, that there is no serial correlation between the residuals. We could of course also use a higher lag order. This would however not change the outcome, as the p-value increases with the number of lags. With 22 lags for example we have a p-value of 0.7043.



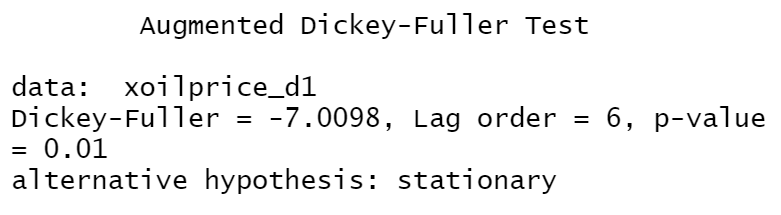
* 1. **Oil price**
     1. **First differences**

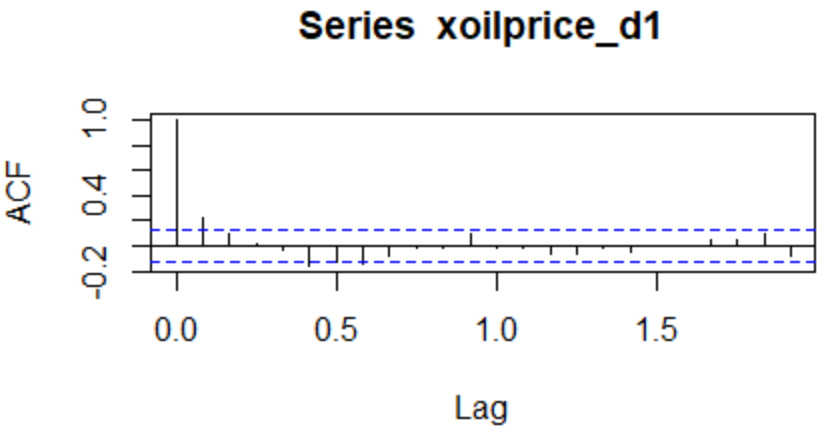
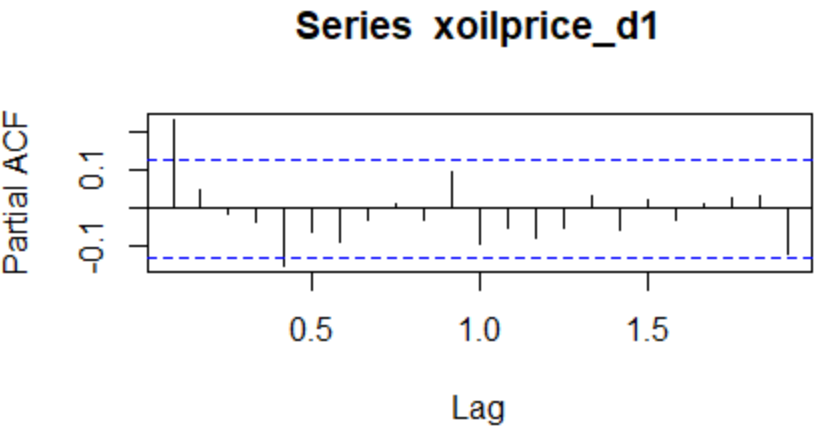
We basically proceed as we did with total production, so we will not go into as much detail as before.

First, we compare the data on oil prices with the corresponding first-differences. It becomes evident that the first-differences are more likely to be stationary so we continue working with this data.



To check for stationarity we again conduct the augmented Dickey-Fuller test. The p-value of 0.01 lets us reject the hypothesis and therefore conclude that this is also a stationary process.

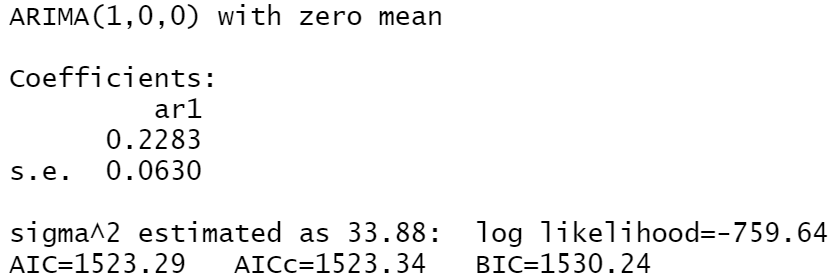


We then take a look at the ACF and the PACF:

Whilst the ACF shows significant lags at the first, second and sixth lag, the PACF dies so at the first and fifth lag. It is therefore difficult to construct an ARMA model from this information.

* + 1. **Model selection**

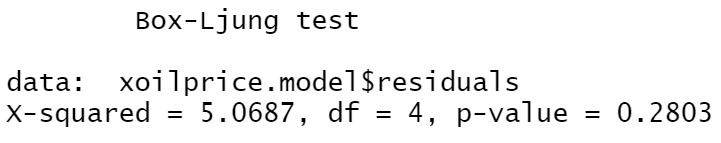
We again use AIC and BIC as information criteria to find an appropriate ARMA model. This time we get the same model, AIC and BIC value from both alternatives, to be exact:



Note that this time the model has a mean of zero.

We therefore conclude that the oil price can be described best by the following model:

Looking at the Box-Ljung test using 5 lags (determined by the (P)ACF), we get a p-value of 0.2803 and again see that the hypothesis of no serial correlation between the residuals cannot be rejected.

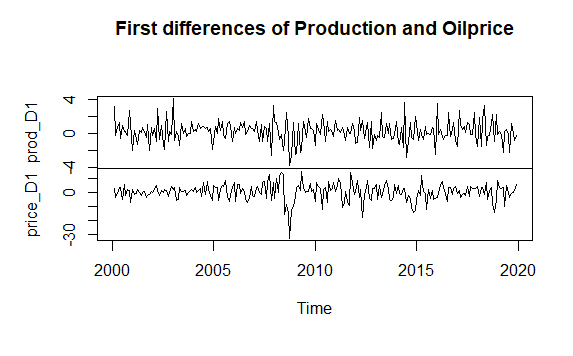
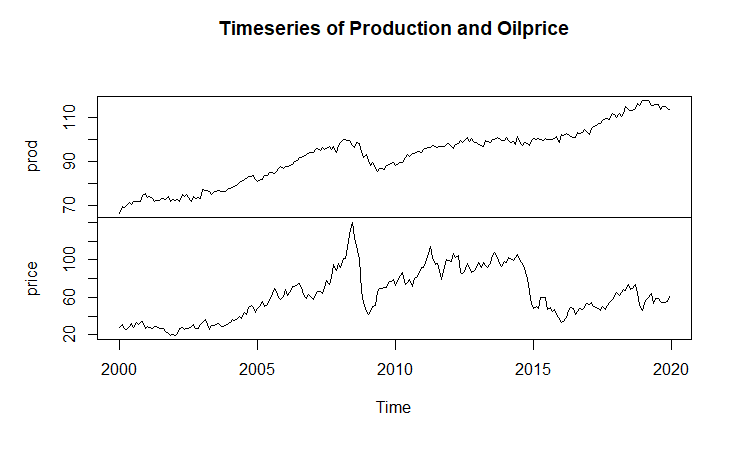


1. **Multivariate Analysis**

To capture the relationship between the oil price and total production of Austria over time we use the Vector autoregression model. As we mentioned before our hypothesis is that Austrian production strongly depends on the oil price.

To conduct the estimation we use the R package *vars*.

First of all, it is important to use stationary data. In section two, we have already tested for stationarity, so we can now continue to use them: the first differences of the total production and the oil price.



Firstly, we use the command *VARselect()* to determine the optimal number of lags.

|  |  |  |  |
| --- | --- | --- | --- |
| **AIC** | **HQ** | **SC** | **FPE** |
| **7** | 1 | 1 | 7 |

We choose to go with Akaike’s information criterion, so our process follows a vector autoregression of order 7.

We use seven lags and estimate the VAR model. The output gives us a total of 28 estimators, 10 of which are significant at a 5% level. Since it makes no sense to interpret these estimators, we will not show them here. It is noteworthy that the R^2 for both dependent variables is around 15%, which could indicate at least some predictive value of the model. We decided to focus on the model diagnostics.

We can see that X(prod) almost only depends on the lagged price, whilst Y(price) depends on both its own lags as well as lagged production.

**3.1 Model Diagnostics**

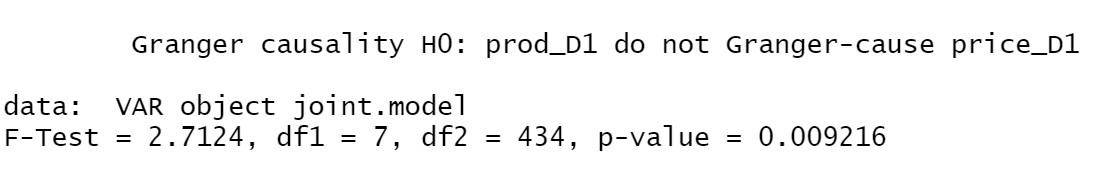
In the VAR model, we assume that the residuals are uncorrelated, which needs to be tested by the *serial.test()* command in R. This function computes the multivariate Portmanteau-test for serially correlated errors.

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung

The H0 for this test is that errors have no serial correlation whereas the H1 is errors have a serial correlation. We fail to reject the null hypothesis and conclude that errors have indeed no serial correlation.

Furthermore, we want to test for Granger causality, which basically tests if using the lagged values of the oil price to forecast production delivers a better forecast than only using the lagged values of the production and vice versa. We use the R command *causality()*, which uses a F-type Granger causality test and a Wald-type (testing for nonzero correlation between the error processes of the cause and effect variables) to test for Granger causality between our variables. We get the following results:



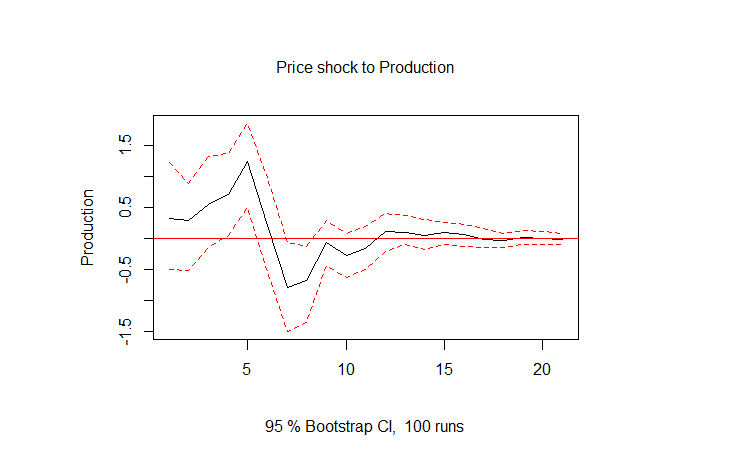
and

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung

We can see that the p-values are significant on a 5 % level, which means that production depends on lagged values of the price and the price depends on lagged values of production.

**3.2 Impulse Response function**

 In the plot below, we can see the impulse response of a positive price shock on production. The dashed-red line represents the 95%-confidence interval. We can see that a positive price shock increases the production for up to 5 periods, after that production stabilizes at the pre-shock level within a year.

1. **Appendix: Var Output**

===========================================================

Dependent variable:

----------------------------

y

(1) (2)

-----------------------------------------------------------

prod\_D1.l1 -0.236\*\*\* 0.202

(0.067) (0.321)

price\_D1.l1 0.031\*\* 0.167\*\*

(0.014) (0.067)

prod\_D1.l2 -0.109 0.454

(0.068) (0.325)

price\_D1.l2 -0.006 0.054

(0.014) (0.068)

prod\_D1.l3 0.065 0.614\*

(0.069) (0.327)

price\_D1.l3 0.011 0.005

(0.014) (0.067)

prod\_D1.l4 -0.013 1.051\*\*\*

(0.068) (0.326)

price\_D1.l4 0.028\*\* -0.017

(0.014) (0.066)

prod\_D1.l5 -0.006 0.234

(0.069) (0.329)

price\_D1.l5 -0.010 -0.151\*\*

(0.014) (0.066)

prod\_D1.l6 0.002 -0.653\*\*

(0.069) (0.329)

price\_D1.l6 0.038\*\*\* -0.063

(0.014) (0.067)

prod\_D1.l7 0.139\*\* -0.609\*

(0.066) (0.316)

price\_D1.l7 0.004 -0.108

(0.014) (0.068)

const 0.193\*\* -0.085

(0.088) (0.420)

-----------------------------------------------------------

Observations 232 232

R2 0.150 0.163

Adjusted R2 0.095 0.109

Residual Std. Error (df = 217) 1.193 5.695

F Statistic (df = 14; 217) 2.734\*\*\* 3.010\*\*\*

===========================================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

1. **References**

Data on Oil Prices: Crude Oil WTI Futures Historical Data, retrieved from investing.com; https://www.investing.com/commodities/crude-oil-historical-data, June 19.2021.

Data on Production: Organization for Economic Co-operation and Development, Production of Total Industry in Austria [AUTPROINDMISMEI], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/AUTPROINDMISMEI, June 19, 2021.