WIRTSCHAFTSUNIVERSITÄT WIEN
Vienna University of Economics and Business

Bachelor Thesis

Titel of Bachelor Thesis (english)	Forecasting of economic data using state-of-the-art methods
Titel of Bachelor Thesis (german)	Prognostizieren von Wirtschaftsdaten mit modernsten Methoden
Author (last name, first name):	El Sirfy, Maroan
Student ID number:	11839876
Degree program:	Bachelor of Science (WU), BSc (WU)
Examiner (degree, first name, last name):	DPhil, Laura, Vana Gür

I hereby declare that:

1. I have written this Bachelor thesis myself, independently and without the aid of unfair or unauthorized resources. Whenever content has been taken directly or indirectly from other sources, this has been indicated and the source referenced.

2. This Bachelor Thesis has not been previously presented as an examination paper in this or any other form in Austria or abroad.

3. This Bachelor Thesis is identical with the thesis assessed by the examiner.

4. (only applicable if the thesis was written by more than one author): this Bachelor thesis was written together with

The individual contributions of each writer as well as the co-written passages have been indicated.

23/09/2021
Date

M. El Sirfy
Unterschrift

Forecasting of Economic Data using State-of-the-Art Methods

Bachelor Thesis

Maroan El Sirfy

Student ID number: 11839876
Subject Area: Information Business
Supervisor: Laura Vana Gür, DPhil
22. September 2021

Contents

List of Figures

List of Tables

Abstract

This thesis compares time series forecasting methods by using quarterly U.S. economic indicators. Three approaches were analysed, the appropriateness tested and then used to forecast data, based on which the models were evaluated: (1) a univariate autoregressive ARIMA approach, (2) a multivariate autoregressive VAR and (3) a multivariate machine learning approach LSTM. The forecasting results presented fewer errors in the LSTM model than the VAR in all and ARIMA models in most cases. The results suggest that the LSTM method yields better predictions for the chosen economic data than an autoregressive approach.

1 Introduction

The forecasting of key macroeconomic indicators such as GPD, debt and consumer expenditure not only take part in broadening the understanding of macroeconomic behaviour based on past events but rather is and will continue to be used to support crucial and far-reaching decision-making processes.

This levitates the importance of exactness and reliability in the methods employed, used to forecast specificities about the indicators. It can often require "discretionary choices about the data and methods" (Hall, 2018). Being able to accurately forecast indicators has, therefore, historically at least partly been based on subjective economic intuition and judgement.

Raising objectivity and therefore the correctness of predictions, however, can be achieved by employing data-driven approaches to mitigate subjective influence on the results. Equally important then, is the ability to judge the accuracy of the machine learning models employed.

Considering the significance of an accurate prediction, this thesis aims to provide an overview of currently available approaches and discusses certain models in detail.

1.1 Research Question

The recent economic state, where a short-term recession has majorly influenced recent economic policies as well as monetary policies in a to-date unprecedented way, shows the volatility of these markets. This thesis will attempt to answer the question if commonly employed forecasting methods can perform well in predicting economic data while taking its recent state into account.

2 Theoretical Background

As this thesis will discuss methods of forecasting for specific types of data, a literature overview of currently available findings will provide the ability to reflect the topic in its current research context. For this, two main areas have to be investigated: time series data and forecasting methods.

2.1 Overview: Time Series Data

Time series are sequentially over time collected observations. The data points in a time series serve two distinct purposes as part of an analysis: studying the dynamic structure of the series and forecasting future data points based on the information available in that series or related ones. This exemplifies a unique feature of time series, where observations from a common population cannot be assumed to be independent. Rather, studying dependencies in time series is a fundamental concept. (Cryer & Chan, 2008, Chapter 1)

Multivariate time series achieve just that. As opposed to univariate series, where a single varying dimension is observed, multivariate time series consider the dependencies between multiple variables. Economic data on a macro scale make assumptions about dependencies between observations likely, given a common population of the observations.

Stationarity in time series is the fundamental assumption about the statistical properties of a time series. To identify if a time series is stationary autocorrelation can be used. A strict stationarity entails, that the mean, variance, and autocorrelation structure do not change over time. Therefore, a series that shows cyclical behaviour is stationary. However, one that shows seasonality or trends is not stationary. Time series are commonly assumed to be weakly stationary. A non-stationary time series can be transformed into a stationary one using a variety of methods, such as differencing. (Hyndman & Athanasopoulos, 2018; Prins et al., 2021, Chapter 6.4; Tsay, 2005)

2.2 Overview: Univariate Methods

Formally the same, and similar in properties to a simple linear regression model, is an autoregressive (AR) model often seen in time series literature (Tsay, 2005, Chapter 2.4). Prins et al. (2021) note that in fact, the "linear regression of the current value of the series against one or more prior values of the series" (Chapter 6.4.4.4.), simply is an AR model as they show through this Equation (1):

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-1} + A_t \tag{1}$$

As Tsay (2005) explains thoroughly AR models, therefore, are weak stationary, which "enables one to make inferences concerning future observations" (p. 25). Moreover, it makes the interpretation of the model straightforward because they can be analyzed using standard linear least squares techniques. They are a common approach to univariate time series analysis.

The errors A_t are ideally homoscedastic and preferably normally distributed. Testing the independence, homoscedasticity and distribution of residuals is an important way to diagnose ARIMA models. (Date, 2021; Yürekli et al., 2005)

Similar to AR models and just as common are simple moving average (MA) models. According to Prins et al. (2021) the difference being conceptually that it is not the linear regression of the current value against one or more prior values but random shocks of prior values (white noise), where these random shocks can be "propagated to future values of the time series" ("MA Models" section). Moving average models can be denoted as in Equation (2):

$$X_t = \mu + A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - \cdots - \theta_q A_{t-q} \tag{2}$$

The interpretation of moving average models is more complicated than for AR models, as non-linear procedures need to be used instead of linear least squares.

Combining the autoregressive approach with the moving average model results in an ARMA model, popularized by the statisticians George Box and Gwilym Jenkins hence often referred to as the Box-Jenkins Approach.

The Equation (3) according to Prins et al. (2021, Chapter 6.4.4.5) for an ARMA model is just a combination of the AR (1) and MA (2)

$$\begin{aligned} X_t \quad = \quad & \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \\ & A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - \cdots - \theta_q A_{t-q} \,, \end{aligned} \tag{3}$$

which assumes a (weakly) stationary time series.

ARIMA, short for autoregressive integrated moving average, is used in cases where the data shows evidence for non-stationarity. In such cases, differencing the time series achieves stationarity, which produces the ARIMA model.

2.3 Overview: Multivariate Methods

The multivariate equivalent of an AR model is a vector autoregression (VAR) model, sometimes also called ARV model. It simply incorporates autoregressive parameters in form of a matrix, where more than one vector of observation can be included in the model. Equation (4) is, therefore, similar to an AR model:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + a_t, \tag{4}$$

where x_t is a vector of observation, ϕ_k is an $n \times n$ matrix of AR parameters and a_t "multivariate white noise" (Prins et al., 2021, Chapter 6.4.5.). Like with ARIMA models, a_t refers to the errors, which are ideally homoscedastic, normally distributed, and not autocorrelated.

There are also vector ARMA or VARMA models, however, Tsay (2005) list a wide range of issues by including the moving average with AR in a vector model such as the "*identifiability* problem", where the VARMA "model for the data is not uniquely defined" (p. 371). Simply put, it can result in a misleading situation of exact multicollinearity (Kim, 2019). Simionescu (2013), having researched VARMA models in the context of economic data comes to the same "conclusion … reached by other researchers" (p. 9), where VARMA models are difficult to apply in practice due to their structure, where for small time horizons, using stable variables and carefully choosing the measure of evaluation, they can yield better forecasting results. Moreover, Tsay states that "VAR … models are sufficient in most [cases]" (p. 372).

2.4 Overview: Long-Short Term Memory

A widely used time series analysis approach is the Long-Short Term Memory (LSTM) neural network. LSTM is a special kind of recurrent neural network, a feature of which is, allowing learned information to persist and utilize it. While recurrent neural networks might have problems learning long-term dependencies, where relevant needed information might be too far back in the learning process, LSTMs are explicitly designed with long-term dependencies in mind. Although LSTMs are not exclusively used for time series data and its use cases, they offer forecasting functionality for both, univariate and multivariate time series. (Olah, 2015)

Recent studies have shown that ARIMA models can outperform widely popular modern neural network approaches, such as LSTM in cases of univariate forecasting, countering the notion that machine learning methods are "the key solution to all predictive modelling problems" (Brownlee, 2019b, para. 1). Brownlee mentions the importance of using classical methods "as a baseline when evaluating any machine learning … methods" (para. 4).

3 Data

Forecasting economic data, by definition, requires data points collected in a timely context. The main data used for this thesis is therefore going to be time-series data.

3.1 Indicators

The thesis focuses on economic indicators of the United States of America. “Classic” economic indicators have been chosen, for the sake of simplicity, the seemingly overarching consensus of non-issues with well-established indicators, and because this thesis focuses on the forecasting approach.

The main advantage for the chosen indicators is, that the data is openly available, and reaches back more than 70 years which provides more data points for later use in the forecasting methods. Furthermore, U.S. data taken from the in section 3.2 mentioned source has the advantage of being closely monitored by publicly and privately competing interests, assuring accuracy. (Fox et al., 2020, Chapter 1, p. 2)

A disadvantage of using the aforementioned indicators is, that they only represent an estimated economic indicator to evaluate the current state of economic activity. Individual indicators (e.g., GDP alone), do not necessitate an accurate depiction of the periodic wealth within an economy (Tzvetkova & Hepburn, 2018).

3.1.1 Gross Domestic Product

According to the U.S. Bureau of Economic Analysis [BEA] (2020), whose data is used in this thesis, the GDP (Gross Domestic Product) is the market “value of the goods and services produced [by labour and property located] in the United States” (para. 1). Meaning, it represents a feature measure of US economic output. It, therefore, supplies a solid basis for the analysis of economic data and its forecasting.

3.1.2 Federal Debt held by Federal Reserve Bank

Federal debt (also called national debt) according to the United States Department of the Treasury [U.S. Treasury] (2010) refers to the sum of all direct liabilities of the U.S. government.

While public debt makes up about 78% of the national debt (U.S. Treasury, 2021b) and includes investors (foreign and national), public debt held by federal reserve banks recently represents a uniquely significant portion therein (U.S. Treasury, 2021a, p. 53). This is largely due to the unprecedented allowance of the Federal Reserve for an unlimited amount of quantitative easing since March 2020 (Board of Federal Reserve, 2020; U.S. Treasury, 2021a, p. 7).

Because of the aforementioned reasons, its recentness and its significant impact on the future economic state it has been chosen as opposed to a total account of accumulated public debt in the U.S.

3.1.3 Personal Consumption Expenditure

As explained in the methodology used by the BEA, Fox et al. (2020, Chapter 5: PCE) define personal consumer expenditure (PCE) as "the goods and services purchased by 'persons' [households and NPO serving households]" (p. 2), which represents the "primary engine that drives future economic growth" (p. 1) because it shows the amount households are spending on consumption as part of their income and how much they save, i.e. have not yet spent on consumption.

3.1.4 (Natural) Rate of Unemployment

The rate of unemployment is the percentage of the labour force that is unemployed.

Cyclical shifts (determined by demand) will continuously cause unemployment to diverge from the "natural rate". Milton Friedman explains this as the hypothetical rate under "full employment" (Nobelprize.org, 1976).

The natural rate of unemployment (also called the noncyclical rate of unemployment), therefore, represents percentage-wise “the rate of unemployment arising from all sources except fluctuations in aggregate demand” (Federal Reserve Bank of St. Louis [St. Louis Fed], 2021). It has been chosen as an indicator to possibly mitigate the high volatility in the rate of unemployment, as well as its importance in estimating potential GDP. (St. Louis Fed, 2021)

3.2 Data Origin

The datasets are available through FRED (Federal Reserve Economic Data), which is an open online database aggregating US national data collected publicly and privately.

Five types of indicators (and therefore datasets) have been chosen for the aforementioned reasons: Gross Domestic Product (BEA, 2021b), Unemployment Rate (U.S. Bureau of Labor Statistics, 2021), Natural Rate of Unemployment (U.S. Congressional Budget Office, 2021), Federal Debt held by Federal Reserve Banks (U.S. Treasury. Fiscal Service, 2021), and Personal Consumption Expenditures (BEA, 2021c).

3.3 Legal and Ethical Framework

As all of the data is openly available through FRED, the Research Department of the Federal Reserve Bank of St. Louis, there are no legal or ethical concerns taken into consideration on this occasion. (St. Louis Fed, n.d., “Terms of Use” section)

3.4 Characteristics of the Data

The main datasets used are time-series datasets, where the data is indexed in time order. The time-frequency of the data points is quarterly. The data types are a date factor and numeric.

To analyse a multivariate time series, combining relevant indicators on an equal time axis is a crucial step to have meaningful input. To alleviate issues in inaccurately grouping or converting monthly data to quarterly or yearly or vice-versa, all datasets chosen are using quarterly data points.

3.4.1 Attributes of the Data

The quarterly available dates for all datasets are 01.01, 01.04, 01.07, and 01.10 of each year respectively. The GDP, as well as the PCE data points, reach from 1947 to 2021. They, therefore, include 297 observations. The debt dataset (FDHBFRB) data points reach from 1970 to 2021. The unemployment data, however, is slightly different, starting from 1948 and 1949 respectively but including estimated future data points which will be excluded in this thesis.

The second attribute of the five datasets is their respective value. GDP, PCE and debt are measured in billions of dollars accurate to three decimals. GDP and PCE are furthermore seasonally adjusted by their annual rate. Both unemployment rates are measured in percentages.

3.5 Preparation of the Data

The original datasets do not need initial data processing for inserting or reading the data. As for the accuracy of the data, it is already processed and verified as even "in a few instances, series are derived by BEA after… reworking of the source data" (BEA, 2021a, "Key source data and assumptions" section).

Combining the different datasets to create a multivariate time series is done by joining them on the dates. If there are unequal starting points in the time series, earlier entries (such as GDP and PCE) will be omitted, to have a common starting point among all 5 datasets, vice-versa with end points. All predictive datapoints of the unemployment datasets are omitted. This results in the joint starting date of 01.01.1070 and end date of 01.01.2021, encompassing 205 observations.

4 Approaches

To identify the accurateness of forecasting methods, three common approaches will be tested in this context of economic data. As discussed in the theoretical overview section 2.4, using classic approaches to evaluate machine learning ones not only provides context but also establishes benchmarks to beat. This thesis will analyse ARIMA, a univariate approach, VAR, a multivariate approach, and LSTM a machine learning approach. The models will be compared based on their predicted vectors with root mean squared errors (RSME). It represents the normalized distance between the predicted and the observed values and is suitable to measure the accuracy of predictions (Hyndman & Koehler, 2006).

4.1 Exploratory Data Analysis

To explore the data and possibly see trend or seasonality, simple plots of the data show the time series structure.

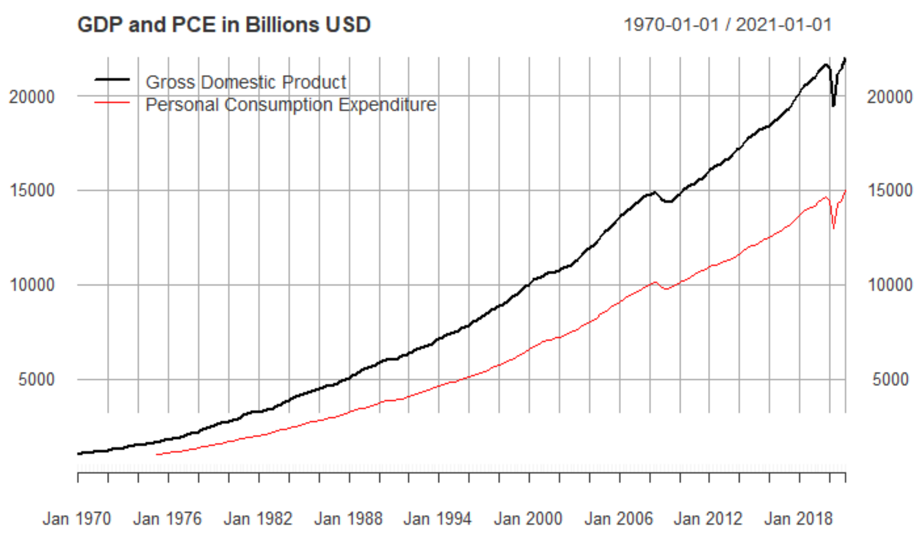

Figure 1: GDP and PCE (USA)

In Figure 1 the structural similarity between gross domestic product and the personal consumption expenditure is notable and a clear trend is visible in both cases.

In Figure 2, a clear trend is visible as well, where fitting a model might be less straightforward. Notable is the recent fast increase in 2021 in debt, which might be difficult for a model to accurately predict.

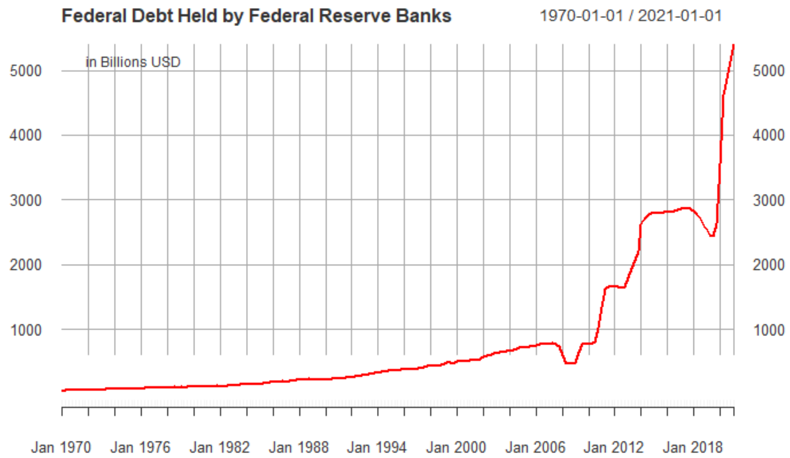

Figure 2: Federal Debt held by Federal Reserve Banks (USA)

The unemployment rate is volatile in contrast to the other time series used, with a recent all-time high increase in 2021 as can be seen in Figure 3. It is likely (weakly) stationary but will have to be analyzed empirically in the next section.

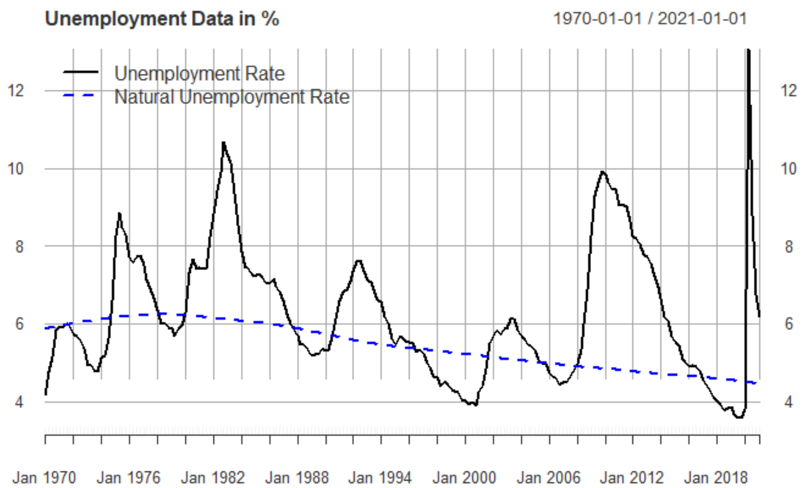

Figure 3: Unemployment Rate (USA)

Table 1: Correlogram of all variables

	GDP	PCE	debt	UR	NUR
GDP	1.000000	0.999814	0.854244	-0.185427	-0.973562
PCE	0.999814	1.000000	0.853427	-0.180416	-0.973641
debt	0.854244	0.853427	1.000000	-0.100349	-0.780960
UR	-0.185427	-0.180416	-0.100349	1.000000	0.229863
NUR	-0.973562	-0.973641	-0.780960	0.229863	1.000000

In Table 1, a strong positive correlation between GDP, PCE and debt can be seen and a negative one between those variables and NUR. The correlations indicate that creating a multivariate series from these variables is justified.

Table 2: Correlogram of all differenced variables

	GDP	PCE	debt	UR	NUR
GDP	1.000000	0.975762	-0.438339	-0.885755	-0.106398
PCE	0.975762	1.000000	-0.428271	-0.865018	-0.104067
debt	-0.438339	-0.428271	1.000000	0.434965	-0.072006
UR	-0.885755	-0.865018	0.434965	1.000000	0.027843
NUR	-0.106398	-0.104067	-0.072006	0.027843	1.000000

For the differenced data in Table 2, the strong positive correlation between GDP and PCE remains but debt is no longer correlated. Differenced UR has a strong negative correlation with GDP and PCE. The negative correlation between NUR and the other variables is gone.

4.2 Time Series Analysis

Before any model-building approach can be conducted, determining the stationarity of a time series is fundamental for the next steps. The stationarity can be analysed by autocorrelating a time series.

4.2.1 Autocorrelation

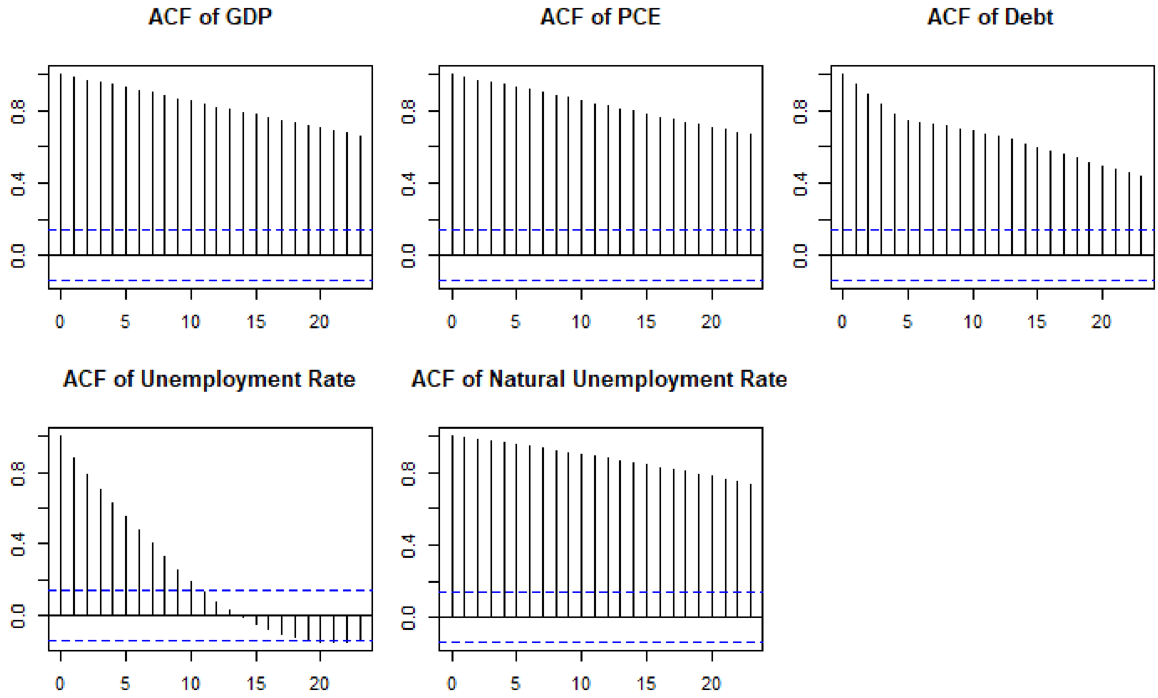

Figure 4: Autocorrelation of the time series
Y-axis is the ACF, X-Axis is the lag value

Every time series in Figure 4: Autocorrelation of the time series has more than the first 10 lag values outside the blue indication, meaning the 95% confidence interval. It indicates, the autocorrelations in each series are significantly different from zero at the 5% level and shows the non-stationarity of the time series. This means while modelling, differencing will have to be considered.

The only exception to this is the shown ACF plot of the unemployment rate in Figure 4. To investigate empirically if this series is stationary, an Augmented Dickey-Fuller (ADF) test can be used (Prabhakaran, 2019).

With this, the alternative hypothesis is tested to see if the series is stationary.

Table 3: ADF Test of Unemployment Rate

```
##  Augmented Dickey-Fuller Test
##
## data:  ts_main$UR
## Dickey-Fuller = -3.49, Lag order = 5, p-value = 0.04494
## alternative hypothesis: stationary
```

We can see, the p-value for the unemployment rate time series is lower than the significance level of 0.05, hence the null hypothesis of a non-stationarity can be rejected. The time series is stationary.

4.2.2 Partial Autocorrelation

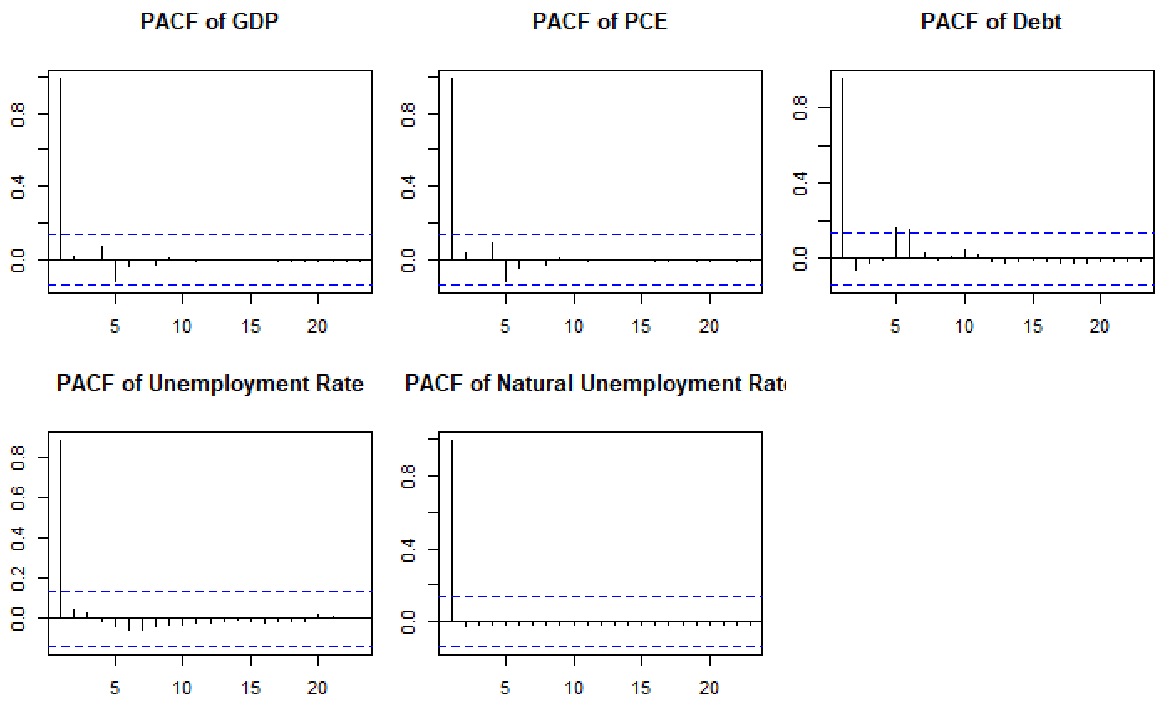

Figure 5: Partial autocorrelation of the time series.
Y-axis is the PACF, X-Axis is the lag value

The partial autocorrelation plots the autocorrelation of the residuals. In this case, the residuals are the effects not already explained by earlier lags. In Figure 5 there are, with the singular exception of the PACF of Debt, no significant correlations found of the residuals.

4.3 ARIMA Approach

As discussed in the overview of ARMIA, it is a development of a common autoregressive model, combining it with the moving average resulting in a Box-Jenkins approach.

Important to note before being able to present results is that ARIMA models are notated by ARIMA(*p,d,q*). In this context, p is the order of the autoregressive part i.e., AR(*p*), d the degree of differencing involved i.e., I(*d*), and q the order of the moving average part of the model i.e., MA(*q*) (Hyndman & Athanasopoulos, 2018, Chapter 8.5).

4.3.1 Estimating Models

ARIMA is a univariate approach, so single indicators will have to be considered. Therefore, each of the five indicators will be considered separately. To select the appropriate values for *p,d,* and *q*, the auto.arima() algorithm will be used for all models (for reference see supplementary code). It automatically selects the order of AR and MA used. For this, the Akaike's Information Criterion is used and minimized as much as possible. However, for selecting the degree of differencing, which is necessary as mentioned in section 4.2 the information criteria usually tend not to be well suited, as differencing changes the data. For estimating equation parameters of the model, the maximum likelihood estimation (MLE) can be used. (Hyndman & Athanasopoulos, 2018, Chapter 8.6)

The estimation of the following models can be seen in the reference code file, the following sections will consider the residuals of the models and their forecasting. As the diagnosis for all models is conducted similarly (see reference code), the steps will be shown in section 4.3.1.1 fully and then only if necessary to avoid repetitiveness. The forecasting will initially be standardized to a 95% confidence level with 10 years. The data used to train the model is split into 80% training and 20% test data against which the forecasting results will be evaluated.

4.3.1.1 ARIMA Model GDP

The resulting model for GDP is an ARIMA(1,2,2) model, with $AIC = 1755.086$. A lower AIC would be considered better, but among the tried configurations, this represents the best. Looking at the autocorrelation of the residuals in Figure 6 one can see that there are no significant autocorrelations left in the residuals. This indicates that the residuals act like white noise, which is ideal. However, the residual plot shows a rather distorted normal distribution, as there are few bins with higher concentration than others. Moreover, the residual variance is heteroscedastic, with a significant spike near the end of the series. Stockhammar and Öller (2012) mention a solution to deal with heteroscedasticity in time series: employing (generalized) autoregressive conditional heteroscedasticity (ARCH/GARCH).

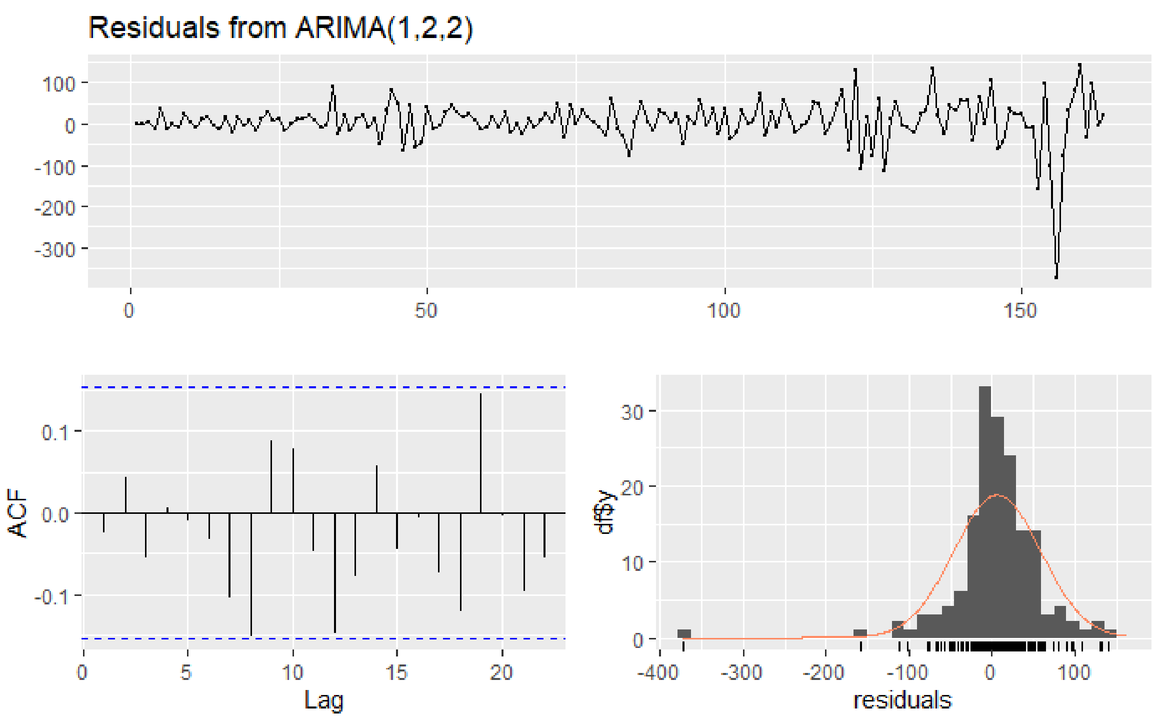

Figure 6: Residuals of the ARIMA(1,2,2) GDP model

Extending the diagnosing of the model, one can test the lack of fit by using the Box-Ljung test, which also tests the autocorrelations of the residuals (Prins et al., 2021). From Table 4 we can see the output of the test for this ARIMA model, with a p-value over the 0.05 significance level. We can conclude that there is no significant lack of fit.

Table 4: Box-Ljung test of the GDP ARIMA(1,2,2) model residuals

```
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,2,2)
## Q* = 9.5281, df = 7, p-value = 0.2169
##
## Model df: 3.   Total lags used: 10
```

Since the model can fit the data, calculating a forecast is the next step. In Figure 7 we can see the GDP ARIMA(1,2,2) model used to predict 10 years quarterly (blue), and the "real" test data (red). The grey outline signifies the confidence interval.

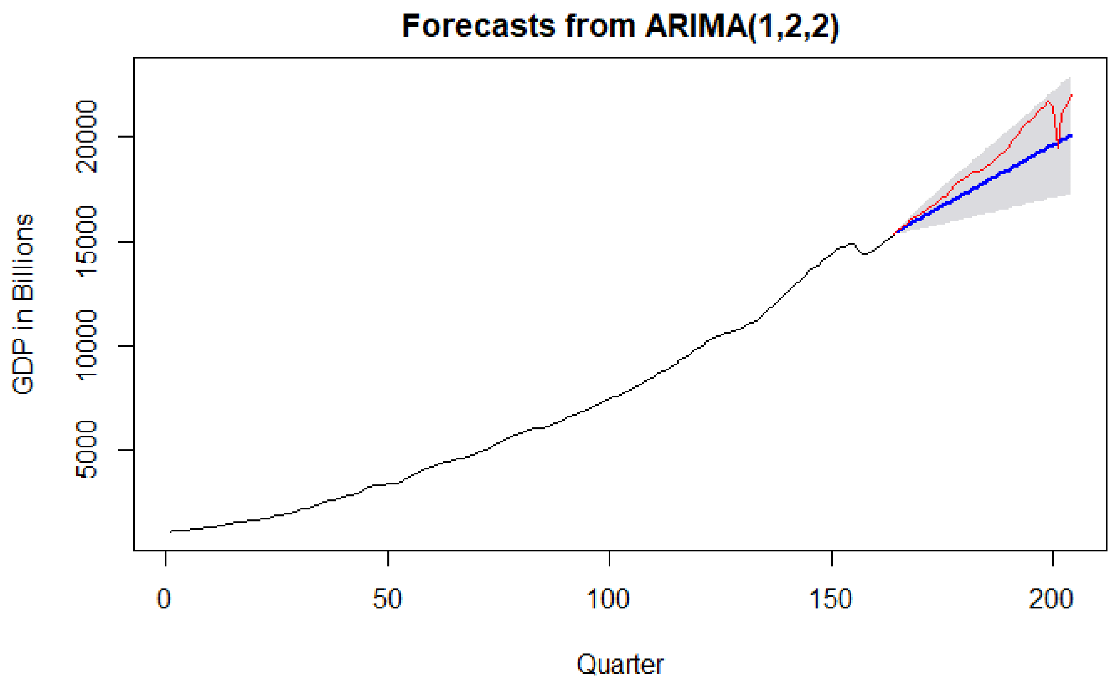

Figure 7: GDP Forecast using ARIMA(1,2,2)

The RSME for this prediction is 982.551, however, one must keep in mind, that the error is calculated in absolute terms and can only be used for comparison within the same variable for different models.

4.3.1.2 ARIMA Model PCE

The ARIMA model used for PCE is similar to that used for the GDP, due to their structural similarity. The algorithm suggests ARIMA(1,2,1) with an AIC of 1636.59. The Box-Ljung test results in non-significance and the residuals are within threshold, hence forecasting can commence. Figure 8 shows the 10-year forecast, including confidence interval and test data to validate. The RSME for this prediction is 584.929.

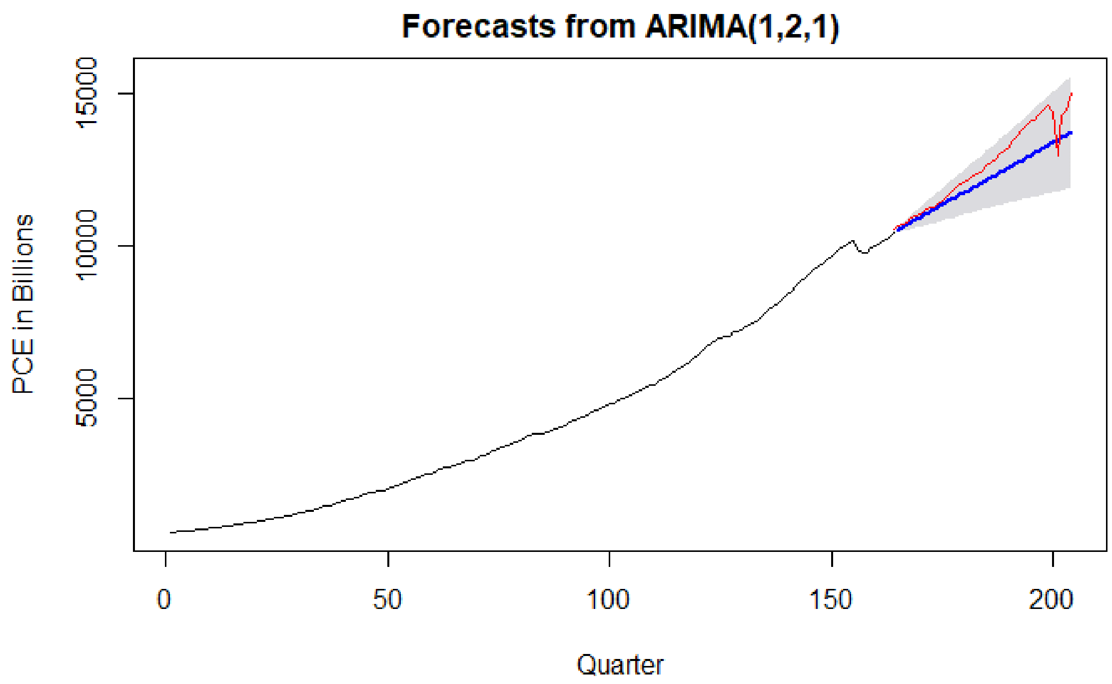

Figure 8: PCE Forecast using ARIMA(1,2,1)

4.3.1.3 ARIMA Model Debt

For the debt time series, minimizing the AIC results in ARIMA(0,1,5) with a drift. The AIC is 1480.275 so quite high, but the Box-Ljung test still results in non-significance for the autocorrelations of the residuals. However, heteroscedastic residuals in Figure 9 indicate a deficient model.

The 10-year prediction can be seen in Figure 10. Due to the significant increase in 2021 as opposed to lagged values, the test data is beyond the significance level threshold and might be impossible to forecast validly.

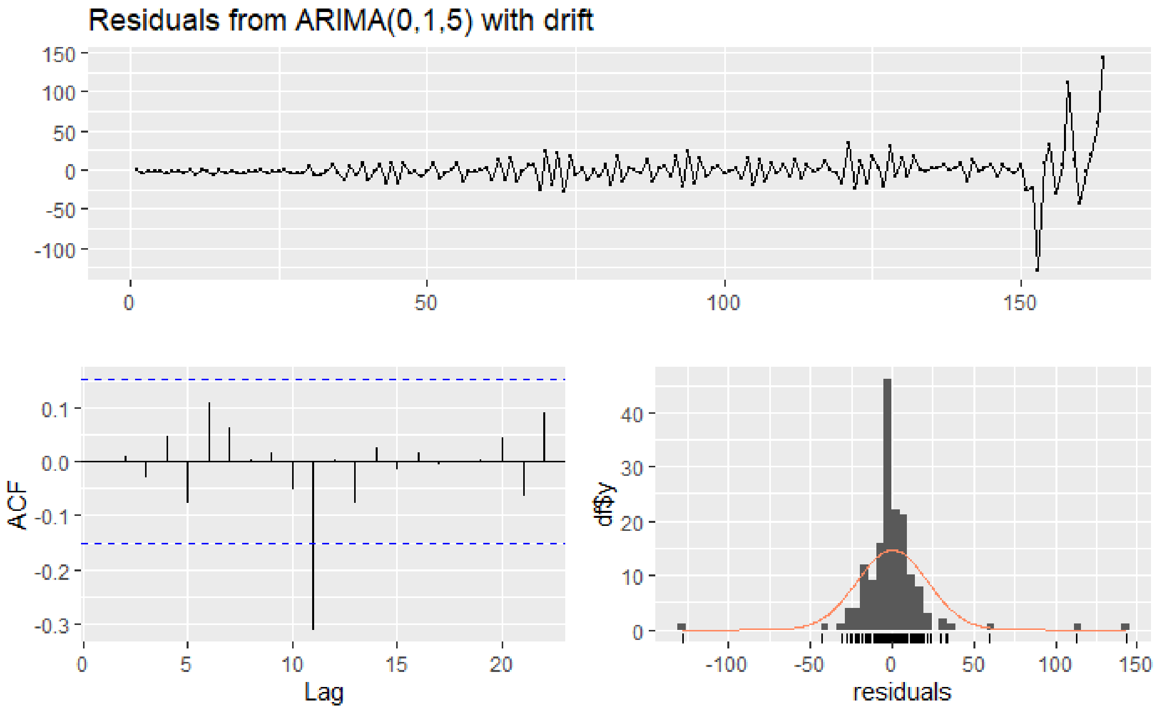

Figure 9: Residuals of the debt model ARIMA(0,1,5) with drift

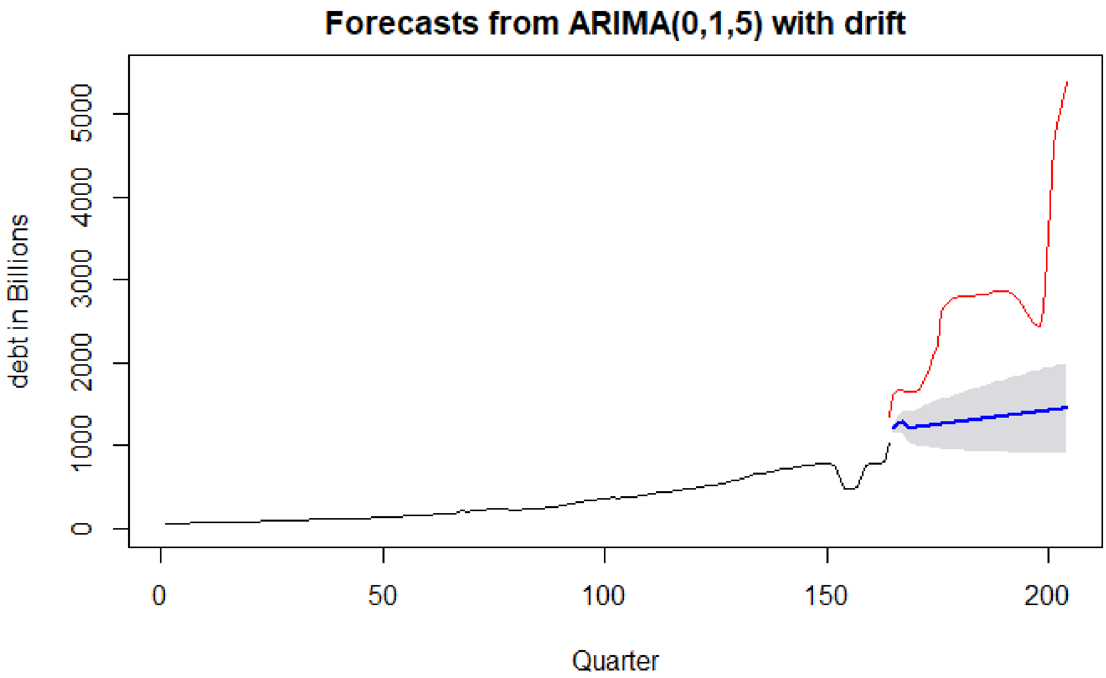

Figure 10: Forecasting Debt using ARIMA(0,1,5) with drift

The RSME for the debt prediction is 1510.12.

4.3.1.4 ARIMA Model Unemployment Rate

The unemployment rate is interesting as the ADF test proved it is a stationary time series. Differencing will therefore likely not be required.

However, for this time series, ARIMA(1,1,0) is used, which is minimizes AIC to 33.39, much lower compared to the previous models. The autocorrelations are within threshold and non-significant, making it a good model. Forecasting the fitted model can be seen in Figure 11. It has a wide confidence interval, making test data to be outside the threshold unlikely.

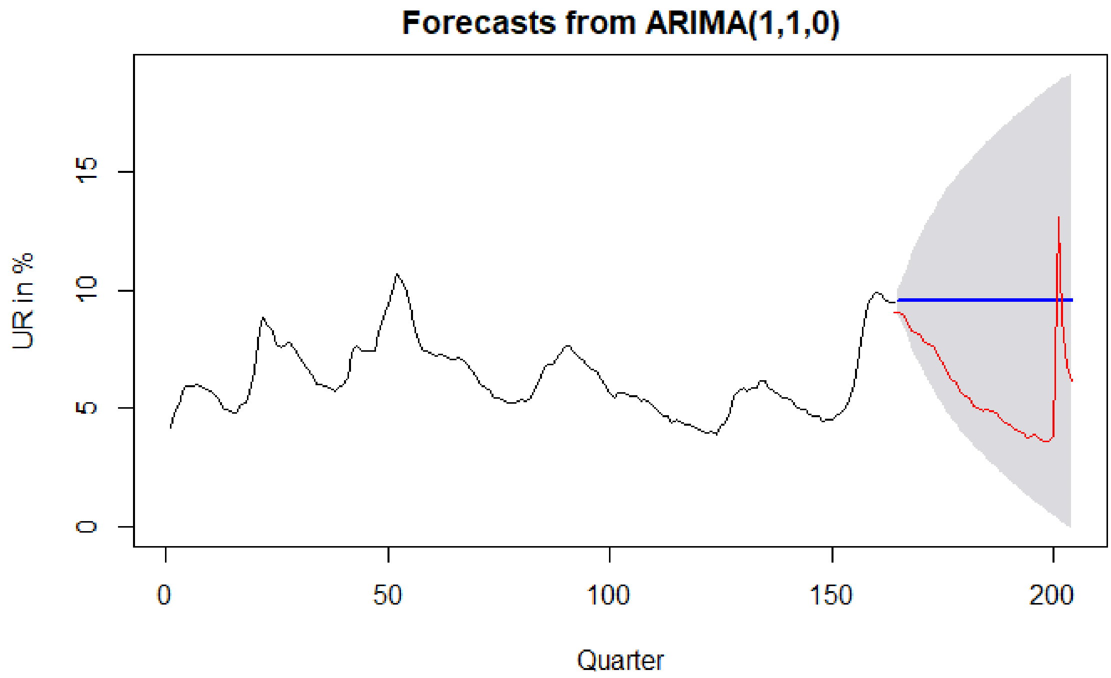

Figure 11: Forecasting Unemployment Rate using ARIMA(1,1,0)

The RSME for this prediction is 4.093.

4.3.1.5 ARIMA Model Natural Unemployment Rate

The results for the natural unemployment rate required manually adjusting the differencing. This is because the suggested ARIMA(2,2,2) model resulted in significant autocorrelations of the model and residuals. Setting $d = 1$, meant minimizing the AIC to -2152.088. Consequently, ARIMA(2,1,2) is used in this case. The no-longer significant autocorrelations can be seen in Figure 12. The residuals still experience a spike near the end of the series, indicating heteroscedasticity.

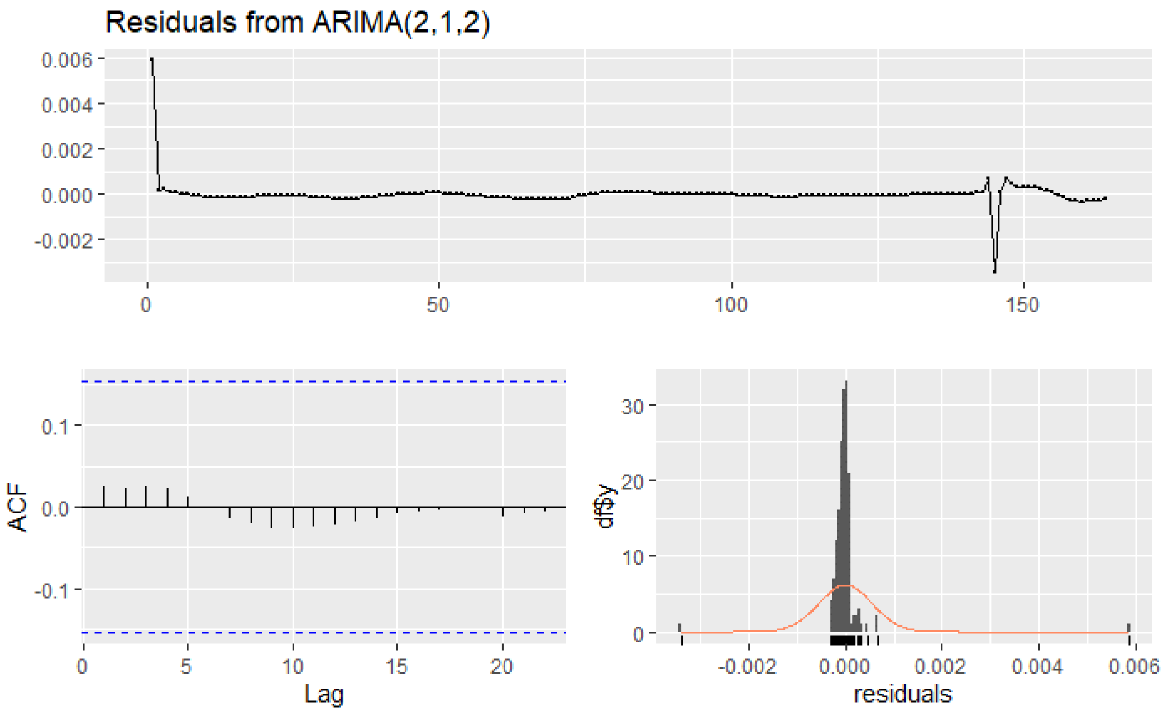

Figure 12: Residuals of the ARIMA(2,1,2) NUR model

Moreover, this meant an improvement of the forecasting model as well, which is shown in Figure 13. The red test line aligns almost perfectly with the forecast (blue). However, the confidence interval is still wide near the end of the 10-year forecast. The RSME for the natural unemployment rate of the ARIMA forecast is 0.0794.

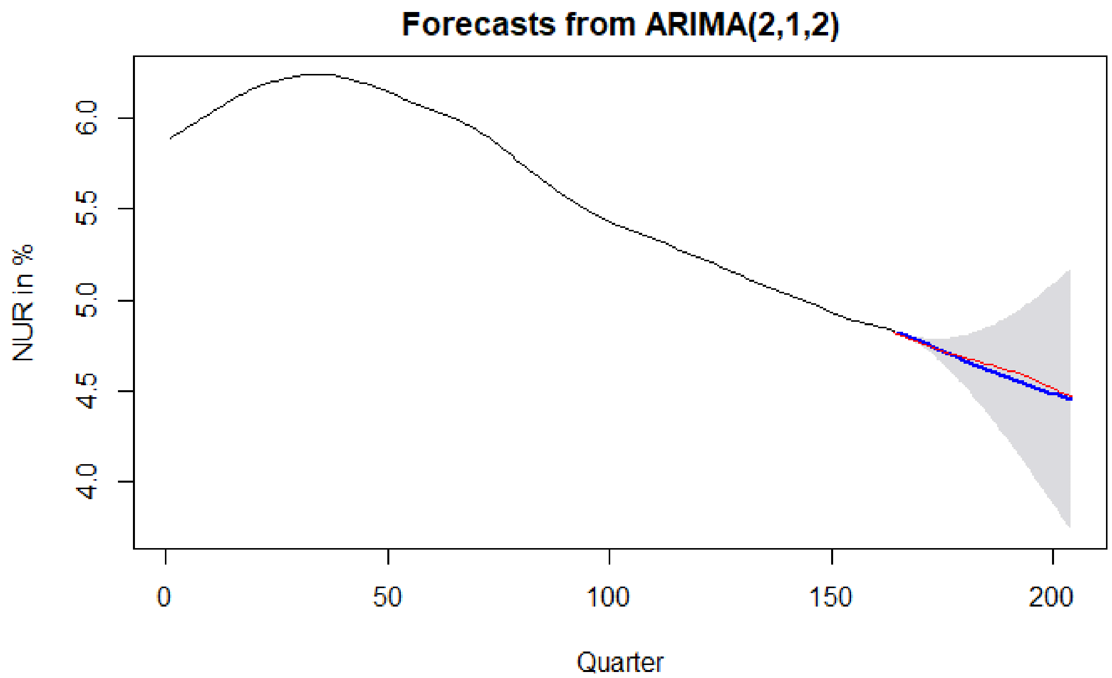

Figure 13: Forecasting NUR using ARIMA(2,1,2)

4.4 Vector Auto Regression Approach

Using a VAR model, different indicators can be included in the model as vectors, resulting in a model where the autoregressive approach of a variable is combined with that of others in the system. The methodology is inspired by Sims (1992). As the autocorrelation has already been analysed in section 4.2, a VAR model will be estimated and built. This will then be diagnosed. If this results in a useable model, three simulations will be tested: Granger Causality, Impulse Response and Forecast Error Variance Decomposition. Lastly, the VAR model will be used to forecast. Like in ARIMA it will be trained with data using the 80:20 split, to be able to compare the results of the models using the RSME.

4.4.1 Estimation

To estimate and build the model, the “vars” package is used (see reference code file). As the algorithm does not automatically difference data, an attempt should be made to estimate a model using differenced values. However, in this case, it results in worse model diagnostics than without differencing. Since the data is non-stationary, inference cannot be conducted.

The first step is to determine the amount of lags p used for the VAR model. Attention has to be raised to the likely problematic scenario, where a VAR model becomes too heavily parameterised. Choosing to include 5 variables, each lag results in a respective multiple of coefficients i.e., 5 variables times 8 lags result in 40 coefficients. Like in the ARIMA case, AIC can be used to determine appropriate lag length values. (Kotzé, n.d.)

Sims (1992) using a VAR model on US data economic data, decided for a lag value of 14, however, in this use case it would result in significant autocorrelations in the residuals. In fact, in this multivariate series, suboptimal autocorrelations in the residuals are seen in all tested lag values, and p tends to be maximized after 12. For the main model lag value $p = 4$ has been chosen, as it fulfilled the criterion for most common and lowest AIC found, as well resulting in other favourable diagnostic results seen in the next section 4.4.2.

4.4.2 Diagnostics

Autoregressive conditional heteroscedasticity (ARCH) is a standard tool used in structural VAR analysis to identify structural shocks, which is often complemented by the identification of the impact effect of the shocks. (Lütkepohl et al., 2020) We can see in Table 5 the H0 cannot be rejected, indicating a stable model, which is one of the reasons why the lag value $p = 4$ can be justified.

Table 5: ARCH result for VAR(4)

```
##  ARCH (multivariate)
##
## data:  Residuals of VAR object var1.model
## Chi-squared = 2220, df = 2700, p-value = 1
```

One can also test the structural breaks in residuals using ordinary least squares (OLS), which can be seen in Figure 14. Based on these results, there is no evidence to suggest structural breaks.

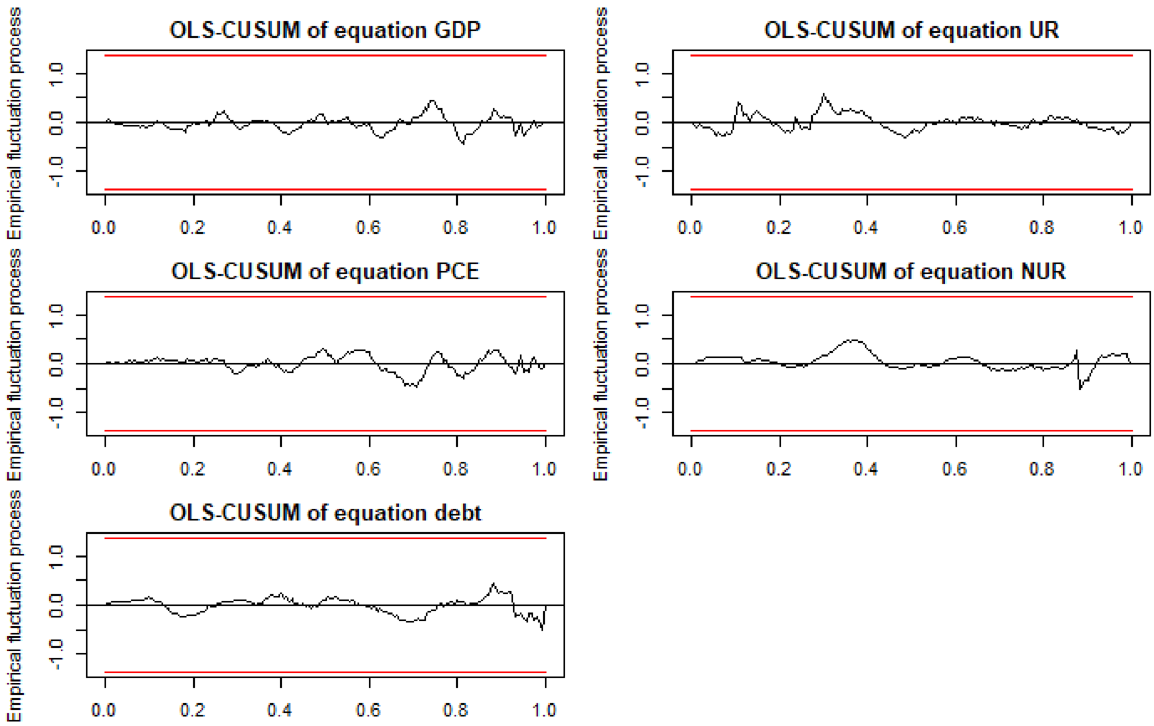

Figure 14: Structural Breaks of VAR(4) residuals

The last analysis used, is the Jarque-Bera test for assessing multivariate normality in terms of skewness and kurtosis. In this case, it is tested on the residuals of the VAR model, with a clear result of rejection to the null hypothesis, and therefore the rejection of a normal distribution in the model.

4.4.3 Simulation

Because the data was not differenced in the model estimation and since the normality of the residuals is compromised, one cannot make a definite inference about causation. Inability to provide identification for dynamic causal effects in vector autoregression models is a much-discussed topic in recent literature (Stock & Watson, 2018; Yan et al., 2021). Moreover, as with any econometric causality test, the Granger causality test is not able to test casual relationships, "because the possibility of a *post hoc ergo propter hoc* fallacy cannot be excluded" (Pfaff, 2008, p. 36). Keeping in mind, true causality may not be inferable, the results of the granger causality test may not be conclusive.

Summarizing the results of the Granger causality test the VAR model produces, considering a 5% confidence interval, the H0 is rejected for all variables and therefore they granger-cause the respective other variables.

An impulse response function can be used to shock the variables and see their behaviour according to the model. Focusing on shocking the debt variable, we get the results in Figure 15, showing 20 periods ahead.

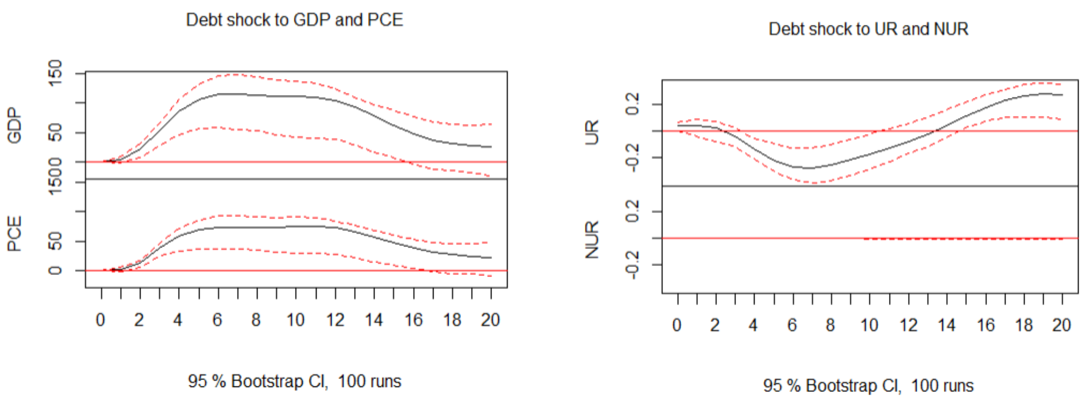

Figure 15: Debt impulse on GDP and PCE (left) and UR and NUR (right)

For an increase in debt, there is a strong increase in the GDP and PCE in the short term. In the long term, they increase moderately. Conversely, the unemployment rate decreases in the short term but then interestingly increases in the long term. The shock to natural unemployment is expectedly not significant.

Forecast error variance decomposition shows the amount of future uncertainty in one time series being a consequence of shocks in each variable in the system (Pfaff, 2008). Exemplifying this in Figure 16 with the FEVD for PCE, one can see in the decomposition, that shocks in the GDP time series alone amounts to more than 30% of the variability of PCE in the short term. This evolves and falls to about 20% in the long run, where debt explains more of the variability. Interestingly, it can also be seen that shocks from other variables in the system to the natural unemployment rate are not very important, especially in the short term.

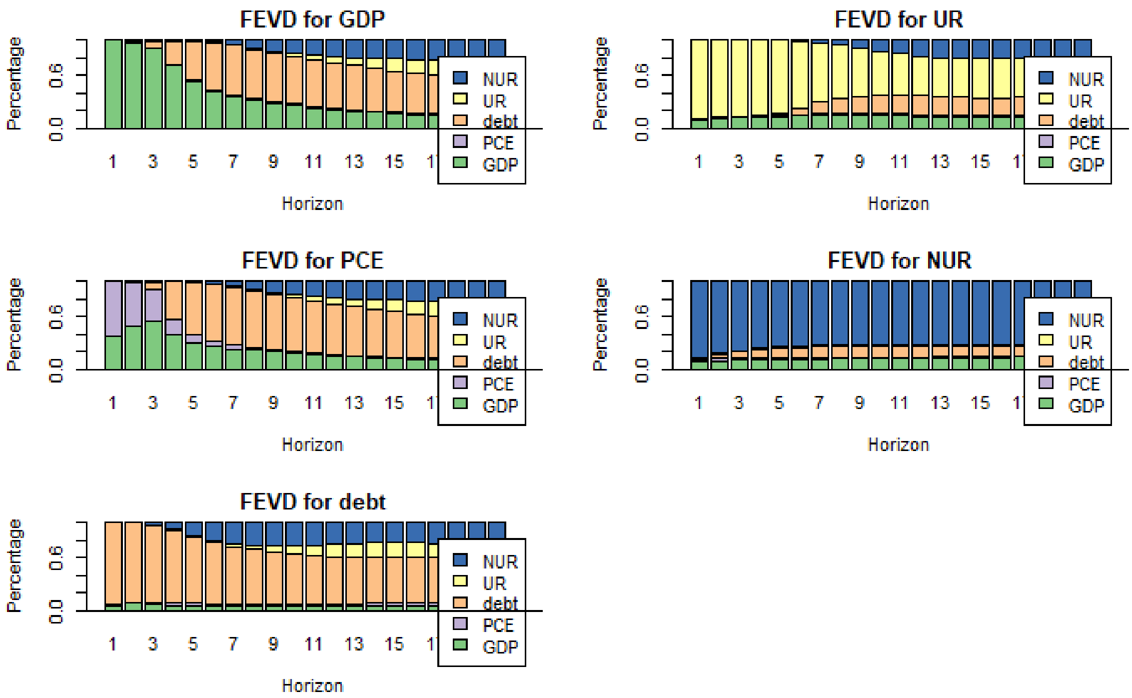

Figure 16: Forecast Error Variance Decomposition

4.4.4 Differencing for VAR

As the model was used to forecast, it presented suboptimal error scores, compared to ARIMA for example and it appeared that tweaking data might be necessary. As the time series are non-stationary, a logical step would be to difference the data for an autoregressive model. Differencing for VAR models is a much-discussed topic in literature (Ashley & Verbrugge, 2009), as it also has consequences on the inference of a model.

In this case, it resulted in similar problems faced by the original model, with suboptimal results in the autocorrelations of the residuals. Moreover, the error rate based on the prediction of the differenced data is worse than using just the training data. This can be seen in Table 6, where the RSME for the model using differenced data is only slightly better for the debt data and worse for the other variables.

Table 6: Comparison of VAR RSME using original and differenced data

	VAR	**VAR Diff**
GDP	2821.388	5693.563
PCE	2046.92	3995.495
Debt	1633.673	1181.603
UR	3.990817	2.6918
NUR	0.1793175	0.3780744

Considering the non-competitiveness of both models, a vector error correction model might be considered in such a case (Lütkepohl, 2004). Such a model tests for cointegration between variables and includes error correction.

To stay in scope for this thesis, the better, not differenced model will be considered for forecasting.

4.4.5 VAR Forecasting Result

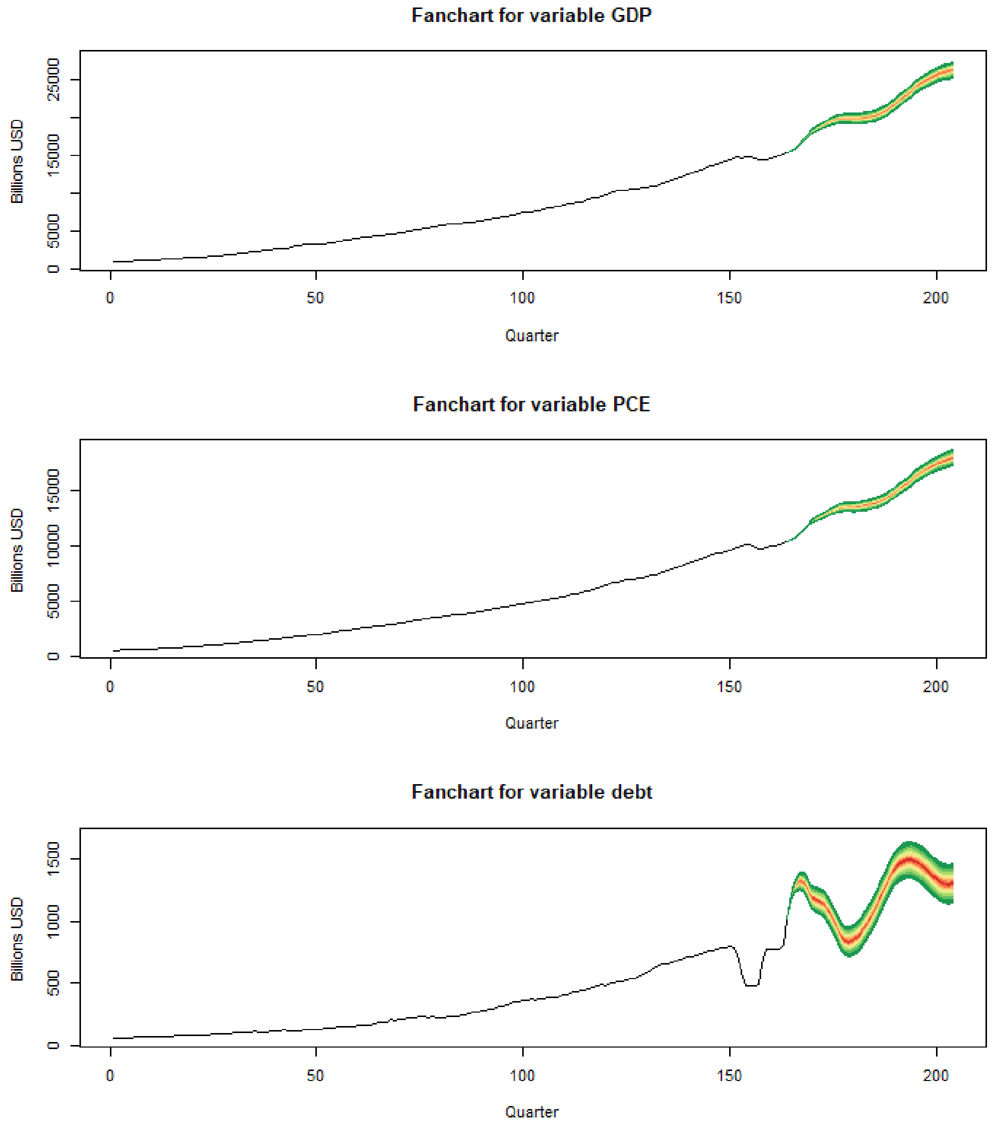

Figure 17: 10-Year Forecast using VAR(4) - GDP, PCE, debt

Here in Figure 17 GDP and PCE can be seen to follow a similar upward trend. Both variables can be seen as relatively stable, as the confidence interval is quite narrow, even near the 200-quarter mark. Interestingly, the forecast assumes volatility of debt held within the federal bank. This is interesting because the federal bank could in theory buy and sell securities to increase or reduce held debt in the short term. On the other hand, it is unlikely, as most of the securities currently held have maturities of five years or less (Board of the Federal Reserve, 2021, p. 4), and the FED is much more likely to simply not reinvest in those maturing assets and taper in a conservative approach (Cox, 2021). This makes volatility in the series improbable. However, this is to be expected, as Simionescu (2013), points out that for quarterly U.S. Treasury data a VARMA model has a higher degree of accuracy.

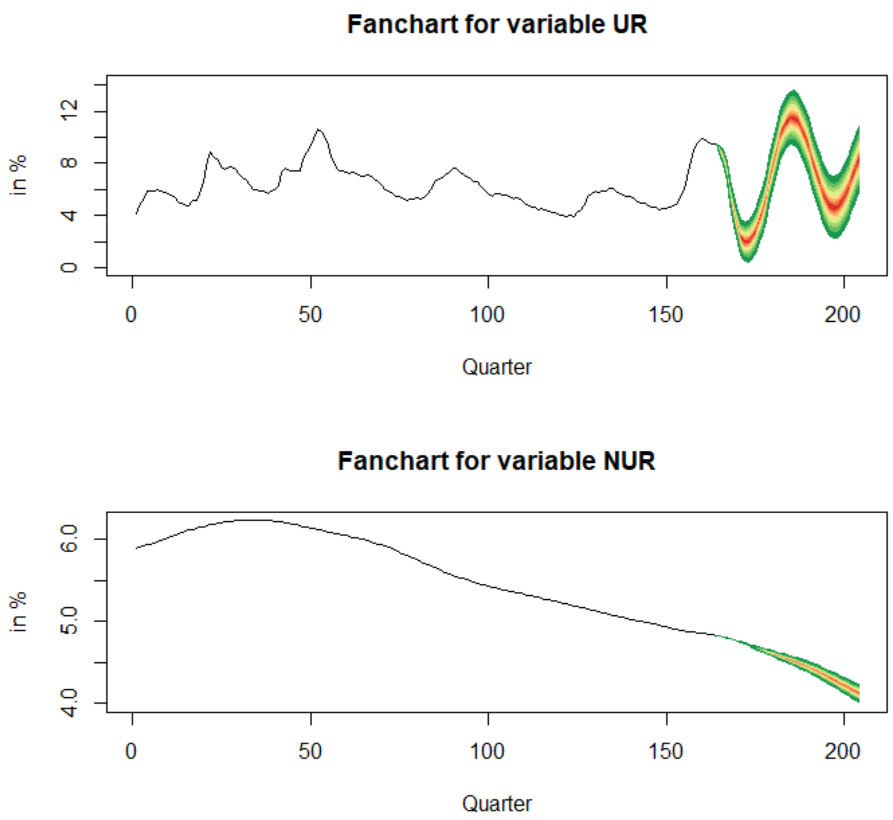

Figure 18: 10-Year Forecast using VAR(4) - UR, NUR

Forecasting the unemployment rate time series with this model also presents itself as problematic, as it is nearly unrealistic that it trends towards non-existence, which can be seen in Figure 18. Even in the upper 5% confidence interval case, it is an unrealistic future scenario where the unemployment rate trends to 0% and then quickly rise towards 10%. The RSME value for all variables, including VAR, can be seen in Table 7: Summary of RSME values, in section 5.

4.5 Long Short-Term Memory Approach

Given the diverse use cases of LSTM for time series data, a clear definition has to be provided. Because the time series are synced to the same collection points, they are parallel and because like the previously analyzed VAR model, an attempt will be made to forecast several quarters at a time, the forecast is going to be multiple steps at a time. Brownlee (2020) describes this as the "Multiple Parallel Input and Multi-Step Output" LSTM model.

As opposed to a regression approach, where the length of the input and predicted data are flexible, a recurrent neural network sequence approach is more challenging. A possible way to achieve this task is employing the LSTM neural network in form of an encoder-decoder model, formally used for language processing. Cho et al. (2014) provide an excellent definition:

> … RNN Encoder-Decoder, consists of two recurrent neural networks (RNN) that act as an encoder and a decoder pair. The encoder maps a variable-length source sequence to a fixed-length vector, and the decoder maps the vector representation back to a variable-length target sequence. (p. 1)

It is suitable, as variable input sequences are used to predict variable output sequences making it a sequence-to-sequence prediction problem (Brownlee, 2017). The result of an LSTM prediction varies as the algorithm and evaluation procedure have a stochastic nature, so each time a model is built and used to predict, the values will change (Brownlee, 2018).

4.5.1 Modelbuilding

The methodological approach has been inspired by Brownlee (2017, 2020) and Jagadeesh (2020), where they provide a widely used approach to multivariate forecasting with LSTM. The main part of the model building is taken from Jagadeesh's approach for an encoder-decoder model, but data preparation, hyperparameters, diagnostics and presentation of results have been changed to accommodate this use case of economic data. The data is, like for the previous models, split into training and test data.

As opposed to the previously used methods, the data necessarily must be scaled before it can be used in the LSTM model and afterwards reverse scaled to present results.

Two concrete models have been tested. The first employs one encoding and one decoding layer. The second model employs two respectively. In the case of two encoding and decoding layers, the output of the first layer is used as input for the second layer. A configuration with multiple "hidden" layers is also called Stacked LSTM (Brownlee, 2018).

Iterative improvements have been made, not only to hyperparameters of the model but also to the training data size, as it influences the amount of validation errors. The specific parameters that were tried and changed in many configurations for the model, were the unit size for each layer, the epoch amount (how many times the learning algorithm goes through the entire training set), and the batch size. The benchmark was the mean absolute error of the models, the amount of under and overfitting, and the RSME of the predictions. The improvements were mainly focused on the stacked model, as it got better results than the single-layer model, however, they are very similar. The mean absolute error and RMSE are calculated for every variable separately.

The final configuration of hyperparameters for the LSTM is as follows: learning rate = 0.0005, 100 cells (units) per layer, 50 epochs with a batch size of 16. The model uses for the training data a rolling window approach where always 2 quarters are used to test the next.

4.5.2 Error Checking

To analyse, if a model is suitable for the data, train and validation errors can be checked. This also enables a comparison between different encoder-decoder models and the ability to iteratively improve the model until underfitting/overfitting can be ruled out. Moreover, RSME is calculated.

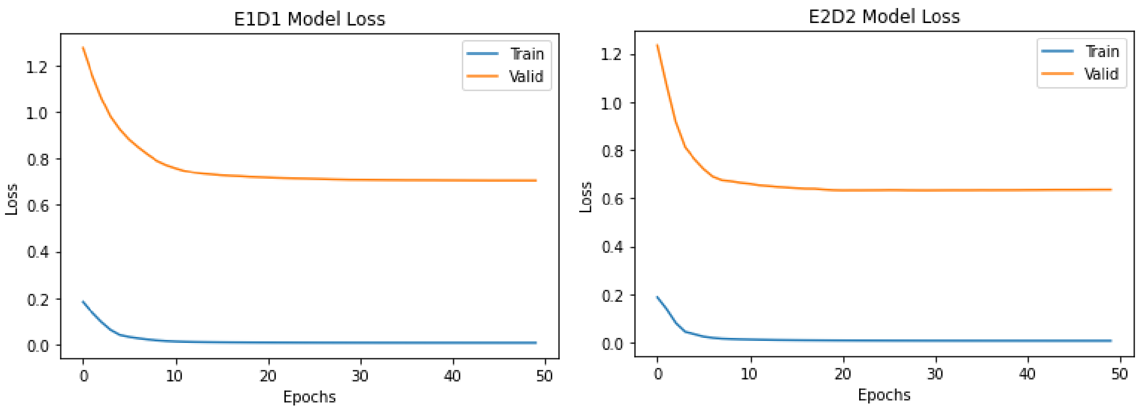

Figure 19: LSTM Encoder-Decoder Loss

The graphs in Figure 19 can be used to determine if the model is under or overfitting. In this case, underfitting can be seen, as an optimal model would be indicated, by the validation line to converge to the training line. This could be solved, theoretically, by including more layers in the model or including more training data. Then, the layers could act as a hyperparameter for under and overfitting as suggested by Jagadeesh (2020). In both model instances, learning levels off at about 10 epochs, so more layers might not be beneficial.

The LSTM Encoder-Decoder model showed high sensitivity to training data size. This is expected as a study (Brownlee, 2019a) showed a significant impact of training and test size on deep learning models. However, more training data was not included, as firstly, comparability between the models had to be ensured and secondly, enough test data needs available to reasonably approximate the performance of the model.

4.5.3 LSTM Result

The model chosen to predict is the stacked two encoder two decoder model. The non-repeatability for the exact values must be mentioned again. As the model uses stochastic procedures the values will change each time the model is built and used, and therefore the RSME value can differ significantly from one model run to the next.

The result of the best run can be seen in Figure 20. Here the model is trained with the data represented in the blue line. The green line is the predicted data, and the orange is the test data. An RSME score for all variables can be seen above their graph.

The predicted values are similar for GDP and PCE. The model expectedly did not predict the rapid increase in debt. The prediction for the natural unemployment rate is accurate, however, a spike at the end can be seen.

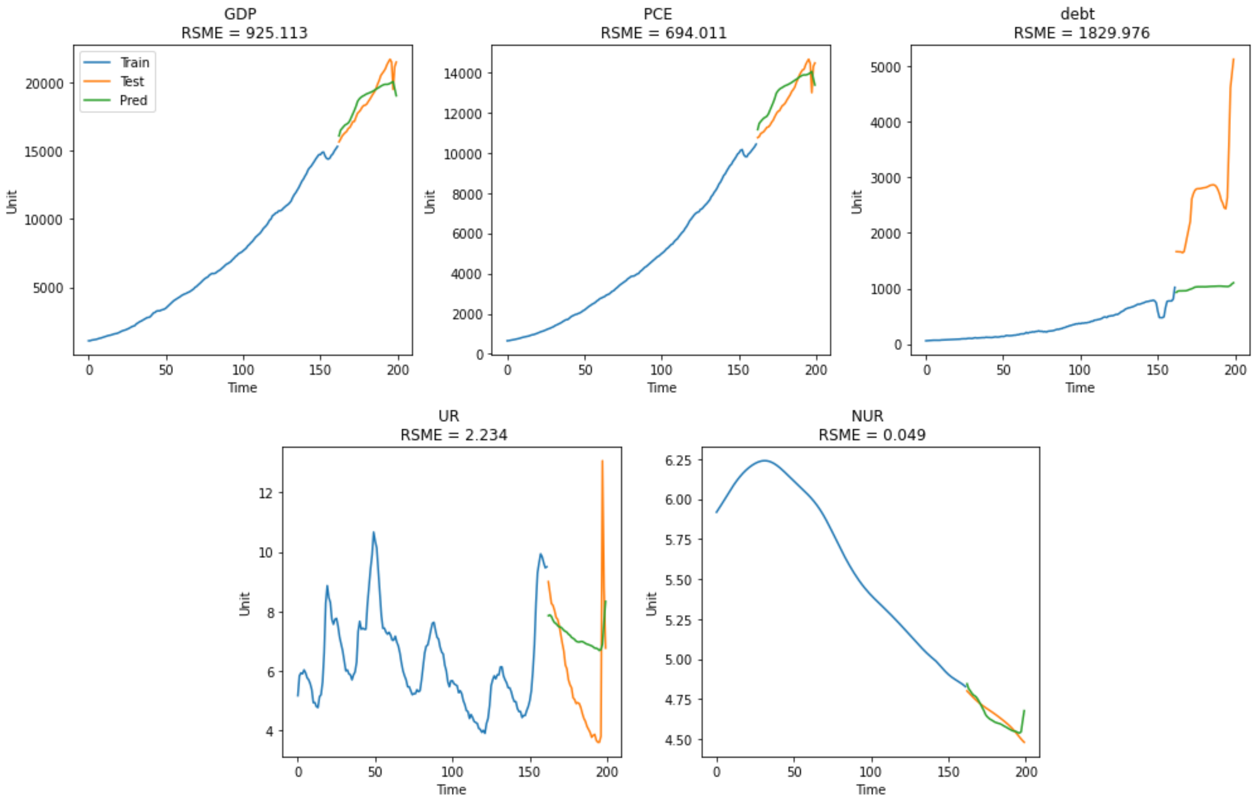

Figure 20: LSTM Summary of Forecasting Results

5 Summary

The data for all models were split into training and test data. For this, the first 164 observations (80%) were used for training and the remaining 20% were used for testing the predictions.

All employed ARIMA models result in non-significant autocorrelations but they all experience heteroscedasticity in their residuals.

Importantly, homoscedastic errors are an assumption for basic least squares models. Variances in the error terms can lead to falsely narrow standard errors and confidence intervals. A solution for this problem would be to employ ARCH or GARCH models, where homoscedasticity of the errors is the focus. (Engle, 2001)

The VAR model could theoretically result in better predictions if the data is differenced. However, as mentioned in this case compromises favourable diagnostic results and leads to similar or worse RSME values. Nevertheless, the model is suboptimal as especially the debt and unemployment rate predictions are erroneous. The VAR model also does not allow to infer causation, as the data is not differenced, and the normality of the residuals is compromised.

Although the stacked LSTM Encoder-Decoder multistep output model seemed to underfit, it provided formidable forecasting results. The validation loss could be improved by employing more layers or providing more training data. A disadvantage is that the model uses stochastic procedures in the machine learning algorithm and evaluation, the exact values are hardly repeatable.

To compare the models based on their RMSE, separate variables have to be considered, as the RSME values for all models are in absolute terms. Table 7 suggests the VAR model as the worst choice among them. This also holds, when only comparing the two autoregressive approaches, where VAR only provides a better result for the unemployment rate, compared to the ARIMA model.

Table 7: Summary of RSME values for all approaches

	ARIMA	**VAR**	**LSTM**
GDP	982.551	2821.388	925.113
PCE	584.9291	2046.92	694.011
Debt	1510.12	1633.673	1829.976
UR	4.093048	3.990817	2.534
NUR	0.07940361	0.1793175	0.049

The LSTM model results in a better prediction of GDP, UR, and NUR, according to these RSME scores. It indicates to be the best model. Comparing the multivariate methods confirms that, with the exception of the debt variable.

6 Conclusion

Using multivariate indicators, in the case of economic data, implies a discretionary choice about fitting related data in a coherent manner. In this case, 5 commonly analysed indicators of the U.S. have been chosen: the gross domestic product, the personal consumption expenditure, federal debt held by the federal reserve bank, the unemployment and natural unemployment rate.

The methods to analyse and forecast time series data are diverse, but three approaches have been utilized and compared based on their accuracy of prediction: a univariate approach in the form of five different ARIMA models, and two multivariate approaches. VAR to enable comparison between autoregressive methods and LSTM to employ a machine learning approach. LSTM has been used in the form of an encoder-decoder model, to enable multivariate input and output of the data. For each approach the 5 indicators were split into training and test data, to enable model evaluation and comparability.

The ARIMA models achieved the lowest RSME for 2 variables PCE and debt. All univariate models, however, indicated variance in their errors, implying the need to adopt better-suited methods for such cases, such as ARCH/GARCH.

VAR got consistently inferior prediction results, not only quantitatively but also qualitatively, with improbable predictions in the debt and unemployment rate time series. Moreover, because a normal distribution of residuals was not given, causal inference cannot be made with the model.

In conclusion, the LSTM model achieved the best error scores among the methods, by forecasting the GDP and both unemployment time series more accurately. The model indicated underfitting and could be improved by including more encoding-decoding layers in the model.

7 Further Research

In the univariate case, improving the ARIMA models could be achieved by utilizing ARCH or GARCH models or using a filter suggested by Stockhammar and Öller (2012).

Exploring the possibility of a vector error correction model (VECM), instead of a VAR could result in better predictions for the autoregressive multivariate approach.

To expand upon the multivariate approach, stacking more LSTM layers of an encoder-decoder model, could result in less validation error, however other hyperparameters would likely have to be adjusted. Given the favourable outcome of a univariate model in forecasting some of the variables, using LSTM for univariate cases, as suggested by Brownlee (2018), could result in improvements of the machine learning approach.

To benchmark different machine learning approaches, another commonly used time series forecasting method is the Prophet algorithm, which could be evaluated against the LSTM approach (di Pietro, 2021).

References

Ashley, R. A., & Verbrugge, R. J. (2009). To difference or not to difference: a Monte Carlo investigation of inference in vector autoregression models. *International Journal of Data Analysis Techniques and Strategies*, *1*(3), 242. https://doi.org/10.1504/ijdats.2009.024295

Board of Governors of the Federal Reserve System. (2020, March 23). *Federal Reserve issues FOMC statement* [Press release]. https://www.federalreserve.gov/newsevents/pressreleases/monetary20200323a.htm

Board of Governors of the Federal Reserve System. (2021, August). *Factors Affecting Reserve Balances of Depository Institutions and Condition Statement of Federal Reserve Banks*. Federal Reserve statistical release. https://www.federalreserve.gov/releases/h41/current/h41.pdf

Brownlee, J. (2017, August 23). *Encoder-Decoder Long Short-Term Memory Networks*. Machine Learning Mastery. https://machinelearningmastery.com/encoder-decoder-long-short-term-memory-networks/

Brownlee, J. (2018, November 14). *How to Develop LSTM Models for Time Series Forecasting*. Machine Learning Mastery. https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/

Brownlee, J. (2019a, January 2). *Impact of Dataset Size on Deep Learning Model Skill And Performance Estimates*. Machine Learning Mastery. https://machinelearningmastery.com/impact-of-dataset-size-on-deep-learning-model-skill-and-performance-estimates/

Brownlee, J. (2019b, August 5). *Comparing Classical and Machine Learning Algorithms for Time Series Forecasting*. Machine Learning Mastery.

https://machinelearningmastery.com/findings-comparing-classical-and-machine-learning-methods-for-time-series-forecasting/

Brownlee, J. (2020, October 20). *Multivariate Time Series Forecasting with LSTMs in Keras*. Machine Learning Mastery. https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–1734. https://doi.org/10.3115/v1/d14-1179

Cox, J. (2021, August 18). Markets already have started adjusting to the Fed's expected policy tightening. *CNBC*. https://www.cnbc.com/2021/08/18/markets-are-adjusting-to-the-feds-expected-policy-tightening.html

Cryer, J. D., & Chan, K. (2008). *Time Series Analysis: With Applications in R (Springer Texts in Statistics)* (2nd ed.). Springer.

Date, S. (2021, June 20). *Introduction to Regression With ARIMA Errors Model*. Time Series Analysis, Regression and Forecasting. https://timeseriesreasoning.com/contents/regression-with-arima-errors-model/

di Pietro, M. (2021, August 28). *Time Series Forecasting: ARIMA vs LSTM vs PROPHET - Analytics Vidhya*. Medium. https://medium.com/analytics-vidhya/time-series-forecasting-arima-vs-lstm-vs-prophet-62241c203a3b

Engle, R. (2001). GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives*, *15*(4), 157–168. https://doi.org/10.1257/jep.15.4.157

Federal Reserve Bank of St. Louis. (n.d.). *Legal Notices, Information and Disclaimers*. FRED. Retrieved 24 August 2021, from https://fred.stlouisfed.org/legal/

Federal Reserve Bank of St. Louis. (2021, February 1). *Noncyclical Rate of Unemployment*. FRED. https://fred.stlouisfed.org/series/NROU

Fox, D. R., McCulla, S. H., Smith, S., & Seskin, E. P. (2020, March). *NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts*. U.S. Department of Commerce, Bureau of Economic Analysis. https://www.bea.gov/resources/methodologies/nipa-handbook/pdf/all-chapters.pdf

Hall, A. S. (2018). Machine Learning Approaches to Macroeconomic Forecasting. *Economic Review, The Federal Reserve Bank of Kansas City*, *Q IV*, 63–81. https://doi.org/10.18651/er/4q18smalterhall

Hyndman, R. J., & Athanasopoulos, G. (2018, April). *Forecasting: Principles and Practice (2nd ed)*. Otext. https://otexts.com/fpp2/

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Jagadeesh, S. (2020, December 14). *Multivariate Time Series Forecasting with LSTMs in Keras*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/10/multivariate-multi-step-time-series-forecasting-using-stacked-lstm-sequence-to-sequence-autoencoder-in-tensorflow-2-0-keras/

Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, *72*(6), 558–569. https://doi.org/10.4097/kja.19087

Koop, G., & Korobilis, D. (2010). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. *Foundations and Trends® in Econometrics*, *3*(4), 267–358. https://doi.org/10.1561/0800000013

Kotzé, B. K. (n.d.). *Vector autoregression models*. Kevinkotze. Retrieved 10 September 2021, from https://kevinkotze.github.io/ts-7-var/

Lütkepohl, H. (2004). Vector Autoregressive and Vector Error Correction Models. *Applied Time Series Econometrics*, 86–158. https://doi.org/10.1017/cbo9780511606885.004

Lütkepohl, H., Meitz, M., Netšunajev, A., & Saikkonen, P. (2020). Testing identification via heteroskedasticity in structural vector autoregressive models. *The Econometrics Journal*, *24*(1), 1–22. https://doi.org/10.1093/ectj/utaa008

Nobelprize.org. (1976, October 14). *The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1976* [Press release]. https://www.nobelprize.org/prizes/economic-sciences/1976/press-release/

Olah, C. (2015, August 27). *Understanding LSTM Networks*. Colah’s Blog. https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time Series with R* (2nd ed.). Springer Publishing. https://doi.org/10.1007/978-0-387-75967-8_2

Prabhakaran, S. (2019, November 2). *Augmented Dickey-Fuller (ADF) Test*. Machine Learning Plus. https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/

Prins, J., Filliben, J. J., Croarkin, C. M., Hembree, B., Guthrie, W. F., Tobias, P., & Heckert, N. A. (2021). *e-Handbook of Statistical Methods* [E-book]. NIST/SEMATECH. https://doi.org/10.18434/M32189

Simionescu, M. (2013). The Use of Varma Models in Forecasting Macroeconomic Indicators. *ECONOMICS & SOCIOLOGY*, *6*(2), 94–102. https://doi.org/10.14254/2071-789x.2013/6-2/9

Sims, C. A. (1992). Interpreting the macroeconomic time series facts. *European Economic Review*, *36*(5), 975–1000. https://doi.org/10.1016/0014-2921(92)90041-t

Stock, J., & Watson, M. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *NBER* | *National Bureau of Economic Research*. Published. https://doi.org/10.3386/w24216

Stockhammar, P., & Öller, L.-E. (2012). A Simple Heteroscedasticity Removing Filter. *Communications in Statistics - Theory and Methods*, *41*(2), 281–299. https://doi.org/10.1080/03610926.2010.521289

Tsay, R. S. (2005). *Analysis of Financial Time Series (Wiley Series in Probability and Statistics)* (2nd ed.). Wiley-Interscience.

Tzvetkova, S., & Hepburn, C. (2018, March 26). *The missing economic measure: Wealth*. Our World in Data. https://ourworldindata.org/the-missing-economic-measure-wealth

United States Department of the Treasury. (2010, October 13). *Web-Archive: Bureau of the Public Debt*. Bureau of the Public Debt. https://web.archive.org/web/20101013020929/http://www.publicdebt.treas.gov/
Web-Archived from original: http://www.publicdebt.treas.gov/

U.S. Bureau of Economic Analysis. (2020, August 18). *Gross Domestic Product [GDP]*. BEA. https://www.bea.gov/resources/learning-center/what-to-know-gdp

U.S. Bureau of Economic Analysis. (2021a, April 29). *Gross Domestic Product (Advance Estimate) and Annual Update*. BEA. https://www.bea.gov/data/gdp/gross-domestic-product

U.S. Bureau of Economic Analysis. (2021b, July 26). *Gross Domestic Product [GDP]* [Dataset]. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/GDP
retrieved July 29, 2021

U.S. Bureau of Economic Analysis. (2021c, July 26). *Personal Consumption Expenditures [PCEC]* [Dataset]. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/PCEC
retrieved July 29, 2021

U.S. Bureau of Labor Statistics. (2021). *Unemployment Rate [UNRATE]* [Dataset]. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/UNRATE
retrieved July 29, 2021

U.S. Congressional Budget Office. (2021, February 1). *Noncyclical Rate of Unemployment [NROU]* [Dataset]. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/NROU retrieved July 29, 2021

U.S. Department of the Treasury. (2021a, March). *Profile of the Economy Financial Operations International Statistics Special Report*. Bureau of the Fiscal Service, Cash Accounting and Reporting Division. https://www.fiscal.treasury.gov/files/reports-statements/treasury-bulletin/2021-1.pdf

U.S. Department of the Treasury. (2021b, August 6). *Debt to the Penny*. U.S. Treasury Fiscal Data. https://fiscaldata.treasury.gov/datasets/debt-to-the-penny/debt-to-the-penny

U.S. Department of the Treasury. Fiscal Service. (2021). *Federal Debt Held by Federal Reserve Banks [FDHBFRBN]* [Dataset]. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/FDHBFRBN retrieved July 29, 2021

Yan, Y., Huang, H. C., & Genton, M. G. (2021). Vector Autoregressive Models with Spatially Structured Coefficients for Time Series on a Spatial Grid. *Journal of Agricultural, Biological and Environmental Statistics*, *26*(3), 387–408. https://doi.org/10.1007/s13253-021-00444-4

Yürekli, K., Kurunc, A., & Ozturk, F. (2005). Testing the Residuals of an ARIMA Model on the C ekerek Stream Watershed in Turkey. *Turkish Journal of Engineering and Environmental Sciences*, *29*(2), 61–74. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.583.2054&rep=rep1&type=pdf