MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"A Comparison of the Forecasting Performance of Vector Autoregression and Neural Networks"

verfasst von / submitted by

Benjamin Drucks, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 913
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Volkswirtschaftslehre
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Robert Kunst

Abstract

Using quarterly time series data on GDP growth, inflation, interest rate and unemployment from 24 OECD countries, I want to compare the forecasting performance of vector autoregressions (VAR) and artificial neural networks (ANN). The dataset is split into an estimation/training and a testing part and for each country a VAR and an ANN is estimated. The estimated models are then employed to forecast the testing period and their performance is evaluated by calculating the root mean squared errors (RMSE). The model type with the smaller RMSE on average yields a superior forecasting performance for the data set at hand.

German Abstract

Ziel der Arbeit ist es unter Verwendung der Zeitreihen von BIP Wachstum, Inflation, Zinssatz und Arbeitslosigkeit für 24 OECD Mitgliedsstaaten die Prognoseleistung von Vektor Autoregression (VAR) und künstlichen Neuralen Netzen (ANN) zu vergleichen. Der Datensatz wird zweigeteilt, in einen Teil zur Schätzung/Training der Modelle und einen Test-Teil um ihre Leistung zu überprüfen. Für jedes Land wird eine VAR und ein ANN berechnet und ausgewählt. Mit Hilfe der geschätzten Modelle wird eine Prognose für den Test- Zeitraum erstellt und die Leistung anhand der Wurzel des mittleren quadratischen Fehlers (RMSE) bewertet. Jenes Modell mit dem durchschnittlich kleineren RMSE hat eine überlegene Prognoseleistung für den vorliegenden Datensatz.

Contents

Chapter 1

Introduction

"It's hard to make predictions, especially about the future." is a quote attributed to Niels Bohr. Even in a world of ever increasing amounts of data available for scientific research and hence for predictions about the future this saying still holds true. It is important to mention though, that the tools developed in the last half century for the purpose of forecasting economic variables provide a solid basis to make predictions about future values. The most important economic variable it seems is still the Gross Domestic Product (GDP) and therefore forecasting future GDP values appears to be a good benchmark to compare different methods of economic forecasting which is the aim of this master thesis.

Ever since a seminal paper by Christopher Sims in 1980 [24] a strong view in the literature has evolved that vector autoregression (VAR) yields a superior forecasting performance compared to economically funded macroeconometric forecasting models. Economically funded in this sense means that the macroeconometric models try to mimic a simplified version of the actual economy whereas vector autoregression is a purely statistical model (albeit the dimension and selection of variables relies also on economic theory). Neural networks (NN) on the other hand gained considerable attention over the last decade especially for their power regarding classification problems (e.g. speech and visual recognition). Therefore it is the aim of this thesis to compare the forecasting performance of a "standard" VAR and a neural network.

Regarding the structure of the paper, a brief discussion of the data used is followed by a review of the existing literature. In the subsequent methods section the time series are checked for their properties and then the VARs and neural networks are estimated. The next chapter yields the results and their robustness is checked and at last the thesis is completed with a conclusion. Some test statistics and the graphs of all forecasts can be found in the appendix.

Chapter 2

Data

2.1 Overview

The data used in this analysis was taken from OECD and contains quarterly time series data for 24 out of 36 member states from Q1 2000 to Q3 2018 for the following variables: GDP growth [17], inflation [18], short term interest rate [19] and unemployment [20]. The data is already deseasonalized hence the main reason for using quarterly data is to have longer time series to provide a large and heterogenous sample of countries. Illustration 1 shows the OECD countries analyzed in the thesis in green and those excluded (due to missing data) in red.

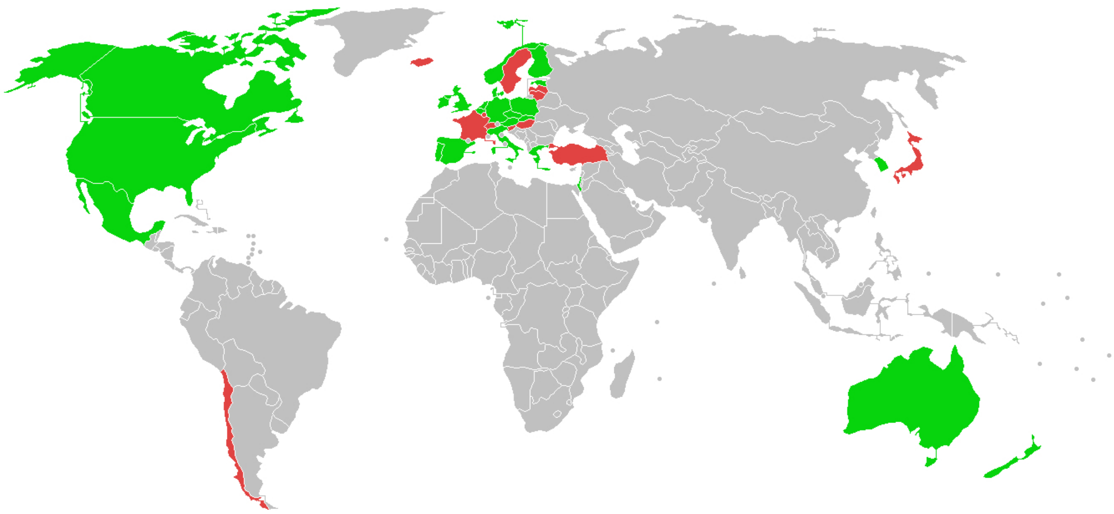


Figure 2.1: OECD members; green: used in the analysis, red: excluded [27]

Futhermore the time series are split into a training/estimation part (2000-2015, 64 observations) and a testing part (2016-2018, 11 obs), to be able to evaluate the forecasting performance of the models.

Chapter 3

Literature Review

The sheer amount of literature regarding the comparison of different forecasting methods is overwhelming. Therefore this section can only provide a rough overview with the focus on certain papers that are most related to this thesis. A direct comparison of the results obtained in this work with the findings presented below is possible only to a limited extent as especially the implementation techniques for the neural networks differ substantially.

In their paper Moshiri and Cameron [15] compare the forecasting performance of neural networks and, among others, vector autoregression to predict the inflation rate. They find that their implementation of a neural network outperforms a VAR only for the shortest of their forecast horizons (3 steps) and that for their longest forecast (12 steps) the VAR is superior to the neural network.

Zhang and Kline [28] estimate an ensemble of different neural networks to test their forecasting performance on several quarterly time series. The most important finding of their research are that NNs perform best on macroeconomic data, that patterns such as trend and seasonality should be removed to improve the forecasting performance of NNs (which is done in this thesis) and that simpler NNs work better than more complicated ones. They also note that their time series (64 obs) might be too short and could produce an overfit.

The aim of Aiken's [2] paper is to apply a neural network to forecast annual US GDP percentage changes. The data is separated into a training and testing part and he then compares the predictions made by the NN with expert consensus and naive estimates (among others the Three-Month-Rule-of-Thumb) and finds that the network provides superior forecasts.

The study by Swanson and White [26] compares the forecasting power of neural networks and vector autoregression for nine different macro economic variables. The authors find that for the majority of the variables under investigation the neural networks outperform the vector autoregression for both forecast horizons (one month and one year).

Chapter 4

Methods

4.1 Stationarity Testing

To determine whether the time series used are stationary the Augmented Dickey-Fuller Test (ADF) was applied to all 96 series. It tests the null hypothesis that a unit root is present in the sample versus the alternative that the time series is stationary. The results of the test are depicted in Table 4.1.

The ADF tests show that for the time series of the short-term interest rates and especially unemployment the null hypothesis can not be rejected at a significance level of 5% and therefore one could conclude that these time series are not stationary. Since the null hypothesis is the presence of a unit root, it has been argued that the ADF test tends to miss stationarity even if the series is in fact stationary (when one cannot reject the null, but the p-value is not that large) and therefore the so-called Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test is applied to the data.
The main difference is that the KPSS test has stationarity as the null hypothesis. Utilizing the ndiffs function (from the forecast package [9] in R [23]), only the suggested levels of integration (at 5% significance level) are presented in Table 3.2.

The ADF test results and the suggested level of integration by the KPSS test lead to the same conclusion: The unemployment time series and to a lesser extent those of the short-term interest rate show signs of non-stationarity. This can lead to problems when estimating a VAR since one of the assumptions is stationarity of the time series involved. The next stept is to check for cointegration. First the Phillips-Ouliaris test is applied which tests whether two or more time series are cointegrated. The null hypothesis of the test is that there is no cointegration. The p-values of the test for each country are printed in Table 3.3.

For 21 out of 24 countries the null hypothesis can be rejected, which leads to the conclusion that cointegration is prevalent. One way to deal with cointegration is to apply the so-called Johansen procedure (named after Søren Johansen [10]) which tests whether the rank of the impact matrix is zero or of higher order. The results are also shown in Table 3.3 and the ranks of the impact matrix are between one and three.

ADF Test Results

Country Code	GDP Growth		Inflation		ST Interest Rate		Unemployment	
	T-stat	p-value	T-stat	p-value	T-stat	p-value	T-stat	p-value
AUS	-5.3533	0.01	-5.4707	0.01	-2.0171	0.568	-1.5606	0.753
AUT	-3.8208	0.0233	-4.0786	0.0119	-2.708	0.2881	-2.8584	0.2272
BEL	-4.1074	0.0106	-4.345	0.01	-2.708	0.2881	-2.5386	0.3567
CAN	-3.551	0.0446	-5.3817	0.01	-3.6333	0.0376	-2.5746	0.3421
CZE	-2.6231	0.3225	-4.2221	0.01	-2.9637	0.1845	-2.9698	0.182
DEU	-3.6479	0.0364	-4.4686	0.01	-2.708	0.2881	-2.8369	0.2358
DNK	-3.309	0.0783	-2.9758	0.1796	-2.2718	0.4648	-2.5532	0.3508
ESP	-2.965	0.184	-3.7961	0.0244	-2.708	0.2881	-2.3269	0.4425
EST	-2.4522	0.3917	-3.8919	0.0201	-2.9763	0.1794	-3.0384	0.1542
FIN	-3.185	0.098	-3.9092	0.0194	-2.708	0.2881	-1.9052	0.6134
GBR	-3.3926	0.065	-1.3383	0.8431	-2.4741	0.3829	-1.2914	0.8621
GRC	-2.4121	0.408	-2.6151	0.3257	-2.8104	0.2466	-2.9832	0.1766
IRL	-1.7771	0.6653	-3.6055	0.04	-2.708	0.2881	-1.8336	0.6424
ISR	-3.9424	0.0179	-3.1229	0.12	-4.2674	0.01	-2.7474	0.2721
ITA	-3.9671	0.0168	-3.1175	0.1222	-2.708	0.2881	-1.8224	0.6469
KOR	-4.7848	0.01	-3.9351	0.0182	-3.3379	0.0737	-2.9153	0.2041
MEX	-3.8079	0.0239	-4.7493	0.01	-5.6995	0.01	-1.9348	0.6014
NLD	-3.202	0.0953	-3.6567	0.0356	-2.708	0.2881	-2.8175	0.2437
NOR	-4.5336	0.01	-4.6859	0.01	-2.5226	0.3632	-2.2931	0.4562
NZL	-3.1189	0.1216	-3.5688	0.0431	-2.1564	0.5116	-1.9922	0.5781
POL	-2.7034	0.2899	-3.9649	0.0169	-6.0231	0.01	-3.1864	0.0978
PRT	-3.3389	0.0735	-4.4846	0.01	-2.708	0.2881	-2.7436	0.2737
SVK	-2.6768	0.3007	-3.9288	0.0185	-2.8671	0.2236	-2.1586	0.5107
USA	-3.1578	0.1058	-3.9308	0.0184	-3.2239	0.0918	-2.2917	0.4568

Table 4.1: ADF Test Results

KPSS Test Results

Country	QGDP	CPI	STINT	UNEMP
AUS	0	1	1	1
AUT	0	0	1	1
BEL	0	0	1	1
CAN	0	1	1	0
CZE	0	1	1	1
DEU	0	0	1	2
DNK	0	1	1	1
ESP	1	1	1	1
EST	1	1	1	0
FIN	1	1	1	0
GBR	0	1	1	1
GRC	1	1	1	2
IRL	0	1	1	2
ISR	0	0	1	1
ITA	0	1	1	2
KOR	0	1	1	1
MEX	0	1	1	1
NLD	0	1	1	1
NOR	0	0	1	0
NZL	0	1	1	1
POL	0	1	2	1
PRT	0	1	1	1
SVK	0	1	1	1
USA	0	1	1	1

Table 4.2: KPSS Test; Level of Integration

PO Test Results, Johansen Procedure

Country	t-stat PO	p-value PO	t-stat JO at r	rank JO
AUS	-64.236	< 0.01	4.65	2
AUT	-36.769	0.02	9.05	2
BEL	-28.723	0.09	17.38	1
CAN	-31.465	0.06	12.49	2
CZE	-19.405	> 0.15	13.03	1
DEU	-38.864	0.02	14.05	1
DNK	-56.756	< 0.01	6.51	2
ESP	-9.264	> 0.15	14.86	1
EST	-52.543	< 0.01	9.85	2
FIN	-52.771	< 0.01	9.84	2
GBR	-24.485	> 0.15	5.28	2
GRC	-57.634	< 0.01	14.61	1
IRL	-75.543	< 0.01	8.50	2
ISR	-34.195	0.04	0.40	3
ITA	-25.247	0.15	13.74	1
KOR	-56.339	< 0.01	12.24	2
MEX	-39.982	0.01	11.13	2
NLD	-46.397	< 0.01	17.91	1
NOR	-85.847	< 0.01	16.21	1
NZL	-47.875	< 0.01	11.01	2
POL	-76.199	< 0.01	12.99	1
PRT	-51.235	< 0.01	18.43	1
SVK	-66.041	< 0.01	8.09	2
USA	-44.653	< 0.01	10.16	2

Table 4.3: Phillips-Ouliaris test; Johansen Procedure

In general, when time series are not stationary (therefore integrated) and cointegration is present, one should use a Vector Error Correction (VEC) model instead of a standard Vector Autoregression (VAR). If making out-of-sample forecasts is the only objective of a VAR model though, Hamilton [6] renders it possible to use a VAR in levels (without integrating non-stationary time series) instead of a VEC model which is also done in one of the papers in the literature review [26]. This approach was first described and performed in a fundamental paper by Engle and Yoo [4]. Both models are hereinafter estimated and applied to make a prediction for each country to compare their performance and assert whether a VECM yields improved forecasts. The one with a smaller RMSE on average will then compete with forecasts made by the Neural Networks.

4.2 Vector Autoregression

4.2.1 VAR

A generalized depiction of the models used in this section can be seen below. It is crucial to determine the number of lags. One way would be to use information criteria (in this case AIC seems the most appropriate) and to calculate the AIC for every possible lag length and choose the one with the lowest value. Since AIC is known to produce an overfit (more lags than necessary) and the goal of this research is to find the forecasting model that has the smallest RMSE (for the variable GDP Growth) over the forecast period another option is to estimate all models and pick the one with the smallest mean squared error.

$$\begin{bmatrix} y_t \\ i_t \\ r_t \\ u_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} + \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ i_{t-1} \\ r_{t-1} \\ u_{t-1} \end{bmatrix} + ... + \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ e_{3,t} \\ e_{4,t} \end{bmatrix}$$

For each of the 24 countries 22 different models were estimated (containing 1-11 lags and with (w/) or without (w/o) a constant).
As an example the calculated RMSEs for Belgium for all models can be seen in table 4.4 (the smallest value highlighted in green). In the case of Belgium the model with the smallest RMSE has three lags and no constant. Then the roots of the characteristic polynomial are calculated. For the VAR(3) with no constant for BEL all roots are smaller than one in the modulus and have the following absolute values:

0.9990	0.9338	0.8672	0.8672	0.5998	0.5208
0.5208	0.4992	0.4992	0.4177	0.3448	0.3448

The fact that all roots are smaller than one in the modulus for the VAR, which is true for all countries, is not sufficient though, to conclude that the VAR(3) is stable and hence stationary. Therefore (and given the results in section 4.1) a VEC model is estimated for each country in the following section. The same procedure is repeated for the remaining 23 countries which yields table 4.5.

BEL	w/ constant	w/o constant
1 lag	0.2262	0.2255
2 lags	0.2074	0.2058
3 lags	0.2025	0.2016
4 lags	0.2114	0.2139
5 lags	0.2579	0.2397
6 lags	0.3015	0.2465
7 lags	0.2991	0.2740
8 lags	0.2594	0.2660
9 lags	0.3237	0.2927
10 lags	0.5356	0.4103
11 lags	0.7903	1.0072

Table 4.4: RMSE, all VAR models BEL

Country	Model	RMSE
AUS	VAR(2) w/o cons	0.3259
AUT	VAR(1) w/ cons	0.2894
BEL	VAR(3) w/o cons	0.2016
CAN	VAR(5) w/o cons	0.3740
CZE	VAR(1) w/o cons	0.6132
DEU	VAR(1) w/o cons	0.3141
DNK	VAR(6) w/o cons	0.9310
ESP	VAR(4) w/ cons	0.2408
EST	VAR(3) w/o cons	0.6198
FIN	VAR(1) w/o cons	0.3143
GBR	VAR(3) w/o cons	0.3062
GRC	VAR(4) w/o cons	0.4675
IRL	VAR(3) w/o cons	3.4951
ISR	VAR(1) w/o cons	0.3376
ITA	VAR(3) w/o cons	0.1875
KOR	VAR(5) w/ cons	0.4194
MEX	VAR(2) w/ cons	0.4211
NLD	VAR(2) w/ cons	0.2739
NOR	VAR(5) w/o cons	0.5592
NZL	VAR(2) w/ cons	0.2246
POL	VAR(1) w/ cons	0.5906
PRT	VAR(7) w/ cons	0.4590
SVK	VAR(1) w/o cons	0.2779
USA	VAR(2) w/ cons	0.1988

Table 4.5: RMSE VAR, all countries

Country	Model	RMSE
AUS	VEC(2) w/o cons	0.3264
AUT	VEC(5) w/ cons	0.2495
BEL	VEC(2) w/ cons	0.2069
CAN	VEC(3) w/ cons	0.4883
CZE	VEC(2) w/ cons	0.5515
DEU	VEC(2) w/o cons	0.3436
DNK	VEC(5) w/ cons	0.3264
ESP	VEC(4) w/o cons	0.1779
EST	VEC(3) w/ cons	0.5669
FIN	VEC(2) w/ cons	0.2890
GBR	VEC(3) w/ cons	0.2169
GRC	VEC(2) w/ cons	0.6772
IRL	VEC(3) w/o cons	3.5219
ISR	VEC(3) w/o cons	0.3319
ITA	VEC(2) w/ cons	0.2147
KOR	VEC(2) w/o cons	0.3821
MEX	VEC(2) w/ cons	0.4995
NLD	VEC(2) w/ cons	0.4039
NOR	VEC(6) w/o cons	0.5833
NZL	VEC(2) w/o cons	0.2206
POL	VEC(2) w/ cons	0.6128
PRT	VEC(3) w/ cons	0.4250
SVK	VEC(2) w/ cons	0.1800
USA	VEC(5) w/o cons	0.2138

Table 4.6: RMSE VEC, all countries

The VAR models that yield the smallest RMSE for the forecast period for each country have between one and seven lags and a majority are models without a constant. The much higher value for the Ireland forecast can be explained by spikes in the data during the forecasting period (see the graphic in appendix B).

4.2.2 VEC

The Vector Error Correction model takes care of the combination of integrated time series and the presence of cointegration. Since the determination of the lag length of the VEC model is found by defining the one of the VAR model, again all possible models are estimated and for each country the one with the smallest RMSE over the testing period is chosen. The results are shown in table 4.6. The results of the VEC and VAR models appear very much alike and one can calculate the mean RMSE for all forecasts performed by each method:

$$\begin{array}{lcr} \text{Mean RMSE VAR} & = & 0.5184 \\ \text{Mean RMSE VEC} & = & 0.5260 \\ \hline & & -0.0076 \end{array}$$

The difference between the two forecasts seems negligible, nevertheless the Diebold-Mariano test is performed to conclude whether the predictions differ statistically.

The null of equal predictive accuracy can be rejected for 13 out of the 24 forecasts (test statistics and p-values can be found in appendix $A.1$) and for these 13 countries the mean of the VAR is again slightly lower than that of the VEC (0.4617 vs. 0.4862). Consequently the forecasts produced by the VAR models will compete with the Neural Networks.

4.3 Neural Networks

4.3.1 General Structure

The history of neural networks extends back to middle of the twentieth century with the initial idea to mimic the workings of the human brain. Practical application though was only possible with the subsequent rise of computing power and starting in the 1990s ever more articles in the field of economics using neural networks were published.
Neural networks can be used for classification, causal modeling and time series forecasting with the latter being the obvious application in this thesis. There are many different types of neural networks with the most common one in forecasting applications being a feedforward neural network where information moves only in one direction (much like Vector Autoregression). The schematic view seen in figure 4.1 shows a so-called multilayer perceptron (which is a term often used interchangeable with feedforward NN), with four input nodes and one hidden layer with three nodes:

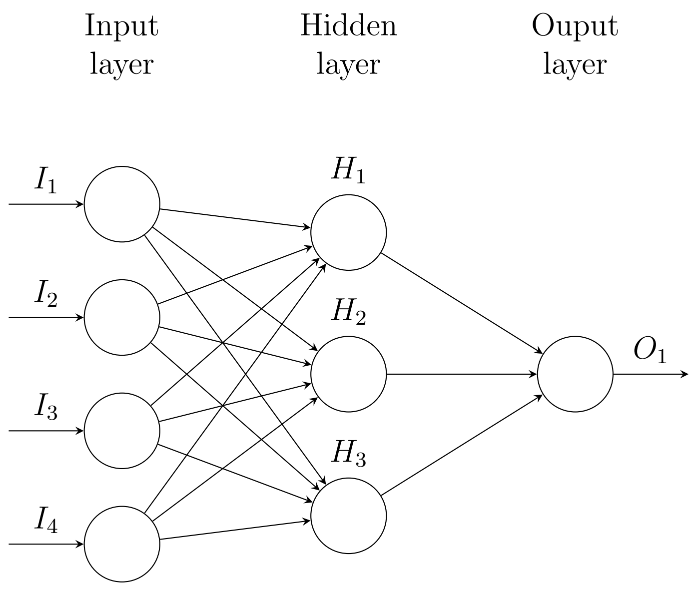


Figure 4.1: Multilayer perceptron

The four input nodes are connected to the nodes of the hidden layer and that again is connected to the node of the output layer. For each node in the hidden layer a linear combination of the input is produced which is then transformed by a non-linear function (the so-called activation function). The final step is to take a weighted average of the outputs to make a prediction.

The main difference between Neural Networks and the Vector Autoregression (and the VEC) is that the latter are linear models whereas the multilayer perceptron is a non-linear one. To enable comparison of the forecasts of the linear and non-linear models the same data sets are used (see data chapter). Therefore the NNs trained in the next section use the time series of GDP growth, inflation, short-term interest rate and unemployment for their four input neurons.
The next step is to determine the number off hidden layers and the number of nodes in each. It is interesting to note that the so-called "deep learning" neural networks, that very recently gained a lot of popularity for their impressive performance in computer vision, speech recognition and natural language processing, apply multiple hidden layers (therefore "deep" learning). According to the literature for regression-type problems a single hidden layer is usually sufficient and all networks constructed and trained for this thesis use only one hidden layer. There is no widely accepted approach to specify the number of nodes in a hidden layer and according to the literature the performance is not very sensitive to the number of nodes. As an activation function a sigmoid logistic or a hyperbolic function is usually used. [22]

4.3.2 Practical Implementation

Various possible implementations of a feed-forward neural network, applied to provide economic forecasts, are conceivable. The basic methodology of neural networks, as outlined in the previous chapter, is made available in the *R* environment [23] by the package `neuralnet` [5]. Only very recently another package called `nnfor` [12] was released with the main purpose of producing economic forecasts by utilizing the artificial intelligence of neural networks which is used for all forecasts produced by NNs in this thesis.

Multivariate NN

To specify and train a multilayer perceptron for each country - within the *R* package `nnfor` - the time series of interest (QGDP) is used as the main input and the other time series are included as additional regressors. As mentioned above all networks use one hidden layer. Attempts to vary the number of nodes in the hidden layer did not yield any improvement regarding the forecasts produced and therefore the standard provided by the *R* package `nnfor` was used (five nodes in the hidden layer). The activation function is a sigmoid one which is predetermined by the `neuralnet` package.
Compared to linear regression models, even simple neural networks (as the ones applied here) have substantially more parameters to estimate which makes them overly flexible and the training of neural networks is a complex nonlinear optimization problem which both makes it more likely to produce an over- or underfit. To circumvent that problem for each country an ensemble of 20 networks (each with slightly different starting weights) is trained and each one produces a forecast. Then the median of those forecasts is taken, so no single over-fitted network is preferred. The table 4.7 shows the root mean squared error of the forecasts of the NN ensembles for each country.
The mean RMSE of the forecasts produced by the multivariate neural networks (as all time series are used as inputs) is much higher at 1.6538 than that of the VAR.

Country	RMSE multiNN
AUS	0.3372
AUT	0.2894
BEL	0.4415
CAN	0.9098
CZE	0.7941
DEU	0.8718
DNK	2.4292
ESP	1.1873
EST	1.7431
FIN	0.6177
GBR	1.9529
GRC	0.9019
IRL	17.6080
ISR	0.6852
ITA	0.1723
KOR	0.5533
MEX	0.5417
NLD	2.2613
NOR	0.6871
NZL	0.2777
POL	0.6371
PRT	0.5576
SVK	0.8821
USA	0.3170

Table 4.7: RMSE multiNN, all countries

This can be partly explained by few very bad forecasts (especially IRL, where the network prediction is extremely cyclical), but also the other forecasts perform worse on average. As mentioned above neural networks are prone to overfitting and it could be that taking the mean of forecasting sample is not enough to correct for that. A different approach with drastically simplified NNs is therefore taken in next section.

Univariate NN

To reduce the chance of producing an overfitted network ensemble a univariate NN is estimated and trained for each country. Here the only input is the QGDP time series therefore the schematic of the network changes (figure 4.2). All other specifications (hidden layers, nodes) remain the same. The RMSE of the forecasts produced by the univariate NNs are depicted in table 4.8.
Again one can calculate the RMSE and compare it with the forecasts produced by the multivariate NN:

$$\begin{array}{rcr} \text{Mean RMSE uniNN} & = & 0.6939 \\ \text{Mean RMSE multiNN} & = & 1.6538 \\ \hline & & -0.9599 \end{array}$$

Clearly the univariate NNs outperform the multivariate ones but to be sure the Diebold-Mariano test is performed (test statistics and p-values can be found in appendix $A.2$) with the result that for 15 of the 24 countries the forecasts differ statistically significant. The mean RMSE of those 15 countries is again much lower for the univariate networks and therefore they will compete with the VAR in the next section.

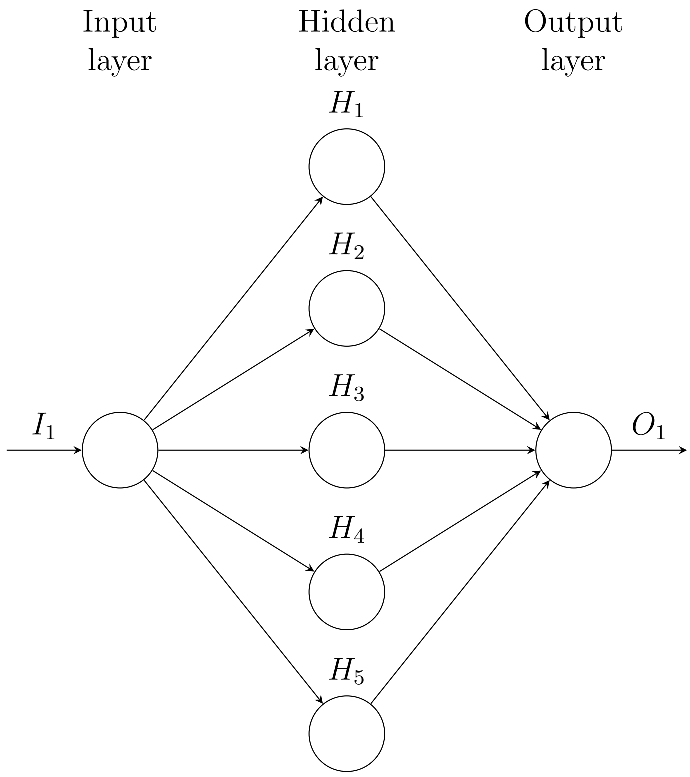


Figure 4.2: Univariate Neural Network

Country	RMSE uniNN
AUS	0.3447
AUT	0.3550
BEL	0.2053
CAN	0.4246
CZE	0.9252
DEU	0.3331
DNK	0.9796
ESP	0.7418
EST	0.9059
FIN	0.3676
GBR	0.3020
GRC	0.8077
IRL	3.5674
ISR	0.4153
ITA	0.7241
KOR	0.4885
MEX	0.6755
NLD	0.9318
NOR	0.7167
NZL	0.2583
POL	0.6053
PRT	0.3532
SVK	0.4942
USA	0.7283

Table 4.8: RMSE uniNN, all countries

Chapter 5

Results

5.1 Results

Combining the results obtained in the sections above one can compare the RMSE of the forecasts produced by the VAR and the univariate NN.

Country	RMSE VAR	RMSE uniNN	uniNN - VAR	Better Perf
AUS	0.3259	0.3447	0.0188	VAR
AUT	0.2894	0.3550	0.0656	VAR
BEL	0.2016	0.2053	0.0037	VAR
CAN	0.3740	0.4246	0.0506	VAR
CZE	0.6132	0.9252	0.3120	VAR
DEU	0.3141	0.3331	0.0191	VAR
DNK	0.9310	0.9796	0.0486	VAR
ESP	0.2408	0.7418	0.5010	VAR
EST	0.6198	0.9059	0.2861	VAR
FIN	0.3143	0.3676	0.0533	VAR
GBR	0.3062	0.3020	−0.0042	uniNN
GRC	0.4675	0.8077	0.3402	VAR
IRL	3.4951	3.5674	0.0723	VAR
ISR	0.3376	0.4153	0.0777	VAR
ITA	0.1875	0.7241	0.5366	VAR
KOR	0.4194	0.4885	0.0692	VAR
MEX	0.4211	0.6755	0.2545	VAR
NLD	0.2739	0.9318	0.6579	VAR
NOR	0.5592	0.7167	0.1576	VAR
NZL	0.2246	0.2583	0.0338	VAR
POL	0.5906	0.6053	0.0147	VAR
PRT	0.4590	0.3532	−0.1058	uniNN
SVK	0.2779	0.4942	0.2163	VAR
USA	0.1988	0.7283	0.5295	VAR
mean	0.5184	0.6938	0.1754	

Table 5.1: RMSE results, VAR and uniNN

With two exceptions the root mean squarred error of the forecasts produced with vector auto regressions are smaller than the ones of the univariate neural networks.

5.2 Robustness Testing

The results presented in the previous section induce the conclusion that for the available dataset the VAR produces superior forecasts compared to the univariate NN. This only holds though if the forecasts produced for each country are actually statistically different, i.e. that their difference is significant when tested.
Various types of tests were proposed in the literature and for this thesis a variation of the Diebold-Mariano (DM) test [3] is applied: Broadly speaking it tests the null hypothesis that the models have the same equal predictive accuracy (EPA) against the alternative (models do not have the same EPA, therefore one performs better than the other). The version used in this chapter (and also in the appendix) uses a finite sample correction, suggested by Harvey-Leybourne-Newbold [14]. The following table shows the test results for each country:

Country	DM t-stat	DM p-value
AUS	−1.8645	0.0918
AUT	−0.3472	0.7356
BEL	−0.8502	0.4151
CAN	−1.2089	0.2545
CZE	0.4974	0.6297
DEU	−0.3552	0.7298
DNK	−1.4546	0.1764
ESP	1.5990	0.1409
EST	−3.1067	0.0111
FIN	−1.1234	0.2875
GBR	0.3307	0.7477
GRC	−4.6189	<0.01
IRL	−2.0298	0.0698
ISR	−0.9469	0.3660
ITA	1.2088	0.2546
KOR	−1.3108	0.2192
MEX	−5.8162	<0.01
NLD	−6.3405	<0.01
NOR	−2.0180	0.0712
NZL	−0.5739	0.5787
POL	−1.5381	0.1550
PRT	7.5657	<0.01
SVK	−1.5202	0.1594
USA	−3.4686	<0.01

Table 5.2: DM-Test VAR/uniNN

For 18 of the 24 countries in this sample the null hypothesis of Equal Predictive Accuracy can not be rejected (with the common significance level of 5%). This means for those 18 countries the forecasts produced by the Vector Autoregression and the Neural Networks are not significantly different and therefore their performance is the same. For the remaining 6 countries the VAR performs better in 5 of the forecasts and the univariate NN in one. Whether these results can support the assessment that the VAR outperforms the NN overall will be discussed in the last chapter.

Chapter 6

Conclusion

In this thesis the forecasting performance of four different models was compared. The data used are quarterly time series of 24 OECD members, namely GDP growth, inflation, short-term interest rate and unemployment spanning the time period Q1 2000 to Q3 2018. The aim was to find the best forecast (measured as the smallest RMSE) over the 11-step forecast period (Q1 2016 to Q3 2018).

For this large, heterogenous sample of countries first a vector autoregression and then a vector error correction model were estimated. For each country all possible models were computed (restricted by the length of the time series and OLS estimation) and the one with the smallest RMSE over the forecast period was chosen. As the vector autoregression slightly outperformed the vector error correction models it was chosen to compete with the neural networks.

As for the neural networks at first a multivariate one was estimated and trained for each country and its performance was compared to univariate one (which only uses the QGDP time series as input). As the univariate networks outclassed the multivariate ones their forecasts entered the final competition. This "forecasting contest" is depicted in the in figure 6.1.

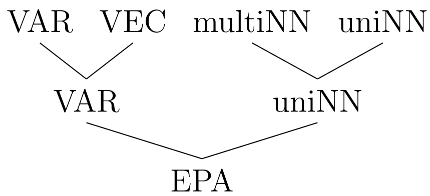


Figure 6.1: "Forecasting competition"

For 18 of the 24 countries the forecasts made by the vector autoregressions and the univariate neural networks have, according to the Diebold-Mariano, equal predictive accuracy (EPA) which means the forecasts are statistically the same. In the remaining cases the test rejects the EPA hypothesis and for five of the countries the RMSE of the VARs is smaller than that of the NNs and in one case it is the other way around.

Concluding it can be stated that for the dataset at hand a VEC model does not

improve on the forecasts of a VAR model and that presumably due to overfitting a multivariate NN is outperformed by a univariate one (an example of the KISS principle).
The main finding of this thesis is though, that neural networks perform almost as well as vector autoregressions in predicting quarterly gross domestic product growth while de facto using only a quarter of the data as input. That underscores their outstanding performance and their possibly increased application in the domain of economic forecasting.

Appendix A

Test Statistics

A.1 Diebold-Mariano Test: VAR vs VEC

Country	DM t-stat	DM p-value	VAR-VEC	Better Perf
AUS	−2.1652	0.0556	−0.0005	VAR
AUT	4.5463	<0.01	0.0399	VEC
BEL	−0.674 17	0.5155	−0.0053	VAR
CAN	−2.8696	0.0167	−0.1143	VAR
CZE	1.9892	0.0747	0.0617	VEC
DEU	−0.659 33	0.5246	−0.0296	VAR
DNK	−4.4904	<0.01	−0.0106	VAR
ESP	1.1253	0.2868	0.0629	VEC
EST	0.745 64	0.4730	0.0530	VEC
FIN	0.5763	0.5772	0.0253	VEC
GBR	1.2100	0.2541	0.0893	VEC
GRC	−2.6692	0.0235	−0.2097	VAR
IRL	−0.4380	0.6707	−0.0268	VAR
ISR	0.1560	0.8791	0.0057	VEC
ITA	−0.9316	0.3735	−0.0272	VAR
KOR	2.4658	0.0333	0.0373	VEC
MEX	−2.5963	0.0267	−0.0784	VAR
NLD	−11.72	<0.01	−0.1300	VAR
NOR	−23.378	<0.01	−0.0242	VAR
NZL	0.1475	0.8856	0.0040	VEC
POL	−2.0765	0.0646	−0.0223	VAR
PRT	4.4463	<0.01	0.0341	VEC
SVK	1.8365	0.0962	0.0979	VEC
USA	−0.6091	0.5561	−0.0149	VAR

Table A.1: DM-test, VAR/VEC

In general a test horizon $h = 3$ was chosen, for three countries it is not possible to calculate the test statistic at that point, therefore adjacent values were chosen ($h = 4$ for AUS, $h = 2$ for GRC and KOR).

A.2 Diebold-Mariano Test: multiNN vs uniNN

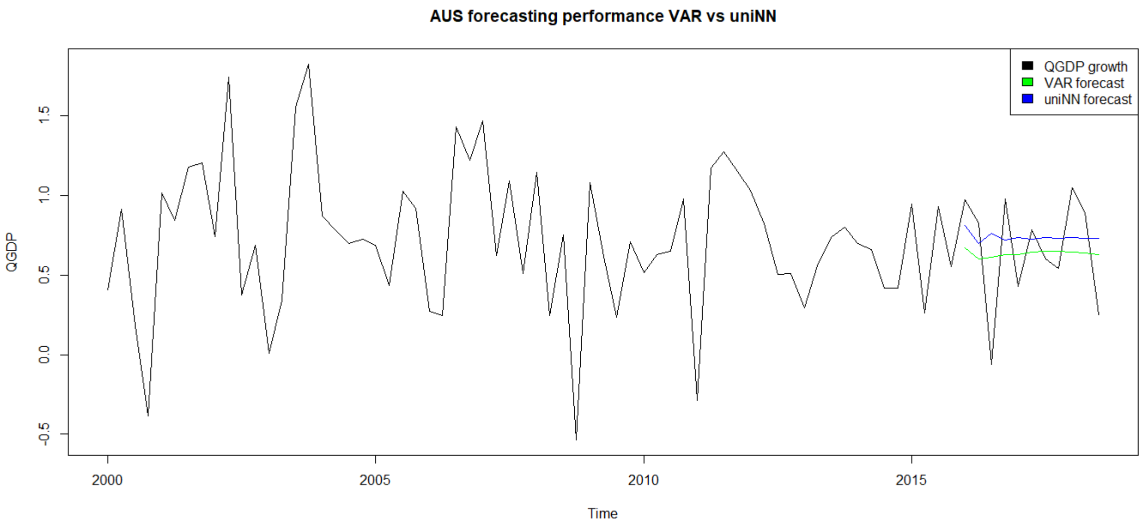
Country	DM t-stat	DM p-value	uniNN - multiNN	Better Perf
AUS	0.8837	0.3976	0.0075	multiNN
AUT	2.6425	0.0246	−1.9684	uniNN
BEL	2.2798	0.0458	−0.2362	uniNN
CAN	2.2614	0.0473	−0.4852	uniNN
CZE	−3.4727	<0.01	0.1311	multiNN
DEU	2.3658	0.0396	−0.5387	uniNN
DNK	1.7308	0.1142	−1.4496	uniNN
ESP	1.1035	0.2957	−0.4455	uniNN
EST	4.7448	<0.01	−0.8371	uniNN
FIN	1.4854	0.1683	−0.2501	uniNN
GBR	1.7469	0.1112	−1.6509	uniNN
GRC	1.4714	0.1719	−0.0942	uniNN
IRL	3.265	<0.01	−14.0406	uniNN
ISR	2.3502	0.0406	−0.2699	uniNN
ITA	−3.4367	<0.01	0.5518	multiNN
KOR	4.2645	<0.01	−0.0648	uniNN
MEX	−2.9074	0.0156	0.1338	multiNN
NLD	8.0661	<0.01	−1.3296	uniNN
NOR	−0.4542	0.6594	0.0296	multiNN
NZL	2.8087	0.0185	−0.0193	uniNN
POL	0.3206	0.7551	−0.0318	uniNN
PRT	1.1763	0.2667	−0.2044	uniNN
SVK	3.6829	<0.01	−0.3879	uniNN
USA	−2.8527	0.0172	0.4114	multiNN

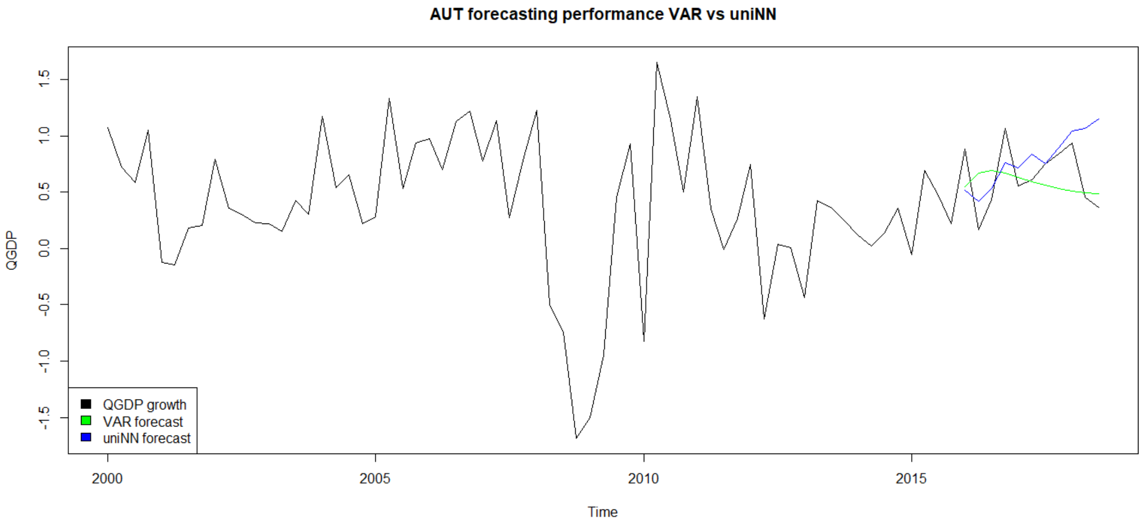
Table A.2: DM-test, multiNN/uniNN

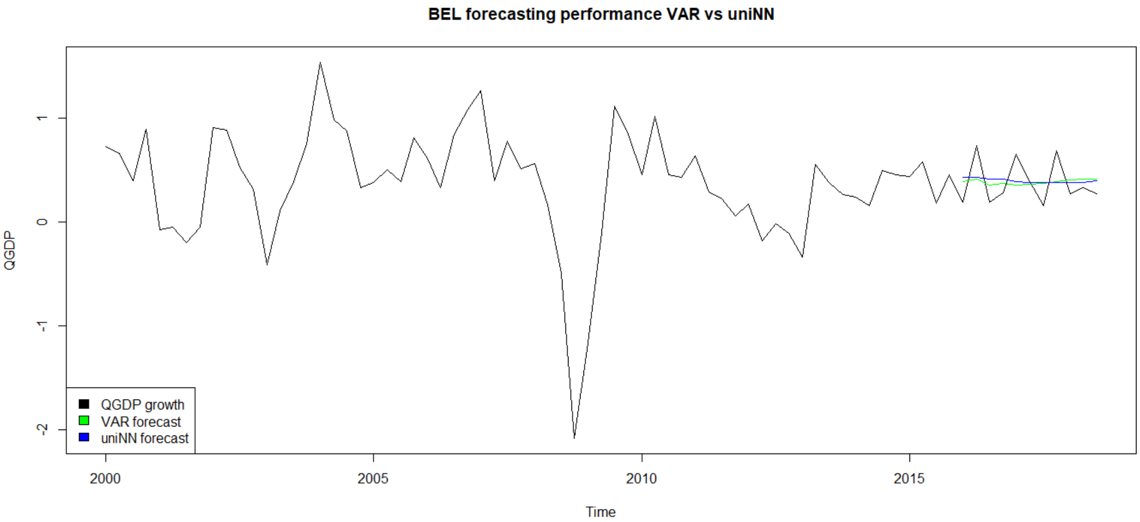
In general a test horizon $h = 3$ was chosen, for two countries it is not possible to calculate the test statistic at that point, therefore adjacent values were chosen ($h = 4$ for GRC, and $h = 2$ for IRL).

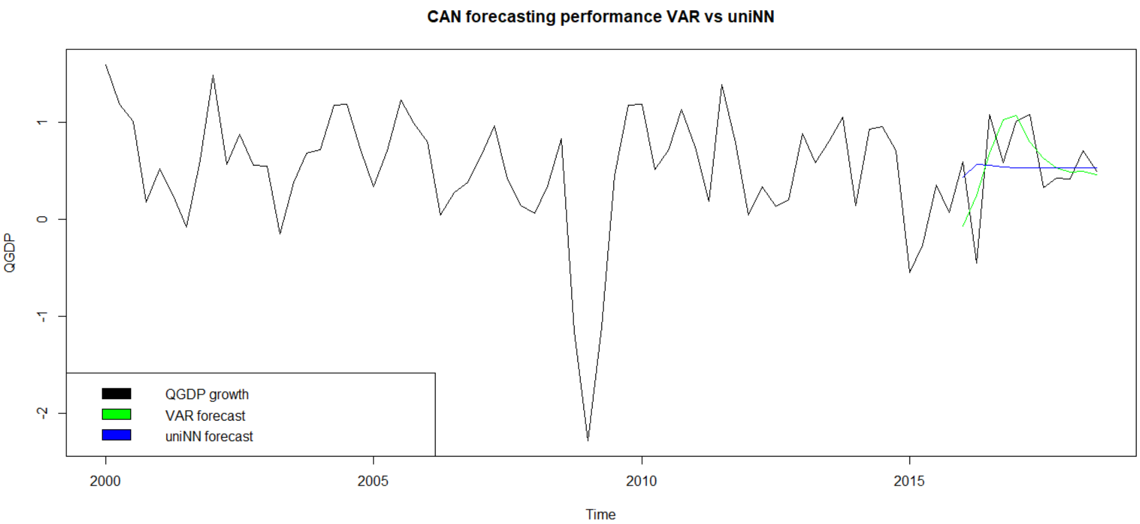
Appendix B

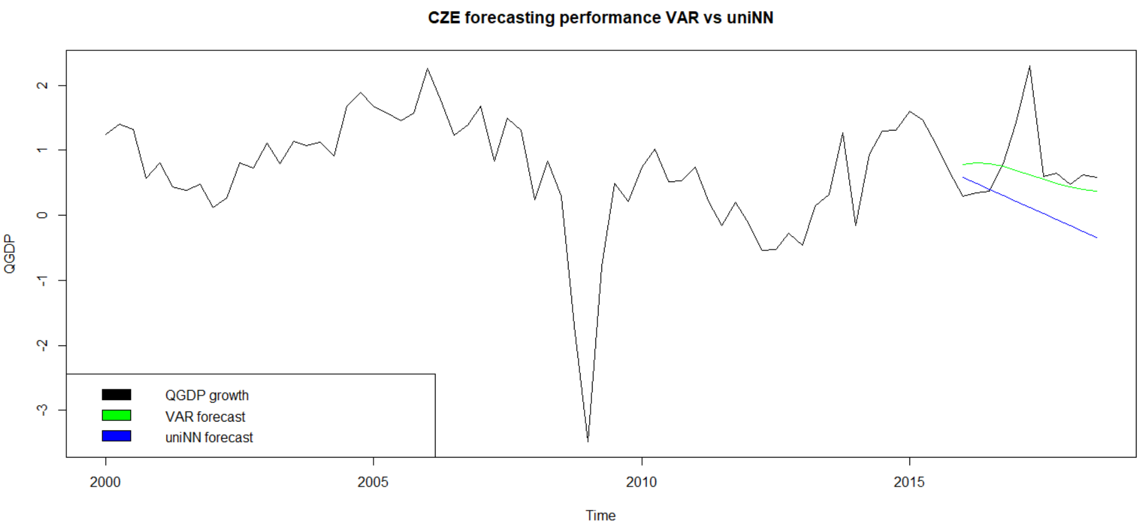
All Forecasts - Country Graphs

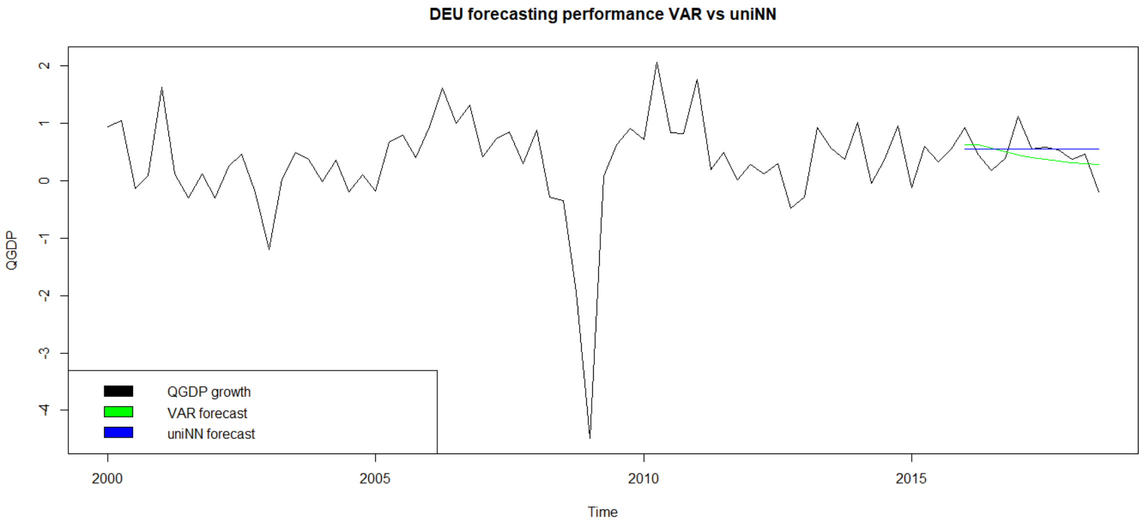


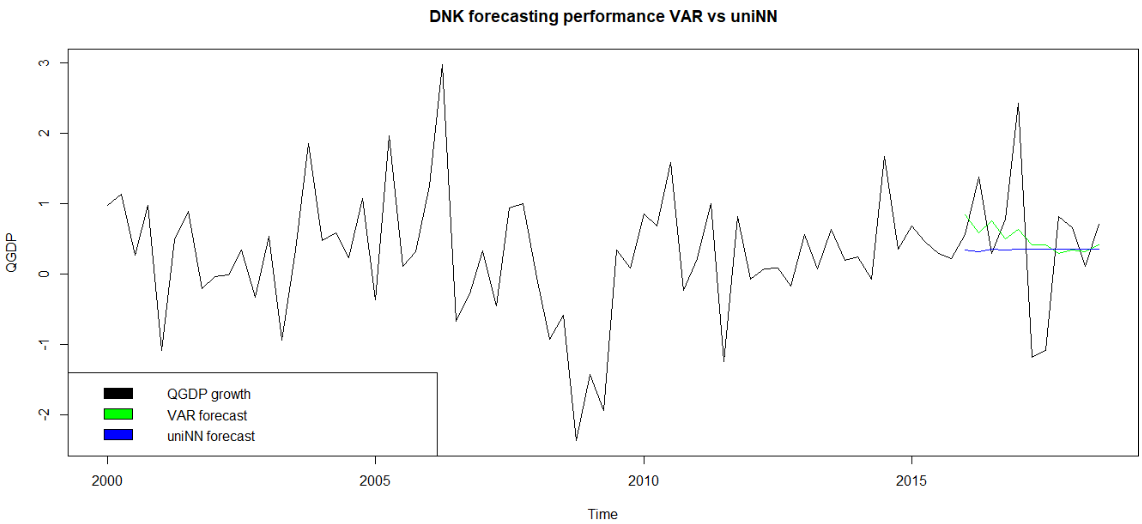


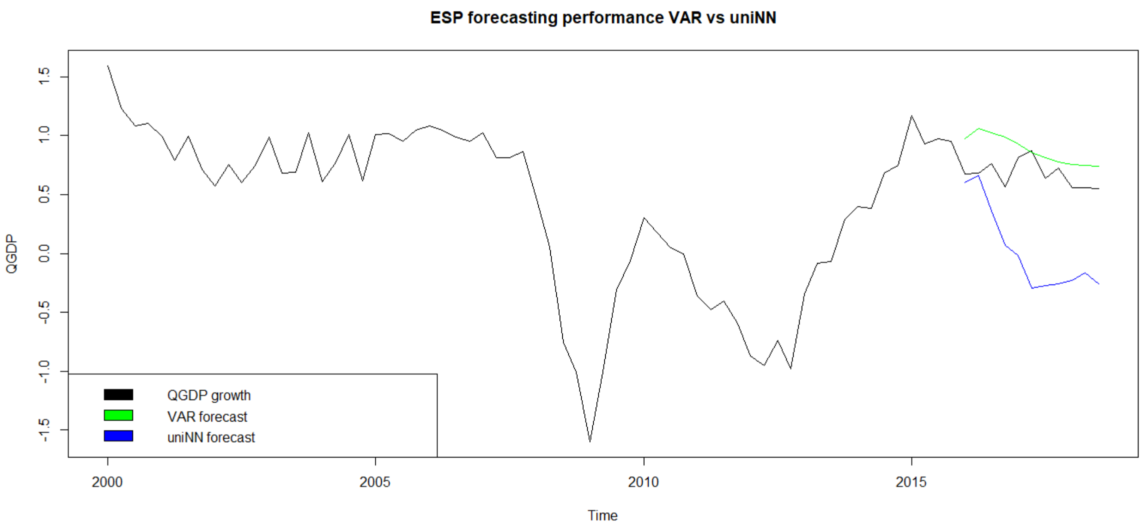

BEL forecasting performance VAR vs uniNN
QGDP
Time
QGDP growth
VAR forecast
uniNN forecast

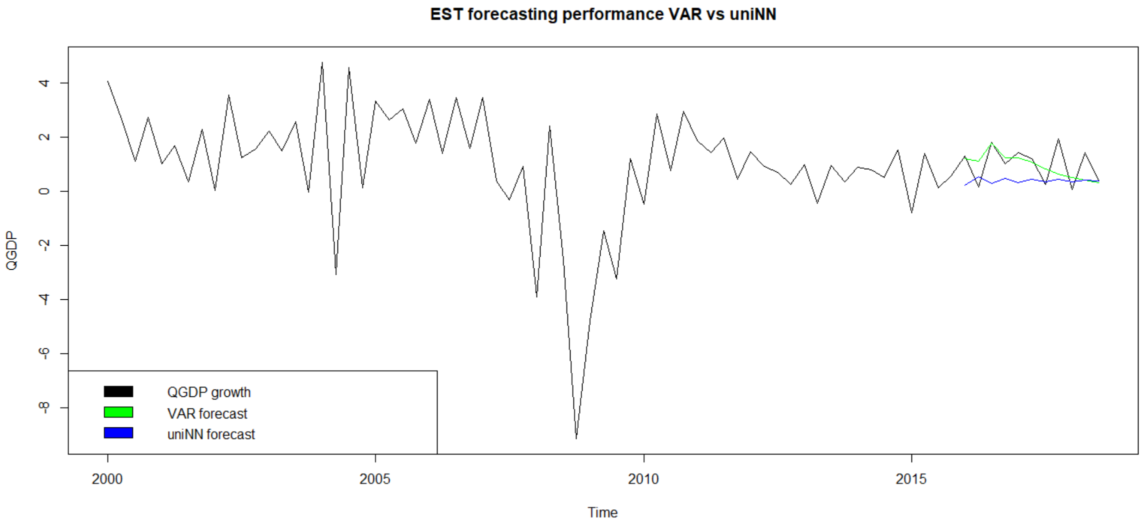

CAN forecasting performance VAR vs uniNN
QGDP
Time
QGDP growth
VAR forecast
uniNN forecast

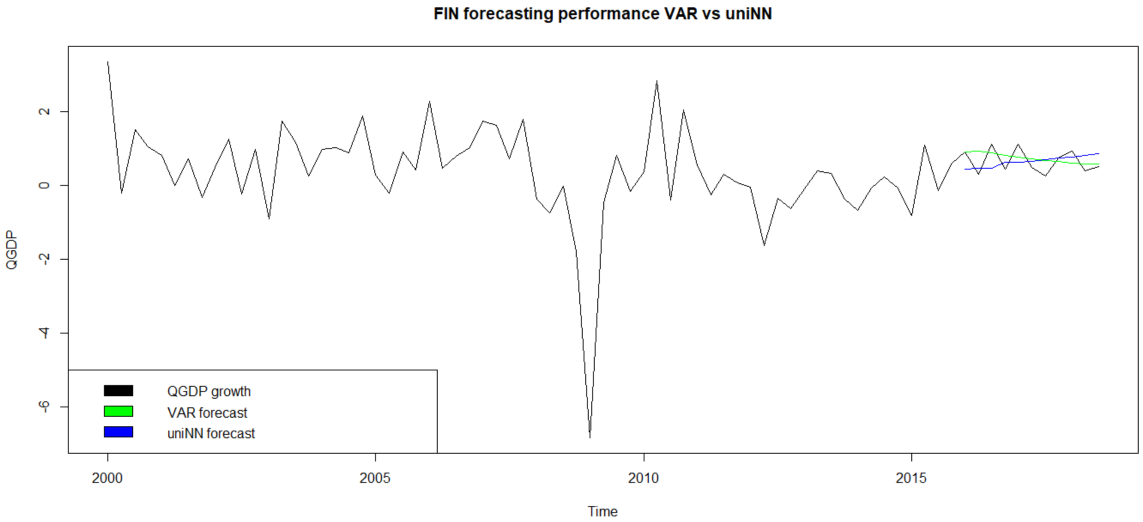

CZE forecasting performance VAR vs uniNN
QGDP
Time
QGDP growth
VAR forecast
uniNN forecast

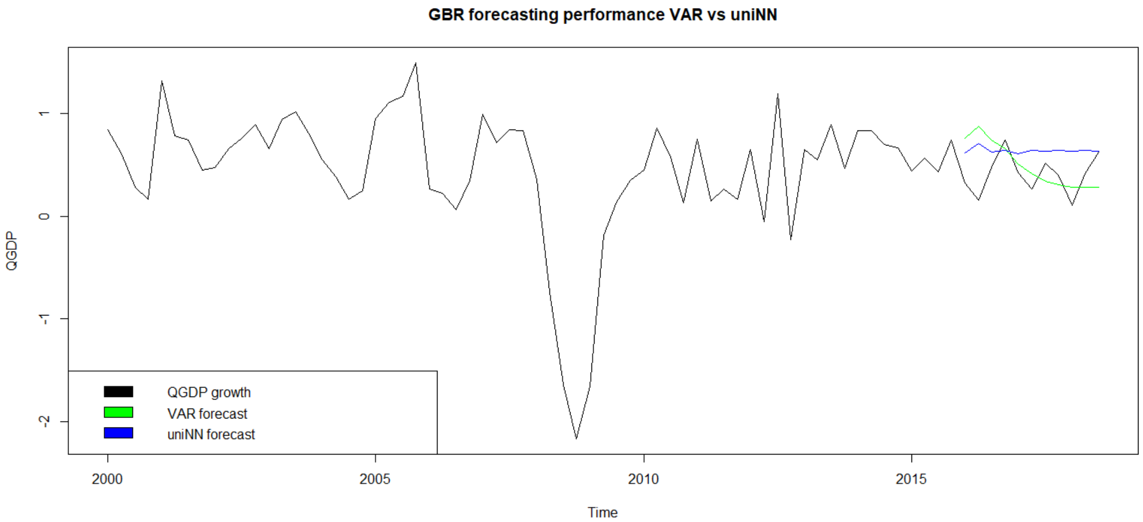

DEU forecasting performance VAR vs uniNN
QGDP
Time
QGDP growth
VAR forecast
uniNN forecast

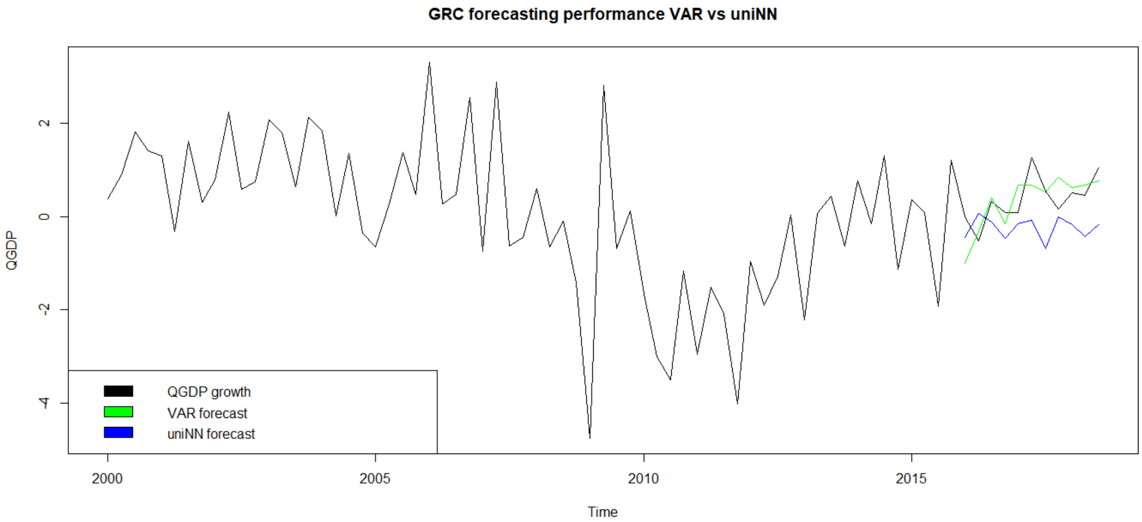

DNK forecasting performance VAR vs uniNN
QGDP
Time
QGDP growth
VAR forecast
uniNN forecast

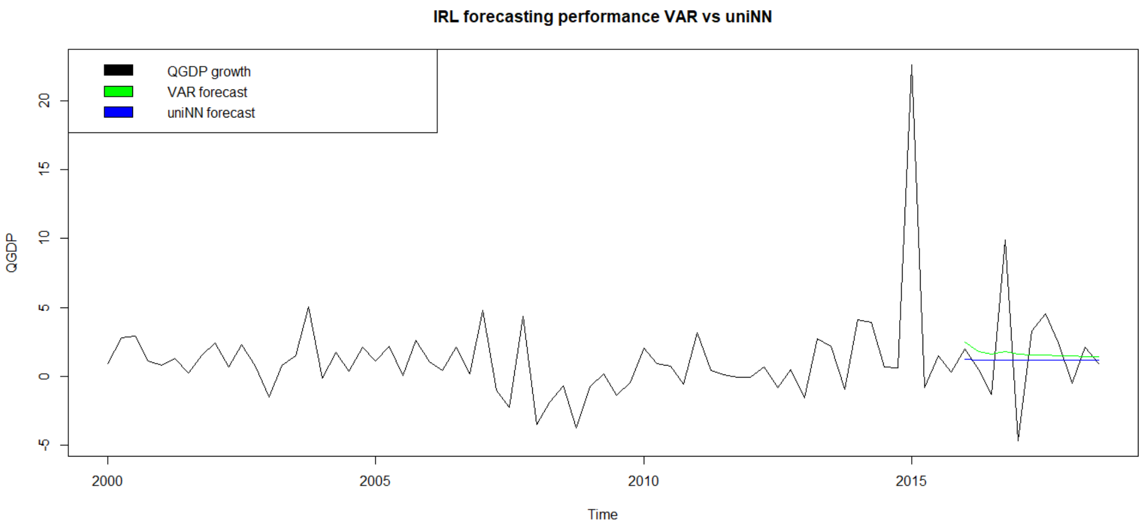

ESP forecasting performance VAR vs uniNN
QGDP
Time
QGDP growth
VAR forecast
uniNN forecast

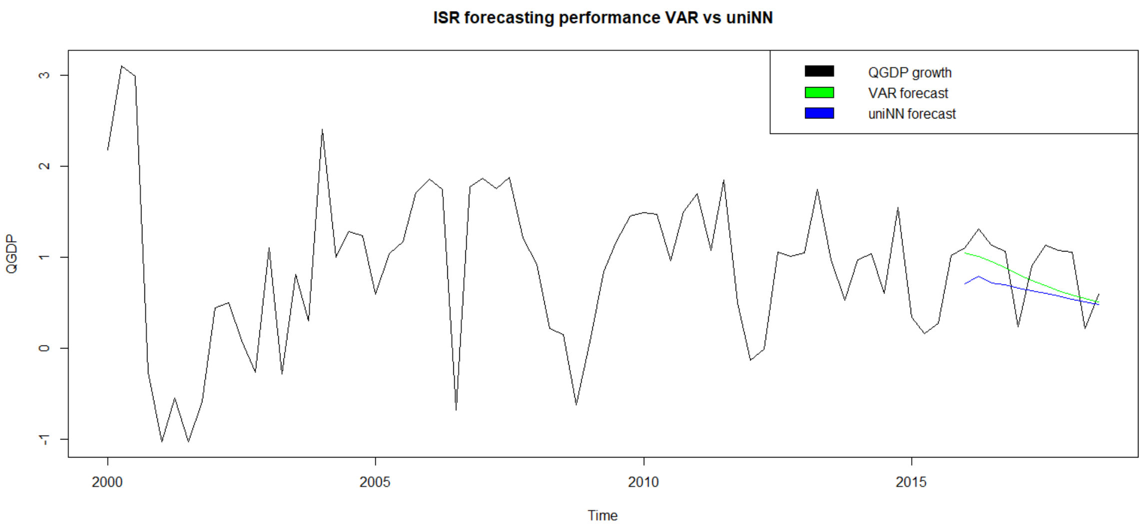

EST forecasting performance VAR vs uniNN
QGDP
Time
QGDP growth
VAR forecast
uniNN forecast
2000
2005
2010
2015

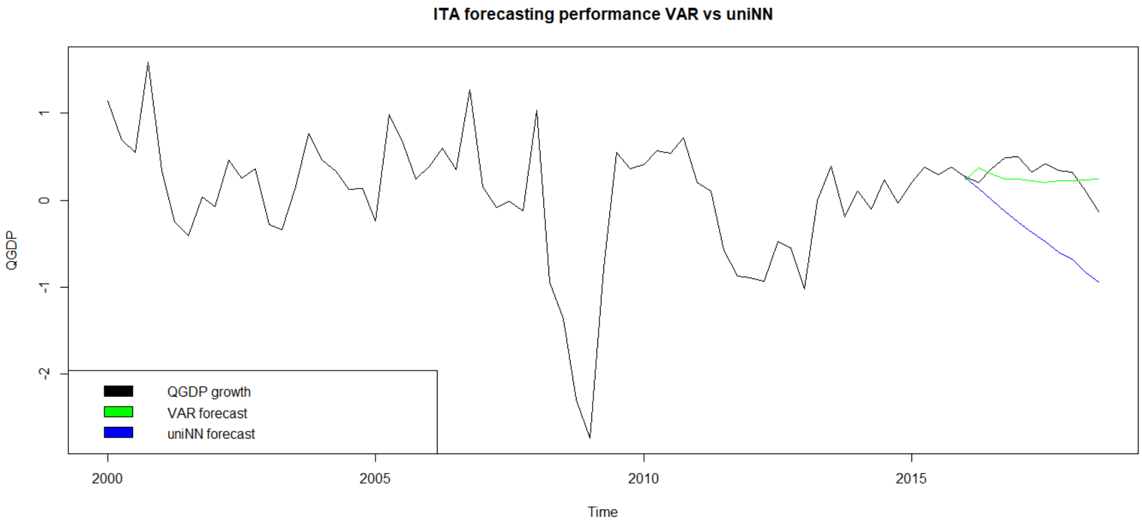

FIN forecasting performance VAR vs uniNN
QGDP
Time
QGDP growth
VAR forecast
uniNN forecast
2000
2005
2010
2015

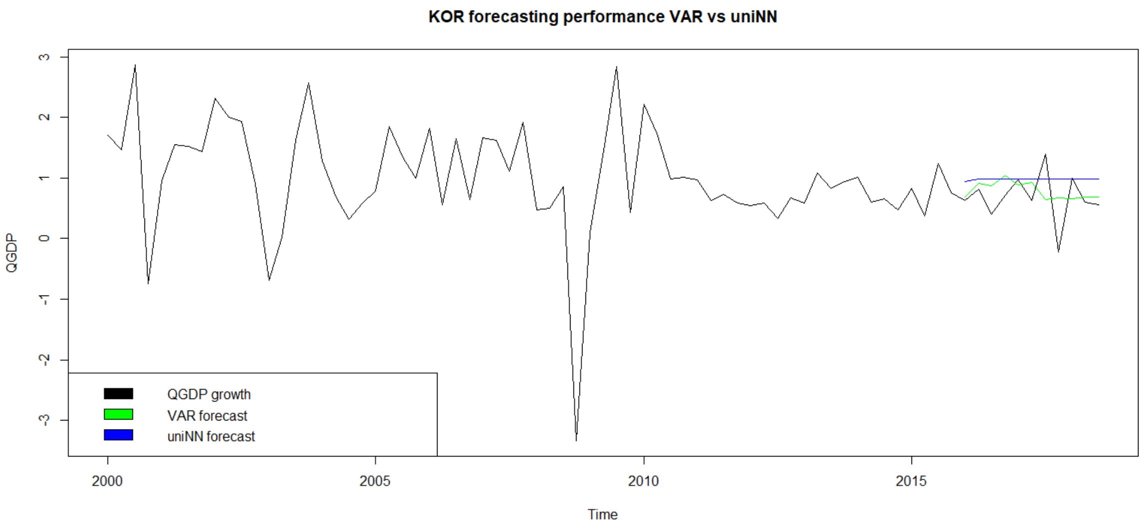

GBR forecasting performance VAR vs uniNN
QGDP
Time
QGDP growth
VAR forecast
uniNN forecast
2000
2005
2010
2015

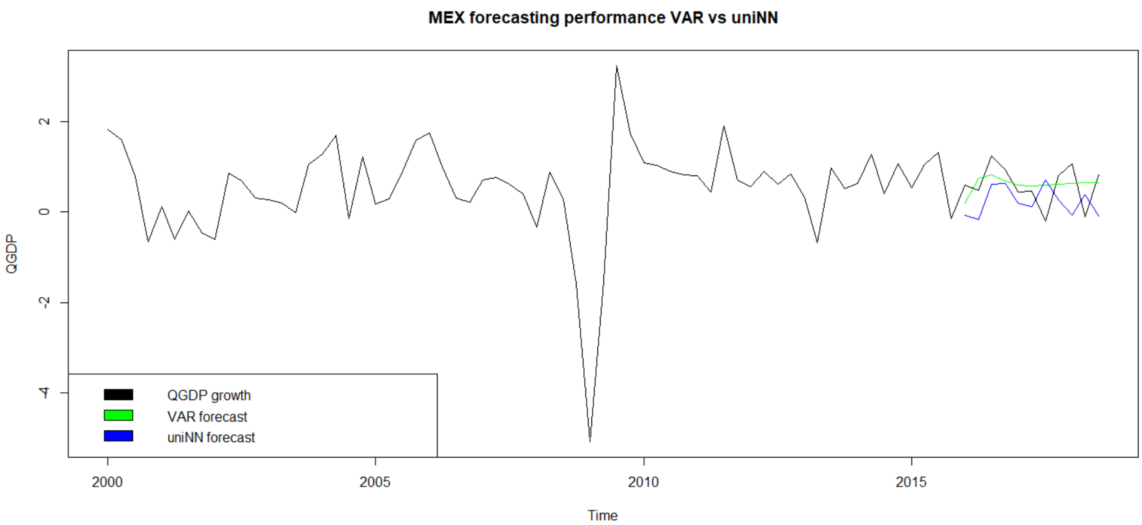

GRC forecasting performance VAR vs uniNN
QGDP
-4
-2
0
2
QGDP growth
VAR forecast
uniNN forecast
2000
2005
2010
2015
Time

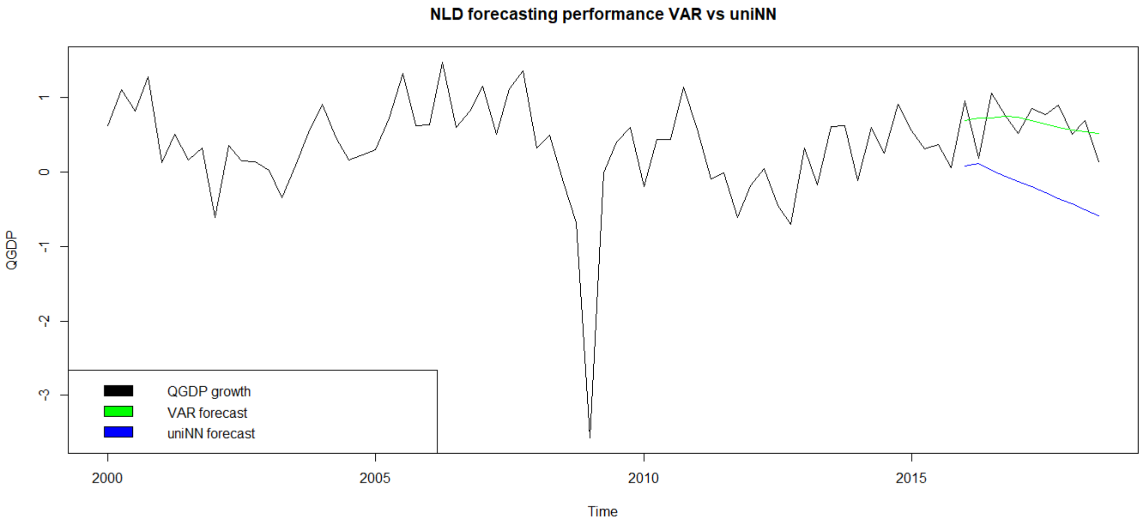

IRL forecasting performance VAR vs uniNN
QGDP
-5
0
5
10
15
20
QGDP growth
VAR forecast
uniNN forecast
2000
2005
2010
2015
Time

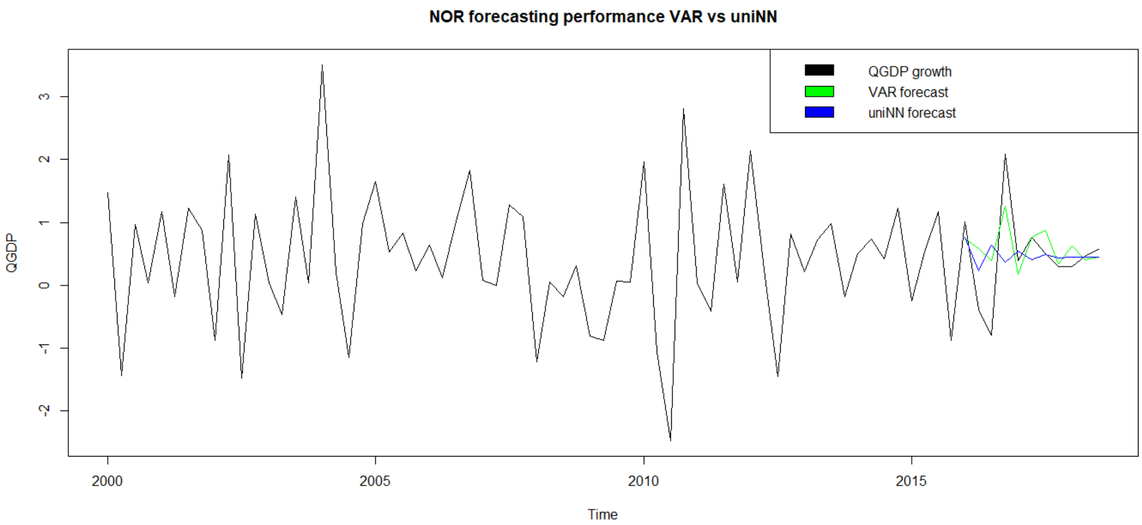

ISR forecasting performance VAR vs uniNN
QGDP
-1
0
1
2
3
QGDP growth
VAR forecast
uniNN forecast
2000
2005
2010
2015
Time

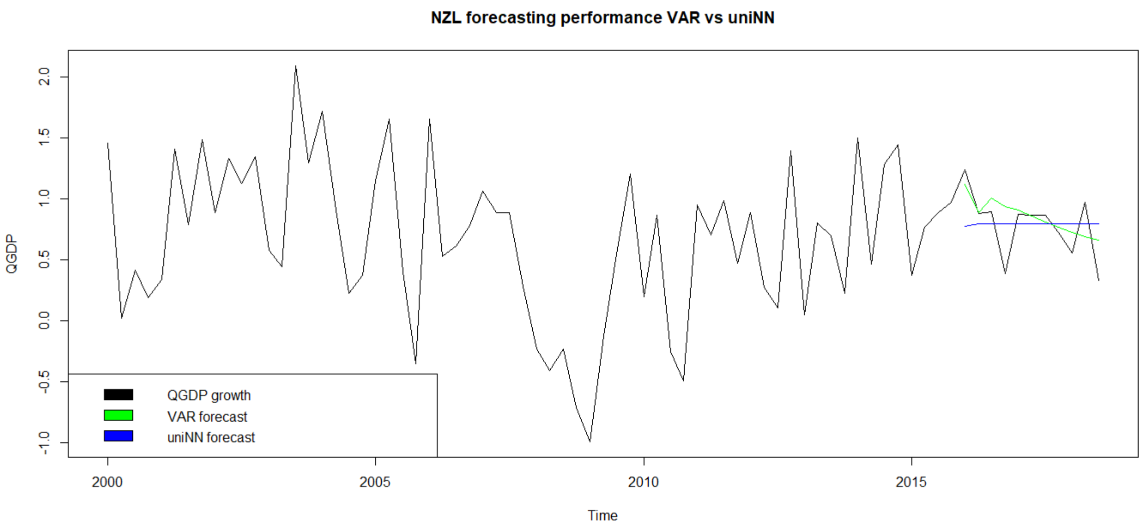

ITA forecasting performance VAR vs uniNN
QGDP
1
0
-1
-2
QGDP growth
VAR forecast
uniNN forecast
2000
2005
2010
2015
Time

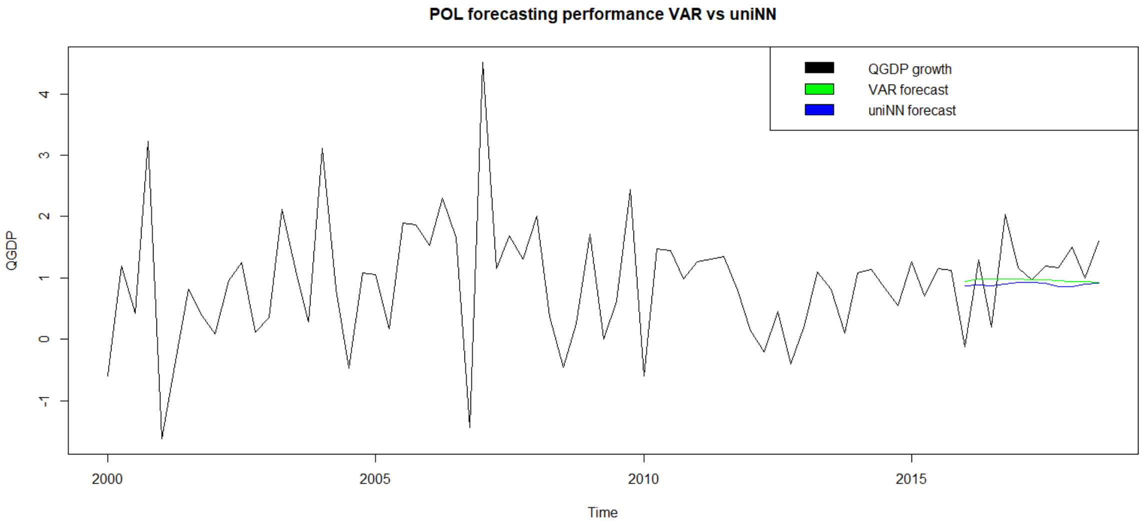

KOR forecasting performance VAR vs uniNN
QGDP
3
2
1
0
-1
-2
-3
QGDP growth
VAR forecast
uniNN forecast
2000
2005
2010
2015
Time

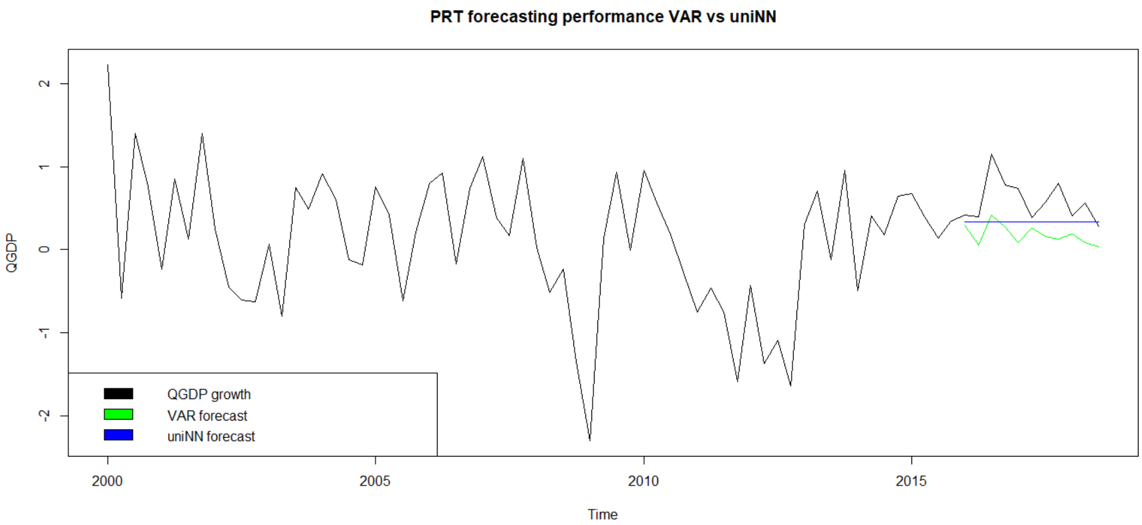

MEX forecasting performance VAR vs uniNN
QGDP
2
0
-2
-4
QGDP growth
VAR forecast
uniNN forecast
2000
2005
2010
2015
Time

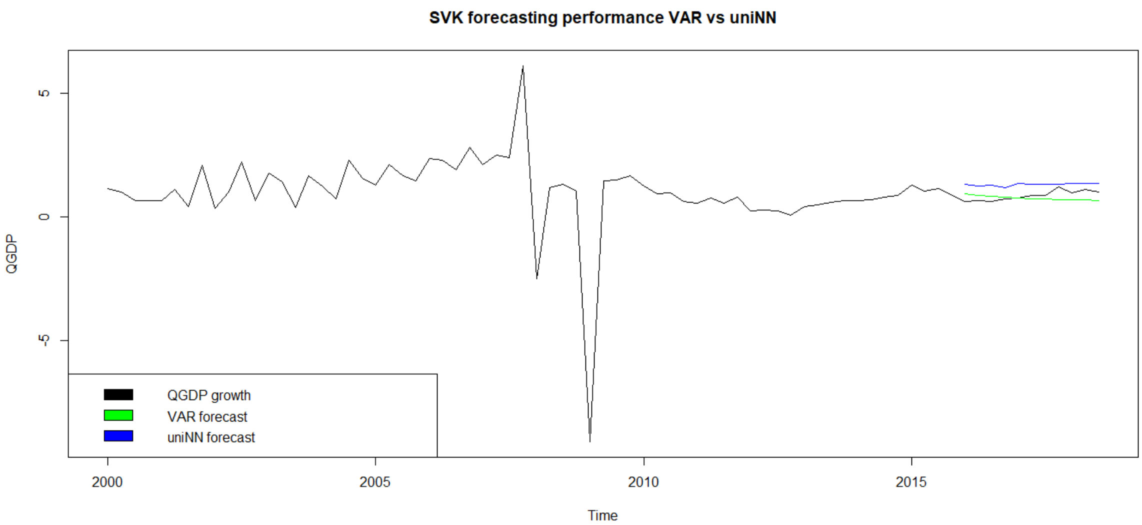

NLD forecasting performance VAR vs uniNN
QGDP
Time
2000
2005
2010
2015
QGDP growth
VAR forecast
uniNN forecast

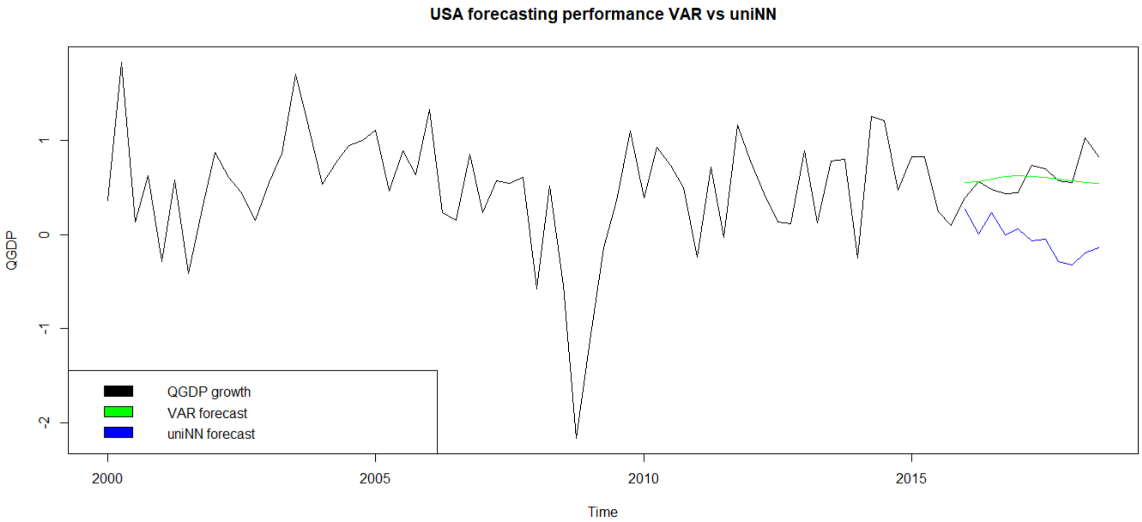

NOR forecasting performance VAR vs uniNN
QGDP
Time
2000
2005
2010
2015
QGDP growth
VAR forecast
uniNN forecast


NZL forecasting performance VAR vs uniNN
QGDP
Time
2000
2005
2010
2015
QGDP growth
VAR forecast
uniNN forecast


POL forecasting performance VAR vs uniNN
QGDP growth
VAR forecast
uniNN forecast
QGDP
Time
2000
2005
2010
2015


PRT forecasting performance VAR vs uniNN
QGDP growth
VAR forecast
uniNN forecast
QGDP
Time
2000
2005
2010
2015


SVK forecasting performance VAR vs uniNN
QGDP growth
VAR forecast
uniNN forecast
QGDP
Time
2000
2005
2010
2015



USA forecasting performance VAR vs uniNN
QGDP
1
0
-1
-2
QGDP growth
VAR forecast
uniNN forecast
2000
2005
2010
2015
Time

List of Tables

List of Figures

Bibliography

[1] Hirotugo Akaike, *A New Look at the Statistical Model Identification*, IEEE Transactions on Automatic Control, Vol. AC-19, No.6, pp. 716-723 1974.

[2] Milam Aiken, *Forecasting the United States gross domestic product with a neural network*, Journal of International Information Management, Vol. 9, Issue 1, pp. 67-75 2000.

[3] Diebold, F.X. and R.S. Mariano, *Comparing predictive accuracy*, Journal of Business & Economic Statistics, Vol. 13, pp. 134-145 1995.

[4] Robert F. Engle, Byung Sam Yoo, *Forecasting and testing in co-integrated systems*, Journal of Econometrics , Vol. 35, Issue 1, pp. 143-159 1987.

[5] Stefan Fritsch, Frauke Guenther and Marvin N. Wright, *neuralnet: Training of Neural Networks*, R package version 1.44.2, https://CRAN.R-project.org/package=neuralnet 2019.

[6] James D. Hamilton, *Time Series Analysis*, Princeton Univers. Press, 1994.

[7] Harvey, D., Leybourne, S., Newbold, P., *Testing the equality of prediction mean squared errors*, International Journal of Forecasting, Vol. 13, pp. 281-291 1997.

[8] Goh Bee Hua, *Residential construction demand forecasting using economic indicators: a comparative study of artificial neural networks and multiple regression*, Construction Management and Economics, Vol. 14, No. 1, pp. 25-34 1996.

[9] Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F, *forecast: Forecasting functions for time series and linear models*, R package version 8.5, http://pkg.robjhyndman.com/forecast 2019.

[10] Søren Johansen, *Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models*, Econometrica, Vol. 59, No. 6, pp. 1551-1580 1991.

[11] Søren Johansen, *Cointegration: an overview*, Department of Applied Mathematics and Statistics, University of Copenhagen 2004.

[12] Nikolaos Kourentzes, *nnfor: Time Series Forecasting with Neural Networks*, R package version 0.9.6, https://CRAN.R-project.org/package=nnfor 2019.

[13] Helmut Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer, Berlin Heidelberg New York, 2005.

[14] R.S. Mariano and Daniel P.A. Preve, *Statistical Tests for Multiple Forecast Comparison*, Journal of Econometrics, Vol. 169, pp. 123-130 2012.

[15] Saeed Moshiri, Norman Cameron *Neural Network Versus Econometric Models in Forecasting Inflation*, Journal of Forecasting, Vol. 19, pp. 201-217 2000.

[16] Guy Nason, *A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series*, Journal of the Royal Statistical Society, Series B, 75, 879–904, 2013.

[17] OECD Database, *Quarterly GDP*, accessed on 25.02.2019 https://data.oecd.org/gdp/quarterly-gdp.htm 2019.

[18] OECD Database, *Inflation (CPI)*, accessed on 25.02.2019 https://data.oecd.org/price/inflation-cpi.htm 2019.

[19] OECD Database, *Short-term interest rates*, accessed on 25.02.2019 https://data.oecd.org/interest/short-term-interest-rates.htm 2019.

[20] OECD Database, *Unemployment rate*, accessed on 25.02.2019 https://data.oecd.org/unemp/unemployment-rate.htm 2019.

[21] Emrah Önder, Fırat Bayır, Ali Hepşen *Forecasting Macroeconomic Variables Using Artificial Neural Network and Traditional Smoothing Techniques*, International Conference on Economic and Social Studies, 2013.

[22] Keith Ord, Robert Fildes, Nikolaos Kourentzes, *Principles of Business Forecasting 2e*, Wessex Press, New York 2017.

[23] R Core Team, *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria, https://www.R-project.org/ 2018.

[24] Christopher A. Sims, *Macroeconomics and Reality*, Econometrica, Vol. 48, No. 1, pp. 1-48 1980.

[25] James H. Stock, Mark W. Watson, *Vector Autoregressions*, Journal of Economic Perspectives, Vol. 15, No. 4, pp. 101-115 2001.

[26] Norman R. Swanson and Halbert White, *A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks*, The Review of Economics and Statistics, Vol. 79, No. 4, pp. 540-550 1997.

[27] Vardion, *BlankMap-World.png*, modified by Benjamin Drucks, https://commons.wikimedia.org/wiki/File:BlankMap-World.png.

[28] G. Peter Zhang and Douglas M. Kline, *Quarterly Time-Series Forecasting With Neural Networks*, IEEE Transactions on Neural Networks, Vol. 18, No. 6, pp. 1800-1814 2007.