Sosyal Bilimler
Enstitüsü

T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İKTİSAT BİLİM DALI

"FORECASTING GDP GROWTH OF 10 EURO AREA COUNTRIES: LSTM vs. NARX Neural Network"

M. A. Thesis

YIDIRESI ALIYAKEZI

İSTANBUL, 2022

T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İKTİSAT BİLİM DALI

"FORECASTING GDP GROWTH OF 10 EURO AREA COUNTRIES: LSTM vs. NARX Neural Network”

M. A. Thesis

YIDIRESI ALIYAKEZI

Advisor: Prof. Sadullah ÇELİK

ISTANBUL, 2022

ABSTRACT

"FORECASTING GDP GROWTH OF 10 EURO AREA COUNTRIES: LSTM vs. NARX Neural Network"

In this thesis, the GDP growth of 10 Euro Area countries (Austria, Belgium, Spain, Germany, France, Finland, Ireland, Italy, The Netherland, and Portugal) are predicted using two of the latest forecasting methods of artificial neural network algorithm. These two forecasting methods are Long Short-Term Memory (LSTM) network and the Nonlinear Autoregressive Exogenous Inputs (NARX) network models which are the most popular techniques in the recurrent dynamic neural network literature. The forecasting performance of LSTM and NARX is compared with the root mean square error, mean absolute error, and r-square values of the tests. While the LSTM network shows better performance for some countries such as Finland, France, Italy, and, Portugal, the NARX method outperforms LSTM for Austria, Belgium, Germany, Spain, and Ireland. In the Netherland case, both techniques produced the nearly same result. The contribution of each variable to the model prediction is evaluated by Shapley Additive Explanations (SHAP) values. In this thesis, according to the SHAP values, the positive and negative impact of each feature varies for countries and the strength of the impact shows variety. We conclude this result as the economic, geopolitical, and demographic features of countries. Besides, the contribution of the exchange rate to the prediction ranked in the lowest order means that the exchange rate slightly contributes to GDP growth of selected euro area countries. The reason lies in the fact that all EA countries are using Euro as local currency. In the global monetary market, Euro has a strong purchasing power, and this makes EA countries to be independent in currency and not get affected much by exchange rate fluctuations.

Key Words: GDP Growth, Leading Indicators, Neural Network Forecasting, LSTM, NARX, Macroeconomic forecasting, Artificial neural networks; Comparative methods

ÖZET

"10 EURO BÖLGESİ ÜLKELERİNİN GSYİH BÜYÜMESİ TAHMİNİ: LSTM ve NARX Sinir Ağı"

Bu tezde, 10 Euro Bölgesi ülkesinin (Avusturya, Belçika, İspanya, Almanya, Fransa, Finlandiya, İrlanda, İtalya, Hollanda ve Portekiz) GSYİH büyümesi, yapay sinir ağı algoritmasının en güncel tahmin yöntemlerinden ikisi kullanılarak tahmin edilmektedir. Bu iki tahmin yöntemi, tekrarlayan dinamik sinir ağı literatüründe en popüler teknikler olan Uzun Kısa Süreli Bellek (LSTM) ağı ve dışsal girdilere sahip doğrusal olmayan otoregresif sinir ağdır (NARX). LSTM ve NARX'in tahmin performansı, testlerin ortalama hata karesi, ortalama mutlak hata ve r-kare değerleri ile karşılaştırılmıştır. LSTM ağı İspanya, Finlandiya, Fransa, İtalya ve Portekiz gibi bazı ülkeler için daha iyi performans gösterirken, NARX yöntemi Avusturya, Belçika, Almanya, İrlanda için LSTM'den daha iyi performans göstermiştir. Hollanda örneğinde, her iki teknik de hemen hemen aynı sonucu vermiştir. Her bir değişkenin model tahminine katkısı Shapley Additive Explanations (SHAP) değerleri ile değerlendirilmiştir. Bu tezde, SHAP değerlerine göre, her açıklayıcı değişkenin ekonomik büyümeye olumlu ve olumsuz etkisi, ülkelere göre değişmektedir ve etkinin gücü de değişkenlik göstermektedir. Bunun nedeninin ülkelerin ekonomik ve demografik özelliklerinden kaynaklandığı sonucuna varılmıştır. Ayrıca, döviz kurunun etkisi modelde en alt sırada yer almaktadır, buna göre seçilen 10 avro bölgesi ülkelerin GDP büyümesi tahmini için döviz kurunun katkısı çok küçüktür. Bunun nedeni, tüm EA ülkelerinin yerel para birimi olarak Euro kullanmasıdır. Küresel para piyasasında Euro'nun güçlü bir satın alma gücüne sahip olması, EA ülkelerinin para biriminde bağımsız olmalarını ve döviz kuru dalgalanmalarından fazla etkilenmemelerini sağlamıştır.

Anahtar Kelimeler: GSYİH Büyümesi, Öncü Göstergeler, Sinir Ağı Tahmini, LSTM, NARX, Makroekonomik tahmin, Yapay sinir ağları; Karşılaştırmalı yöntemler

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my primary supervisor, Prof. Sadullah ÇELİK, who guided me throughout this research. I also would like to extend my sincere thanks to my defense committee for their invaluable feedback and expertise during the presentation.

I am extremely grateful to my dearest husband Hakan DEMİR for his financial and moral support during this journey. Thanks should also go to my family and friends. Their belief in me has kept my spirits and motivation high during this process.

Lastly, I wish to extend my special thanks to dear Lina BAROKAS, who provided her knowledge and proficiency in neural network forecasting and python.

YIDIRESI ALIYAKEZI

Istanbul, 2022

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN: Artificial Neural Network

AUT: Austria

BEL: Belgium

DEU: Germany

EA: Euro Area

ESP: Spain

FIN: Finland

FRA: France

GDP: Gross Domestic Product

IRL: Ireland

ITA: Italy

LSTM: Long -Short Term Memory

MAE: Mean Absolute Error

NARX: Nonlinear Autoregressive Exogenous Input

NLD: the Netherland

PRT: Portugal

RMSE: Root Mean Square Error

SHAP: Shapley Additive Explanations

1. INTRODUCTION

This study aims to forecast the GDP growth of 10 Euro Area countries using the advanced artificial neural network methodology of Machine Learning. The idea of the topic comes from the necessity of economic forecasting using complex non-linear artificial neural network methods in machine learning. Economic structures are complex and most of the relationships between macro and microeconomics features are non-linear rather than linear. In the latest studies in economics, researchers show the non-linearity of economic variables with neural network models and compare the model performance with the other traditional linear or non-linear regression models. In this thesis, we will use two of the most powerful neural network methods named the Long-Short Term Memory (LSTM) network model and Nonlinear AutoRegressive eXogenous inputs (NARX) model for forecasting GDP values of 10 Euro Area countries.

Forecasting can be defined as a process of estimating events in the near or far future by using data we obtained from the past or present (Hyndman & Athanasopoulos, 2018). Companies, investors, engineers, researchers, etc. need to forecast something they need to know before it happens. As economists, we also need to predict economic phenomena or variables in short-medium and long-term periods. Although the research on macroeconomic forecasting started years ago, the methods and accuracy are could not reach the desired level. However, the latest developments in the forecasting techniques especially the algorithms under the neural networks and random forests have been moving forward the forecasting results to a certain level.

The artificial neural network has many types. Currently, six types of ANN are commonly being used in machine learning. They are Feedforward Neural Network, Radial basis function Neural Network, Kohonen Self Organizing Neural Network, Recurrent Neural Network (RNN), Convolutional Neural Network, and Modular Neural Network. This Study is going to implement the recurrent neural network model with Long Short-Term Memory (LSTM) network, and Nonlinear AutoRegressive eXogenous inputs (NARX) network. Recurrent neural networks, LSTM, and NARX will be introduced in detail in the following sections.

Many economic researchers proved that ANN models have advantages in predicting economic variables in the long run. This study will examine the forecasting power of two machine-learning techniques with ten leading indicators for GDP growth of 10 Euro area countries for the 1990-2020 periods. The rest of the study will be organized as follows: in the next section, deep literature reviews relative to GDP forecasting and artificial neural network applications will be presented. In the third section, the research data will be introduced, and descriptive statistics of the data will be given. In the fourth section, LSTM and NARX architectures will be introduced. In the fifth section, the empirical finding will be discussed, and relative test results will be interpreted. The conclusion and further discussion will be placed in the final sections. After the References section, the Appendix section will include relevant figures, tables, and other outcomes of the test.

Achievements in forecasting output, output gap, potential growth, inflation, and other major macroeconomic variables encourage this study. In this part, before presenting the relative literature review of the study, a short introduction about GDP growth and artificial neural networks will be given.

1.1. Gross Domestic Product (GDP)

Gross Domestic Product (GDP) is defined as the market value of total final goods and services produced within a country’s boundaries in a year by both its citizens and foreigners. GDP growth indicates the annual or quarterly average rate of change of the gross domestic product (GDP) at market prices based on constant local currency, for a given national economy, during a specified period of time. GDP is one of the most crucial macroeconomic variables that help politicians’ economic decisions and political strategies. There are too many factors and variables affecting GDP growth inside and outside of the economy. The determinants of GDP include Inflation, Foreign direct investment, Exchange rate, M2 to the M3 money supply, Unemployment, Wage, The cost of production, Productivity, Consumption, Private and Government investment, and The Balance of imports & export, etc. These variables vary depending on the country's structure, socioeconomic and demographic background, and developmental status.

Policymakers need an on-time and reliable forecast of some macroeconomic variables such as GDP growth, inflation, productivity, and exchange rate to be able to take the right and necessary actions for the monetary and fiscal policy. However, the data publication of some indicators could delay one or two periods, whereas economic policies should be applied one or two periods before the activities happen. Therefore, using immediate data to forecast the next period of activity is important.

GDP growth has been predicted in plenty of ways since macroeconomics was born. The determinants of GDP growth are varying by country. In this study, we take ten Euro Area countries and ten macroeconomic variables as input variables affecting the GDP growth of these countries. The selected ten EA countries are Austria, Belgium, Spain, Germany, Finland, France, Ireland, Italy, the

Netherland, and Portugal. Selected input variables are long-term interest rate, short-term interest rate, unemployment rate, exchange rate, Inflation, government spending, investment, consumer confidence index, current account balance, and a composite leading indicator. As GDP is affected by many other factors, however, due to the difficulties of including all the relevant variables in the model, we fixed the explanatory variables in these ten.

1.2. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a data processing paradigm under Artificial Intelligence technology. It was developed with inspiration coming from the biological neural system such as the human brain. The human brain collects the data, processes it, converts it to useful information, and offers humanity to use it. The Artificial Neural Network is working with the same logic. It collects the data and transforms them into information through various complex processes just as the brain does for us (Dongare et al. 2012).

1.2.1. The Biological Neuron

The biological neuron has three main actors which help the brain to fulfill its duty. The system consists of dendrites, axons, and synapses. A cell body of a neuron receives the signal via dendrites, transmits the signals to other neurons through the axon, and completes the connection by using the synapse. This is a basic process of how brain neurons work (Cavuto, 1997).

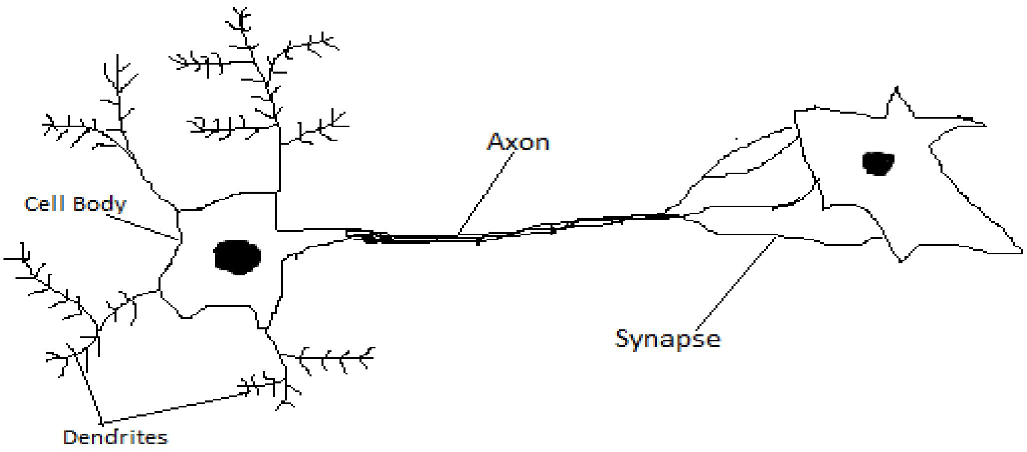


Figure 1.1. Biological Neuron

Source: Cavuto D. J. (1997).

1.2.2. The Artificial Neuron

Artificial Neuron uses the working principle of the brain. As in the biological neuron system, there are three main layers: Input layer, hidden layer, and output layer[1]. The input layer can be thought of as the dendrites of the neuron, the hidden layer represents the axon, and the output layer is like the final signal received by another cell via synapse (Kukreja et al. 2016). With a more basic explanation, we can think of the neural network as a nonlinear regression function of the dependent and independent variables (Chung and Tung, 1995). Due to its successful imitation of the human brain, the artificial neural network is widely used today in weather forecasting, economic forecasting, pattern recognition, and other robotic technologies (Kukreja et al. 2016).

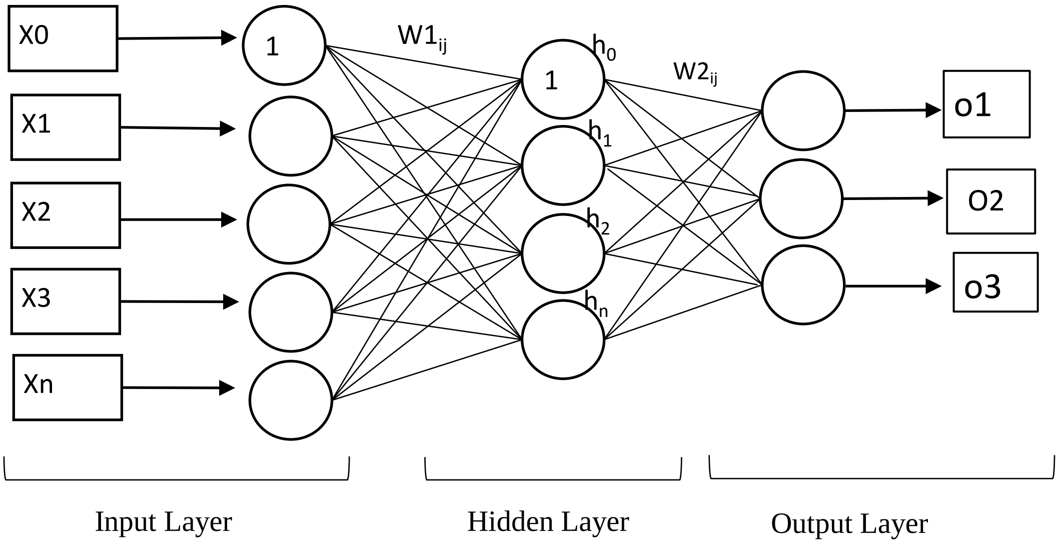


Figure 1.2. Simple Structure of an Artificial Neuron

Source: Drawn by the author.

When an artificial neural network executes the calculation process, the input unit dispatch the signals to the neuron "Hidden Units" by weights, the signals arrived at the hidden units are first gathered in the "sum" unit. (see figure 1.3). In the next step, the sum of the signals will be transformed by a chosen activation function. Lastly, the transformed activation unit is forwarded to the output unit (Swanson and White, 1997).

---

[1] The most simple neural network does not include a hidden layer, it can be thought of as linear regression (Hyndman & Athanasopoulos, 2018).

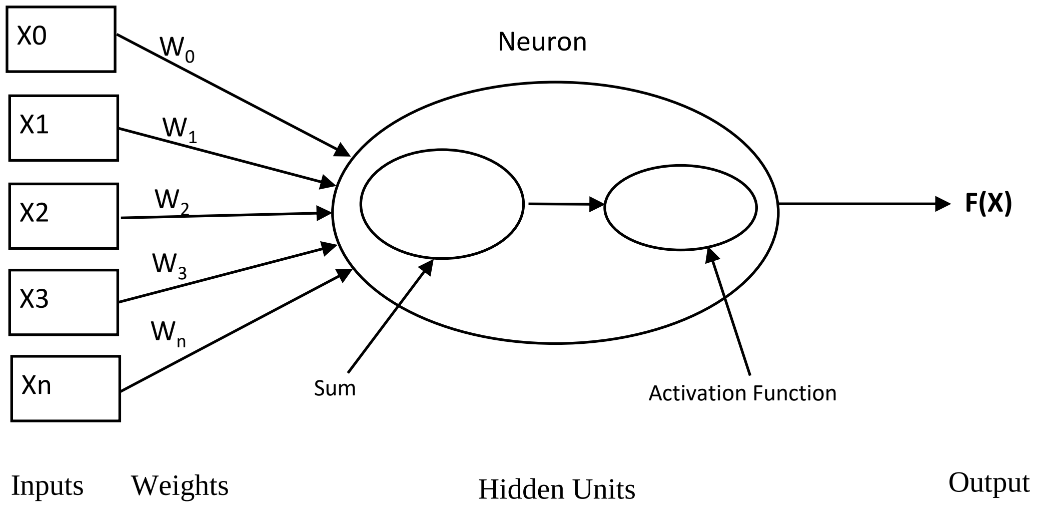


Figure 1.3. Basic Mathimatical Structure of ANN

Source: Drawn by authors.

2. LITERATURE REVIEW

Macroeconomic forecasting plays an important role in policymaking and economic decisions. An accurate and significant prediction can serve valuable information for policymakers and economists even for non-expert agents. Granger and Newbold (2014, p.150) said "*the particular important question to economists is when will the next turning point in the economy occur?*" The GDP Growth and its determents are the leading factors in evaluating the turning point of the economy in the past, current, and future time dimensions. Therefore, researchers have tried to estimate key macroeconomic variables in several ways (see Chung and Tung (1995), Swanson and White (1995), Stock and Watson (1999), Qi. (2001), Heravi et al. (2003), Marijana et al. (2009)). A comprehensive literature review is presented as follows.

According to Maasoumi et al. (1994), although the simple linear models with unit roots are still dominating the econometric modeling literature, a new era of economic and econometric modeling has started thanks to the development of artificial intelligence and machine learning. As a useful output of artificial intelligence, ANN is providing quick and accurate calculations for statistical analysis. Maasoumi et al. analyze US macroeconomic variables such as interest rate, unemployment, S&P 500 index for stock prices, real GDP, and real wages employing ANN. As in other studies, they employed a single hidden layer "feedforward" neural network. One suggestion of the study is the model fit could be improved by raising the number of training periods. Another useful suggestion is a good combination of activation function, lagged values of a dependent variable, and cost functions can give a better fit for the model.

Swanson and White's researches are valuable examples of applications of neural networks. They concentrated on the model selection approach in forecasting macroeconomic time series and used neural networks and linear models simultaneously. In two of their study, they forecasted nine macroeconomic variables with a model selection approach using neural network models. In 1995, they evaluated adaptive models and non-adaptive models in forecasting performance. The adoptive models they mentioned are the artificial neural network models and as the result, adoptive models always outperform nonadoptive models in forecasting the levels and directions of macroeconomic variables. In their study in 1997, they compared flexible, nonflexible, linear, and nonlinear econometric models in real-time. In the study, they focused on model selection and performance. Four important results concluded as follows: first, flexible specification nonlinear models outperform fixed specification linear models in forecasting complicated macroeconomic phenomena varying over time. Second, they propounded the first evidence that more rigid fixed specification models poorly explain time-varying coefficients than flexible specification models. The third contribution is that the study presented the inability of the Schwarz information criterion (SIC) in model selection. They suggested using cross-validation techniques or multiple hidden layers neural network models, especially in out-of-sample model selection criteria. Fourth, they concluded that level models are more suitable for forecasting rather than random walk and ARIMA models.

Chung and Tung (1995) forecasted the exchange rate of the US dollar, British Pound, Canadian Dollar, Deutsche Mark, Japanese Yen, and Swiss franc using two different neural network models. This was a quite initial empirical study of neural networks and economic forecasting. They performed feedforward neural network and recurrent neural network models for 1245 observations. The focus of the study is to investigate whether neural networks can provide superior out-of-sample forecasts to linear models. They found that different neural networks perform differently for their series. The Predictive Stochastic Complexity (PSC) test is a useful tool in selecting suitable neural network models. However, in their conclusion, they indicated that the results of the empirical study provide limited evidence in encouraging the effectiveness of neural network models.

Estrella and Mishkin (1998) examined financial variables as leading indicators of US crises. Financial leading indicators such as interest rate and spread, stock prices, and monetary aggregates are evaluated separately and compared with other financial and non-financial indicators. The analysis examines the out-of-sample performance of leading indicators from one to eight quarters. The result of the research has revealed that stock prices are a good indicator in 1-3 quarters. The study concluded that financial variables could complement macroeconomic models and other predictions by providing them with the ability to control quickly and reliably. This study sheds light on this thesis on choosing variables.

Stock and Watson (1999) forecast the inflation of the US economy for one year. They performed out-of-sample forecasting based on the Philips Curve. They forecasted not only a bivariate model but also a multivariate forecasting model with leading indicators. As a summary of the study, forecasting inflation using the Philips Curve outperforms forecasting using macroeconomic variables such as interest rate, money, and commodity prices.

Pons (2000) evaluated the GDP forecasting accuracy of IMF and OECD for G7 countries. He comments on the accuracy in terms of the biasedness and efficiency of OECD and IMF current-year forecast and year-ahead forecast. He compared the root mean square error (RMSE), mean absolute error, and Thiel’s U of the OECD and IMF forecasts. According to the results, the OECD forecast shows better accuracy than the IMF forecast, and the current-year forecast is greater than the long-term year-ahead forecast. Regarding efficiency, both IMF and OECD forecasts are efficient except in the case of Japan. He also examined the directional accuracy by measuring the x2 test and both organizations exhibit similar accuracy of directions.

Min Qi (2001), in his study, the US great recession was forecasted with the new neural network method. He stated that the business cycle was asymmetric and it could not be predicted accurately with linear constant parameter single-index models, so complex nonlinear models such as the new neural network models could predict the recession more efficiently. In the study, he used four interest rate and spread variables, three stock price indices, eight monetary aggregates, nine individual macro indicators, and three indices of leading indicators as input variables. He tested the relevancy of each variable group with the US recession both by themselves and suitable combinations of the variables. As a result of this research, the interest rate spread was determined as the only indicator that best predicts the US great recession among the 27 variables for the next 2-6 quarters. Rests of the variables improve the result when they are combined with interest rate spread.

Tkacz (2001) forecasted Canadian GDP growth using neural networks. To measure the forecasting performance of the neural network, he applied three-time series models, one linear model, and neural network models with univariate and multivariate bases simultaneously. To see the difference between these three models, he conducted one-quarter and four-quarter tests separately. In the conclusion of his study, in the short run, linear models and time series models show better performance while neural network models capture success in the long run forecasting.

As we included some macroeconomic variables as explanatory variables of GDP growth, it is worth looking at studies related to the relationship between these variables and GDP. Aron and Muellbauer (2002) demonstrate the effect of interest rates on the GDP growth of South Africa. They apply multistep forecasting techniques to forecast the GDP growth of South Africa and draw attention to the effect of interest rate on GDP. In their summary, they indicate that interest rate affects GDP growth with the effect of monetary policy regime and exogenous shocks. We will see how interest rates are related to the GDP of the 10-euro area countries in the empirical finding part of the study.

The impact of macroeconomic variables such as inflation, interest rate, and exchange rate on GDP is examined in the studies of Jilani et al. (2010) and Semuel and Nurina (2015). The result of the study by Jilani et al. shows a significant relationship between inflation, interest rate, and exchange rate with GDP for the economy of Pakistan. The multivariate regression analysis of the study indicates there is a negative relationship between inflation and interest rate with GDP, while the exchange rate has a positive effect on the output. Contrary to the conclusion of the study of Jilani et al. (2010), Semuel and Nurina (2015) found in their study that, inflation, and the interest rate has a significant and positive impact on the GDP of the Indonesian economy, while the exchange rate has no significant effect on GDP.

Allende et al. (2002), compared the forecasting performance of ANN and the traditional statistical time series method. They mentioned the advantages and disadvantages of both the multilayer feedforward ANN and ARIMA time series model. According to the results of their application, ANN performs well in predicting complex datasets. It does not need too large data to train the algorithm and does not require process knowledge in model specification while the statistical method requires much more dataset and probability distribution in the training. . In ANN time series forecasting, it can predict multiple outputs and there is a possibility of developmental design, but the ARIMA model is only able to predict single or few outputs and developmental designs have not been used yet in this method. On the other hand, in some cases the ARIMA forecasting method outperforms ANN. For instance, when the dataset is smaller, and the model just needs to estimate fewer parameters, a statistical method will give a better fit for the estimation. Also, if measures of uncertainty are required, statistical methods should be used. However, when the relationship between the dependent and independent variables is non-linear, there are limited applications in the statistical method and the ANN should be preferred in this case.

Heravi et al. (2003) forecasted European industrial production using a comparative analysis of linear and neural network models. They evaluated the root mean square error of the models and concluded that linear models show more significant performance than neural network models in the one-year horizon, but neural network models outperform linear ones in forecasting the direction of variation as shown in Swanson and White (1995). They emphasized the importance of nonlinear models in predicting macroeconomic variables.

Banerjee et al. (2005) examined the forecasting power of a large set of leading indicators and US macroeconomic variables for the Inflation and GDP growth of the Euro Area by using three different approaches. They believed that the development of the Euro Area is following the development of the US economy and the ECB policy goes parallel with the FED. According to their empirical finding, univariate leading indicator models systematically outperform the ex post autoregression models for both inflation and GDP growth, but the best indicator changes over time.

Besides, they indicate the average best indicators as world GDP and demnad growth, interest rate, public expenditure and labor market variables. Önder et al. (2013) applied a similar method for Turkey's GDP growth and inflation forecasting for 2023 and used exponential smoothing techniques, decomposition method, and neural network method. They argued that the continuous graphs of ANN estimation reflect the real situation and show more logical trends than other estimations.

Marijana et al. (2009), forecast the GDP growth of 27 EU member countries with linear regression models and neural network models. They performed pooled OLS models and fixed-effect models for using them as reference models to be able to evaluate the performance of neural network models. They choose a feedforward neural network model and backpropagation algorithm to train the data. At the end of the research, the investigators suggest that it is useful to combine the linear regression models and neural network models to forecast macroeconomic variables such as GDP growth. Furthermore, their results show that adding financial variables enhances the forecasting performance of the model.

Jahn (2018) predicted 15 developed economies' GDP growth from the 1996 to 2016 period by using the artificial neural network model. In this study, he took three simple linear regression models as a reference model to interpret his ANN regression model. He conducts three different tests on his data and chooses the model which gives the best result. He applied the fixed effect model, dynamic panel model, and pooled OLS model and chose the pooled OLS model to be a reference model for the comparison of the ANN model because the pooled model does not require any conversion of the input data. The author used the Levenberg- Marquardt algorithm to test the performance of ANN. He selected a random 80% of the observation for training the algorithm and used 20% for the validation. From the result of his study, he concludes that the ANN model gives more significant results than the reciprocal linear models and captures the time trends very well. Unlike the linear regression model, the marginal effect of the time variable on the GDP growth rate is non-constant and explains the GDP growth more reasonably. The RMSPE of the ANN forecasting is much smaller than the linear model forecasting.

3. RESEARCH DATA

The data used in this study are collected from OECD and Eurostat databases. The data Period is January 1990- to December 2020. The dependent variable to be forecasted as output is GDP growth for 10 Euro Area countries, which are Austria, Belgium, Germany, Spain, Finland, France, Ireland, Italy, the Netherland, and Portugal. Despite there being 19 countries in the Euro Area, due to the lack of data, nine EA member countries have not been included in the study. The independent variables to be used as input in the model are consumer confidence index (CCI), composite leading indicator (CLI), current account balance (CurrAcc), the exchange rate (ExRate), government spending (Gov_spending), inflation (Inf_CPI), Investment, long-term interest rate (Long_intrate), short-term interest rate (Short_intrate), and unemployment rate (Unem_Rate). These variables have been selected according to the relative literature on GDP forecasting and neural network methodologies.

Figure 3.1 illustrates the annual GDP growth in the selected countries from 1990 to 2020. GDP growth of euro area countries generally moves in the same direction except for Ireland. During the economic crises in the world, western countries were also influenced by the negative effect of the recessions; thereby GDP growth fell in Europe. There was an early 1990s recession during the 1990 and late 1993, 2001 economic recession, the financial crisis of 2007-08 and European debt crises during 2008-2012, and finally, the Covid-19 pandemic caused the economic downturn in the world indeed in Europe.

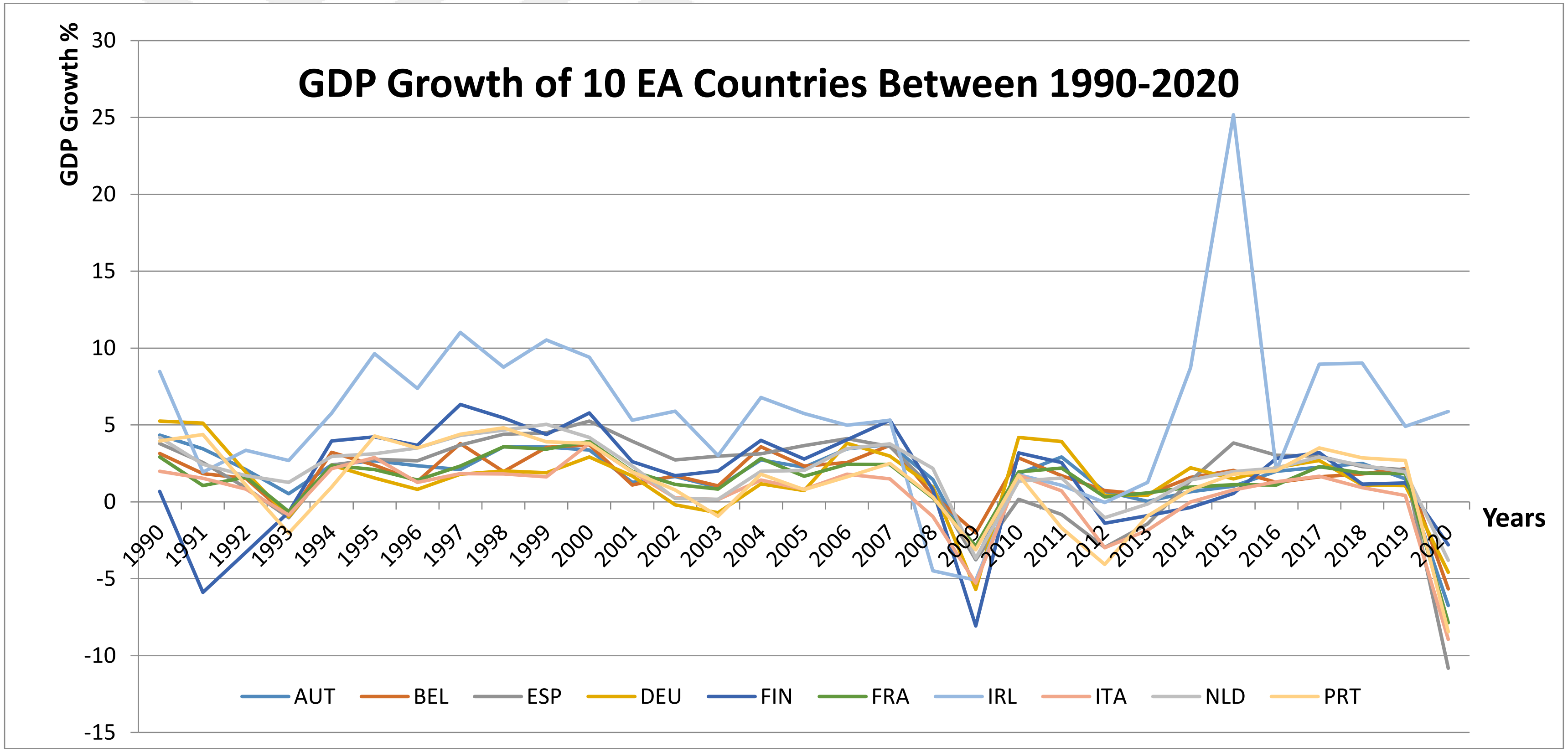
From the graph, we can see the dropdown in the growth rate with the lagged impact of the recession periods, especially in 1993, 2009 and 2020 growth rate dropped dramatically and some countries such as Germany, Finland, Ireland, and Italy experienced negative growth around -5% to -8%. Finland experienced a deep depression from 1991 to 1993 because of some external shocks coming from the Soviet Union and sharp cycles in the OECD area. Another factor that triggered the great recession was the political mistake by the Finland government. The poorly designed financial policies caused a credit crunch and accelerated unemployment. The GDP growth sharply decreased by 13% and the unemployment rate rose from 3% to almost 20% in two years. Other eight countries except for Ireland are also deeply affected by crises even though not as much as Finland.

Unlike other countries, Irish GDP grew by over 5% from 1993 to 2008 and after falling slightly between 2008-2013, it exploded with 25% growth in 2015 and continues to grow despite the covid-19 pandemic. OECD reported that the main reason for the particularly high Irish GDP growth rates lies in the low corporation tax rates employed by the Irish government attracted a large part of multinational corporations and led them to reallocate their investment in both economic activities and intellectual properties to Ireland.

With the covid-19 pandemic, the economy worsened and almost all countries grew negatively. At the beginning of 2020, coronavirus spread widely and western countries including Italy, Germany, and France were severely injured by the virus. Governments struggled with the lockdown and both private and public sectors faced contraction or bankruptcy. As we can see from the figure, all countries except Ireland went into negative growth, and the economy shrunk by more than 10%. IMF described the contraction as the worse since the Great Recession of the 1930s (BBC). During the pandemic recession, unemployment raised along with the financial shrinkage in real sectors, restricted trade, and tourism, raised mortality and labor shortage due to the illnesses, declined tax and increased government expenditure triggered the economic downfall and as a result of these pressures downturns in the GDP growth became inevitable.

As shown in figure 3.1 from late 2019, the GDP growth of selected countries started to fall and came to zero level in early 2020, afterwards the growth rate declined sharply. Within a few months growth rate of Italy, Spain, Belgium, France, and Portugal saw the lowest rate after the great recession in 2009. Austria, Finland, Germany, Ireland, and the Netherland also had a negative growth rate in those years.

Figure 3.1. GDP Growth of 10 EA Countries between 1990-2020



Source: Drawn by author.

3.1. Data Transformation

In the dataset of this study, some variables are originally monthly, and some are collected on an annual and quarterly basis. We transformed the annual and quarterly data to monthly in Eviews by employing the low-frequency to high-frequency quadratic match average data conversion method. We completed the missing values with the mean values of the available series. In table 3.1 we summarize the data with variable names, start and end date of each variable, number of observations in the series, original frequency, and converted frequency of the set.

Table 3.1. Data Information

Variable	Start Date	End Date	# of Observations	Frequency	Converted
GDP Growth	1960-Q2	2021-Q4	108	Quarterly	Monthly
Long-term interest rate	1990-M01	2021-M09	369	Monthly	-
Short-term interest rate	1989-M06	2022-M02	377	Monthly	-
Unemployment Rate	1993-M01	2022-M01	324	Monthly	-
Exchange Rate (EUR/ECU)	1975-M01	2015-M12	480	Monthly	-
Inflation (CPI)	1960-M04	2021-M12	728	Monthly	-
Consumer Confidence Index	1986-M06	2022-M02	426	Monthly	-
Composite Leading Indicator (CLI)	1987-M01	2022-M02	420	Monthly	-
Current Account Balance	2003-Q2	2021-Q3	73	Quarterly	Monthly
Investment	1994-Q1	2021-Q2	109	Quarterly	Monthly
Government Spending	1995	2020	25	Annual	Monthly

In table 2, we describe the data with full terminology, the abbreviation we used during the test, and the source of the variable.

Table 3.2. Explanation of variables

Variable	**Explanation**	**Source**
GDP_growth	Growth rate of gross domestic product	OECD
Long_intrate	Long-term interest rate	OECD
Short_intrate	Short-term interest rate	OECD
Unem_Rate	Unemployment rate	OECD
Exrate	Exchange Rate (EUR/ECU)	Eurostat
Inf_CPI	Inflation measured by consumer price index	OECD
CCI	Consumer confidence index	OECD
CLI	Composite leading indicator	OECD
CurrAcc	Current Account Balance of payment	OECD
Investment	Investment measured by gross fixed capital formation	OECD
Gov_spending	General Government spending	OECD

4. THE MODELS

4.1. Recurrent Neural Network

The recurrent neural network is a supervised machine learning method with one or more feedback loops. The working principle of the recurrent neural network is saving the output of one layer and feeding this output as an input to the next layer. In each loop, each neuron remembers information from the previous layer and saves this information for later use. In this study, we implemented the two most commonly used recurrent neural network models.

4.2. LSTM Architecture

Hochreiter & Schmidhuber introduced a long short-term memory (LSTM) network in 1997. They developed this new recurrent network in conjunction with a suitable gradient-based learning algorithm. LSTM was designed specifically to solve the vanishing gradient and error back-flow problems that other recurrent networks could not address. Additionally, it can apprehend long-time dependencies in the sequence and leads to many successful runs, learns fast, and can handle complex, artificial long-time lags tasks.

A layer of LSTM architecture comprises a set of connected blocks named memory blocks. These blocks contain one or more memory cells which can be thought of as a computational unit and three gates which are the input gate, forget gate, and the output gate. Gates are the weighted functions that manage the information flow in the cell. A memory cell has three weight parameters in the input time steps. Input weights are the weight of the inputs in the current time step. The internal state is for calculating the output for the current time step. An output weight is the weight of output in the last time step.

The basic structure of an LSTM can be expressed as:

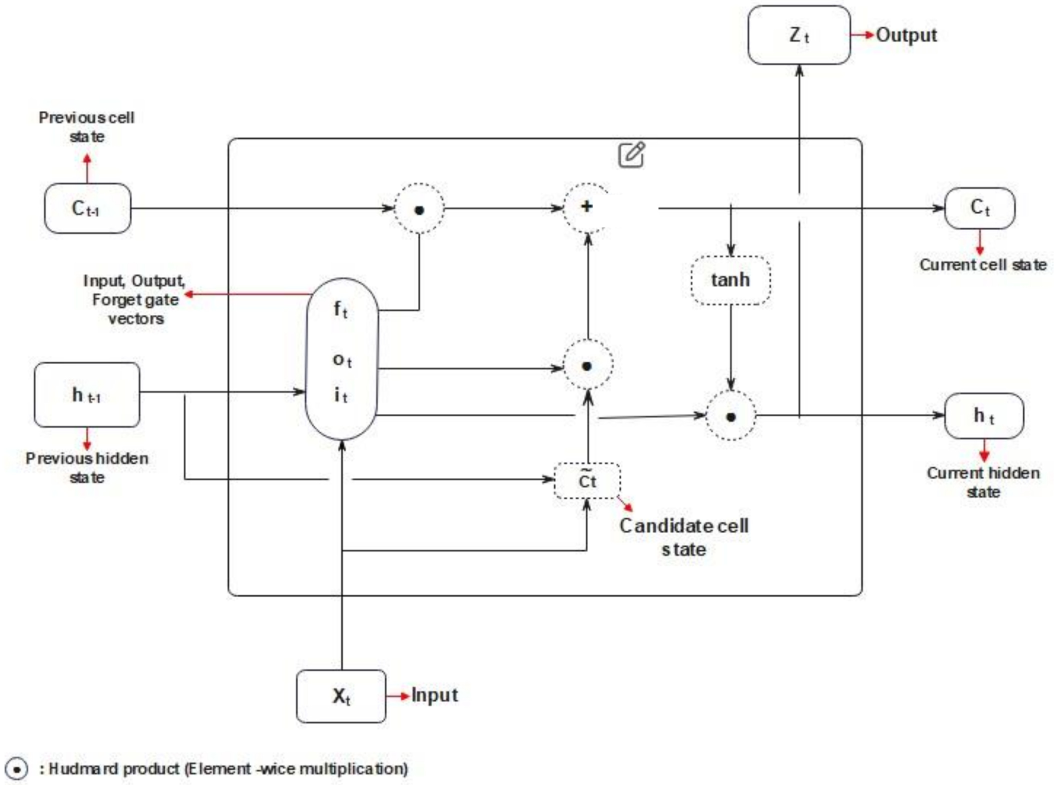


Figure 4.1. General Structure of Long-Short Term Memory Network

Source: Author’s drawing based on Jason Brownlee (2017).

LSTM network has different models to address different forecasting cases. Researchers can choose the best model according to the aim of the study and the structures of the data. The most common types of LSTM architecture are Vanilla LSTMs, Stacked LSTMs, CNN LSTMs, Encoder-Decoder LSTMs, Bidirectional LSTMs, and the Generative LSTMs. Each of these models has a particular structure with layers, input, and output features. In our model, we use Vanilla LSTMs which is the most general form of LSTM network.

4.2.1. Vanilla LSTMs

Vanilla LSTMs have a simple structure rather than other complicated LSTM models. It has the input layer, fully connected hidden layer, and fully connected output layer. It gives successful results in small sequence forecasting problems mostly. Vanilla LSTMs can classify sequence depending on the multiple distributed input time steps, from thousands of time steps, filters the most relevant inputs, and memorize them. Has the ability to predict the next sequence as a function of previous time steps.

The graphical and functional architecture of a Vanilla LSTMs:

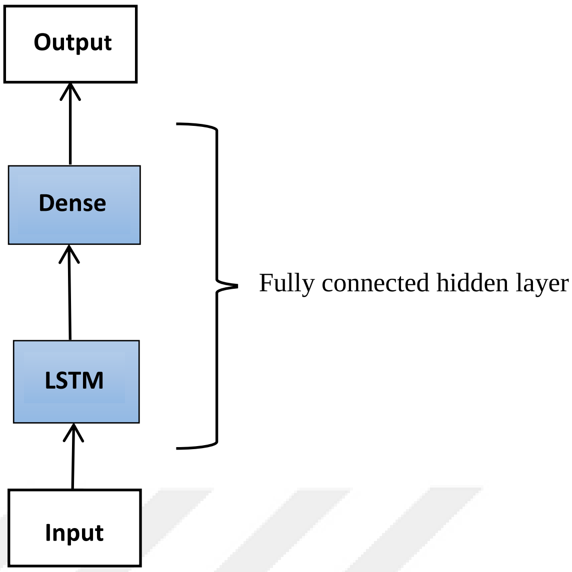


Figure 4.2. Structure of Vanilla LSTMs

Source: Drawn by authors

The Functional State of Vanilla LSTMs:

$$i_t = \sigma(W_i \cdot h_{t-1} + V_i \cdot x_t + P_i \cdot C_{t-1} + b_i) \qquad (a)$$

Where i_t is the input gate at time t, σ is the sigmoid activation function of the input gates with the output range between [0, 1]. W_i is the weight matrices of the input gates and multiplied with the previous hidden state of the cell h_{t-1}. V_i Expresses the weight matrices of the current input of the cell and multiplied by the input x_t at time step t. P_i represents the weight matrices of input and multiplied by the previous cell state C_{t-1}. b_i is the bias vector.

$$o_t = \sigma(W_0 \cdot h_{t-1} + V_0 \cdot x_t + P_0 \cdot C_t + b_0) \qquad (b)$$

Where o_t describes the output gate vectors. In this layer, the previous hidden state h_{t-1} is multiplied by the weight matrices of the output gate. V_0 and P_0 denote the weight matrices of the current input and output gates, respectively. The weight matrices of the output gates multiplied by the current cell state C_t rather than the previous one as in the function (a) and bias vectors b_0 are added.

$$f_t = \sigma\left(W_f \cdot h_{t-1} + V_f \cdot x_t + p_f \cdot C_{t-1} + b_f\right) \qquad (c)$$

f_t represents the forget gate vectors. Forget gate decides which data from the previous step will be used in the next step and which will be forgotten in this gate. The results from the forget gate are between 0 and 1. “0” means the data have been forgotten and “1” means the data have been selected for use in the next step. Forget gate can be thinking as a filtration system of the LSTM layers. Other parameters are namely same as the other parameters mentioned above and are only performing in the forget gate.

$$\widetilde{C_t} = \tanh(W_c \cdot h_{t-1} + V_c \cdot x_t + b_c) \tag{d}$$

The candidate cell state $\widetilde{C_t}$ combines the filtered information that comes from the forget gate with the input gate to set up the new cell state. This cell state passes through a hyperbolic tangent activation function and resulted in the value between (-1,1).

$$C_t = i_t \odot \widetilde{C_t} + f_t \odot C_{t-1} \tag{e}$$

$\mathrm{C_t}$ is the current cell state (Long term memory) and $\mathrm{C_{t-1}}$ is the previous cell state. Cell state takes updated information from the input gate in each time step and gives the output of the hidden state. This hidden state output diversifies into an output for the current time step and a hidden state for the next time step. The cell state is renewed in each time step as the feature vector of the current input.

$$h_t = o_t \odot \tanh(C_t) \tag{f}$$

$$Z_t = h_t \tag{h}$$

Here h_t is the output (Short term memory) of the memory cell at time t which goes to both the output layer and hidden layer in the next time step. This gate aims to separate the final memory from the hidden state. C_t includes lots of information that are not useful to be kept in a hidden state. Therefore, the output gate decides which information coming from C_t will be forwarded to the hidden state h_t with the point-wise multiplication of o_t and tanh of the C_t.

In the LSTM network, the output Z_t is equal to the hidden state h_t .

4.3. NARX Neural Network

The nonlinear AutoRegressive exogenous inputs (NARX) recurrent neural network model was implemented by Lin et al. (1996) to address the long-term dependencies problem in gradient-descent learning. The most important feature of NARX is that it uses tapped delay feedback for the output and by doing this; it outperforms other neural network architectures with only single or without output delays. NARX network is less sensitive to vanishing gradient and long-term dependencies problems and able to predict nonlinear time series much more efficiently and robustly than a conventional recurrent neural network Siegelmann et al. (1997).

The algebraic state of NARX is:

$$y(t) = f\left[u(t - D_u), \dots u(t - 1), u(t), y\left(t - D_y\right), \dots, y(t - 1)\right] + \ \varepsilon_t \tag{1}$$

Where $u(t)$ and $y(t)$ are the input and output of the network at time t, respectively. D_u and D_y represent the input and output order. ε_t is the error term between the actual and predicted values.

According to the input variable (t), the hidden layer output at time t can be obtained as:

$$H_i(t) = f_1\left[\sum_{r=0}^{D_u} w_{ir} u(t - r) + \sum_{l=1}^{D_y} w_{li} y(t - l) + a_i\right] \tag{2}$$

Where w_{ir} is the connection weight between the input neuron $u(t-r)$ and i^{th} hidden neuron. w_{li} is the connection weight between the ith hidden neuron and output feedback neuron $y(t-l)$. a_i is the bias of the ith hidden neuron and f_1 is the hidden layer activation function.

The final estimation can be calculated with the combination of hidden layer output:

$$\widehat{y_j}(t) = f_2\left[\sum_{i=1}^{n_h} w_{ji} H_i(t) + b_j\right] \tag{3}$$

Where w_{ji} is the connection weight between the ith hidden neuron and jth predicted output n_h. b_j is the basis of the jth predicted output. n_h is the number of hidden neurons and f_2 is the output layer activation function.

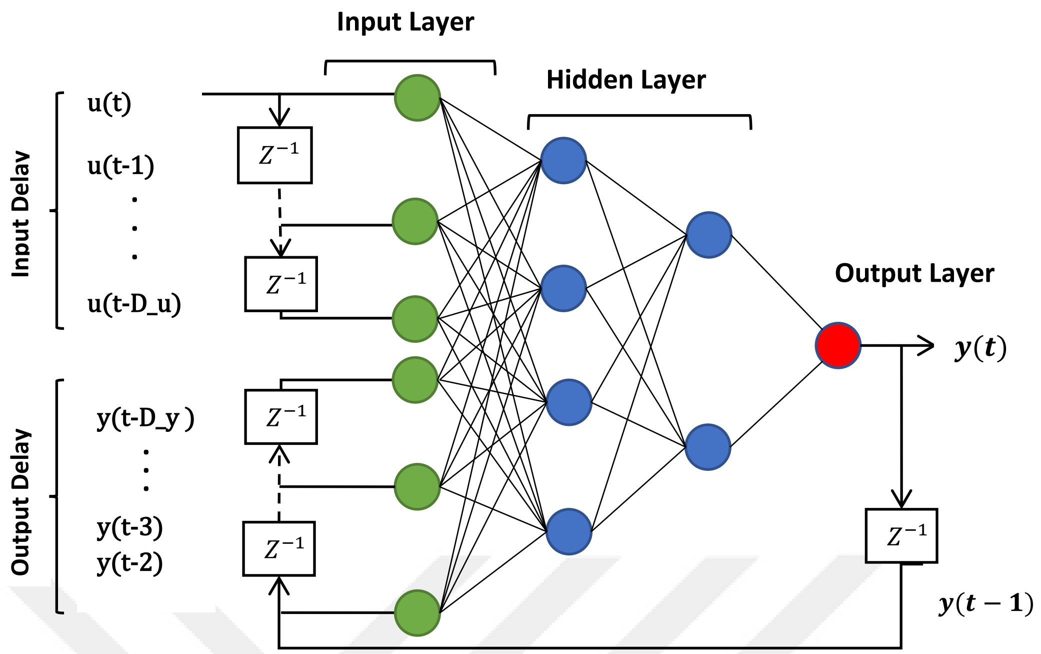


Figure 4.3. General NARX Neural Network Architecture

Source: Drawn by the Author.

4.4. Implementation of the Models

To implement the LSTM and NARX neural network to the data set, we used the 4.2.5 version of Spyder deep learning framework of Python programming language. In this section, we will introduce the implementation process, data preparation, batch size, epoch size, and other associated manipulations applied in this study.

4.4.1. Data Preprocessing

- Min-Max Feature Scaling

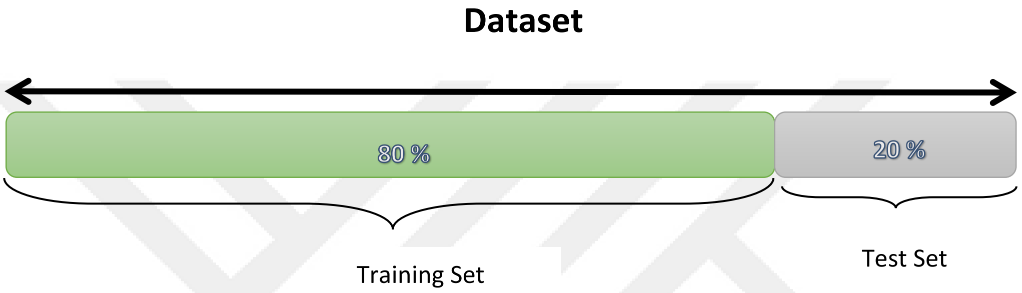
Before running the program, we normalize the data using min-max feature scaling. Since the variables are measured at different scales, they will not contribute to the model equally and might cause a bias problem. Therefore, a suitable normalization makes the training process more efficient and robust. In the min-max scaling method, all input and output variables are scaled to the range [0,1] which means that the new minimum and maximum values of a variable are going to be 0 and 1, respectively.

The formulation of Min-Max Scaler:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (i),$$

- Dataset Splitting

According to the estimation logic of machine learning, the algorithm learns from a part of data and trains itself to give an estimation, and then uses the rest of the part of data to test the accuracy of the performance. We split 80% of the dataset as a train set and 20% of it as a test set to evaluate the forecasting performance of the algorithm.



Source: Drawn by the author.

4.4.2. Hyper-parameter Setting

Hyper-parameters are different from the model parameters. “Hyper” refers to the top-level here and hyper-parameters are the parameters that control the learning process of the algorithm. Defining the optimal size of hyper-parameters is crucial to the forecasting result. The algorithm uses the value of the hyper-parameters and gives a direction to the training procedure. Our LSTM and NARX neural network require an optimal batch size, an optimal number of epochs, the value of learning rate, and the number of hidden layers. Below are the short definitions and values we used in the training.

- Batch Size

Batch size is the number of training samples that will be propagated in each epoch of the network. The most preferred batch size is 32 in general. It means that the algorithm takes the first 32 samples in the training set and trains them in the first round of the learning process. After this training, the algorithm will take the next 32 samples and train them accordingly. This loop will be repeated until all samples in the training set run out.

The batch size can take the value between 1 and the total dataset, however, in our model, we defined 3 different batch sizes to see the best result from the training. We have 360 observations in total and 80% of them, which corresponds to 288 samples, are used for training. Our batch size is 16, 32, and 64 separately. The algorithm will run the program according to these three batch sizes and we will see three forecasting results from each learning in the empirical finding section. We aimed to decide the best batch size of our sample data by doing this separation.

- Number of Epochs

The number of epochs determines the number of times that algorithm passes through the entire training dataset. An epoch is made by one or many batches. The learning algorithm runs as many times as the number of epochs and continuous running until the error terms have been decreased to the minimum level. According to the size of our observations, we defined the number of epochs as 200. It means the entire training set will loop 200 times with the samples, which is defined by the batch size. During the testing, we tried epochs with 100 and 200 to see the difference between them. In the end, the best outcome has been observed with the 200 numbers of epochs.

- Learning Rate λ

The learning rate decides how fast the model learns the problem. We determined four learning rates for the training. They are 0.0001, 0.0003, 0.001, and 0.003 respectively. A lower learning rate needs more training epochs while a higher learning rate requires fewer epochs. We used both low and high learning rates to evaluate the response of the model to the corresponding learning rate with 200 epochs.

4.4.3. Hidden Layers & Dense Layer

The hidden layer stands between the input and output layers as shown in figures 1.2 and figure 1.3. In the hidden layer, there is a set of mathematical operations conducted which use the activation function and transmit the weighted input to the output layer. In this study, the LSTM and NARX hidden layers run with 50 memory cells.

5. EMPIRICAL FINDINGS

We obtained 12 combinations concerning four learning rates and three epoch sizes for each country. The calculation results of LSTMs and NARX have been discussed in this section according to the root mean square error (RMSE), the mean absolute error (MAE), and the R-square values of countries shown in table 5.1 and table 5.2. The distribution of the estimated loss of two models has been illustrated in figures A.1-A.20 in the Appendix section.

As a whole, the LSTM network outperforms NARX for Finland, France, Italy, and Portugal while NARX captures better accuracy for Austria, Belgium, Germany, Spain, and Ireland. Both models give similar estimation results for the Netherland. Specifically, For Austria, the best prediction result is obtained by employing 16 batch size and 0.001 learning rate in NARX, where RMSE and MAE are predicted as 0.0044 and 0.0293 respectively with 0.7582 R-square. For Belgium, the best prediction result is also observed by employing 16 batch size and 0.001 learning rate in NARX, where RMSE and MAE are predicted as 0.0004 and 0.0144 respectively with 0.8956 R-square. For Germany, the best prediction accuracy is obtained by employing 16 batch sizes and a 0.003 learning rate in NARX, where RMSE and MAE are predicted as 0.0005 and 0.0173 respectively with 0.8845 R-square. For Spain, the highest prediction accuracy is obtained by employing 32 batch sizes and a 0.001 learning rate in NARX, where RMSE and MAE are predicted as 0.0010 and 0.0238 respectively with 0.6416 R-square. We did not encounter any abnormal distribution or outlier in Spain’s dataset. However, we cannot say the estimation result of both LSTM and NARX are successful in the model because 0.6416 is not a desirable value of R-square. The reason behind this could be the unique economic infrastructure of Spain’s economy. For Spain economic growth, different analyses may be conducted, or the model could be improved by including more explanatory variables which better reflect the structure of Spain’s economy. For Finland, the best prediction accuracy is obtained by employing 32 batch sizes and a 0.001 learning rate in LSTM, where RMSE and MAE are predicted as 0.0002 and 0.0105 respectively with 0.9432 R-square. We obtained the highest estimation for France in LSTM by employing 16 batch sizes and a 0.0003 learning rate, where RMSE and MAE are predicted as 0.000517 and 0.017 respectively with 0.9706 R-square. For Ireland, the best prediction result is observed by employing 32 batch sizes and a 0.003 learning rate in NARX, where RMSE and MAE are predicted as 0.0012 and 0.0192 respectively with 0.8968 R-square. For Italy, the best accuracy occurred in LSTM with 16 batch size and 0.001 learning rate, where RMSE and MAE are predicted as 0.0005 and 0.0171 respectively with 0.9159 R-square.

For the Netherland, both LSTM and NARX showed similar performance when looking at the R-square values, which are 0.88904 and 0.8818 respectively. However, the corresponding batch sizes and learning rates differ from the corresponding RMSE and MAE. In LSTM, the best result occurs in 64 batch sizes and 0.001 learning rate with 0.0005 RMSE and 0.0143 MAE. In NARX, the best result occurs in 16 batch sizes and 0.0003 learning rates with 0.0005 RMSE and 0.0133 MAE. Finally, for Portugal, LSTM outperforms NARX by employing 64 batch sizes and a 0.003 learning rate, where RMSE and MAE are predicted as 0.0071 and 0.0351 respectively with 0.8206 R-square.

Table 5.1. The Estimation Results of LSTM network

		Australia			Belgium			Germany			Spain			Finland		
Batch Size	**λ**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**MSE**	**MAE**	**R2**	**MSE**	**MAE**	**R2**
16	0.0001	0.0093	0.0733	0.4828	0.0061	0.0680	-0.6871	0.0036	0.0415	0.1854	0.0025	0.0409	0.1502	0.0021	0.0382	0.2896
16	0.0003	0.0060	0.0436	0.6665	0.0005	0.0160	0.8641	0.0011	0.0260	0.7583	0.0023	0.0398	0.2067	0.0005	0.0199	0.8208
16	0.001	**0.0045**	**0.0284**	**0.7521**	0.0074	0.0738	-1.0396	0.0009	0.0224	0.7927	0.0019	0.0355	0.3569	0.0005	0.0188	0.8413
16	0.003	0.0051	0.0398	0.7156	0.0040	0.0494	-0.1141	0.0009	0.0208	0.7936	0.0116	0.0979	-3.0120	0.0011	0.0285	0.6389
32	0.0001	0.0184	0.1054	-0.0162	0.0076	0.0810	-1.0856	0.0067	0.0615	-0.5270	0.0036	0.0492	-0.2443	0.0028	0.0448	0.0638
32	0.0003	0.0093	0.0663	0.4863	0.0040	0.0565	-0.1175	0.0010	0.0219	0.7784	0.0018	0.0327	0.3840	0.0014	0.0315	0.5472
32	0.001	0.0049	0.0350	0.7285	0.0010	0.0227	0.7206	0.0008	0.0229	0.8074	0.0020	0.0403	0.3006	**0.0002**	**0.0105**	**0.9432**
32	0.003	0.0051	0.0372	0.7201	0.0012	0.0289	0.6646	0.0009	0.0239	0.7867	0.0018	0.0345	0.3962	0.0003	0.0124	0.9143
64	0.0001	0.0200	0.1114	-0.1042	0.0017	0.0328	0.5204	0.0094	0.0707	-1.1399	0.0045	0.0547	-0.5669	0.0076	0.0811	-1.5393
64	0.0003	0.0224	0.1124	-0.2389	0.0138	0.1061	-2.8010	0.0044	0.0462	-0.0115	0.0039	0.0510	-0.3292	0.0017	0.0340	0.4275
64	0.001	0.0065	0.0465	0.6419	**0.0004**	**0.0153**	**0.8778**	0.0011	0.0290	0.7396	0.0050	0.0604	-0.7138	0.0003	0.0151	0.8951
64	0.003	0.0053	0.0429	0.7062	0.0006	0.0172	0.8403	**0.0008**	**0.0235**	**0.8257**	**0.0015**	**0.0348**	**0.4683**	0.0009	0.0282	0.7155

		France			Ireland			Italy			The Netherland			Portugal		
Batch Size	**λ**	**MSE**	**MAE**	**R2**	**MSE**	**MAE**	**R2**	**MSE**	**MAE**	**R2**	**MSE**	**MAE**	**R2**	**MSE**	**MAE**	**R2**
16	0.0001	0.0067	0.0585	0.62133	0.0059	0.0680	0.5072	0.0022	0.0375	0.6095	0.0020	0.033	0.5762	0.0247	0.1178	0.3754
16	0.0003	**0.0005**	**0.0167**	**0.97063**	0.0024	0.0377	0.7977	0.0010	0.0248	0.8268	0.000668	0.019	0.855	0.0099	0.0518	0.7502
16	0.001	0.0019	0.0314	0.89249	0.0055	0.0650	0.5479	**0.0005**	**0.0171**	**0.9159**	0.004164	0.063	0.0968	0.0084	0.0519	0.7877
16	0.003	0.0021	0.0313	0.87989	0.0050	0.0577	0.5823	0.0057	0.0706	0.0034	0.004383	0.06	0.0492	0.0105	0.0764	0.7342
32	0.0001	0.0055	0.0608	0.68755	0.0105	0.0928	0.1289	0.0115	0.0960	-1.014	0.0040	0.057	0.1239	0.0399	0.1596	-0.006
32	0.0003	0.0036	0.0453	0.79707	0.0046	0.0600	0.6174	0.0012	0.0259	0.79	0.001588	0.034	0.6556	0.0139	0.0833	0.6497
32	0.001	0.0008	0.0219	0.95258	0.0021	0.0353	0.8286	0.0012	0.0266	0.7944	0.000586	0.013	0.8729	0.0095	0.0506	0.7595
32	0.003	0.0028	0.0357	0.84142	0.0029	0.0431	0.7558	0.0007	0.0203	0.8762	0.000684	0.015	0.8516	0.0078	0.0405	0.8026
64	0.0001	0.0147	0.0947	0.16509	0.0147	0.1115	-0.218	0.0068	0.0761	-0.189	0.006537	0.075	-0.418	0.0371	0.1495	0.0639
64	0.0003	0.0093	0.0741	0.47039	0.0075	0.0821	0.3784	0.0024	0.0342	0.5894	0.001173	0.028	0.7455	0.0282	0.1326	0.2873
64	0.001	0.0013	0.0277	0.92604	**0.0020**	**0.0334**	**0.8363**	0.0014	0.0313	0.7502	**0.000512**	**0.014**	**0.889**	0.0098	0.0583	0.7526
64	0.003	0.0008	0.0227	0.95479	0.0021	0.0359	0.8269	0.0014	0.0319	0.7589	0.001186	0.024	0.7428	**0.0071**	**0.0351**	**0.8206**

Table 5.2. The Estimation Result of NARX Network

		Australia			Belgium			Germany			Spain			Finland		
Batch Size	**λ**	**RMSE**	**MAE**	**R^2**	**RMSE**	**MAE**	**R^2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R^2**	**RMSE**	**MAE**	**R^2**
16	0.0001	0.0077	0.0515	0.5726	0.0021	0.0338	0.4210	0.0061	0.0435	-0.3870	0.0038	0.0477	-0.3140	0.0015	0.0308	0.4872
16	0.0003	0.0058	0.0297	0.6802	0.0013	0.0230	0.6480	0.0013	0.0280	0.7114	0.0020	0.0336	0.2967	0.0006	0.0184	0.7883
16	0.001	**0.0044**	**0.0293**	**0.7582**	**0.0004**	**0.0144**	**0.8956**	0.0014	0.0293	0.6742	0.0016	0.0302	0.4623	0.0003	0.0125	0.9062
16	0.003	0.0114	0.0842	0.3717	0.0044	0.0592	-0.2209	**0.0005**	**0.0173**	**0.8845**	0.0012	0.0277	0.5902	0.0023	0.0433	0.2456
32	0.0001	0.0145	0.0919	0.1988	0.0041	0.0546	-0.1452	0.0025	0.0340	0.4268	0.0059	0.0669	-1.0422	0.0048	0.0555	-0.6116
32	0.0003	0.0080	0.0616	0.5560	0.0008	0.0186	0.7927	0.0037	0.0437	0.1520	0.0029	0.0383	-0.0064	0.0006	0.0205	0.8147
32	0.001	0.0049	0.0287	0.7282	0.0008	0.0182	0.7705	0.0019	0.0355	0.5735	**0.0010**	**0.0238**	**0.6416**	0.0005	0.0160	0.8474
32	0.003	0.0073	0.0546	0.5964	0.0007	0.0168	0.8031	0.0024	0.0466	0.4559	0.0010	0.0276	0.6398	**0.0003**	**0.0141**	**0.9063**
64	0.0001	0.0355	0.1427	-0.9664	0.0189	0.1154	-4.2072	0.0159	0.0830	-2.6255	0.0077	0.0654	-1.6674	0.0060	0.0597	-1.0092
64	0.0003	0.0064	0.0506	0.6437	0.0218	0.1223	-5.0170	0.0112	0.0800	-1.5521	0.0114	0.0883	-2.9387	0.0032	0.0396	-0.0694
64	0.001	0.0182	0.0970	-0.0052	0.0006	0.0158	0.8391	0.0005	0.0158	0.8752	0.0015	0.0300	0.4948	0.0008	0.0212	0.7448
64	0.003	0.0053	0.0225	0.7072	0.0019	0.0240	0.4771	0.0011	0.0249	0.7577	0.0014	0.0219	0.5317	0.0003	0.0124	0.9046

		France			Ireland			Italy			The Netherland			Portugal		
Batch Size	**λ**	**RMSE**	MAE	**R^2**	**RMSE**	MAE	R2	**RMSE**	MAE	**R^2**	**RMSE**	MAE	**R^2**	**RMSE**	MAE	**R^2**
16	0.0001	0.0145	0.0865	0.1747	0.0056	0.0513	0.5319	0.0089	0.0769	-0.5579	0.0061	0.0672	-0.3242	0.0157	0.0849	0.6049
16	0.0003	0.0039	0.0329	0.7805	0.0036	0.0355	0.6999	0.0117	0.0576	-1.0378	**0.0005**	**0.0133**	**0.8818**	0.0101	0.0493	0.7464
16	0.001	0.0028	0.0461	0.8426	0.0049	0.0641	0.5967	0.0009	0.0216	0.8407	0.0015	0.0178	0.6659	0.0102	0.0466	0.7425
16	0.003	0.0012	0.0230	0.9303	0.0068	0.0699	0.4387	0.0009	0.0251	0.8516	0.0029	0.0452	0.3784	0.0097	0.0575	0.7540
32	0.0001	0.0282	0.1206	-0.6000	0.0034	0.0479	0.7144	0.0066	0.0566	-0.1591	0.0031	0.0499	0.3250	0.0277	0.1119	0.2998
32	0.0003	0.0057	0.0518	0.6749	0.0204	0.1154	-0.6913	0.0037	0.0402	0.3607	0.0017	0.0301	0.6221	0.0109	0.0759	0.7240
32	0.001	0.0014	0.0210	0.9181	0.0040	0.0409	0.6707	0.0030	0.0335	0.4798	0.0290	0.1666	-5.3007	0.0087	0.0404	0.7797
32	0.003	**0.0009**	**0.0219**	**0.9485**	**0.0012**	**0.0192**	**0.8968**	**0.0006**	**0.0175**	**0.8967**	0.0007	0.0147	0.8480	0.0096	0.0600	0.7572
64	0.0001	0.0370	0.1788	-1.1031	0.0719	0.2148	-4.9599	0.0118	0.0790	-1.0594	0.0040	0.0405	0.1274	0.0340	0.1514	0.1412
64	0.0003	0.0104	0.0891	0.4119	0.0134	0.0728	-0.1078	0.0080	0.0748	-0.4032	0.0019	0.0273	0.5958	0.0242	0.1192	0.3887
64	0.001	0.0052	0.0506	0.7055	0.0093	0.0752	0.2287	0.0024	0.0312	0.5798	0.0008	0.0129	0.8370	**0.0076**	**0.0312**	**0.8077**
64	0.003	0.0012	0.0219	0.9303	0.0018	0.0280	0.8521	0.0014	0.0223	0.7566	0.0008	0.0152	0.8343	0.0089	0.0586	0.7759

Source: Estimated by the author.

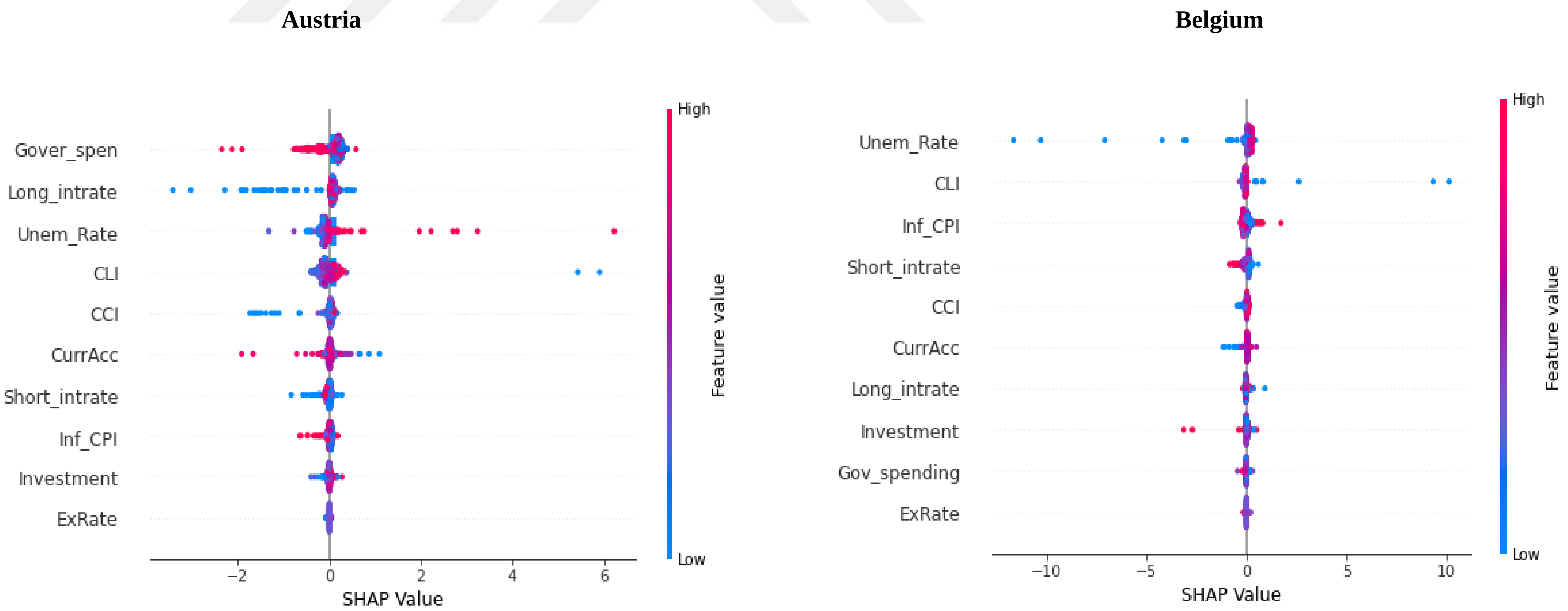
In this study, estimation of the feature importance is also included in the model. Feature importance indicates how a variable contributes to the model estimation either negatively or positively. Feature importance is measured by Shapley Additive exPlanations (SHAP) value, which is developed by Lundberg and Lee (2017). A positive SHAP value represents a positive impact and a negative SHAP value represents the negative impact of a variable on the prediction. SHAP value gives us not only the strength of the contribution but also gives the direction of the impact. The estimation results are shown in figure 5.3 and figure A.21 – A.30 in the Appendix section.

Figure 5.3 depicts the directional impact of each variable on the prediction by country. The X-axis is the SHAP value and Y-axis has all the variables. Each dot on the figure represents one SHAP value for the prediction and variable. Blue dots mean a lower value and pink dots indicate a higher value of a variable. As a whole, the distribution of the blue and pink dots gives us the directional impacts of a feature. For Austria`s GDP growth, government spending and long-term interest rate have the highest impacts and the exchange rate has the smallest contribution to the prediction. Moreover, government spending, current account, and inflation are negatively correlated with GDP growth as the pink dots are distributed on the negative side of the X-axis. The unemployment rate contributes positively to the prediction. The distribution of long-term interest rate, consumer confidence index, and short-term interest rate have negatively low impacts on the GDP growth prediction, but the strengthens of the long-term interest rate is higher than the other two variables. Rests of the variables have a smaller contribution to the prediction because the distribution of dots of them piled up around the zero value of the x-axis. For Belgium`s GDP growth, the unemployment rate has a negative but strong impact on the prediction while composite leading indicators have a strong and positive contribution to the prediction. Other variables seemed to contribute less to the estimation because of the accumulated distribution of the dots around zero value.

For Germany`s GDP growth, the composite leading indicator, the short-term interest rate, and unemployment rate have a negative impact on the model, and inflation, government spending, and long-term interest rate have a positive impact. Besides, investment and current account balance are highly and negatively correlated with GDP growth. The rest of the variables has a small impact on the prediction. For Spain`s GDP growth, consumer confidence index, long-term, and short-term interest rates are the most important features and correlated negatively to the model. For Finland`s GDP growth, the unemployment rate is the most critical feature for the prediction, and this impact is followed by the consumer confidence index and current account. For France's GDP growth, the most critical features for the prediction are investment and inflation. Long-term interest rate and unemployment rate negatively correlated to France`s GDP growth. Short-term interest rate, current account, composite leading indicators, government spending, and exchange rate have a relatively small impact on the prediction. In Ireland's GDP growth prediction, the current account plays the most critical role with a highly and positively correlated distribution. The investment and consumer confidence index follows the current account also with a high and positive impact. The long-term interest rate seems negatively attributed to the prediction. The rest of the variables contribute less despite the intensities nodding around the zero value of the x-axis. In the estimation of Italy`s GDP growth, investment ranked at the top with a high but negative impact. Composite leading indicators ranked in second order with a low but positive impact on the prediction. The rest of the variables do not seem to have a noteworthy impact however we obtained a high R-square score in both LSTM and NARX estimation for Italy. The Netherlands’ GDP growth, consumer confidence index, inflation, short-term interest rate, unemployment, and current account balance contribute to the prediction of high ranks. Composite leading indicators, long-term interest rate, government spending, investment, and exchange rate contribute less to the estimation. Finally, for Portugal`s GDP growth, long-term interest rate, consumer confidence index, and short- term interest rate, and composite leading indicator take place on the top which as the most relevant features for the prediction, the rest of the variables contribute to the estimation slightly.

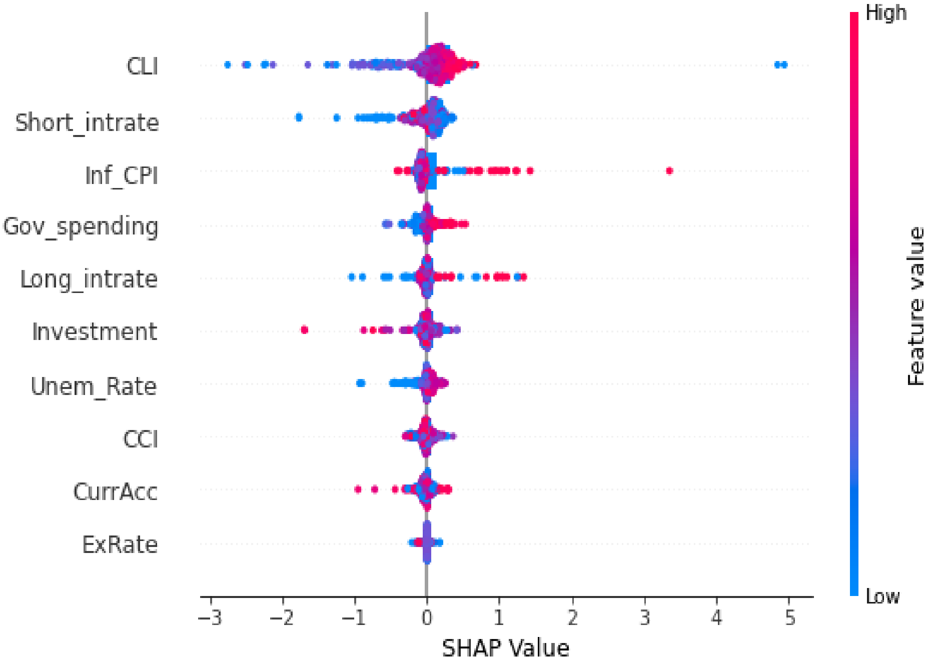
The features are ordered by how much the model prediction is influenced by them in Figure A.23-A.30 in the appendix section. The x-axis represents the average of the absolute SHAP value of each feature. For example, for Austria`s GDP prediction, government spending is the most important feature and is followed by long-term interest rate, unemployment, composite leading indicators, consumer confidence index, and so on. The importance of the variables is varying for each country in our study, this may be caused by the differences in the demographic and economic infrastructure of the countries. Despite the distinction in results, we observed some general outcomes from the prediction. For eight countries, the exchange rate takes position in the 10th order, and for two countries, it states in the 7th order. We evaluated this as all the countries are using the common currency unit Euro in the area, therefore the exchange rate indexed with their local currencies has a very small impact on the GDP prediction.

Figure 5.1: The Feature Importance for EA Countries

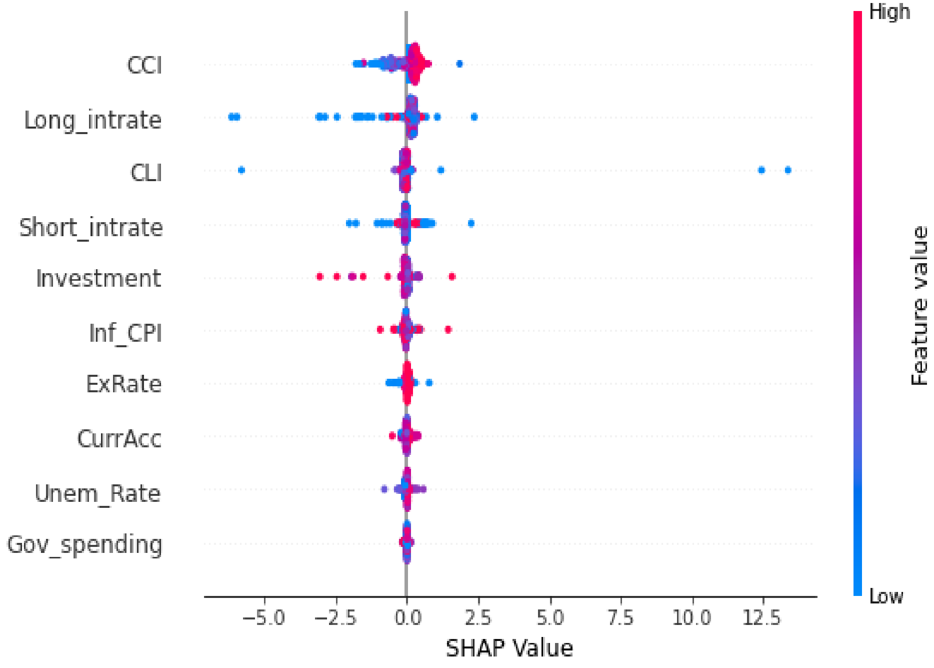

Austria

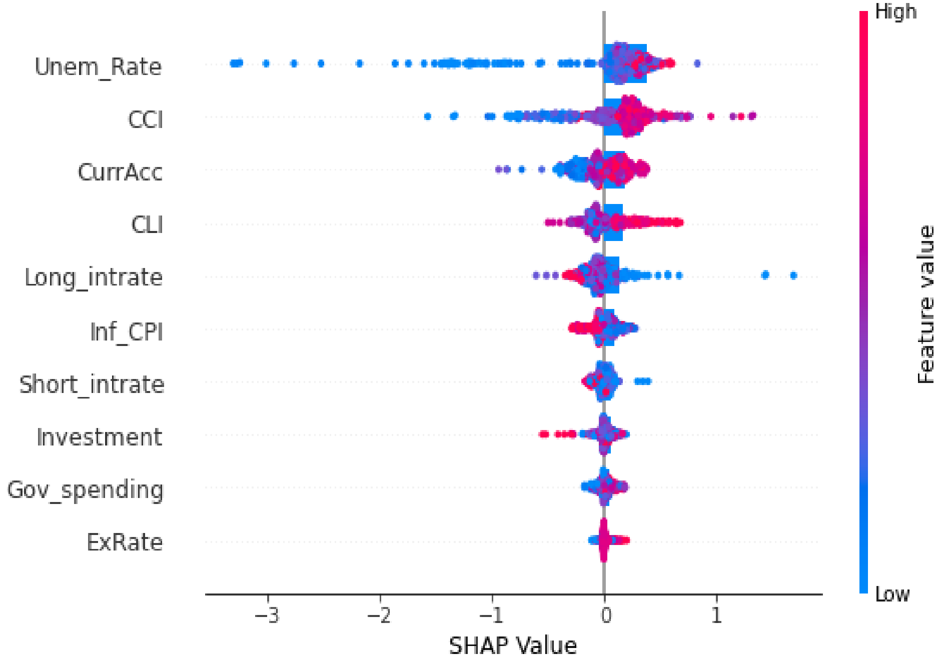
Belgium

Germany

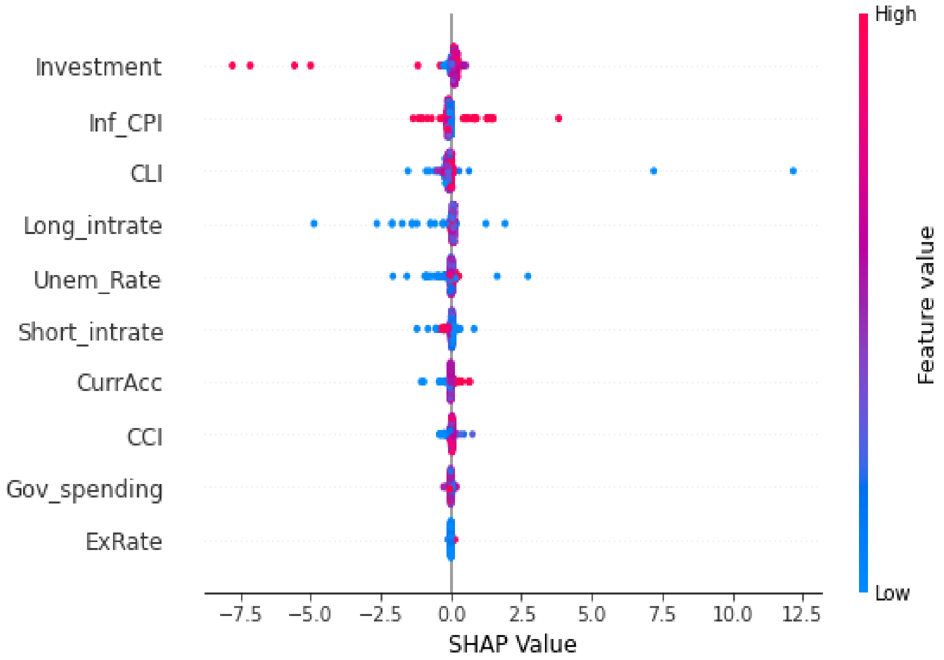


Spain



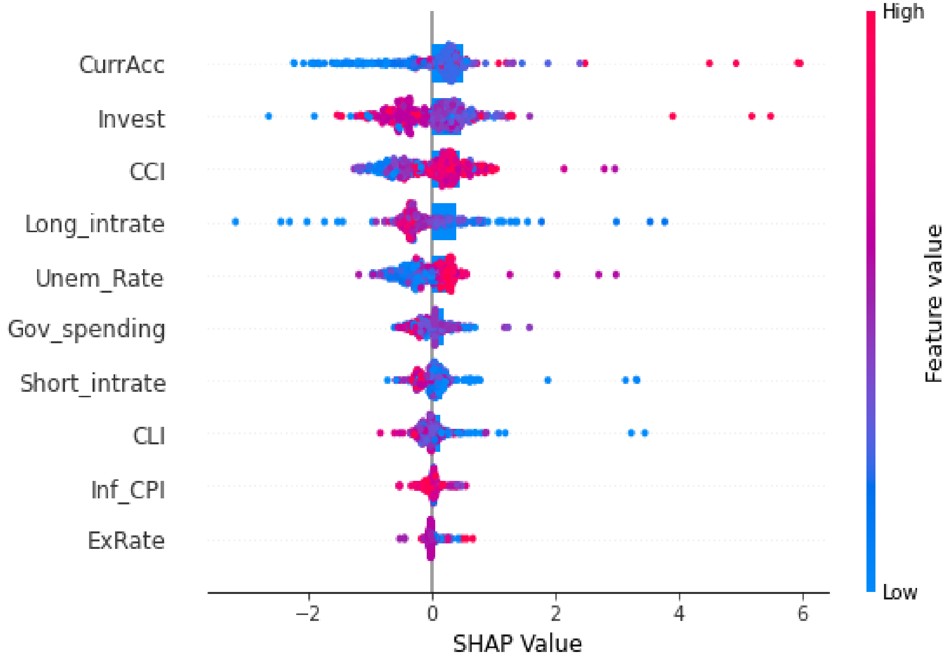


Finland

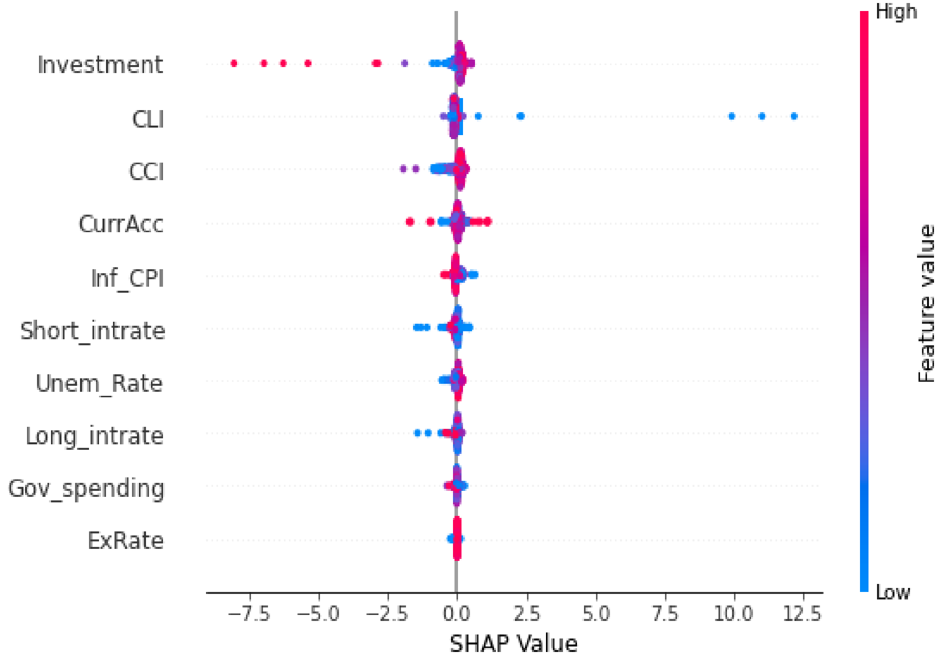


France

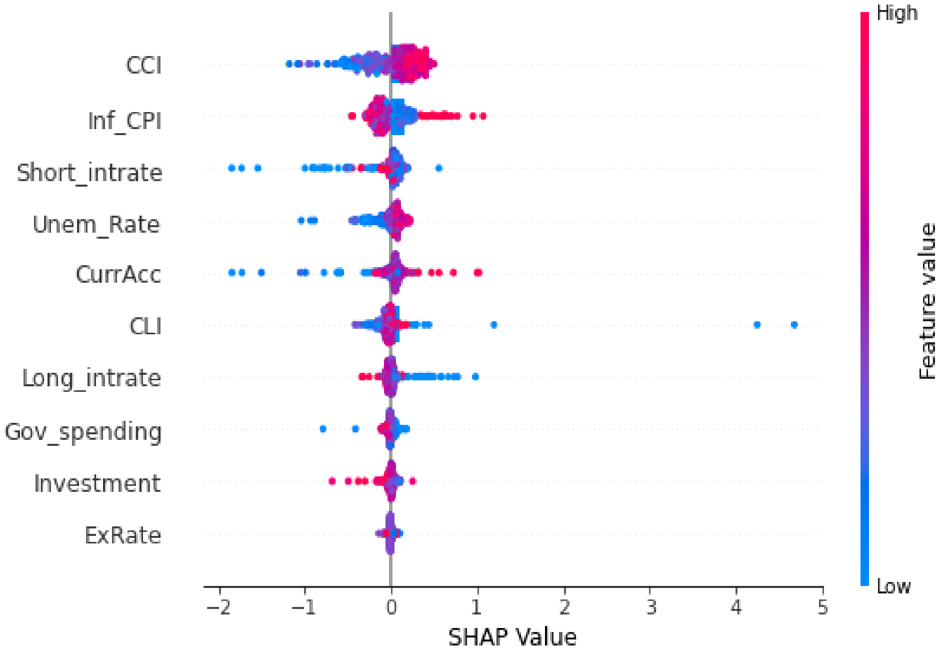
Ireland



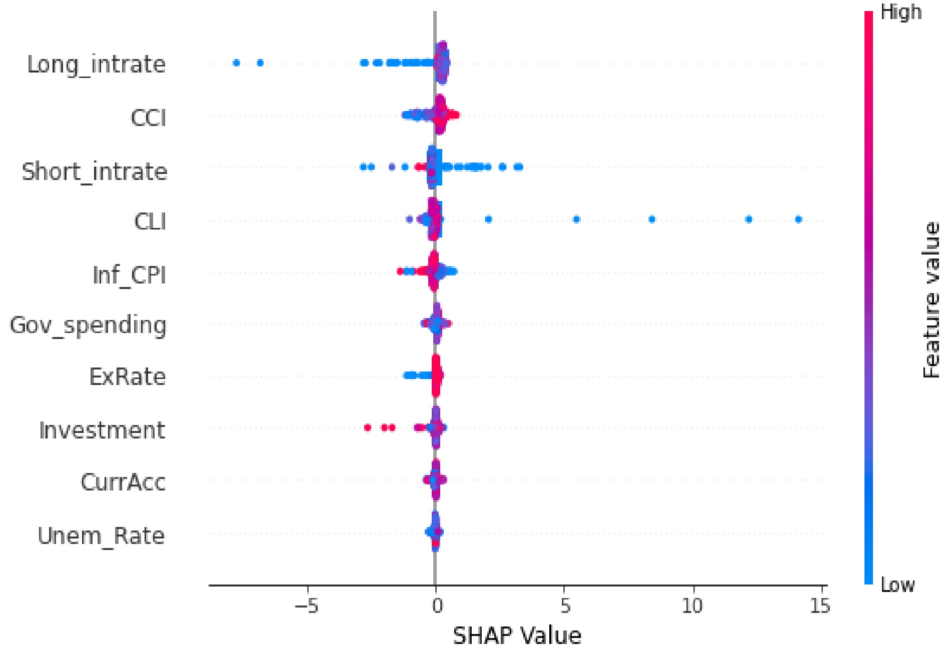
Italy



The Netherland



Portugal



Source: Estimated by authors

6. CONCLUSION

In this study, we have aimed to forecast the GDP growth of 10 EA countries by employing two latest artificial neural network methods, which are LSTM and NARX neural networks. These two models can define a long-term, nonlinear, and complex relationship between dependent and independent variables. We employed LSTM and NARX in our dataset separately and obtained different results according to the size of the selected parameters. We worked with the 30 years of OECD monthly data from 1990 to 2020. We used 288 observations for training the model and used 72 observations for testing the forecasting result. The highest test result was obtained in LSTM for Finland, France, Italy, and Portugal. We observed higher accuracies in NARX for Austria, Belgium, Germany, Spain, and Ireland. Both techniques predicted the Netherlands' GDP growth similarly. We also interpreted the SHAP value, which is demonstrating the contribution of each explanatory variable to the forecasting. We observed that government spending, long-term and short-term interest rate, inflation, consumer confidence index, composite leading indicator, unemployment, and current account balance contributed differently to the estimates for each country, only the exchange rate shows almost no effect on the GDP forecast of the 10 countries. The reason lies in the fact that all EA countries are using Euro as local currency. In the global monetary market, Euro has a strong purchasing power, and this makes EA countries to be independent in currency and not to be affected much by exchange rate fluctuations.

As political implications, we will suggest different strategies for each country due to the different test results of them. As in the study of Chung and Tung (1995), the forecasting results of LSTM and NARX are not the same for countries. Therefore, it is hard to select only one model or recommend only one neural network as the best model. We agree with Banerjee et al. (2005) in selecting the leading indicator. As they mentioned in their study, the economic leading indicator of a country is changing over time. In our model, the level of importance of each variable is not the same for countries. For Austria, it can be suggested that government spending may be adjusted to a certain amount as government spending plays a crucial role in the country's GDP growth according to our result. For Belgium and Finland, it seems unemployment and composite leading indicator affect the growth mostly. Therefore, the government may act on reducing unemployment and inflation policies. For the German economy, composite leading indicator and short-term interest rate seem the most important two factors. In this case, stability of the business cycle and incentive short-term interest rate policy may support the country's economic growth. For Spain and the Netherland economy, the consumer confidence index is at the top, which means that consumer attitude and behavior play a critical role in the country's economy. In these two countries, policies such as fiscal policy, tax rate, and government spending can be followed to increase consumer confidence (Omohundro, 2017). For France and Italy, it can be suggested that the government may take actions to encourage investment. Low interest and tax rates, creating a reliable investment atmosphere can be helpful in GDP growth. In Ireland's GDP growth forecasting, the current account balance stays on top as the most contributed variable to the model, therefore the balance between imports and export is noteworthy in Ireland. For Portugal, the interest rate in both the long and short run is critical for the economy, besides, consumer confidence is another main factor in GDP growth in Portugal.

Countries' economies have been hit hard by the covid-19 outbreak significantly and the effects of the pandemic are continuing. The war in Ukraine also triggers a prospective food crisis and the world seems to be faced with a serious supply shortage of food products. Since economic growth is a consequence of political decisions, economic and social activities, and internal and external balances, it may be difficult to predict growth for the upcoming period. Even World Bank and IMF have changed their growth prediction several times for 2022. Unexpected external shocks have commuted the path of economic growth; therefore, traditional growth theory and existing literature may not be sufficient to forecast the growth of the new world. Federal Reserve Bank, Bank of England, and European Central Bank also failed to predict leading parameters of growth. In further studies, new methods may be developed or dummy variables representing shocks such as war or unexpected crises can be included in the model.

REFERENCES

Allende H., Moraga C., & Salas R. (2002). Artificial Neural Networks in Time Series Forecasting: A Comparative Analysis. *Kybernetik A — Volume 88 (2002),* Number 6. 685-707. http://dml.cz/dmlcz/135496.

Aron J. and Muellbauer J. (2002). Interest Rate Effects on Output: Evidence from a GDP Forecasting Model for South Africa. IMF Staff Papers, 2002, Vol. 49, IMF Annual Research Conference (2002), pp.185-213. https://doi.org/10.2307/3872476.

Banerjee A., Marcellino M., & Mastens I. (2005). Leading Indicators for Euro-area Inflation and GDP Growth. *Oxford Bulletin of Economics and Statistics, 67,* Supplement (2005) 0305-9049. https://doi.org/10.1111/j.1468-0084.2005.00141.x.

Brownlee, J. (2017). *Long short-term memory networks with python: develop sequence prediction models with deep learning*. Machine Learning Mastery.

Cavuto D. J. (1997). *An Exploration and Development of Current Artificial Neural Network Theory and Applications with Emphasis on Artificial Life*. The Cooper Union Albert Nerken School of Engineering, Master dissertation of Engineering, May 6, 1997.

Chung M. K. and Tung L. (1995). Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks. *Journal of Applied Econometrics, Vol. 10,* 347-364. https://doi.org/10.1002/jae.3950100403.

Dongare A. D., Kharde R. R., & Kachare A. D. (2012). Introduction to Artificial Neural Network. *International Journal of Engineering and Innovative Technology (IJEIT) Volume 2,* Issue 1, July 2012. ISSN: 2277-3754. 189-194. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1082.1323&rep=rep1&type=pdf

Estrella A. and Mishkin F.S. (1998). Predicting U.S. Recession: Financial Variables as Leading Indicators. *National Bureau of Economic Research.* 45-61. https://doi.org/10.1162/003465398557320.

Granger, C. W. J., & Newbold, P. (2014). *Forecasting economic time series*. Academic Press.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780. Doi: 10.1162/neco.1997.9.8.1735.

Heravi S., Osborn D. R., & Birchenhall C.R. (2003). Linear versus Neural Network Forecasts for European Industrial Production Series. *International Journal of Forecasting Volume 20,* Issue 3, (2004). 435 – 446. https://doi.org/10.1016/S0169-2070(03)00062-1.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. O Texts.

Jilani S., Cheema F. E. A., & Asim M. (2010). Exploring Impact of Macro Economic Variables on GDP of Pakistan. *Journal of Management and Social Sciences.* Vol. 6, No. 2, (Fall 2010) 65-73. http://dx.doi.org/10.46745/ilma.jbs.2013.09.01.04.

Jahn, M. (2018). Artificial neural network regression models: Predicting GDP growth. *HWWI Research Paper, No. 185,* ISSN 1861-504X. http://hdl.handle.net/10419/182108.

Kukreja H., Bharath N., Siddesh C. S., & Kuldeep S. (2016). An Introduction to Artificial Neural Network. *IJARIIE-ISSN (O)-2395-4396, Vol-1* Issue-5. https://www.researchgate.net/publication/319903816.

Lin, T., Horne, B. G., Tino, P., & Giles, C. L. (1996). Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks,* 7(6), doi:1329-1338. 10.1109/72.548162.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems,* 30. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

Maasoumi E., Khotanzed A., & Abaye A. (1994). Artificial neural networks for some macroeconomic series: A first report. Econometric Reviews, 13:1, 105-122. https://doi.org/10.1080/07474939408800276.

Marijana Ć., Poposki K., & Ivan Ć. (2009). Forecasting economic growth using financial variables - Comparison of linear regression and neural network models. *Conference on Mathematics and Computers in Business and Economics. ISSN: 1790-5109, ISBN: 978-960-474-063-5.* 255-260. https://www.academia.edu/download/85584202/MCBE43.pdf.

Min Q. (2001). Predicting US recession with leading indicators via neural network models. *International Journal of Forecasting 17,* 383-401. https://doi.org/10.1016/S0169-2070(01)00092-9.

Omohundro, S. (2017). Effects of Fiscal Policy on Consumer Confidence. *Issues in Political Economy,* Vol 26(1), 2017, 162-180. https://digitalcommons.butler.edu/ugtheses/354.

Önder E., Bayır F., & Hepşen A. (2013). Forecasting Macroeconomic Variables using Artificial Neural Network and Traditional Smoothing Techniques. *Journal of Applied Finance & Banking,* vol. 3, no. 4, 2013, 73-104. http://dx.doi.org/10.2139/ssrn.2264379.

Pons J. (2000). The Accuracy of IMF and OECD Forecasts for G7 Countries. *Journal of Forecasting. 19,* 53-63 (2000). https://doi.org/10.1002/(SICI)1099-131X(200001)19:1<53::AID-FOR736>3.0.CO;2-J.

Siegelmann, H. T., Horne, B. G., & Giles, C. L. (1997). Computational capabilities of recurrent NARX neural networks, Systems, Man, and Cybernetics, Part B: Cybernetics. *IEEE Transactions on,* 27(2), 208-215. Doi: 10.1109/3477.558801.

Stock J.H., & Watson M. W. (1999). Forecasting Inflation. *Journal of Monetary Economics 44*, (1999) 293-335. https://doi.org/10.1016/S0304-3932(99)00027-6.

Swanson, N. R., & White, H. (1995). A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. *Review of Economics and Statistics*, 79(4), 540-550. https://doi.org/10.1162/003465397557123.

Swanson, N. R., & White, H. (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, 13(4), 439-461. https://doi.org/10.1016/S0169-2070(97)00030-7.

Semuel H. & Nurina S. (2015). Analysis of the Effect of Inflation, Interest Rates, and Exchange Rates on Gross Domestic Product (GDP) in Indonesia. *Proceedings of the International Conference on Global Business, Economics, Finance and Social Sciences.* ISBN: 978-1-941505-22-9. http://repository.petra.ac.id/id/eprint/17008.

Tkacz G. (2001). Neural network forecasting of Canadian GDP growth. *International Journal of Forecasting Vol. 17* (2001), 57–69. https://doi.org/10.1016/S0169-2070(00)00063-7.

https://academic.oup.com/economicpolicy/article-abstract/14/29/400/2366371

https://www.bbc.com/news/business-51706225

APPENDIX

Figure A.1: The Estimated Loss for Austria in LSTM

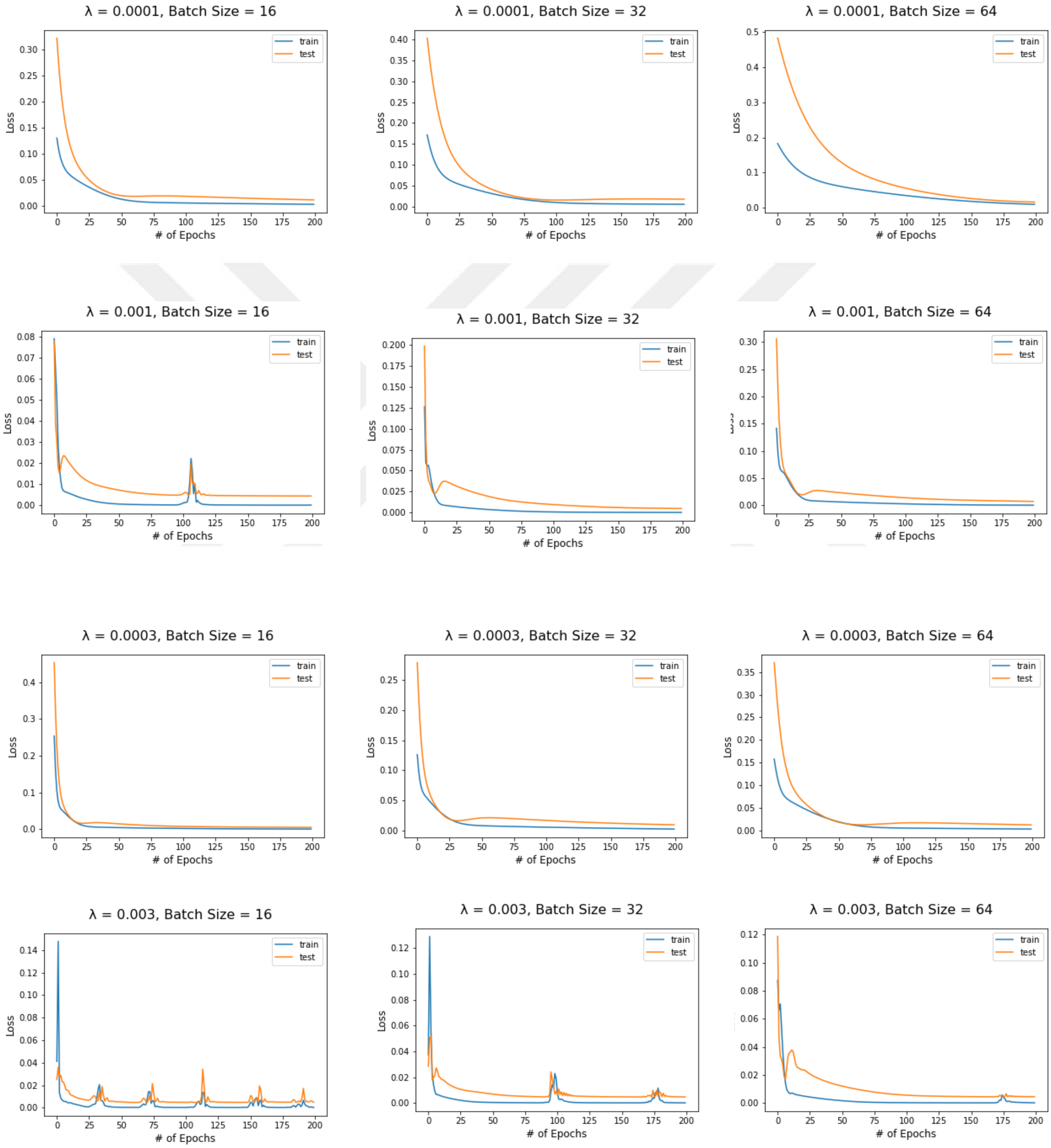


Figure A.2: The Estimated Loss for Belgium in LSTM

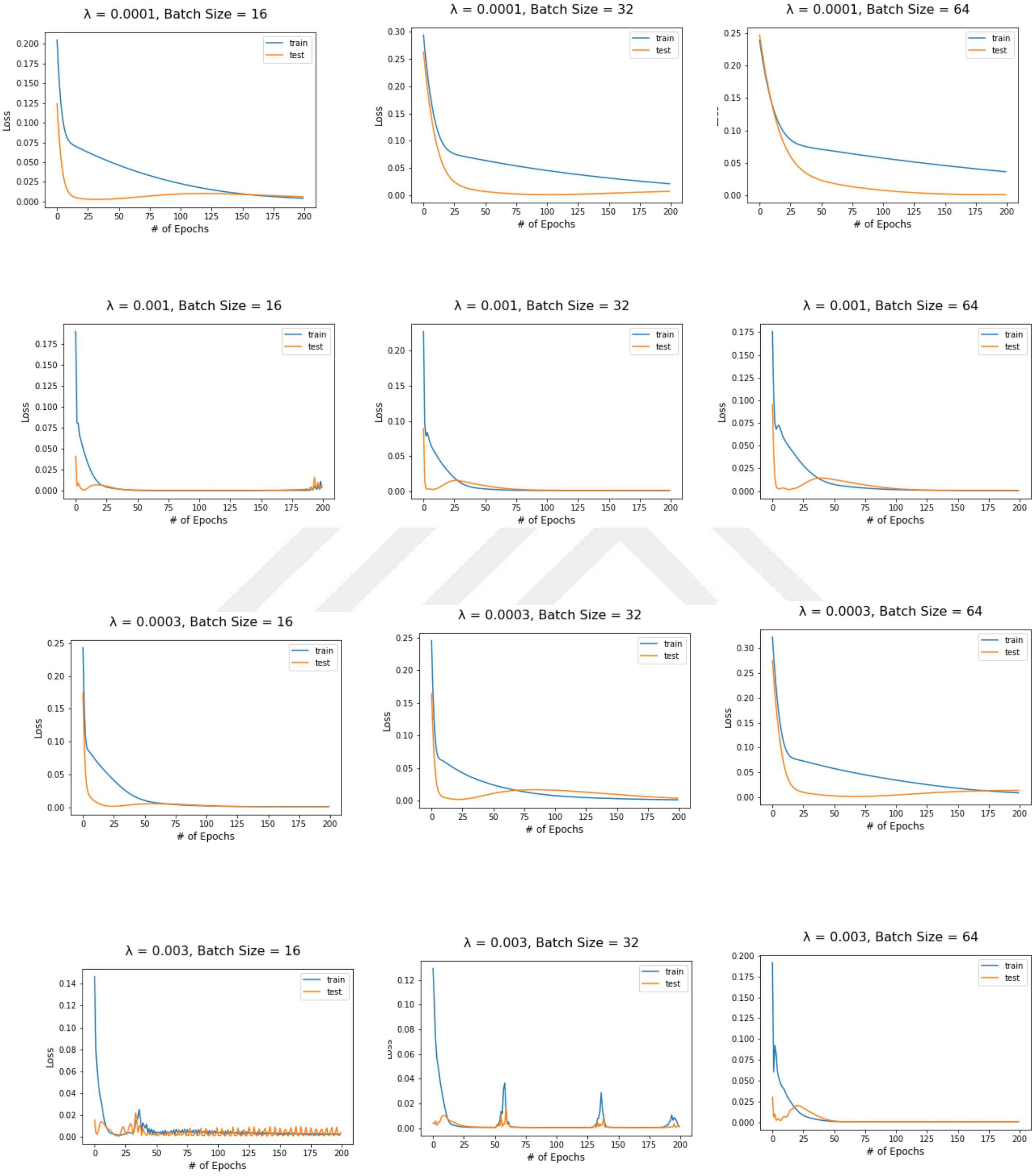
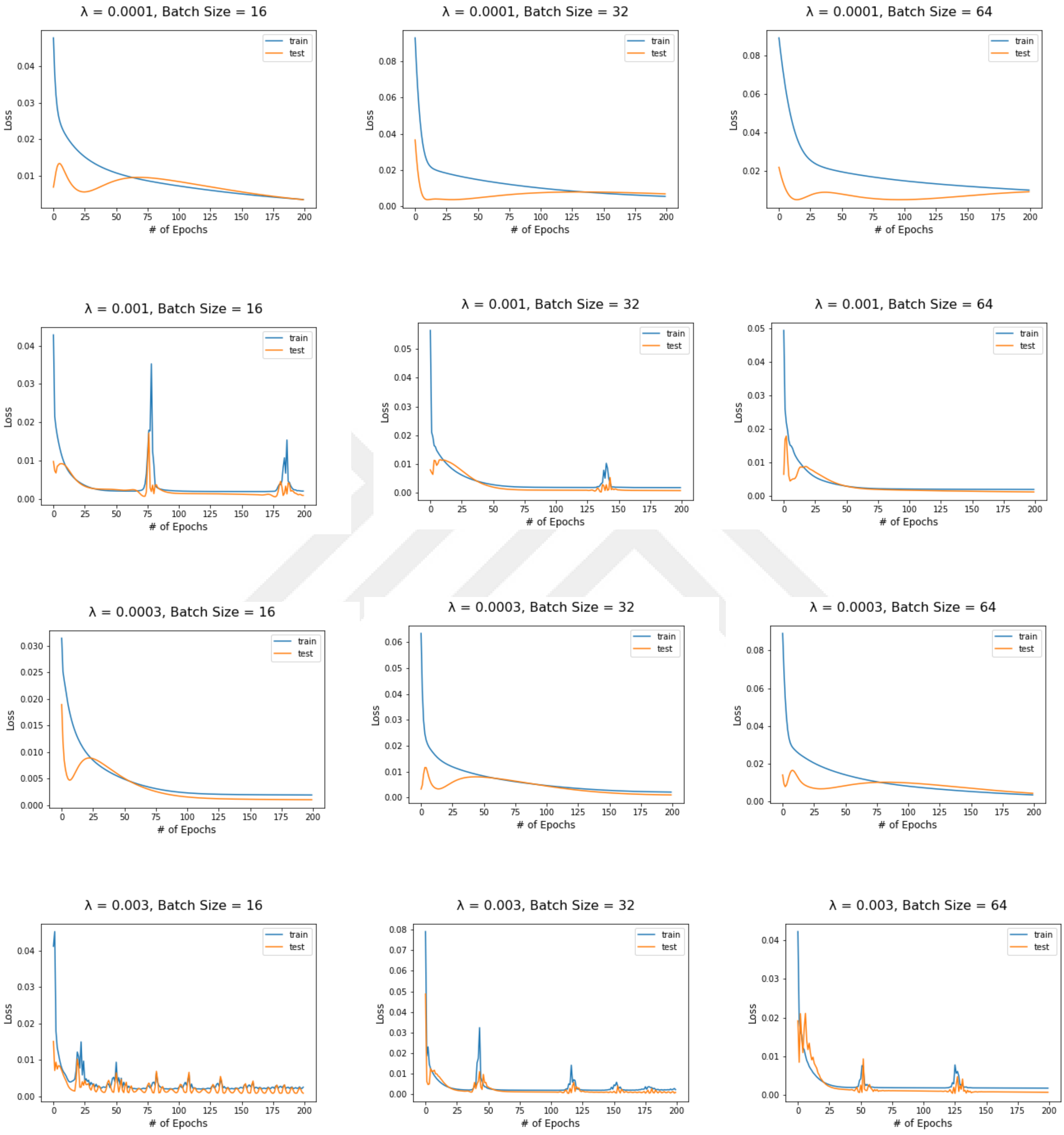


Figure A.3: The Estimated Loss for Germany in LSTM



λ = 0.0001, Batch Size = 16

λ = 0.0001, Batch Size = 32

λ = 0.0001, Batch Size = 64

λ = 0.001, Batch Size = 16

λ = 0.001, Batch Size = 32

λ = 0.001, Batch Size = 64

λ = 0.0003, Batch Size = 16

λ = 0.0003, Batch Size = 32

λ = 0.0003, Batch Size = 64

λ = 0.003, Batch Size = 16

λ = 0.003, Batch Size = 32

λ = 0.003, Batch Size = 64

Figure A.4: The Estimated Loss for Spain in LSTM

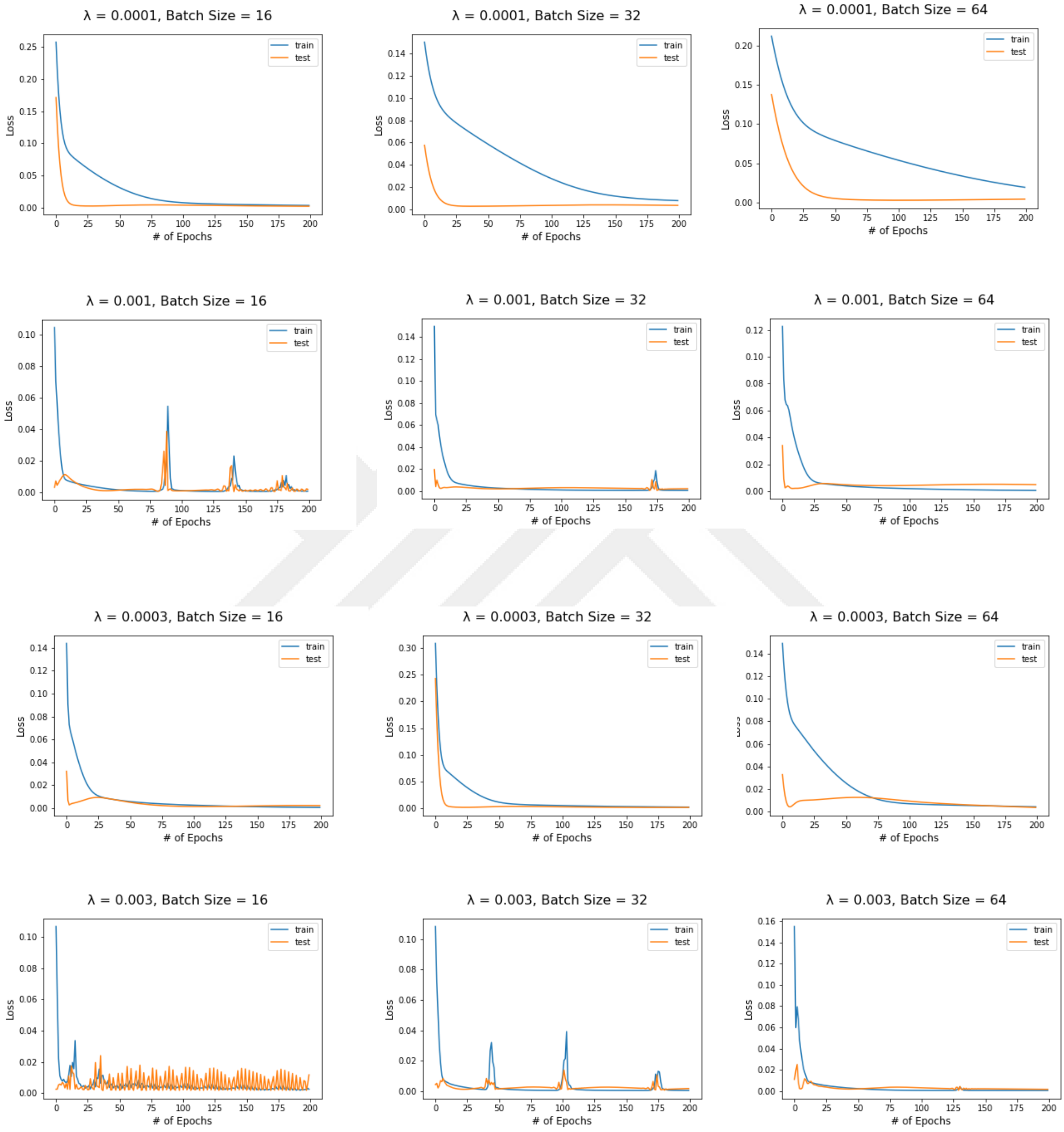
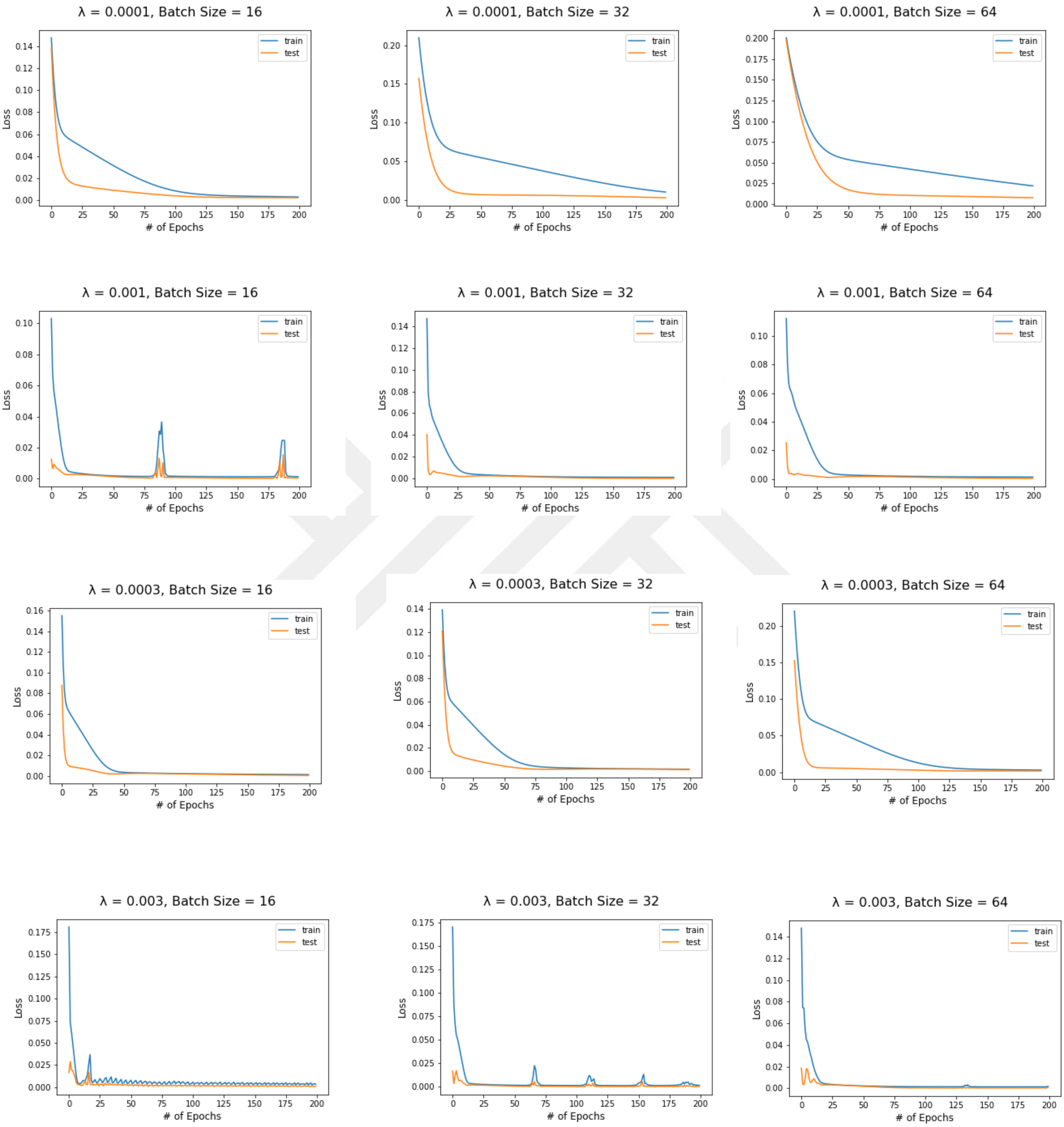


Figure A.5: The Estimated Loss for Finland in LSTM



λ = 0.0001, Batch Size = 16

λ = 0.0001, Batch Size = 32

λ = 0.0001, Batch Size = 64

λ = 0.001, Batch Size = 16

λ = 0.001, Batch Size = 32

λ = 0.001, Batch Size = 64

λ = 0.0003, Batch Size = 16

λ = 0.0003, Batch Size = 32

λ = 0.0003, Batch Size = 64

λ = 0.003, Batch Size = 16

λ = 0.003, Batch Size = 32

λ = 0.003, Batch Size = 64

Figure A.6: The Estimated Loss for France in LSTM

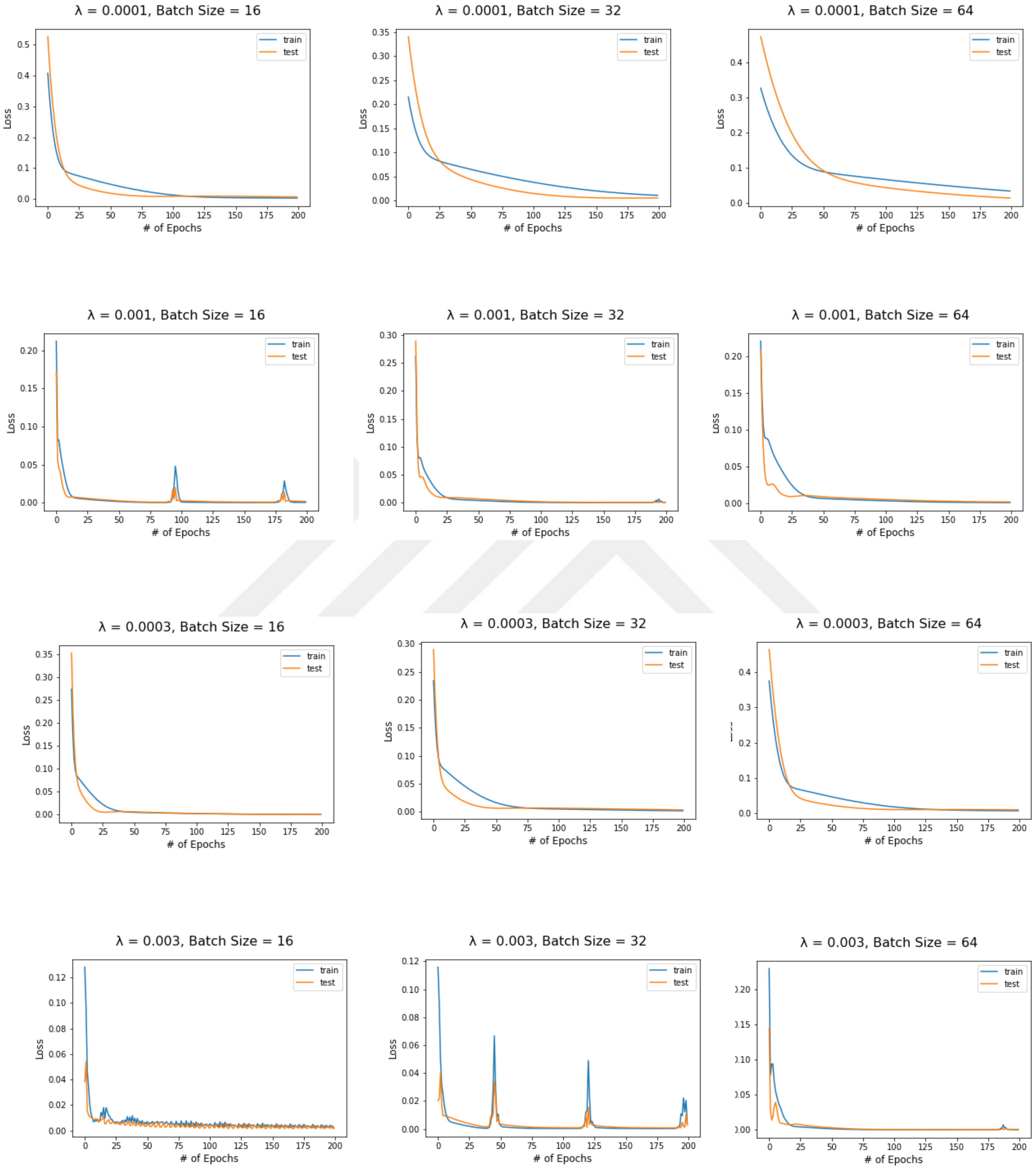


Figure A.7: The Estimated Loss for Ireland in LSTM

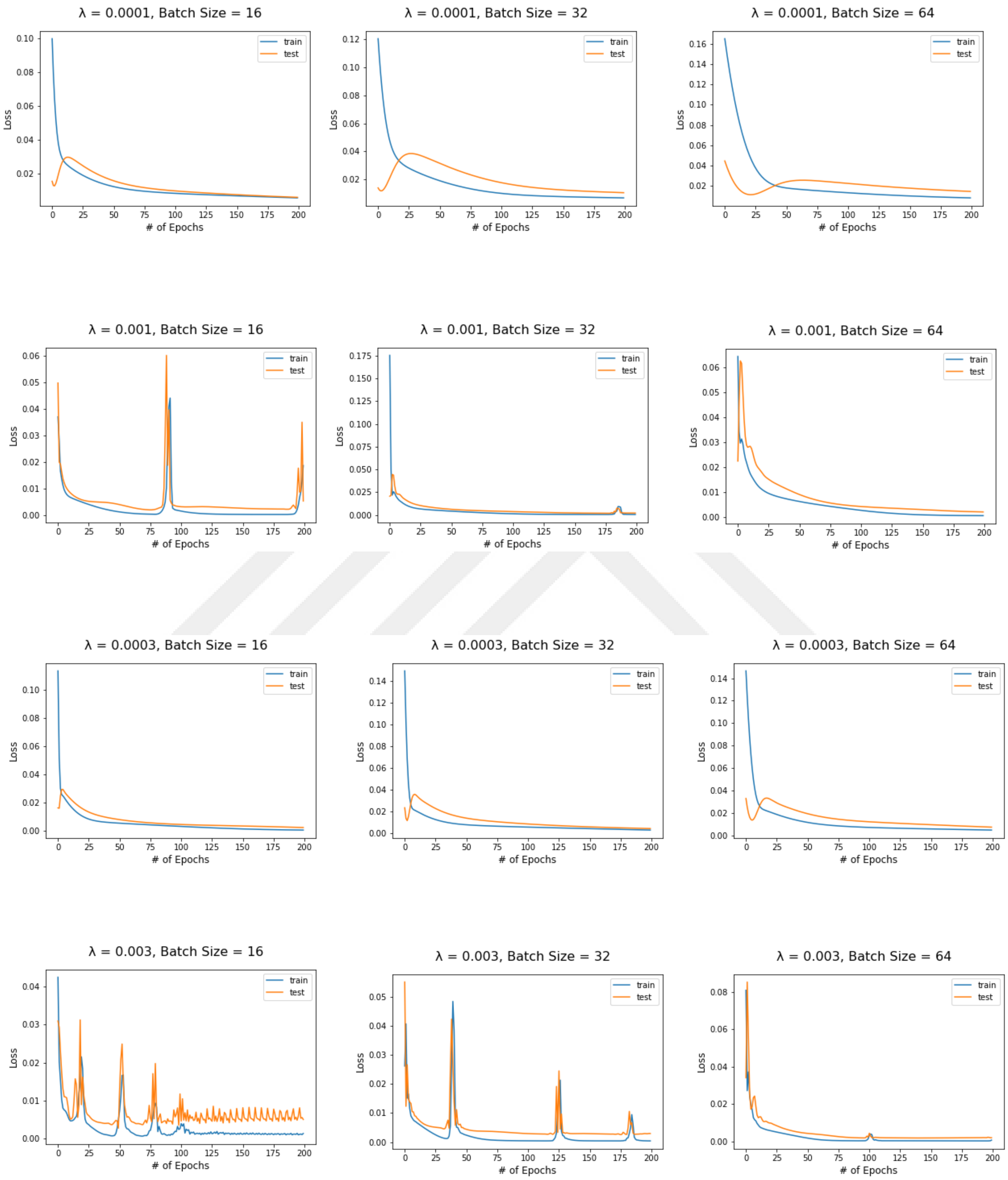
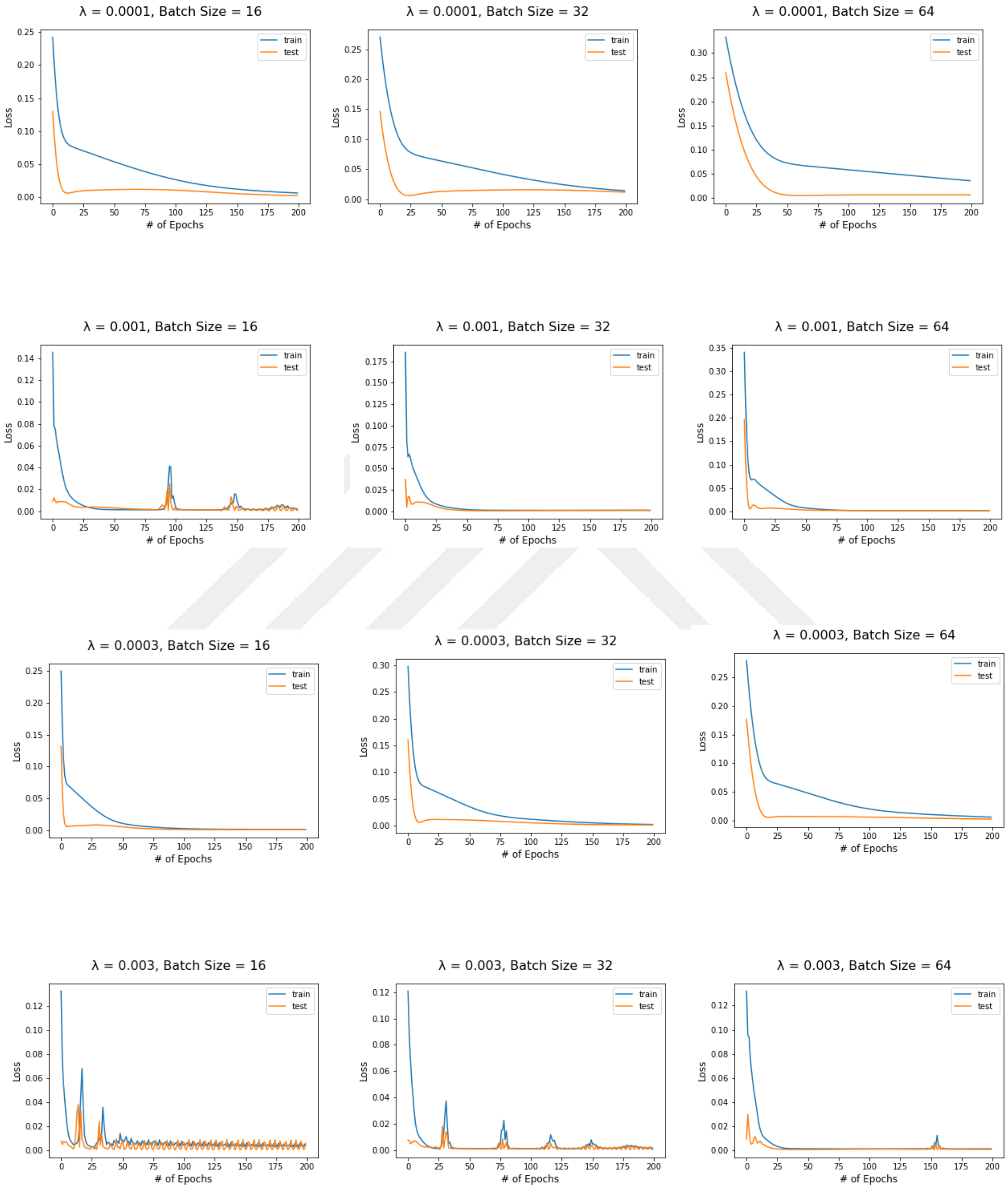


Figure A.8: The Estimated Loss for Italy in LSTM



λ = 0.0001, Batch Size = 16

λ = 0.0001, Batch Size = 32

λ = 0.0001, Batch Size = 64

λ = 0.001, Batch Size = 16

λ = 0.001, Batch Size = 32

λ = 0.001, Batch Size = 64

λ = 0.0003, Batch Size = 16

λ = 0.0003, Batch Size = 32

λ = 0.0003, Batch Size = 64

λ = 0.003, Batch Size = 16

λ = 0.003, Batch Size = 32

λ = 0.003, Batch Size = 64

Figure A.9: The Estimated Loss for the Netherland in LSTM

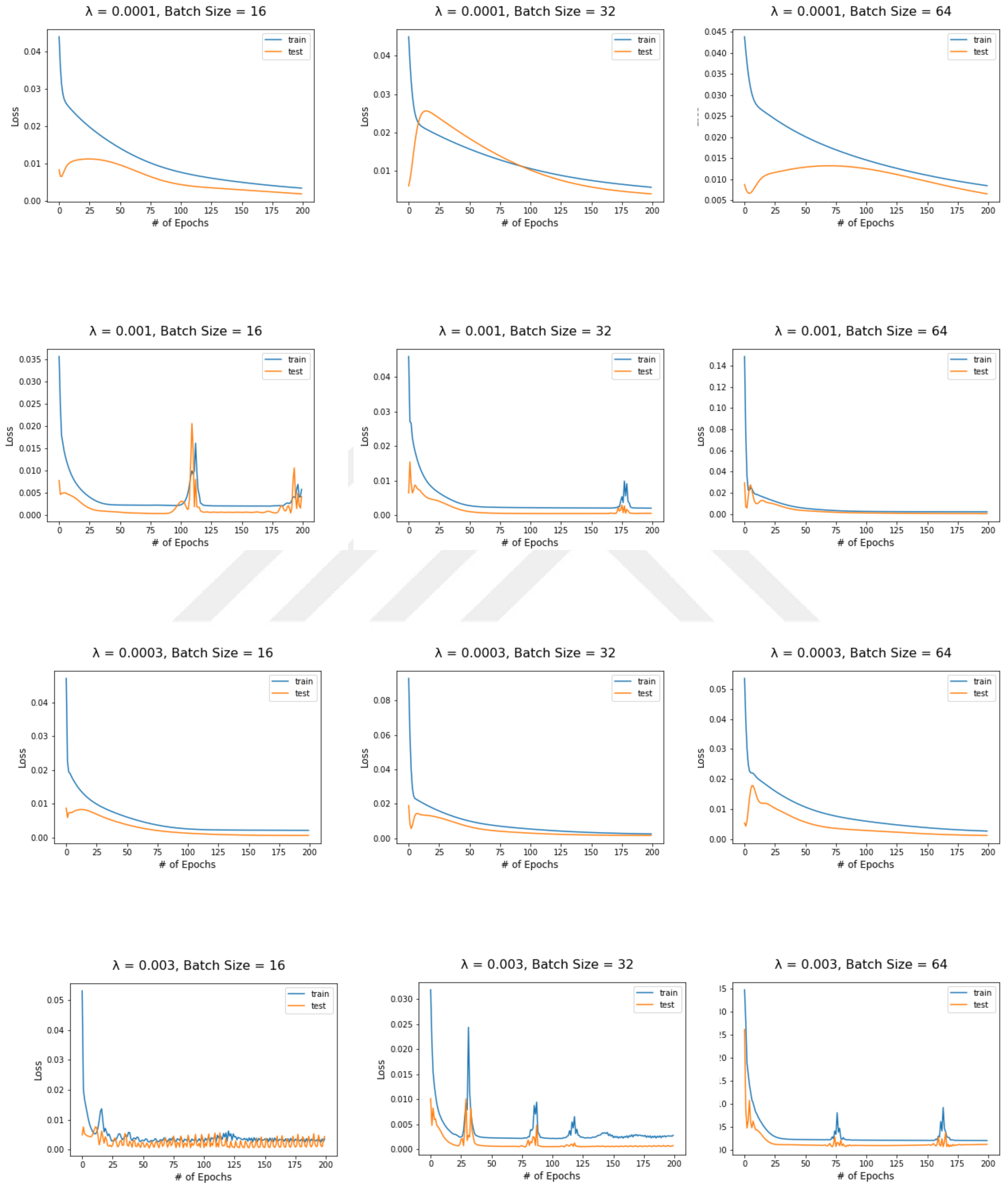
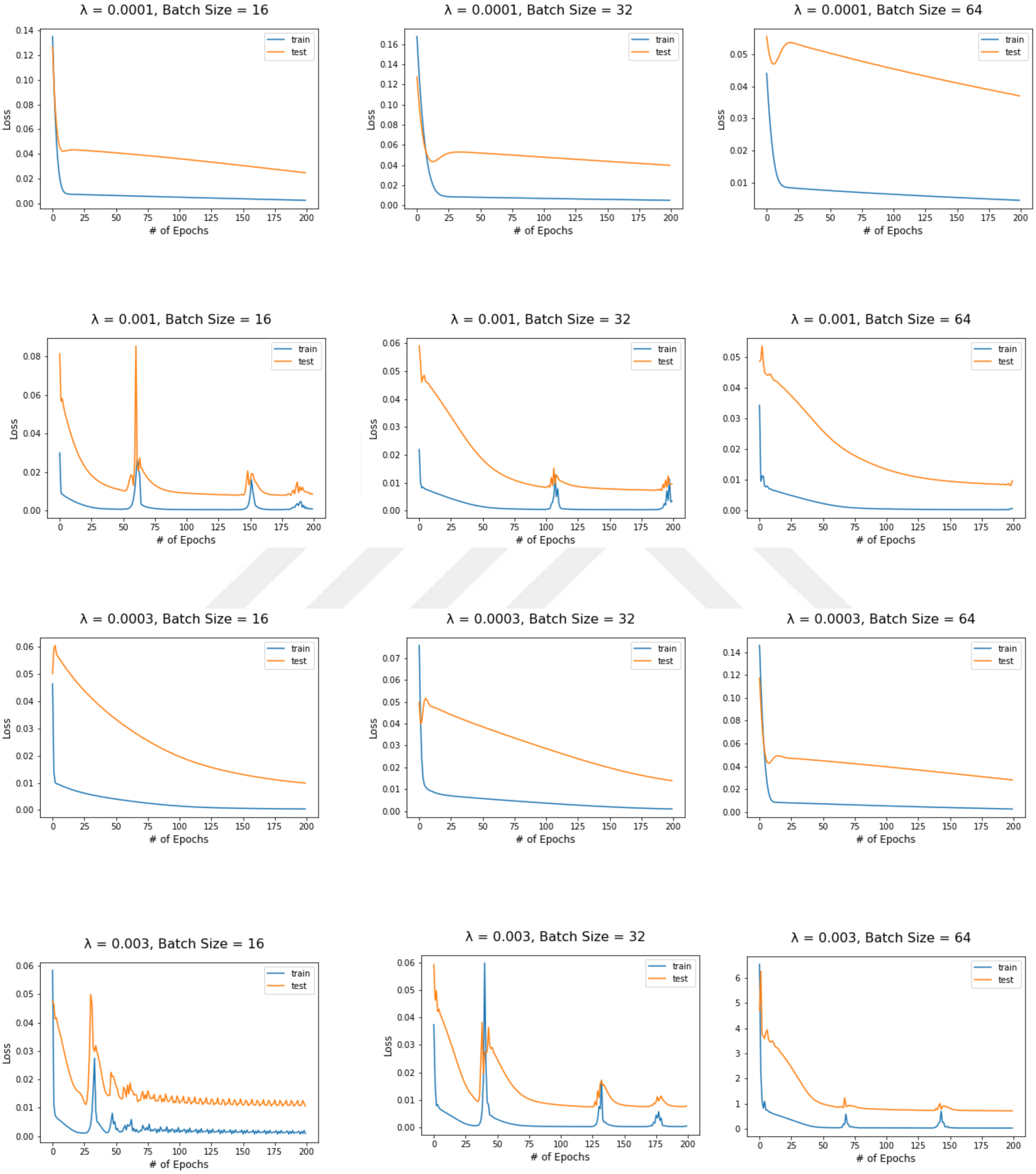


Figure A.10: The Estimated Loss for Portugal in LSTM



λ = 0.0001, Batch Size = 16

λ = 0.0001, Batch Size = 32

λ = 0.0001, Batch Size = 64

λ = 0.001, Batch Size = 16

λ = 0.001, Batch Size = 32

λ = 0.001, Batch Size = 64

λ = 0.0003, Batch Size = 16

λ = 0.0003, Batch Size = 32

λ = 0.0003, Batch Size = 64

λ = 0.003, Batch Size = 16

λ = 0.003, Batch Size = 32

λ = 0.003, Batch Size = 64

Figure A.11: The Estimated Loss for Austria in NARX

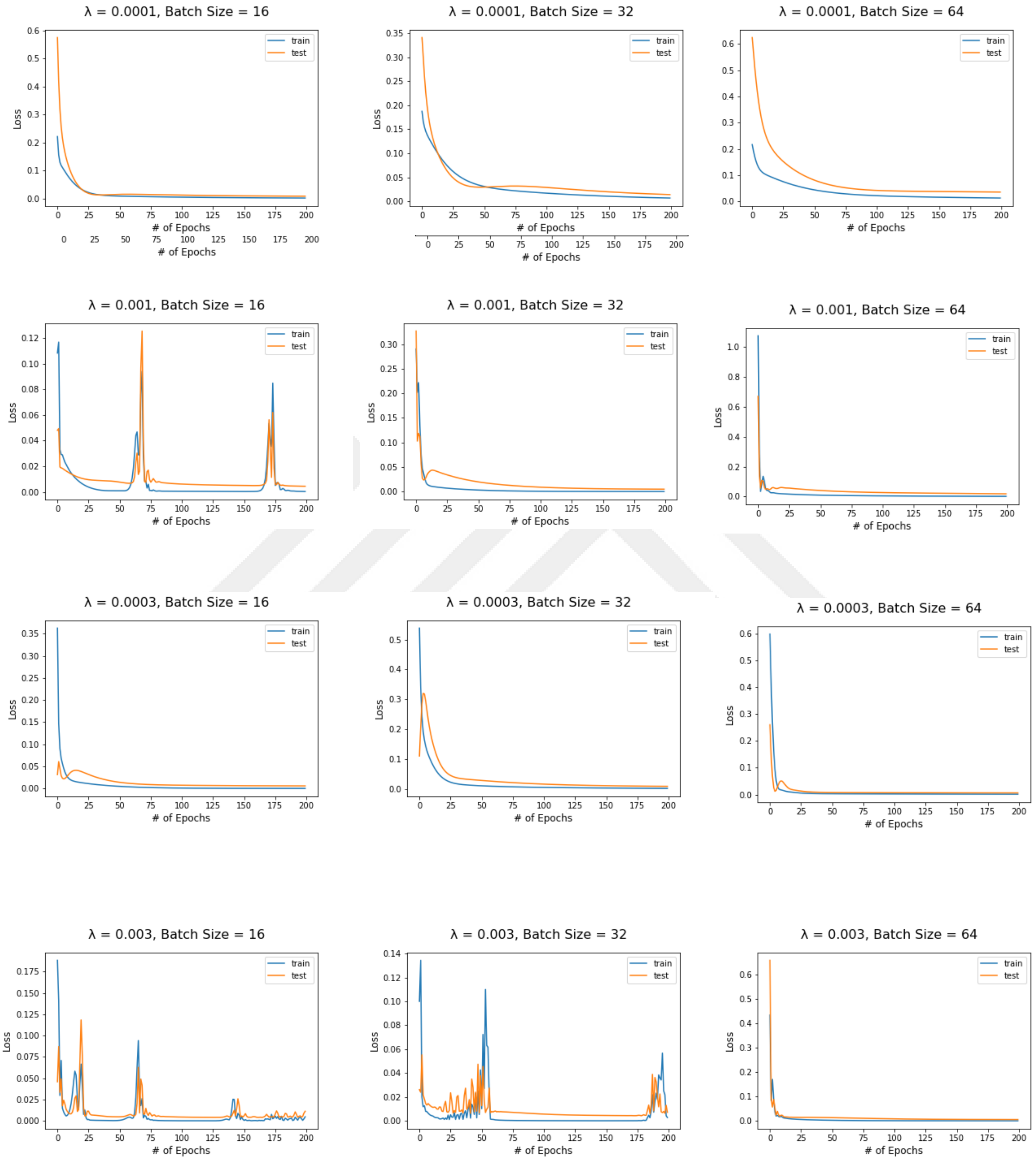


Figure A.12: The Estimated Loss for Belgium in NARX

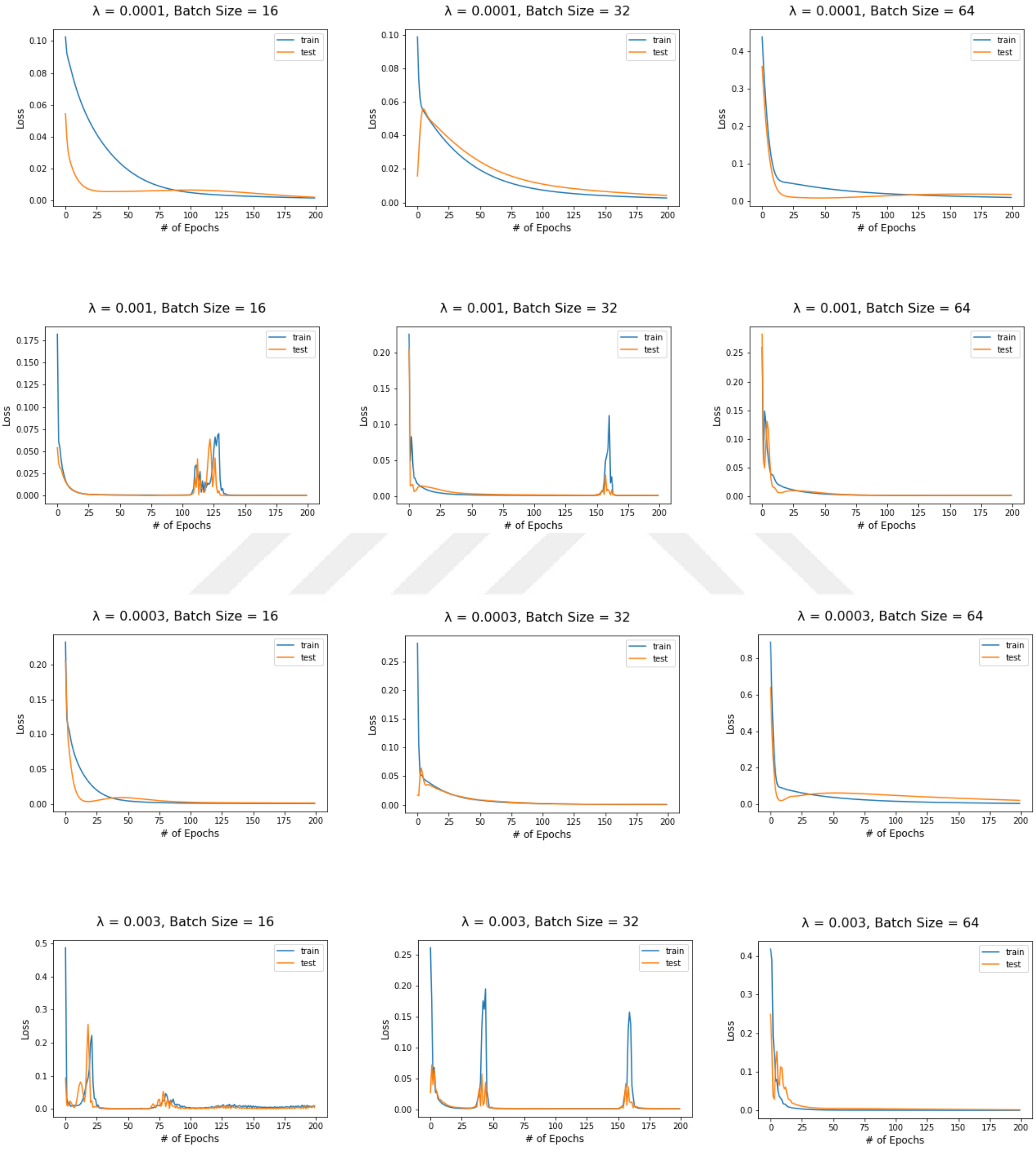


Figure A.13: The Estimated Loss for Germany in NARX

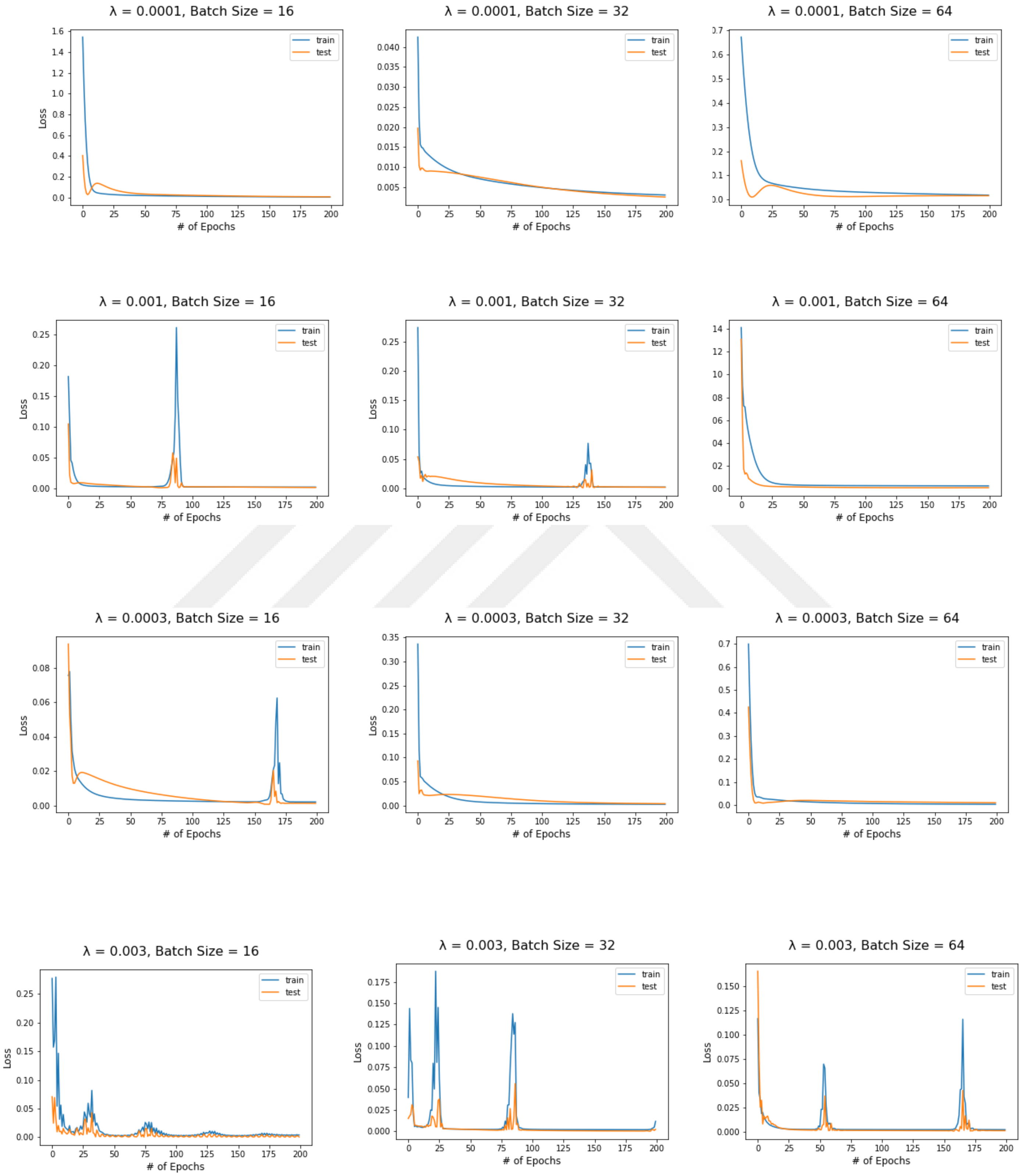
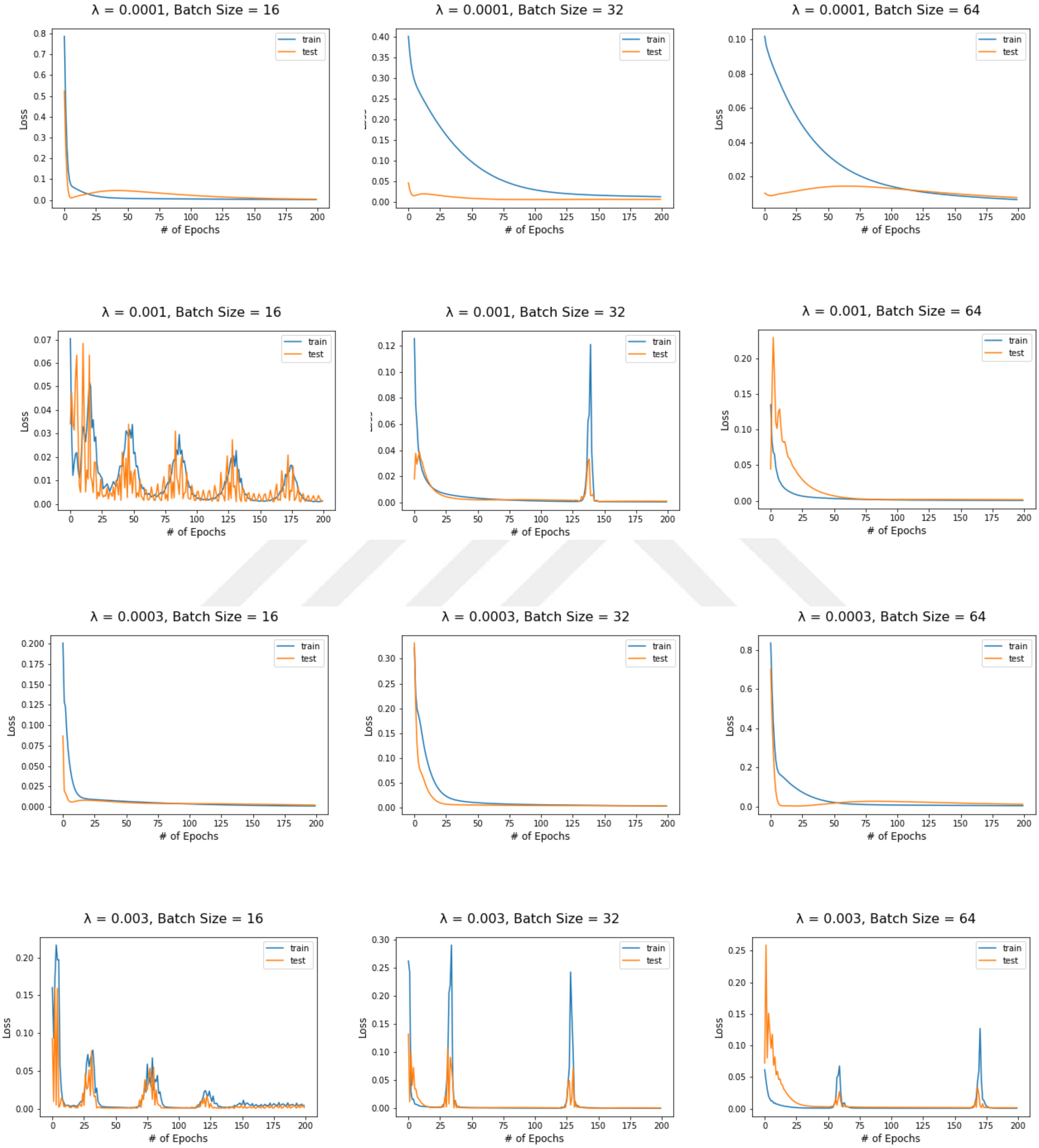


Figure A.14: The Estimated Loss for Spain in NARX



λ = 0.0001, Batch Size = 16

λ = 0.0001, Batch Size = 32

λ = 0.0001, Batch Size = 64

λ = 0.001, Batch Size = 16

λ = 0.001, Batch Size = 32

λ = 0.001, Batch Size = 64

λ = 0.0003, Batch Size = 16

λ = 0.0003, Batch Size = 32

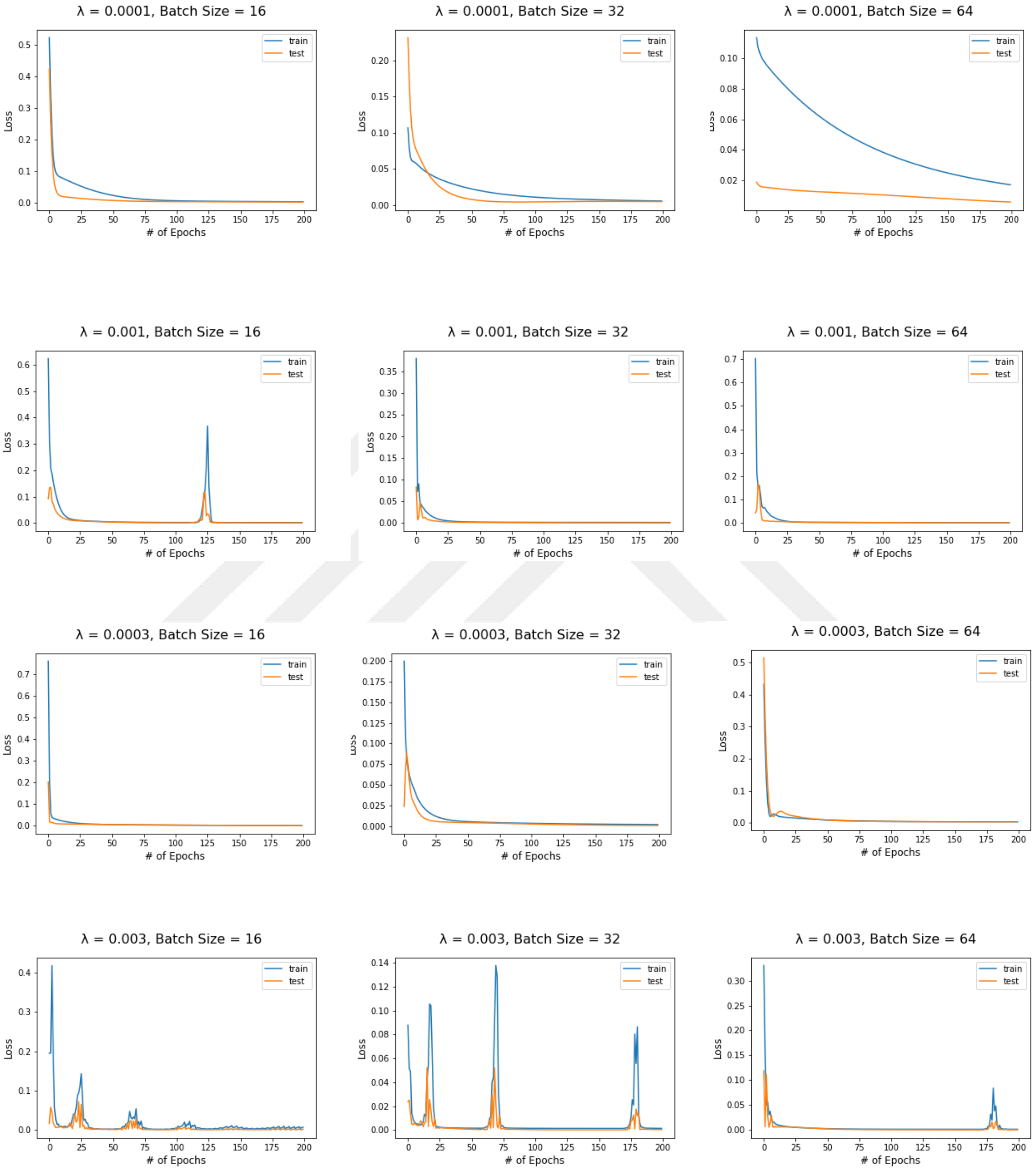
λ = 0.0003, Batch Size = 64

λ = 0.003, Batch Size = 16

λ = 0.003, Batch Size = 32

λ = 0.003, Batch Size = 64

Figure A.15: The Estimated Loss for Finland in NARX



λ = 0.0001, Batch Size = 16

λ = 0.0001, Batch Size = 32

λ = 0.0001, Batch Size = 64

λ = 0.001, Batch Size = 16

λ = 0.001, Batch Size = 32

λ = 0.001, Batch Size = 64

λ = 0.0003, Batch Size = 16

λ = 0.0003, Batch Size = 32

λ = 0.0003, Batch Size = 64

λ = 0.003, Batch Size = 16

λ = 0.003, Batch Size = 32

λ = 0.003, Batch Size = 64

Figure A.16: The Estimated Loss for France in NARX

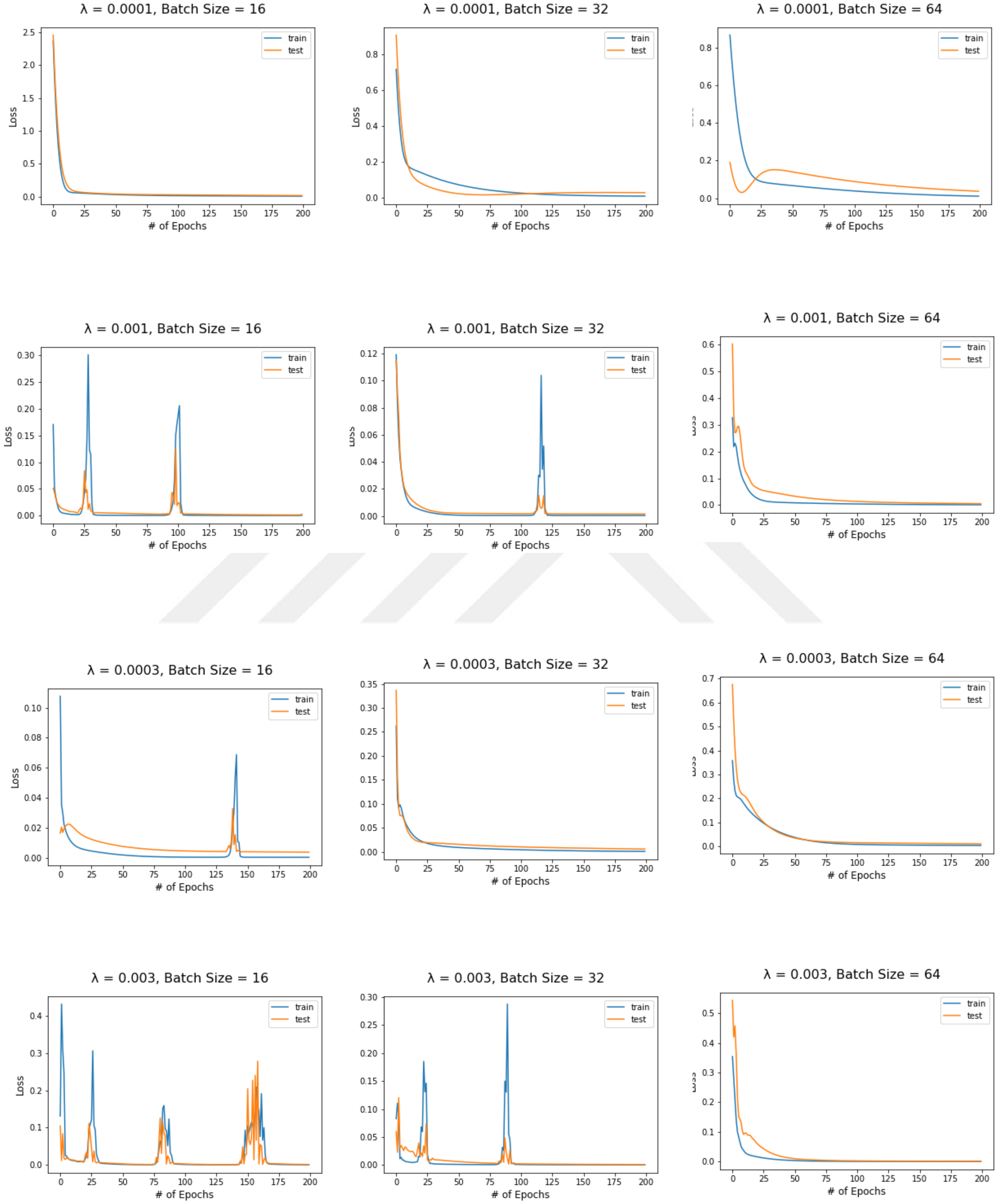


Figure A.17: The Estimated Loss for Ireland in NARX

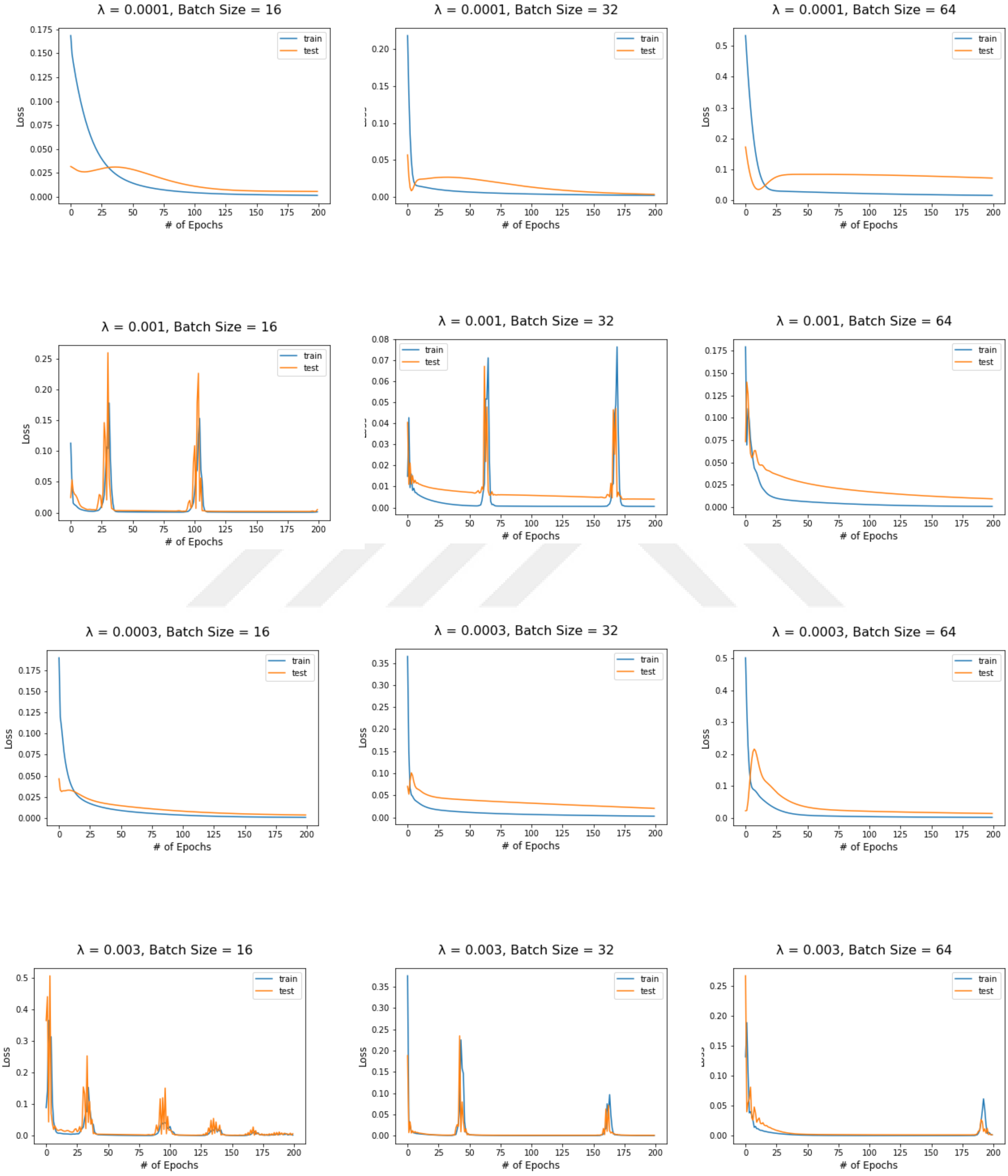


Figure A.18: The Estimated Loss for Italy in NARX

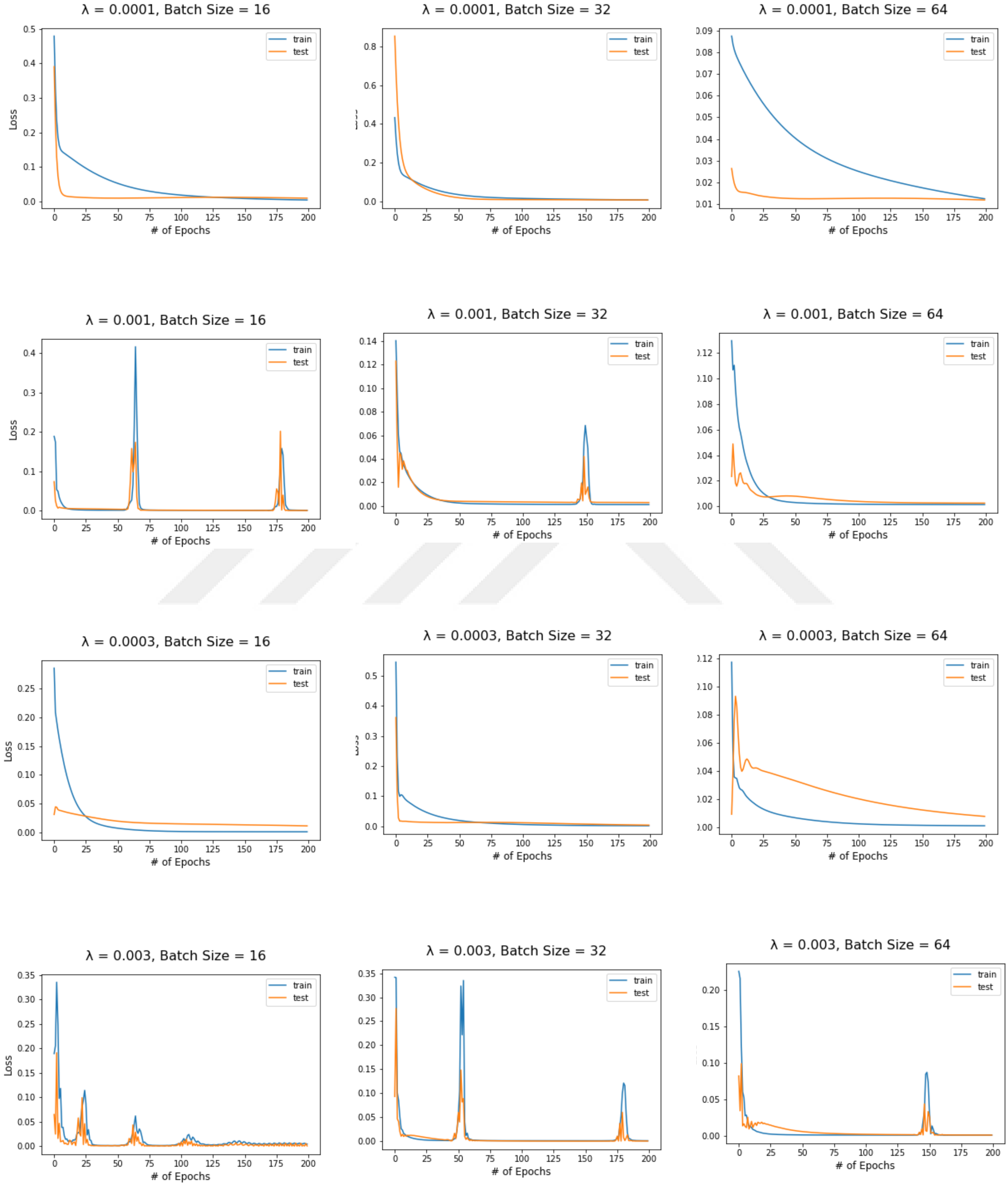


Figure A.19: The Estimated Loss for the Netherland in NARX

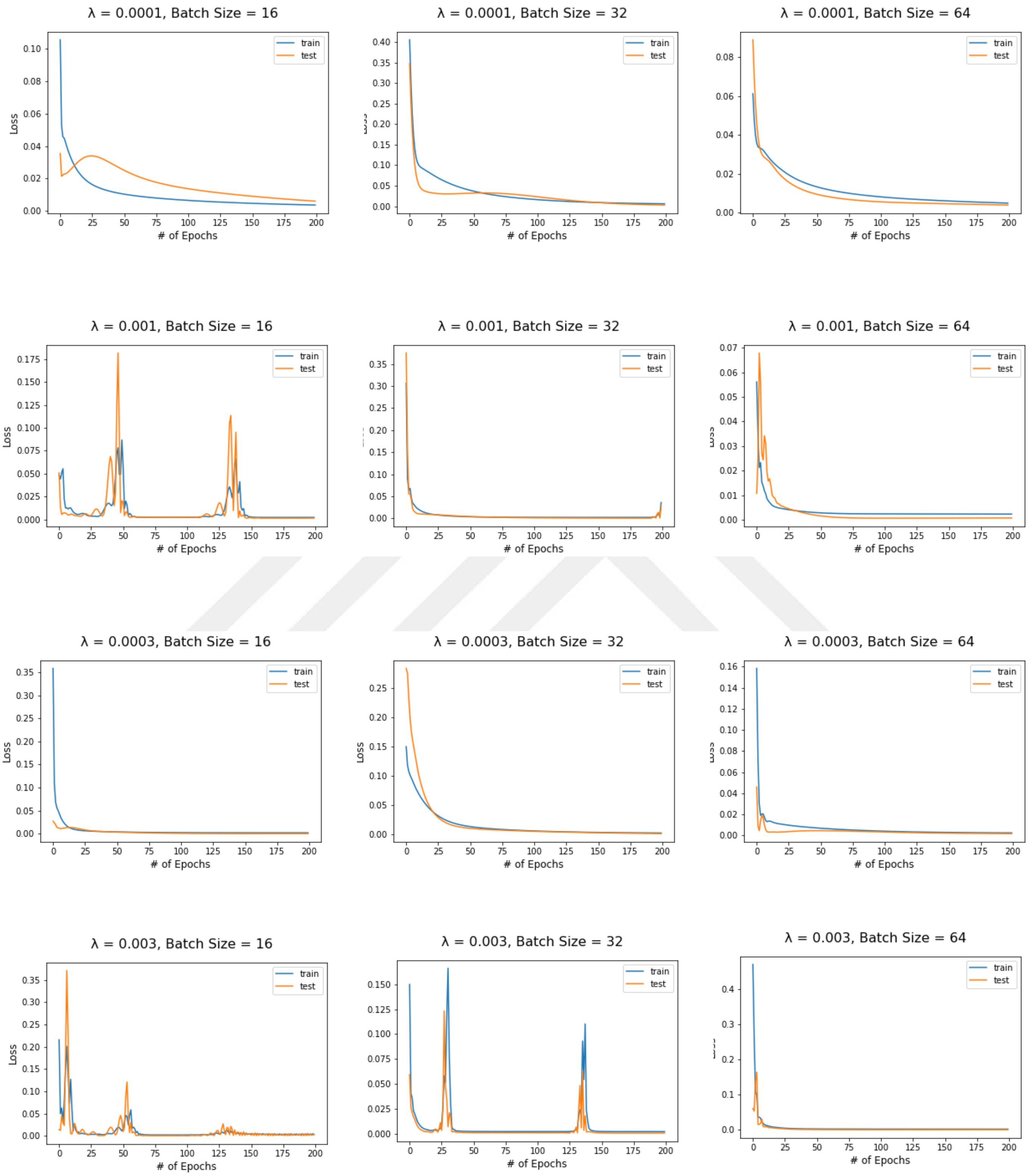


Figure A.20: The Estimated Loss for Portugal in NARX

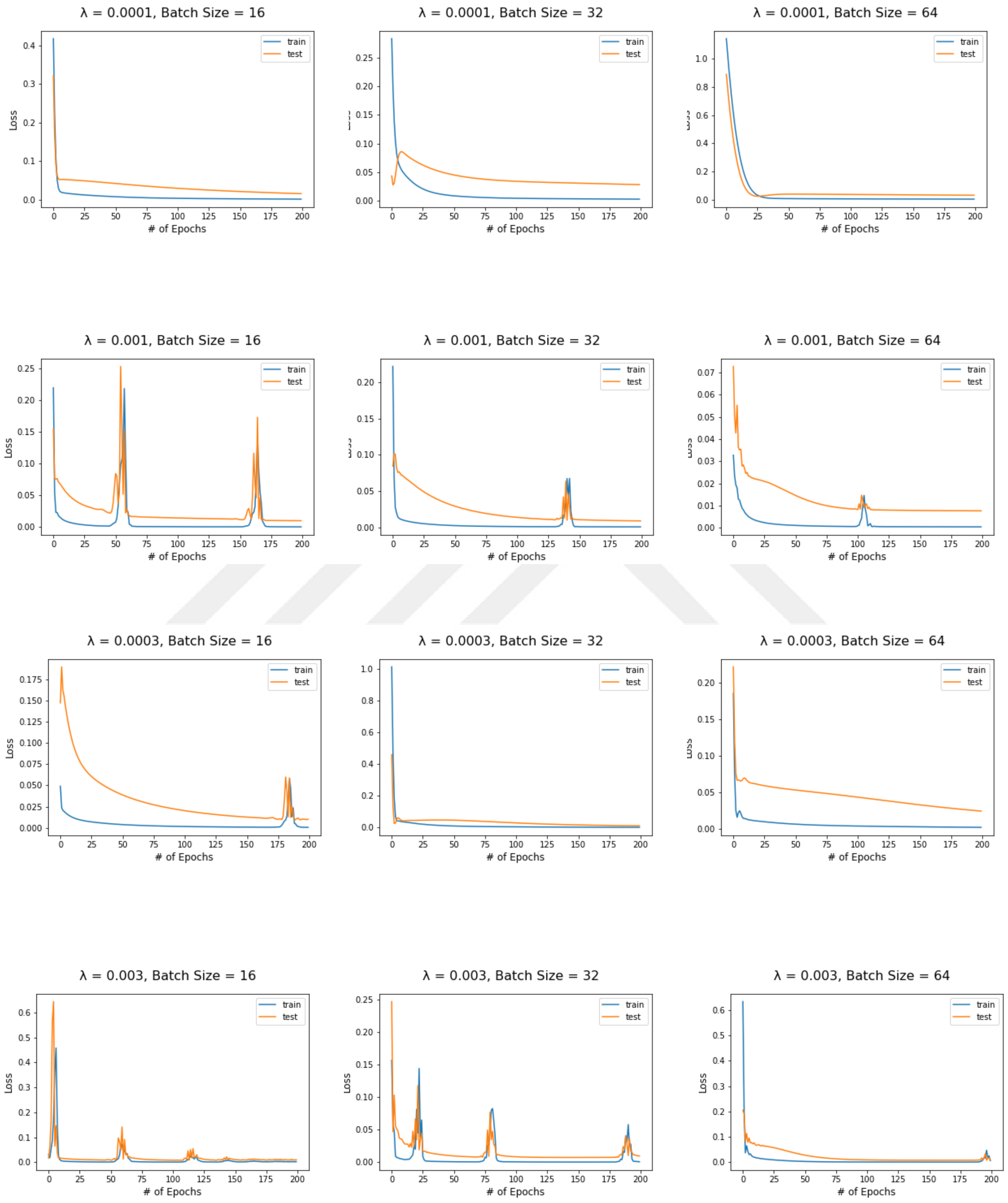


Figure A.21: SHAP Value for Austria in Feature Importance

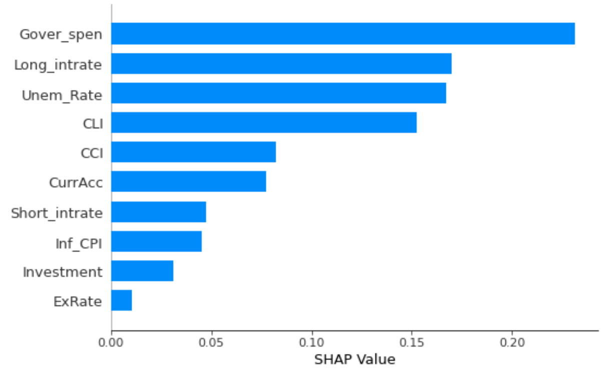


Figure A.22: SHAP Value for Belgium in Feature Importance

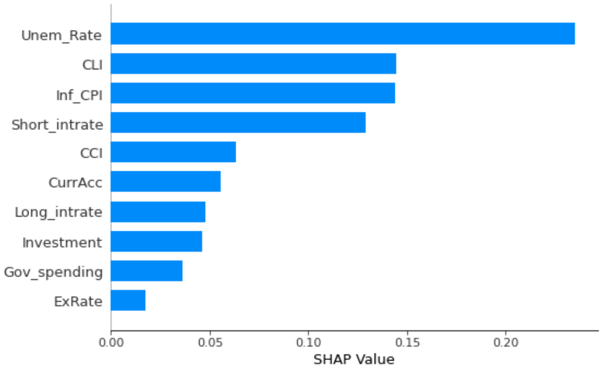


Figure A.23: SHAP Value for Germany in Feature Importance

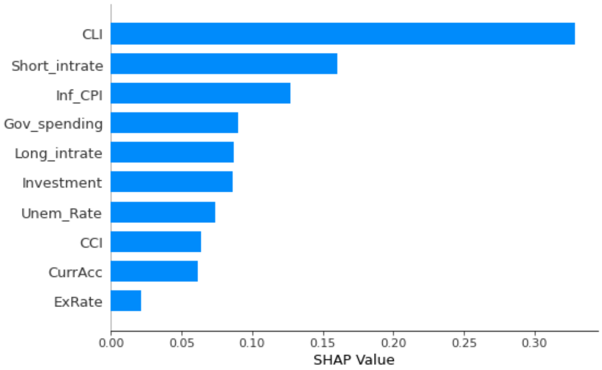


Figure A.24: SHAP Value for Spain in Feature Importance

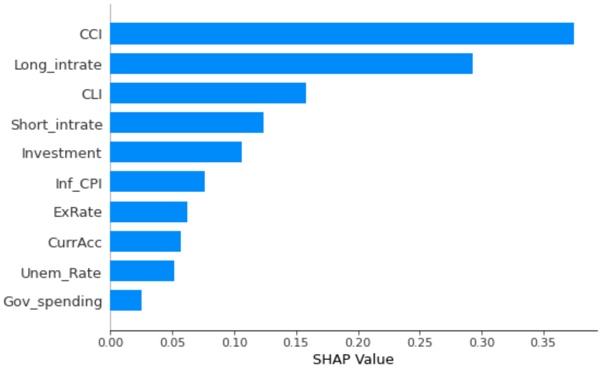


Figure A.25: SHAP Value for Finland in Feature Importance

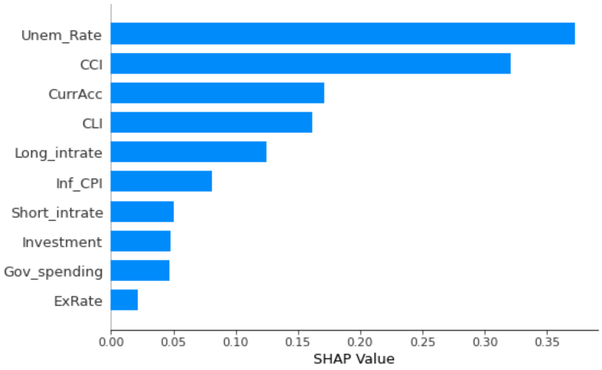


Figure A.26: SHAP Value for France in Feature Importance

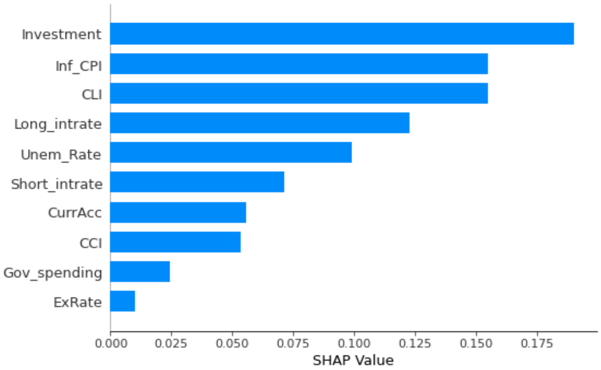


Figure A.27: SHAP Value for Ireland in Feature Importance

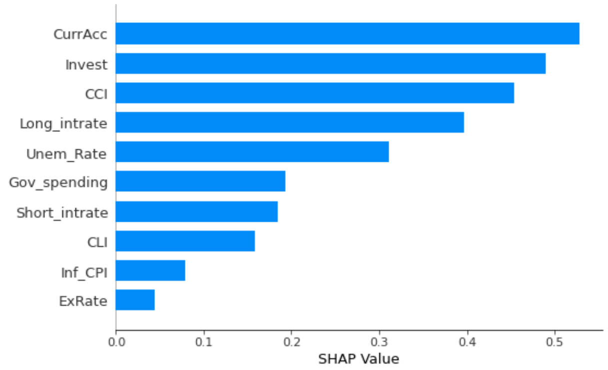


Figure A.28: SHAP Value for Italy in Feature Importance

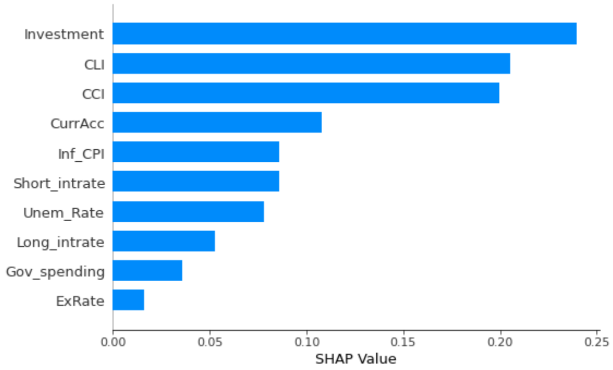


Figure A.29: SHAP Value for the Netherlands in Feature Importance

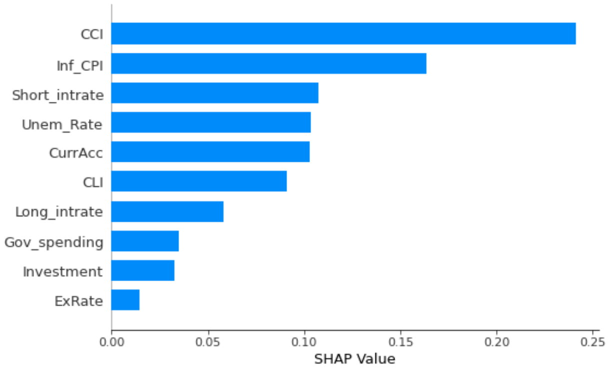


Figure A.30: SHAP Value for Portugal in Feature Importance