Erasmus University Rotterdam

Linear models and neural networks for panel data: Forecasting economic growth for European NUTS 3 regions using hybrid models

Master's Degree Thesis

Econometrics and Management Science: Business Analytics and Quantitative Marketing

Cristian Cracan - 474521

Supervisor: Prof. Dr. Wendun Wang

March 2021

Acknowledgements

I would like to address special thanks to Professor Dr. Wendun Wang (Erasmus University Rotterdam) for his supervision and support when writing this research paper. His knowledge of the field helped me identify the correct approaches from the start of the research process. Moreover, his willingness to provide very quick feedback has greatly improved the quality of the paper.

Abstract

This research introduces two new hybrid methods that can be used to produce forecasts for panel data. The two models implement neural network corrections, by using Multilayer Perceptron (MLP) and Long Short Term Memory (LSMT) on top of predictions produced by a linear model. The used practical setting is a sample of 1066 regional units from 25 countries in the European Union over a period of 16 years. This way, it becomes easier to devise a way to combine the models and to test the quality of the final predictions. In terms of results, the forecasts do not yield a high accuracy suggesting that the models do not capture the patterns in economic growth on regional levels as efficiently as hoped. Nevertheless, some results suggest that using the neural network corrections can be useful in certain scenarios.

Table of Contents

1 Introduction

This academic paper will explore the feasibility of applying two hybrid models that combine linear models with neural networks for forecasting using panel data. The proposed methodology is tested on an empirical setting of annual regional economic growth on the European level with measures of geographic proximity being considered as well. The first component of the hybrid approach is the linear panel data model that incorporates the dynamic relationships in the GDP growth values. The second component is the artificial neural network model that should identify and incorporate the non-linear patterns in the disturbances of the linear model. The relevance of applying this type of model to the proposed task can be explained through several arguments.

First of all, panel data analysis is becoming an important research field in econometrics and related fields of study (Hsiao, 2007). Given that panel data contains both cross-sectional and time series dimensions, its main advantage is the high sample variability. Models built on this type of pooled data have the capability of capturing individual heterogeneities, which can result in both more accurate estimates and forecasts of the target variable. This means that the impact of omitted variables can be controlled more appropriately. Advantages can also include simplified computational complexity and statistical inference given the larger amount of contained information. This can be useful also in the context of applying machine learning models, such as neural networks, that typically require more training observations than traditional econometric models. Hence, pooling data across both cross-sectional and time dimensions can result in more satisfactory results in terms of both inference and forecasting tasks.

With a higher availability of data in panel format and the development of analysis methods, it has become possible to estimate models for panel datasets. The literature about using linear panel data models for forecasting is also well developed. Most of them are centered around the use of dynamic relations in the dependent variables as time series usually have specific structures and patterns that can be used for both inference and forecasting. Economic growth is normally similar across multiple periods and these potential dynamic relations can be used to forecast the values of the following periods. Nevertheless, as panel models are built around incorporating heterogeneous effects, special attention needs to be given to the specification of the type of effects. Assuming fixed or random effects depends on the studied context. Moreover, the correctness of the model depends on the format of the panel as short and long panels may require different specifications.

Spatial variables based on proximity between the units in the cross-section can be used with satisfactory results based on the literature that explores variables with a geographic design. Additional explanatory variables can also be incorporated to improve the forecasting performance. Here attention needs to be paid to the feasibility of using lags of explanatory variables to build accurate forecasts in the dependent variable of interest. There is also the risk of creating a too complex linear model, which might affect the overall forecasting performance. In general, for prediction tasks, it is appropriate to test different specifications and decide based on the selected performance measure.

Furthermore, there are also limitations to what can be achieved using panel data models. The present literature is mostly focused on linear approaches that assume either fixed or random effects. For non-linear approaches, there exist multiple problems that need to be considered (Honoré, 2002). Non-linear panel data models are based on strong assumptions about the distribution of the residuals and there is no definite approach to estimating non-linear effects. In case of fixed effects, an additional limitation can be the Incidental Parameter Problem, which may lead to biased and inconsistent estimates. When this problem arises, the forecasting accuracy can suffer as forecasts are built based on biased models. Hence, additional consideration to be made on this research topic is the capacity to accurately incorporate non-linear patterns in models with panel data structure.

Compared to the field of linear panel data models, the literature about applying non-linear machine learning methods in panel data context is considerably more limited. Applications are mostly limited to building models with non-complex architectures that use lags of the dependent variable and time features like trends and dummies as features. More advanced machine learning architectures potentially can have highly accurate results if used on panel data sets. Compared to regular time series or cross-sectional data, panel data sets have two advantages that can make machine learning especially useful. Given that panels are made of two dimensions, there is a high probability of more complex and non-linear relationships existing in the data. Machine learning methods such as neural networks demonstrate high accuracy especially thanks to their capacity to build non-linear functions between a high number of explanatory variables and the dependent variables (Ahmed et al., 2010). Moreover, with two dimensions, panel data sets also have larger data set sizes. Given that machine learning models require more data for training, applying this type of methods in this scenario can produce more precise functions resulting in accurate predictions.

Moreover, hybrid models that combine the linear approach of existent dynamic panel data models with non-linear machine learning methods such as neural networks for panel data are not discussed in the literature to the best of my knowledge. In regular time series settings, hybrid models with traditional linear models followed by corrections provided by neural networks have been implemented in other research. Hence, it is relevant to test the feasibility of such hybrid applications in a panel setting. Panel data sets have attractive properties that can make machine learning applications suitable. Neural network corrections on the initial forecasts of the linear approach can provide significant improvements given the size and the complexity of panel dimensions. As cross-sectional heterogeneity and irregular patterns in the individual time series intensify, classical dynamic panel models may fail to reproduce the real world dynamics accurately. It is therefore likely that the structure of the errors of the standard econometric model would include the non-linear relationships found in the data. Therefore, it is appropriate to correct the errors in the linear predictions using a second method focused on irregular patterns. Given that machine learning methods can produce more flexible functions between variables, they can be a natural choice in providing the corrections based on the non-linear relationships left in the errors of the linear model.

In terms of the machine learning approach that can be used in such a setting, the neural networks model can potentially produce significant improvements in forecast accuracy. The method has been used extensively for time series forecasting tasks. It has been shown that even the regular architectures show satisfactory performance when working with time series patterns for macroeconomic data (Hill et al., 1996). Moreover, feedforward neural networks have also been used for prediction tasks in panel data applications with satisfactory performance (Longhi et al., 2005). Even though not used specifically for producing error corrections in a hybrid setting, the method has also been applied for macroeconomic data in panel structure. Therefore, this research focuses on applying a feedforward neural network for its first hybrid specification.

Moreover, there is a lack of applications with more complex architectures of neural network such as recurrent types. Thanks to their architecture made for learning from information of the previous time steps, these models can identify patterns efficiently in data with time series structure (Siami-Namini and Namin, 2018). In data that has patterns across time such as the panel data sets, it is important to build models that incorporate them efficiently. The recurrent neural network is applied to fit more complex relations that cannot be explained using linear models. With the limited literature on using recurrent networks for panel data, it is also fitting to test whether using these types of models can be managed in such a context with empirical economic data. Therefore, it is relevant to also use a recurrent neural network as a component of a second hybrid model.

Some limitations of both the feedforward and recurrent neural networks include a propensity to overfit in case of a large number of variables, complex architectures, or inadequate cross-validation techniques, but also potentially high computational complexity resulting in long training times. Its main limitation though is its very low interpretability. Even though good for forecasting tasks, neural networks lack robust ways of explaining the relationships between the variables. Hence, in this application the focus is strictly on forecasting accuracy, with a general discussion in the conclusion that will focus on reasons behind the performance of the proposed methods.

Lastly, it is fitting to build accurate and widely applicable forecasting tools in the context of economic forecasting for regional units across multiple countries. Given the increasingly close economic ties between countries, it is becoming possible to study variables such as economic growth in pooled models. In the context of European countries, the dynamics of economic growth for European Union members and their neighbors may be strongly related. Increasing economic and political connectedness is bound to create common growth patterns and shocks across borders and regions. Hence, spatial variables that reflect geographic and political links are relevant to be used to predict economic growth. Moreover, variables that account for membership to the common European currency or to special economic agreements can be used as predictors.

Therefore, the main research goal of this paper will be to study the feasibility of applying hybrid models consisting of dynamic linear panel models and neural networks. For this, the empirical context of forecasting regional economic growth is considered. The main research question that will be investigated is:

Do the hybrid models combining linear techniques with feedforward neural network and recurrent neural network have good forecasting performances in the context of regional economic growth in Europe?

This research paper also studies the feasibility of using the dynamic linear model with additional spatial connectedness measures for predicting the economic growth values. Hence, a sub-question is given by:

Does the dynamic panel data model with additional measures of spatial connectedness between European regions provide adequate forecasting performance of economic growth?

Moreover, given that machine learning algorithms are computationally more expensive and require a more complex model selection process, it is relevant to study whether applying the second step can be justified by the performance of the combined framework. Hence, another sub-question to be studied is:

Do the hybrid models provide a sufficient improvement in forecasting performance when compared to the forecasting performance of the individual linear model in the same context?

Last research sub-question is related to the comparison between the two hybrid models and determining whether using a feedforward or a recurrent neural network is more appropriate:

Is there a significant difference in forecasting performance between the two hybrid models in the studied panel data context?

As mentioned beforehand, the proposed methods will be applied in the context of forecasting economic growth for a large group of European countries. The measure of economic growth used in this research will be the annual year-on-year change in GDP values. The used test period will be the last three years with available data, specifically 2016, 2017, and 2018. Forecasts are built both one-step ahead and multi-step ahead to test the feasibility of predicting several years in the future. Therefore, compared to research in this field that deals with nowcasting and using very recent economic indicators to predict growth in the immediate quarter, the focus will be rather on building a model that is based on past values of growth and measures of interdependence between the regions. Amidi and Fagheh Majidi (2020) show that bilateral trade and geographic proximity are significant factors that can explain economic growth.

All sections are discussed from this point of view. First, the present literature is discussed with the focus on several subtopics. Then, the data section presents the main details about the setting and the countries of the sample. The description of the methodological process is structured in several smaller sub-sections according to the parts of the hybrid models. The results are presented in several parts with focus on the overall sample and on smaller subgroups. Finally, the paper concludes with a discussion on findings, limitations and suggestions for further research.

2 Literature

The literature related to the methods is presented in this section.

2.1 Dynamic linear panel models for forecasting

Baltagi (2008) summarizes the body of research about forecasting using panel data and presents the Best Linear Unbiased Predictor (BLUP) in the error component model, with details on how to incorporate serial correlation and spatial correlation into the structure of the errors. Moreover, incorporating heterogeneity into the model is discussed. This paper discusses several forecasting applications using panel data models. For example, Baltagi and Griffin (1997) find that by pooling the data, the forecasting performance is better than in the scenario with heterogeneous estimators.

The same findings can be seen in several other research papers working with various types of economic data, with pooled models providing considerably better forecasting performance than time series or cross-section models. In general, it is found that pooling offsets the negative effects of the biases produced by the heterogeneity especially in more long-term out-of-sample forecasting.

Additionally in another paper, Baltagi et al. (2012) describes a way to analyze spatial panel data, where there is a network dependence between units, and performs a wide study using different models. It is found that models that ignore spatial autocorrelation structures and heterogeneity in the data generating process fare worse, but this study does not take into account dynamic explanatory variables or endogenous explanatory variables. Anselin et al. (2008) discuss the increasing use of spatial panel models in econometric research and present several specifications that can be used for different scenarios, but the researchers also note a lack of implementations at that point in time.

Baltagi et al. (2014) introduce a dynamic spatial model that can be applied with good forecasting performance in a panel data context. It mixes non-spatial and spatial elements while also using instruments from previous research on dynamic GMM methods. The spatial GMM estimator introduced by the researchers considers lags of the dependent variables, spatial variables based on proximity between units in the panel set, and additional exogenous variables. The proposed model GMM-SL-SAR-RE incorporates the spatial dependence both in the regressors and in the disturbances, while also accounting for heterogeneity and endogeneity. It has the best forecasting performance both on the Monte Carlo application and on the empirical study with European regional macroeconomic variables.

Hoogstrate et al. (2000) perform a forecasting study on economic growth in a group of developed countries in a model which uses lags of regressors and other economic variables. The researchers find that the forecasts of the pooled model are more accurate than those of the individual models even if pooling restrictions are not always satisfied. Arkadievich Kholodilin et al. (2008) present a study of forecasting GDP growth using regional data from Germany. With a dynamic panel model which takes into account the spatial interdependence, the improvement in forecasting performance is shown to be significant.

In a more recent paper, a new technique for forecasting using panel data is discussed (Liu et al., 2020). The researchers introduce a method that is capable of incorporating correlated random effects in forecasts. This is done by using a two-step procedure. First, a method is used to create consistent estimates of the effects of the used explanatory variables. Then, a second method is used to create accurate estimates of the heterogeneities related to the cross-sectional units. Therefore, the proposed Empirical Bayes approach is capable of creating forecasts that incorporate consistent and accurate estimates of both pooled and heterogeneous effects. Moreover, this method is shown to be feasible for short panels with a large number of cross-sectional units and few time series observations.

2.2 Neural Network forecasting

Given their increasing popularity in both research and business related fields, Neural Networks have more applications for forecasting tasks. Specifically of interest to this research paper is the forecasting using time series and panel data sets. Zhang et al. (1998) present a wide survey of applications with time series up to that point in various research fields. The authors describe the method's capacity to incorporate nonlinear patterns efficiently and have good forecasting performance. Nevertheless, there are limitations associated with using Neural Networks. Implementing them individually in cases without much evidence of non-linearity will cause them to have worse performance than linear models. Moreover, the trial-and-error approaches to selecting the models, the required computational time, and overfitting are problems that need to be considered.

Zhang and Qi (2005) find that preprocessing of the data in the form of detrending or deseasonalization can improve the forecasting performance of the models. Moreover, Zhang and Kline (2007) present a study about forecasting quarterly time series variables using feedforward Neural Networks with various architectures and data processing approaches. They find that the models with more basic architectures and data structures perform better in general.

In terms of economic forecasting, several applications can be noted. Tkacz (2001) use Neural Network models to forecast the GDP growth in Canada and find that they have the best forecasting for year-over-year growth compared to linear models. Huang et al. (2007) explores the feasibility of using Neural Networks for forecasting financial and economic variables. They note that using Neural Networks for predicting variables like economic growth is problematic due to the limited size of the data for individual countries. A more recent paper finds that Neural Networks can have good forecasting performance for predicting economic growth when using a larger sample of countries (Jahn, 2018).

Besides traditional feedforward Neural Network architectures, there is research that studies other forms of Neural Networks such as the Recurrent types. In this family of models, the most well-known is the Long Short-Term Memory (LSTM) model. Thanks to its architectures and variations, the LSTM is not affected by the limitation of exploding or vanishing gradient as in the case of simple Recurrent Neural Networks (Greff et al., 2016). The model is capable of capturing temporal dependencies in time series data and it has been applied with good resulting performance in various areas of study, including finance and economics. A study about using LSTM for forecasting different time series shows that it has a relatively good capacity in addition to it being relatively easy to implement (Yunpeng et al., 2017). There is evidence showing that the LSTM has better forecasting performance than linear ARIMA models (Siami-Namini et al., 2018), with a more complex architecture BiLSTM having a better forecasting performance on time series (Siami-Namini et al., 2019).

Concerning forecasts with the use of panel data and Neural Networks, there is considerably less research. Nonetheless, there are several applications with feedforward Neural Networks that can be discussed. Galeshchuk (2016) considers a small panel data of exchange rate variables and uses a feedforward Neural Network to produce out-of-sample forecasts which yields a good accuracy. Longhi et al. (2005) use Neural Networks to produce forecasts of employment figures in a panel of regions in Germany with accurate results on a short-term forecasting window, but the models rely on the assumption of constant parameters over time. In another paper on forecasting economic variables in a panel context of European countries, namely bilateral trade values, Neural Networks are proved to have satisfactory forecasting performance (Nuroğlu, 2014). Lastly, Jahn (2020) study the forecasting capacity of feedforward Neural Networks on a panel data set of economic growth of 24 countries. They show that there are nonlinear patterns in the studied variables and that Neural Networks can adequately incorporate them in the respective forecasts.

2.3 Hybrid models

The use of hybrid models concerns the combination of models or approaches in one framework. In this research paper, the used hybrid models combine components from a linear approach with components of a non-linear approach. This is inspired by similar applications for time series analysis, namely those that combine linear autoregressive models with Neural Networks. Zhang (2003) propose a hybridization using an ARIMA and feedforward Neural Network for time series forecasting, where the nonlinear component is trained on the residuals of the linear one. It is found that including the Neural Network component improves the forecasting accuracy of the predictions. Khashei and Bijari (2011) improve the model by using lags of both residuals and fitted values as features for training the Neural Network. It is shown that this model has good forecasting performance in several empirical settings. Choi (2018) presents a similar approach using the LSTM method instead of feedforward Neural Networks and finds that the performance of the hybrid model on stock prices data is satisfactory. Good performance is also found when applying the model on sales data, but given the high accuracy of the autoregressive model, the LSTM improvements are marginal in some cases (Cracan, 2020).

3 Data

This section presents the main data and variables that are used to perform the research. Background information into the list of countries from which the territorial units are considered and the reasoning behind using them are presented.

As mentioned beforehand, the main application of this research paper will be the forecasting of GDP growth at the level of territorial units from a large sample of European countries. The countries alphabetically are: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden. Hence, the list includes 26 countries. All countries in the list are member states of the European Union.

The selected level of classification of regions is NUTS 3 (Nomenclature of Territorial Units for Statistics). This is the lowest level geocode standard used across European countries to denominate territorial units for which statistics and data is provided. The NUTS 3 denomination is used by official governmental and European Union mechanisms. Hence, it's a very reliable tool for measuring statistics at local level. The number of territorial units depends on the size and on the population numbers of each country. Therefore, the numbers of territorial units across countries are not balanced.

The panel data set can be obtained from Eurostat. There is information available at NUTS 3 level for Gross Domestic Product at current market prices for the time period from 2002 upto 2019, but for the majority of countries there is no data available for the last year. Hence, the considered time period is 2002-2018. As mentioned before, France has limited data. The only available GDP values for the French territorial units are from 2015 to 2019, meaning that there is considerably less information for France than for other countries. Therefore, the French territorial units are not included in the final dataset. Another issue to consider is that for two Irish regions, namely South-West and Mid-West, there is no data after 2015. Hence, these two territorial units will not be considered.

For some territorial units, the data on last two years is not yet official but estimated. In general, the estimates are close to the true values and the assumption in this study is that they can be used as replacement for values where final data is not yet available. In the figures below, maps at the start of the considered time period, namely 2002, and in the last year of the time period, 2018, can be studied.

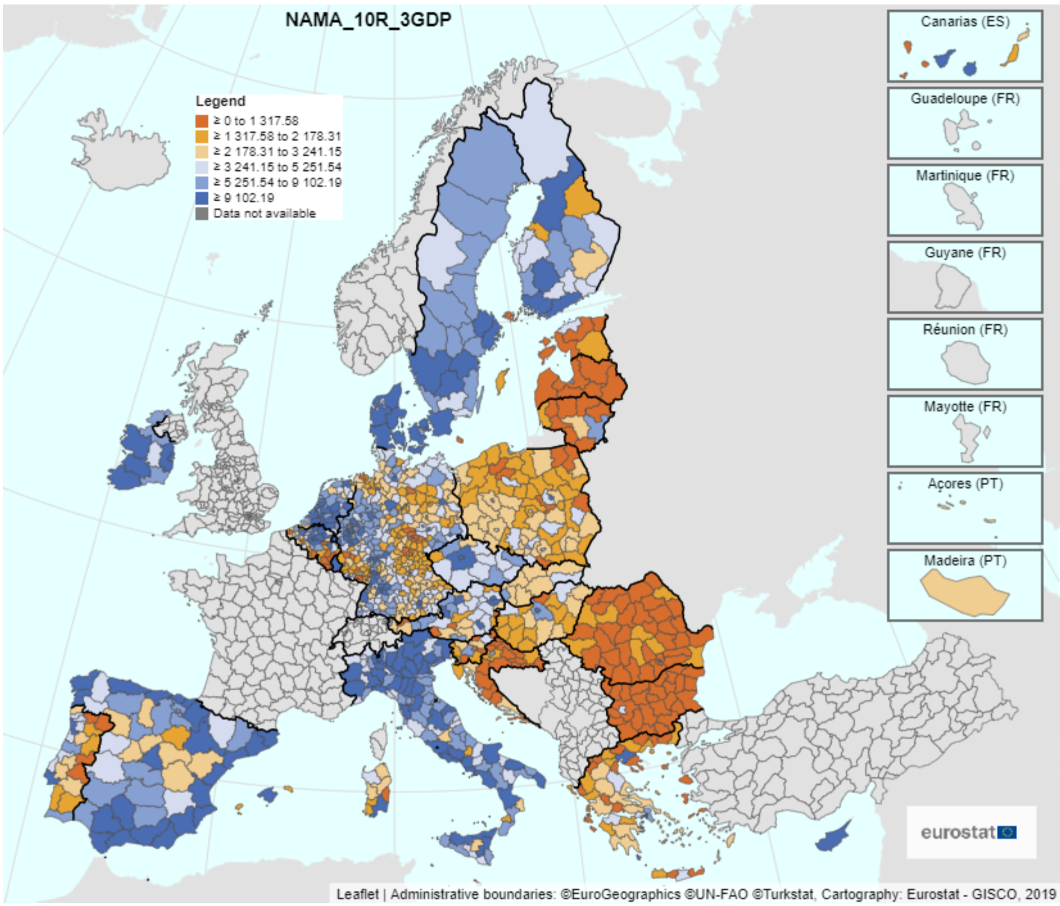

Figure 3.1: GDP values for NUTS 3 territorial units in 2002

Figure 3.1 displays the nominal GDP values in 2002 for the studied territorial units. In general, as can be seen from the map, most units are concentrated in Western and Central Europe in regions with high population density. Specifically, a high fragmentation can be seen in Western Germany and Benelux region. These regions also display the highest GDP values alongside other regions such as Northern Italy and Scandinavia. Given that this map displays GDP values without taking into consideration the population numbers, larger territorial units or for those that incorporate cities have higher GDP values.

Most territorial units in Central and Eastern Europe display lowest values based on the colour scheme. The main reason behind this is that these countries were either still in or just finished the transition period from planned socialist types of economy. Moreover, in 2002, Estonia, Latvia, Lithuania, Poland, Czechia, Slovakia, Hungary, Romania, Croatia, Slovenia, and Bulgaria were not yet official member states of the European Union. Hence at this point, they did not get the benefits of the free market and the additional investments that eventually lead to catch up growth in the following decades.

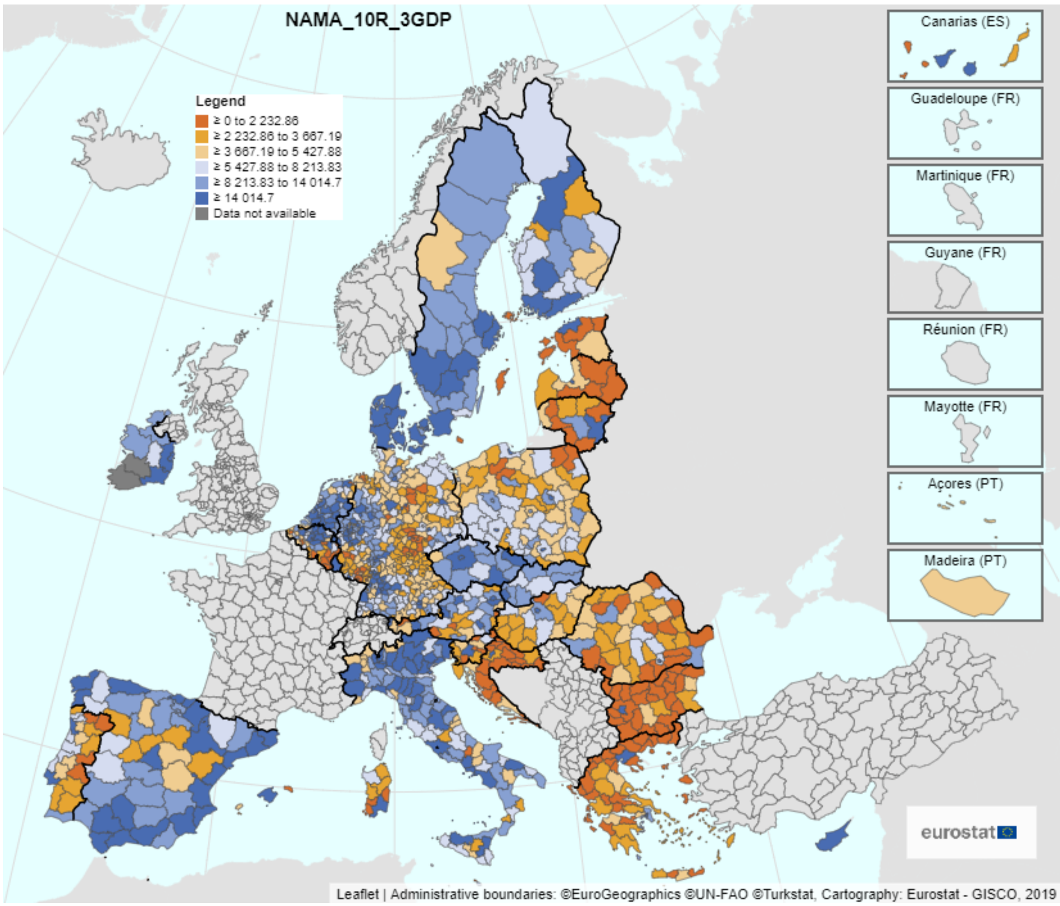

Figure 3.2: GDP values for NUTS 3 territorial units in 2018

Figure 3.2 displays the GDP values in the last year of the data set in 2018. As it can be seen based on the colour scheme, noticeable changes in the economic values across Central and Eastern European countries need to be mentioned. For example, Poland has seen an improvement in economic values especially in the Western and South Western parts of the country relative to the entire European Union. This points to higher growth than the average in these territorial units over the course of the studied period of time. A quick comparison of real GDP per capita in 2004 between countries already part of the Union and the new post-socialist countries shows that average economic demand per capita was nearly 4 times smaller (33,763 and 8,923 million EUR). In 2019, the gap was smaller with the average GDP per capita in these countries being less than 3 times smaller (39,161 and 13,804 million EUR). This means that growth in the new members was larger on average with a 'catch up effect' being observed.

In terms of EU membership status, most of the countries in the data set have been official EU members for the entire time period. In the group of studied countries, 10 have officially joined the Union in 2004, namely: Estonia, Latvia, Lithuania, Poland, Czechia, Slovakia, Hungary, Slovenia, Malta, and Cyprus, two have joined in 2007, namely Romania and Bulgaria, and finally Croatia in 2013 was the last country to join the EU as an official member. Except for Malta and Cyprus which are located in the Mediterranean, all the other countries that joined in the last decades are located in Central and Eastern Europe and have had a socialist type of economy in the past century. Including the information about EU enlargements in the models can be of interest when analyzing the rate of economic growth especially for those countries that became members during the studied period. Another thing from the dataset that needs to be considered for modelling are the downturns for nearly all countries during the period of the financial crisis and the subsequent Eurozone crisis starting from 2008. Including a form of variable that controls for the recession years can be used to prevent possible biases in the estimates from happening.

To study the relative values of economic expansion for the regions, real growth rates are constructed for the panel. To do this, first the nominal GDP values per NUTS 3 territorial units should be converted to real GDP values. This can be done using the GDP deflators on national level. Hence, the real GDP values can be obtained using this formula:

$$GDP_{real} = \frac{GDP_{nom}}{Defl} \cdot 100 \tag{3.1}$$

After obtaining the real GDP values for the entire panel of territorial units, growth values can be computed. This can be done by obtaining the % growth of GDP values compared to the previous year. Therefore, after taking the difference, there are series of 17 time values for each cross-sectional unit in the panel. When considering the territorial units for which there is data for the entire time period, the panel data set consists of $N = 1066$ cross-sectional units with $T = 17$ time series units each. Hence, the data set contains 18122 observations that can be used to train and test the models for this research. This is a big dataset that can be used for applications with Machine Learning tools such as Neural Networks, which typically require a large number of observations.

4 Methods

This section is split up into descriptions of the separate components of the hybrid model and description of the way the two are combined in the hybridization. First, the description of the linear dynamic model used for the forecasting task is presented.

4.1 Linear dynamic panel model

Given that the panel used for this research has a short format with many cross-sectional units and few time series units, this study considers the approach proposed by Liu et al. (2020). It combines several approaches to come to an accurate forecast that incorporate approximations of individual effects based on Bayesian theory. This subsection describes the initial ideas about creating the linear dynamic panel model in accordance to the method proposed by the researchers.

4.1.1 Obtaining consistent estimates

As a first step of their approach, the researchers propose some methods for obtaining consistent parameter estimates in the model, which ultimately is also key to obtaining accurate estimates of the individual effects and to building accurate forecasts for the research. They propose either Generalized Method of Moments (GMM) techniques or the Quasi-Maximum Likelihood Estimator. Given the proven success of GMM models for forecasting in panel context, this type of approach is considered in this research.

Some panel GMM models that are mentioned in the research paper and that can be considered for this type of problems are models proposed by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998). These methods are based on the idea of creating accurate estimates in panel setting with dynamic effects in the dependent variable. Given that including dynamic effects violates the exogeneity of the independent variables, using fixed effects model for either estimation or forecasting can result in inconsistent results. The Arellano-Bond estimator was developed on the idea of Anderson and Hsiao (1981), known as Anderson-Hsiao estimator, by creating an efficient estimator which uses differencing to remove the individual effects and using lags of dependent variables as instrumental variables.

The Blundell-Bond estimator improves on the initial GMM model proposed by Arellano-Bond. It is made for larger autoregressive effects and a larger variance of the individual effects relative to the variance of the errors. The Blundell-Bond estimator makes use of moment conditions, where both lagged differences and lagged levels are used as instruments for the level equation and the difference equation respectively. The creators of the estimator argue that it has less downward bias than the Arellano-Bond model. Additionally, this estimator is made specifically for panel data that contains a fixed small amount of time observations and a large number of cross-sectional units. Hence, this research considers the Blundell-Bond method for obtaining consistent estimates of the parameters in the model.

One important thing to note for the Blundell-Bond estimator and similar approaches, is the requirement that there is no serial correlation in the idiosyncratic errors. In the case with serial correlation, the moment conditions used for obtaining the parameter estimates are not valid and the model might be misspecified. One way to test this is by using the Arellano-Bond test for serial correlation in the first-differenced errors. The test checks the null hypothesis of no serial correlation. Nevertheless, at order 1, we expect to reject the null hypothesis, given that the first difference of independent identically distributed errors is autocorrelated, but rejecting at higher orders means that the model might be misspecified and the moment conditions are invalid. In this case, the model needs to be changed in order for the estimates to be consistent as required. Hence, the choice of explanatory variables is based on satisfying this requirement. Otherwise the model estimates might not be consistent and lead to innacurate forecasting results.

The main component of the dynamic linear model will be the lagged GDP growth values. This means that values of economic growth of one year ahead for each estimated period are considered in the model as the dependent variable. Given that the model is focused on dynamic relations in the variables of economic growth, the main explanatory variables will be the lagged variables of growth. The appropriate levels of the lags are considered mainly from the point of view of having consistent parameter estimates. Moreover, having more lags as explanatory variables which reduces the size of the panel is not considered a problem in this case as the chosen estimators are suited for data with a short time dimension.

As mentioned beforehand, this research also introduces a measure of influence of geographic proximity between the territorial units. This is implemented by creating an additional explanatory variable of average growth in neighbouring territorial units for each unit. Hence by using this variable, it is possible to create a direct relationship between the growth of each unit and its surroundings. For example, it would be expected that average positive growth in the neighbours would also influence the positive growth in the considered territorial unit. To incorporate this, geographic data is studied to manually create a matrix of dummy variables that indicate whether two territorial units are neighbors. Therefore, the spatial explanatory variables are built by making a weight matrix W_s of size NxN which determines whether two territorial have directly shared borders or are very close, such as Nordsjælland region in Denmark and the Skåne region in Sweden, which are separated by the Øresund strait. The elements in W_s corresponding to two neighboring territories will have values equal to 1, with other values being set at 0 including the diagonal values. This can be used to create average growth of neighboring units in the previous year for each NUTS 3 component. This measure of growth of nearby regions can be used to capture the interdependence between regions. The main assumption is that clusters of components with higher or slower growth will influence growth in nearby territories too.

In addition to these variables, this research also considers variables that incorporate periods of recession or sluggish economic growth across the EU. These periods can be reproduced by using dummy variables for the years affected by crises. In the studied period, there are two such periods. First, the financial crisis of 2008 and 2009 affected most European countries. Second, the Eurozone crisis which affected mostly Eurozone countries but had negative effects on other non-Eurozone EU members. This prolonged period of sluggish growth can be captured by creating a dummy variable for years 2011-2014. After multiple checks using the Blundell-Bond, only when using the Eurozone crisis dummy, the test results show that the consistency requirement is satisfied.

During the model selection process, it was found that the best choice of maximum lag level was 5, which managed to maintain the requirement of no serial correlation in the errors, thus keeping the model estimates consistent. Taking this into consideration, the model to be estimated using the Blundell-Bond estimator is given by the following specification:

$$y_{i,t} = \sum_{p}^{5} \rho_p y_{i,t-p} + \beta_1 \frac{1}{n_i} \sum_{j=1}^{N} w_{s,i,j} y_{j,t-1} + \beta_2 EZCrisis_{i,t} + \lambda_i + \epsilon_{i,t} \tag{4.1}$$

where n_i is the number of neighbours that territorial unit i has, $w_{s,i,j}$ is the dummy denoting whether the regional units are neighbours, and λ_i are the individual effects that will be estimated using the second part of the linear approach. The model is estimated with a robust technique for the variance–covariance matrix. The Arrelano-Bond test for serial correlation in the idiosyncratic errors as expected rejects the serial correlation hypothesis at first order and does not reject it at order 2 with a p-value of 0.3396. Therefore, it is assumed that the Blundell-Bond estimates are consistent and can be used in the second part of the approach proposed by Liu et al. (2020).

4.1.2 Forecasting individual effects

The main idea behind the the approach proposed by Liu et al. (2020) is to estimate accurately the heterogeneous effects represented by λ_i's. The assumption adopted by the researchers is that λ_i's are correlated random effects. This means that the individual effects are random but there are still possible correlations with other explanatory variables in the model. This is assumed to hold for this study, as individual effects are bound to be correlated especially with the lagged economic growth values. The implementation described in this paper follows this main assumption by combining several methods to obtain consistent estimates for all effects and the probability distribution of the heterogeneous coefficients λ_i's. The introduced empirical Bayes approach takes advantage of several specifications to obtain accurate estimates.

Building on the approach used to find consistent estimates of parameters, namely the Blundell-Bond estimator in this case, we can use the empirical Bayes formula for obtaining the estimates of the individual effects for each territorial unit in the panel. After obtaining the consistent estimates for the effects of the explanatory variables it is possible to produce a form of distribution for the heterogeneous effects λ_i's.

For this, the researchers propose to use the Tweedie's formula, which produces an approximation for the posterior distribution of λ_i's using a sufficient statistics. The formula for obtaining an estimate of the individual effect for each individual unit i is given by:

$$E^{\lambda_i}_{\theta,\pi,Y_i} = \hat{\lambda}_i(\rho) + \frac{\sigma^2}{T}\frac{\partial}{\partial\hat{\lambda}_i(\rho)}\ln p(\hat{\lambda}_i(\rho), y_{i,0}) \tag{4.2}$$

where $\hat{\lambda}_i(\rho)$ is a sufficient statistic computed as:

$$\hat{\lambda}_i(\rho) = \frac{1}{T}\sum_{t=1}^{T}(y_{i,t} - \hat{y}_{i,t}) \tag{4.3}$$

where $\hat{y}_{i,t}$ is the estimate of economic growth using the Blundell-Bond estimated parameters. The sufficient statistic serves as an average of the distance between the true values of the dependent variable and the estimated values during the studied period for each individual unit.

Hence, the entire formula can be seen as a way to correct the forecasts for each territorial unit. On one hand, the first part of the sum that contains the average difference between true and estimated values for the training sample is created to incorporate a measure of correctness of estimates for each individual unit. On the other hand, the second part of the sum takes the role of a correction which depends on the overall density of individual effects. By adding a weighted derivative of the density of approximated effects, the resulting values also take into account the dependence and relations between the units in the panel dataset.

Nevertheless to obtain values for Tweedie's formula, a way to compute the $p(\hat{\lambda}_i(\rho), Y_{i0})$ is required. For this, the researchers propose a kernel method to estimate the density $p(\lambda_i(\rho), Y_{i0})$. Namely, Liu et al. (2020) use a leave-one-out kernel of the form:

$$\hat{p}^{(-i)}(\hat{\lambda}_i(\rho), y_{i,0}) = \frac{1}{N-1}\sum_{j\neq i}\frac{1}{B_N}\phi(\frac{\hat{\lambda}_j(\rho)-\hat{\lambda}_i(\rho)}{B_N})\frac{1}{B_N}\phi(\frac{Y_{j,0}-y_{i,0}}{B_N}) \tag{4.4}$$

This kernel density is then used to obtain the final forecast formula for the test period. The forecast for next period is given by:

$$Y_{i,T+1} = [\lambda_i(\rho) + (\frac{\hat{\sigma}^2}{T} + B_N^2)\frac{\partial}{\partial\hat{\lambda}_i(\rho)}ln\hat{\rho}^{-i}(\hat{\lambda}_i(\rho), Y_{i0})]^{C_N} + \hat{\rho}Y_{iT} \tag{4.5}$$

where $[x]^C = sgn(x)min(|x|, C)$ is a function that controls the value of the individual effect in the case where the estimated values of the individual effects are either too large or too negative.

To obtain the forecast, there are three steps that still need to taken. First, the gradient of the logarithm of the kernel density needs to be computed. Using the formula for kernel, it is possible to derive the exact formula. This can be obtained as follows:

$$\frac{\partial}{\partial\hat{\lambda}_i(\rho)}ln\hat{\rho}^{-i}(\hat{\lambda}_i(\rho), Y_{i0}) = \frac{1}{\sum_{j\neq i}\frac{1}{B_N}\phi(\frac{\hat{\lambda}_j(\rho)-\hat{\lambda}_i(\rho)}{B_N})\frac{1}{B_N}\phi(\frac{Y_{j,0}-y_{i,0}}{B_N})}\cdot$$

$$\cdot\sum_{j\neq i}\frac{1}{B_N}\phi(\frac{\hat{\lambda}_j(\rho)-\hat{\lambda}_i(\rho)}{B_N})\frac{1}{B_N}\phi(\frac{Y_{j,0}-y_{i,0}}{B_N})\frac{(\hat{\lambda}_i(\rho)-\hat{\lambda}_j(\rho))}{B_N^2} \tag{4.6}$$

By using the estimates from the Blundell-Bond estimator, it is possible to calculate the value of this gradient for each cross-sectional unit in the panel dataset.

Two parameters to calibrate in this formula are the bandwidth B_N and the truncation value C_N. In terms of bandwidth, the researchers propose a value based on the Silverman's rule-of-thumb bandwidth choice. This value is given by:

$$B_N^* = (\frac{4}{d+2})^{\frac{1}{d+4}} max(\frac{1}{N^{\frac{1}{d+4}}}, \frac{1}{(lnN)^{1.01}}) \tag{4.7}$$

where d is the dimension of the density, which is 1 in this scenario. The actual bandwidth is given by $B_N = bB_N^*$, where b comes from a grid of values. The researchers use a grid spacing of 0.1 on the interval $[1, 3]$ and optimize for the lowest loss. In this study due to the high computational requirements associated with the size of the panel, the grid spacing is 0.5 and the bandwidth is selected based on minimizing the Mean Squared Error on the training sample and not on the last forecasting observation. This is done to create a more robust selection of bandwidth.

The truncation parameter C_N is used to exclude cases where the estimates of the individual effects are abnormally large or small. If these abnormalities are not treated, then the mean error might not reflect accurately the accuracy of the models. For this study, the truncation parameter is set as a constant that reflects the highest allowed change in the growth estimate.

4.1.3 Obtaining the linear model forecasts

The approaches described above are combined to create forecasts for both one-step ahead and multi-step ahead approaches. The steps for creating the one-step ahead forecasts are:

1. Use Blundell-Bond estimator to create three sets of model estimates for the three samples in accordance with the forecasts for the years 2016, 2017, and 2018.
2. Create estimates of the individual effects for each set of estimates.
3. Use equation 4.5 to create forecasts for each individual test year.

The steps for creating multi-step ahead forecasts for the test period are given by:

1. Use Blundell-Bond estimator to create one set of model estimates for the training sample up to year 2016.
2. Create estimates of the individual effects.
3. Use equation 4.5 to create forecasts for each individual test year, where lags are replaced by forecasts from the previous years.

By using the method proposed by Liu et al. (2020), it is therefore possible to obtain accurate forecasts in panel context using consistent estimators for both explanatory and heterogeneous effects. Moreover, this approach works for panel with many cross-sectional units and short time series.

4.2 Neural Networks

The artificial neural network is a Machine Learning method used mainly for forecasting tasks given its efficiency of incorporating non-linear patterns in data. The architecture of neural networks includes several layers of nodes with weighted transformations. This allows one model to build a complex non-linear function to represent the target variable usually by using a high number of explanatory variables. Hence, it can capture effects that cannot be incorporated into standard linear models.

The biggest strength of the neural networks also produces its main limitations. This model has a higher computational complexity relative to linear models. In one training process, the optimizer must fit tens or hundreds of weights like in the case of deep architectures. Moreover, the highly complex architecture can result in a high chance of overfitting. This means that the model might fit too well on training data and have sluggish performance on the test data.

By making use of the non-linear strengths of neural networks, this research hopes to create an effective combined model that corrects for the shortcomings of the linear panel model. Nevertheless, the usual pitfalls of the highly complex model need to be considered to create an effective final model. Hence, appropriate tuning processes need to be implemented for each model.

This research makes use of two types of neural networks. The first one is the MultiLayer Perceptron (MLP), which feeds the input data through the network and creates the forecasted values. The second type is a recurrent neural network, specifically the Long Short-Term Memory (LSTM) model, which is proven to have good forecasting performance on time series data.

In the following subsections, the details about the used neural networks are presented. Special subsections are also used to describe how the tuning process is considered for each technique with special considerations to the computational capabilities that make the training feasible.

4.2.1 MLP

The most well-known architecture of a neural network is a feedforward one or otherwise called Multilayer Perceptron (MLP) in research. In this structure, numeric input information is passed through an architecture with multiple layers of nodes (neurons), which always includes an input, an output layer, and a chosen number of hidden layers that connect the two.

Sets of weights that have a similar role to that of coefficients in classical regression analysis connect the layers in the network. The number of weights in one network depends on the chose number of input variables, of hidden layers and of neurons on each layer, and finally also on the number of output variables. Given the high number of such weights in just one network, their main role is to reflect the importance of the relationships found in different variable combinations for creating the predictions in the target.

The feedforward structure of the network implies that input data is passed through the layers of the entire network. These inputs are transformed using the weights that connect the layers, summed, and transformed according to the selected activation function. Moreover, the number of nodes in each layer can have a big effect on the complexity of the approximated function. Specifically, a large number of nodes on the hidden layers will produce a deep network that can approximate very well the training data, but overfitting can become a significant problem in this scenario. Activation functions are used to help the model learn more complex relationships between the inputs and outputs. There are some widely used activation functions, but there is also the option of maintaining just the linear combinations.

In general, an example of the MLP with one hidden layer can be summarized mathematically in this way:

$$y_t = \alpha_0 + \sum_{i=1}^{N_2} \cdot \alpha_i \cdot f_1(\beta_{i,0} + \sum_{j=1}^{N_1} \beta_{j,i} \cdot x_{j,t}) + \epsilon_t \tag{4.8}$$

where N_1 is the number of features without including the bias term, and N_2 is the number of neurons in the hidden layer without accounting for the bias node.

To optimize the weights that connect the nodes in all the layers of the neural network, the MLP weights are fit with the use of the backpropagation algorithm. This backpropagation technique works based on the gradient of the loss function. With regression tasks concerning numeric outputs, most usually measures such as Mean Squared Error (MSE) or the Mean Absolute Error (MAE) are used for the optimization process. The weights in the networks are fit according to the minimization of the selected error function to create an accurate fit on the data. The name of the algorithm itself comes from the fact that the weights are optimized starting from the output layer after one feedforward pass and after predictions are obtained.

The feedforward-backpropagation mechanism is repeated for a specified number of epochs or until a stopping condition is met. To optimize the weights in the approximated function and fit the network, several optimization algorithms based on gradient descent can be used in the applications with neural networks. One of the most used optimization algorithms especially for Deep Learning applications is the Adam optimizer.

4.2.2 LSTM

The Long Short-Term Memory (LSTM) model is part of the family of recurrent neural networks (RNN) with a basic representation of an RNN in Figure 4.1. The visuals and explanations are based on the article by Olah (2015). The difference between the MLP and the RNN is the capacity of the latter to train on outputs from previous iterations. The main limitation of working with the basic RNN structure is the vanishing or exploding gradient when considering outputs from several steps before. Long-term dependencies might not be considered properly for training, which can lead to suboptimal results.

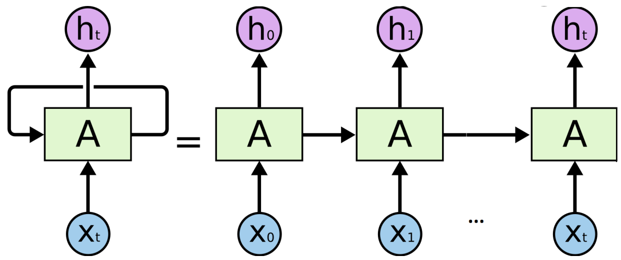

Figure 4.1: Basic RNN architecture

The LSTM model incorporates results from the past using a special cell architecture. One cell of the model has several parts which work together to diminish the risk of the exploding or vanishing gradient problem found in applications with basic RNN architectures. Figure 4.2 presents a simplified version of the cell for a better interpretation of the model, which is used to explain the individual parts.

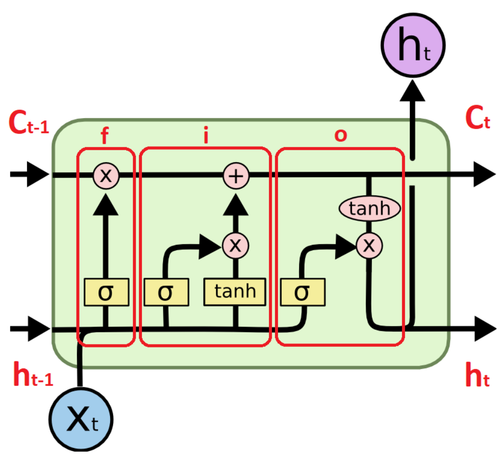

Figure 4.2: LSTM cell architecture

The LSTM is efficient when working with time series data thanks to its special architecture with multiple types of components. The cell state C_t connects the states and can be modified with each time stamp. Hidden state h_t is the output of the cell. C_t and h_t have size equal to the number of hidden nodes, the choice of which affects the level of complexity of the network.

First, the **forget section** (f part in the Figure 4.2) determines which part of the information from the previous cell is excluded. Values from previous cell h_{t-1} and new data x_t are passed into a neural network layer with a sigmoid activation function. The output of the layer is multiplied with the cell state and only a part of the previous cell state is preserved.

Next, the **input section** (i part) adds new information to the cell state. h_{t-1} and new data x_t are passed through a neural network layer with a $tanh$ function and then passed through a forget gate. Preserved values are ultimately added to the cell state C_t.

The **output section** creates the connection to the next cell and the output of the cell. A copy of C_t is transformed using the $tanh$ function and then it's passed through another forget gate determined by h_{t-1} and x_t. h_t is computed and passed to the next cell as hidden state and as output.

The process can be expressed mathematically as following:

$$\text{Forget: } C_t = C_{t-1} \cdot sigmoid(w_{f,0} + w_f \cdot [h_{t-1}x_t]) \tag{4.9}$$

$$\text{Input: } C_t = C_t + tanh(w_{i2,0} + w_{i2} \cdot [h_{t-1}x_t]) \cdot sigmoid(w_{i1,0} + w_{i1} \cdot [h_{t-1}x_t]) \tag{4.10}$$

$$\text{Output: } h_t = tanh(C_t) \cdot sigmoid(w_{o,0} + w_o \cdot [h_{t-1}x_t]) \tag{4.11}$$

where w_f, w_{i1}, w_{i2}, w_o and the biases are the weights to be updated.

The training process in LSTM is more complex than for the MLP, but it follows a similar process with a forward pass through the cells and backpropagation. The loss function is the sum of losses for all time stamps which is done by applying the Chain Rule and summing the gradients. The weights are updated using the optimization algorithm of choice with Adam being often used for Deep Learning applications.

4.2.3 Neural Networks cross-validation

The design of the training process for both the MLP and LSTM is created in order to reduce the risk of overfitting and to maximize the forecast accuracy while using the heterogeneity from the sample of regions. Given that the neural networks in this research are based on the results and errors of the linear panel model, the tuning is process is based on the residuals obtained from the first method. There is a difference in the process between the one-step and the multi-step ahead approaches.

The one-step ahead approach should work according to the results obtained in the linear model for one-step ahead forecasts. Hence for each of the three years in the test set, there is a separate set of residuals that should be used for training the model. Hence, for each of the three sets of residuals from the linear model, the neural network is tuned and subsequently trained with the best specification to make the prediction of the error in economic growth for the next year. In case of the multi-step ahead forecasts, the process only makes use of the first set of residuals, given that only information before the entire test set period can be considered. The neural network is again tuned and trained on this set, and subsequently dynamic predictions for all three years are produced using the trained model.

This research also uses a modified version of cross-validation that makes use of the heterogeneity in the sample of territorial units when tuning the model. This increases the probability of creating a generalizable model with satisfactory outcomes on the out-of-sample data. In settings with panel datasets such as in this research, the cross-validation method needs to take into account the continuity in both cross-sectional and time series dimensions. Hence, selecting random samples of observations in the set to create test sets is not feasible in the panel setting. To consider both the cross-sectional and time series dimensions, we use a modified version of a 4-fold cross-validation. A random selection of 25% of the cross-sectional units is selected to be in the test set in each iteration. In the following iteration, the units that have already been used in the previous iterations cannot be selected again. Therefore, the cross-validation method uses all the units in the panel. Moreover, for those units selected in the test set, only the residuals in the last two years are used for testing. For example for the first set of linear model residuals, the test set consists of the residuals for the years 2014 and 2015 for the selected test panel units and the training set consists of all the other observations.

This type of cross-validation is preferred to other types of cross-validation based on rolling window that are used for time series data. The main reason behind this is that with limited amount of data points per individual cross-sectional unit, this method allows us to test the model specification on data samples of different territorial units without breaking the structure of the data. Ideally, to create a model that is more resilient to dynamics in the out-of-sample window, it would be appropriate to cross-validate on several windows for each unit. Hence, this is a limitation of the methodology used in this research.

Based on this cross-validation procedure which uses different sets of residuals from the linear model, the best specifications for the neural networks are found and subsequently used to forecast the errors of economic growth in the actual out-of-sample years. Overall, given that the models are tuned and tested on random selections of territorial units, the methods should produce unbiased models with good overall forecasting performance of the error corrections.

4.2.4 Selection of Features and Hyperparameters

To apply both neural network specifications for this application using panel data, several considerations need to be made. First, the set of input features needs to be determined. Given that the neural networks in this research will work as a second step for the forecasts of the linear model, the considered input data consists mainly of the residuals of the first model. Hence, the main objective of the function fit by the network will be to capture the remaining patterns in the time series for each territorial unit in the dataset.

Based on the assumption that there are non-linear patterns in the time dimension of economic growth data, this research expects that the neural networks will be able to incorporate valuable patterns in the residuals by including lags as input variables. Based on tests using different types of inputs, this research identifies that the most appropriate number of lags is four. Even if this means that there are less years available for the training set, this model has the highest prediction accuracy on the training set.

In addition to the residual lags, the model also considers the variables from the linear model as features. Hence, additional variables to be used as features in the neural network models are the dummy variables for the years of the Financial and Eurozone Crises. These are used to capture any non-linear influence of the two crises on the economic growth in those periods. Another feature is also the effect of neighbours on the economic growth for each territorial. Compared to the linear model, the neural networks use the average residual level corresponding to the neighbours of each considered region. Moreover, given that their architecture can handle a large number of features, this model also includes dummy features for all countries in the sample. In the panel format used in this research, the country dummies are supposed to incorporate linear or non-linear effects of the country-specific dynamics in economic growth. For example on the European level, there are big differences in economic growth between countries in the initial group of EU members and the members in the Central and Eastern Europe who joined in later stages. The list of features is displayed in the table below.

After determining the final model to be used in the neural networks, a hyperparameter tuning process is required to achieve a good performance. This research will focus on identifying the appropriate architecture for both the MLP and the LSTM. For the latter, the architecture choice refers to the neural network layer used in each cell of the model. Following several trials based on the resulting RMSE of forecasts with at least two hidden layers, it is found that the most feasible number of layers is one. With more than one layer there is a high number of possible architectures

Table 4.1: Table of features used in the neural networks

Feature
Residuals with lags of 1 to 5
Average lagged residual in neighbouring territorial units
Dummy for Eurozone Crisis
Dummy for Financial Crisis
Dummy for country of territorial unit

to consider without a significantly big improvement in accuracy. Hence, optimizing the architecture of neural networks with one hidden layer also satisfies the limited computational capabilities of this research. Therefore, the focus will be specifically on optimizing the hyperparameters of the nodes in the hidden layer and of the types of activation function in the hidden layer for the MLP. For the LSTM, the only hyperparameter to tune is the number of nodes in the layer as the types of activation functions are already fixed in the architecture of the neural network. For the output layer, the activation function is fixed to be a modified version of the tanh function given by:

$$tanh5(x) = 5\frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{4.12}$$

This function restricts the predicted values of the network to be in the interval between -5 and 5 to restrict the risk of very large error corrections that would bias the final combined forecasts. After tests, this limit was found to produce the most reliable results for the entire set.

In terms of the learning process, some parameters that can also be tuned like the number of epochs, the loss function, and the gradient descent procedure are kept fixed due to computational limitations. Hence, the training always uses the fixed 50 epochs, the Root Mean Squared Error loss function, and the Adam procedure for optimizing the weights in the network.

For finding the best specifications for the neural network architectures, this research looks at two metrics. The first one is the error metric. Chosen in this context is the Root Mean Squared Error (RMSE). This metric is used thanks to its high interpretability in this context, as it displays the average deviation from the true value of economic growth in the dataset. A smaller RMSE indicates that the model specification has a better forecasting performance on average. The second metric for choosing the right specification is how well a model can capture the sign of the residual. Most residuals from the linear model are highly volatile with both positive and negative values.

Hence, it is very important that the neural network models are capable of predicting the sign of the error in order to provide an improvement in forecasting performance over the linear model. For this, another metric to look at when choosing the final neural network specification is the ratio of predictions in the test set that have the correct sign. A higher ratio means that the model is better at identifying the direction of the residuals, which means that we can expect an improvement in the overall forecasts over the forecasts of the linear model.

The tuning process results for the **MLP** indicate that the best hyperparameters are given by the combination of 50 nodes on the hidden layer with a tanh activation. Such an architecture even though less complex by architectures with more nodes and several hidden layers provides satisfactory results both in terms of training times and prediction error accuracy. The full results of the hyperparameter tuning process can be seen in the Appendix.

The tuning process results for the **LSTM** indicate a very similar result to the MLP. The most appropriate number of nodes in the hidden layer for which both used metrics produce good results is 50. More detailed results about the tuning results are presented in the Appendix.

4.3 Hybrid model

The proposed hybrid model will follow the main lines of hybrid models that have been applied to time series forecasting. In particular, it will build on the performance of the dynamic linear model by considering lags of residuals as training features for the non-linear Machine Learning component, which is either the feedforward neural network or the LSTM in this research. The algorithm of the hybrid approach can be written as follows:

1. Build the linear dynamic spatial model on the GDP growth values.
2. Fit the linear model. Obtain residuals for the entire training data set and forecasts for the out-of-sample window.
3. Tune the MLP/LSTM with the residuals as training sample, while maintaining the pooled structure for the data set.
4. Train the tuned model architecture and obtain forecasts of the errors for the out-of-sample window.
5. Add the forecasts of the linear model with the forecasts of the errors from the neural networks. The final forecasts are given by: $\hat{y} = \hat{L} + \hat{N}$.

The main idea of the hybrid approach is to study the data twice using two models that would cover both linear and non-linear patterns. If the proposed linear specification successfully incorporates the linear dynamic relations and the linear relations covered by geographic and economic connectedness, then only the non-linear patterns are left to be incorporated by the neural networks. The role of the second component is hence to correct the forecasts of the linear component using predicted residuals. The final predictions should therefore reflect both the linear and non-linear dynamics of the studied variables. One thing to note about the hybrid model is that its two parts are estimated/tuned separately. This is mainly due to the difference in modelling using dynamic panel estimators and neural networks. As mentioned before, the used linear estimator requires several assumptions in order to obtain biased and consistent final estimates. Hence, there is less space to modify the model specifications according to a cross-validation technique and only one model satisfying all assumptions is used. The subsequent steps using neural networks are based on the set of residuals obtained from this dynamic panel model.

4.4 Evaluation of forecasting performance

The models are evaluated in two scenarios. For one-step ahead forecasts, information on the previous year is assumed to be known. For multi-step ahead forecasts for three years ahead, dynamic forecasts are built based on predictions of the previous period where the same trained neural network model is reused without training for each year, meaning that an indirect forecasting method is used. As mentioned before, the forecast period covers the years of 2016, 2017, and 2018. The performance of the models is measured using two metrics. As in the case of the tuning process for the neural networks, the main metric to compare forecasting performance will be the Root Mean Squared Error (RMSE), which provides a convenient relative measure of accuracy when working with regression problems. The second metric will be the share of correctly predicted signs in economic growth. Given that in terms of economic growth, it is important to know whether a downturn or positive growth will occur in the next periods, a good model needs to predict the direction of economic growth as closely as possible.

As mentioned in the research questions, this paper will study the feasibility of applying the hybrid approach to the forecasting task on the European level. To do so, three levels for comparing the accuracy of the models are considered. First, the overall performance of the three considered models is compared on the two prediction horizons in terms of average errors and average ratios of correctly predicted signs. Second, the performance of the models is compared across countries and across geographic groups in Europe. This should allow us to compare the relative errors of predictions in countries with varying economic growth numbers. Third, this research considers the level of correction that the neural network models provide to the linear model forecasts. This is done by measuring the average improvement in prediction accuracy across regional units and across countries of the dataset. In doing so, it would also be possible to find out whether certain elements of the models can be improved or added to have better predictions.

5 Results

For obtaining the results of the research, namely the final predictions of economic growth, several software packages are used. For the process of model selection and obtaining the coefficients of the appropriate linear model, Stata packages are used. Python is used for creating the subsequent forecasts and the individual effects used in the linear forecasts. Subsequently, the Keras package in Python which is built on tensorflow, is used for training the neural networks and obtaining the final forecasts and the metrics discussed in this research. The required data preparation and transformation tasks are also performed using Python.

5.1 Results on the entire set

This subsection concerns the forecasts obtained for the entire set of 1066 territorial units on the out-of-sample years: 2016, 2017, and 2018. The forecasts of economic growth for the units in the considered European countries are discussed using the metrics mentioned in the methods section earlier. To discuss the overall error, the average RMSE values for the three models are presented with a dinstinction between the one-step ahead and the multi-step ahead techniques in the table below.

Table 5.1: Average RMSE on the entire sample

	One-step ahead	Multi-step ahead
Linear model	3.764	3.695
Hybrid-MLP	**3.532**	3.808
Hybrid-LSTM	3.831	**3.677**

The results in the Table 5.1 show that overall the models have a high RMSE on the out-of-sample economic growth values. The linear model has a high deviation from the true values of economic growth in both one-step ahead and multi-step ahead. If the RMSE would be translated to % values in economic growth, then the model would on average be off by 3.76%, which would be a significant error when forecasting the yearly economic growth. The hybridization using the MLP produces better forecasts than the linear model in the one-step ahead context, but fails to do so for the multi-step ahead forecasts. This indicates that the MLP corrections are more accurate for short term predictions. The hybrid model with LSTM follows more closely the values of the linear model and it only marginally improves the forecast accuracy of the linear panel model. Hence, the overall results show that the there is a relatively high forecast error and the corrections provide marginal improvements in prediction accuracy with the MLP being more suitable in the short term technique while the LSTM provides a positive correction for predictions in long term.

Table 5.2: Ratio of correctly predicted signs on the entire sample

	One-step ahead	Multi-step ahead
Linear model	71.3%	72.9%
Hybrid-MLP	**74.9%**	66.1%
Hybrid-LSTM	69.8%	**75%**

In terms of the capacity of the models to capture the correct direction of economic change, the Table 5.2 displays a similar picture to the one presented by the ranking based on the average RMSE. In terms of one-step ahead forecasts, the hybrid-MLP model has the highest percentage of forecasts with a correctly predicted economic growth sign. This means that the MLP corrections not only improve the overall error, but also capture the correct economic change signs for most regional units. In terms of the multi-step ahead technique, the hybrid-LSTM provides the best ratio of predictions with correctly predicted signs. The hybrid model with the MLP in this scenario has the worst performance, meaning that its corrections prove to actually have a negative effect on the forecasts of the linear model.

5.2 Country results

As introduced, the dataset contains data from territorial units from European countries with different development and economic levels. For example, some countries in the Central and Eastern Europe have higher economic growth values than the members in Western Europe. Given the variance in economic growth values for both training and test sets along national lines, it is also relevant to compare error metrics between the territorial units based on national entities. Hence, for better interpretability of the results, the forecast metrics are presented in tables according to the countries that they are part of.

Table 5.3: One-step ahead forecast metrics on national level

	Linear model		Hybrid-MLP		Hybrid-LSTM	
Country	RMSE	Ratio	RMSE	Ratio	RMSE	Ratio
Austria	2.44	83%	2.34	83%	2.56	81%
Belgium	2.04	74%	1.90	81%	2.12	74%
Germany	3.59	73%	3.37	80%	3.61	70%
Ireland	11.09	44%	9.55	44%	11.19	28%
Luxembourg	3.06	100%	1.57	100%	3.69	33%
Netherlands	2.60	88%	2.37	88%	2.62	87%
Cyprus	3.75	100%	5.54	67%	3.64	100%
Greece	3.66	41%	3.67	50%	3.92	46%
Portugal	3.08	85%	3.04	76%	2.95	83%
Spain	2.76	89%	2.81	85%	2.88	86%
Italy	2.23	72%	2.00	59%	2.33	70%
Malta	9.46	0%	4.97	100%	9.99	0%
Bulgaria	6.21	51%	5.52	67%	6.12	52%
Czechia	3.36	98%	3.92	95%	3.44	98%
Estonia	4.47	87%	4.84	87%	4.53	87%
Croatia	3.69	76%	4.98	57%	3.77	75%
Hungary	4.45	38%	2.99	87%	4.44	40%
Lithuania	2.51	67%	2.62	73%	2.59	73%
Latvia	4.31	44%	4.03	67%	4.23	44%
Poland	4.58	71%	4.19	70%	4.64	72%
Romania	5.97	64%	5.78	80%	6.21	58%
Slovenia	2.54	100%	2.03	97%	2.70	92%
Slovakia	7.29	67%	4.87	67%	7.77	71%
Denmark	2.90	67%	2.56	88%	3.09	70%
Finland	2.92	72%	3.28	65%	3.20	74%
Sweden	4.47	51%	4.66	44%	4.38	51%
Average	4.21	69%	3.82	75%	4.33	66%

In terms of the one-step ahead forecasts, the hybrid-MLP model provides on average the best results in terms of both the average RMSE and the ratio of correctly predicted signs. Hence, as shown before for the entire sample of territorial units, the hybrid model with MLP corrections produces the most accurate forecasts when predicting one year ahead. The hybridization with LSTM corrections fails to improve the prediction metrics of the linear model on average. Given the high heterogeneity in economic growth values across the countries of this data set, the prediction metrics are also highly contrasting when comparing between different nations. Some countries with notably high forecast accuracy for the hybrid-MLP models are Belgium, Luxembourg, Italy, Slovenia, and Austria. Apart from Slovenia, these countries have a highly developed economy and have lower values of growth than the emerging markets. On the other side of the spectrum, Ireland, Cyprus, Malta, Bulgaria, and Romania, have the highest forecast errors for the hybrid model. This can be explained by the more volatile growth of territorial units in these countries in the considered period. In general, countries whose regions experienced higher values of growth also display more forecast volatility even one-step ahead. When comparing to the base linear model, the hybrid-MLP model provides an improvement for 17 countries in terms of average RMSE and an improvement for only 10 countries in terms of the ratio, which means that the average results are heavily skewed by a few countries with big improvements such as Luxembourg and Malta. Therefore, the improvements provided by the MLP corrections are not as consistent as hoped with further analysis required.

For multi-step ahead forecasts, the average of results on national level point to the hybrid-LSTM model having the best forecast accuracy in terms of the average RMSE and the ratio of correctly predicted signs. Nevertheless, the differences between the hybridization and the base linear model for this forecast dimension are more negligible than in the case of the one-step ahead predictions. An interesting observation is that the hybrid-MLP produces worse accuracy metrics than the initial model. This means that its corrections on the three-year ahead dimension fails to produce improvements. For the hybrid-LSTM model, the countries with the lowest average errors are Belgium, Italy, Lithuania, Cyprus, Luxembourg, Austria, and the Netherlands. Therefore, the model has good results on countries with differing characteristics, but its metrics are very similar to the ones of the linear model as it fails to produce big improvements. Same can be said about the countries with the biggest average errors as the errors per country are similar across the models. Overall, the LSTM corrections produce an improvement for 18 countries of the sample in terms of the average RMSE and for 12 countries in terms of the ratio. Even though, the numbers are comparable to the improvements of the hybrid-MLP model in the case of one-step ahead forecasts, the improvements are generally marginal for most countries, with the largest one being 0.6 for Luxembourg. These results point to the fact that hybrid models fail to deliver significantly better results than the linear model predictions on a horizon of three years ahead.

Table 5.4: Multi-step ahead forecast metrics on national level

	Linear model		Hybrid-MLP		Hybrid-LSTM	
Country	RMSE	Ratio	RMSE	Ratio	RMSE	Ratio
Austria	2.43	82%	2.83	65%	2.37	86%
Belgium	2.06	75%	1.94	77%	2.02	82%
Germany	3.55	74%	3.42	74%	3.50	79%
Ireland	11.23	50%	9.64	56%	11.37	39%
Luxembourg	2.92	100%	1.40	100%	2.32	100%
Netherlands	2.55	87%	2.81	73%	2.39	90%
Cyprus	2.43	100%	5.95	67%	2.27	100%
Greece	3.57	40%	4.44	44%	3.59	44%
Portugal	3.01	84%	3.93	47%	2.99	84%
Spain	2.70	90%	3.38	60%	2.64	90%
Italy	2.17	73%	2.34	44%	2.13	66%
Malta	7.17	33%	3.05	100%	6.92	50%
Bulgaria	6.13	56%	5.53	65%	6.06	62%
Czechia	3.38	98%	5.16	74%	2.81	98%
Estonia	4.69	87%	5.08	87%	4.30	87%
Croatia	3.38	86%	6.42	29%	3.46	87%
Hungary	4.28	47%	3.11	88%	4.17	53%
Lithuania	2.34	67%	2.39	77%	2.20	73%
Latvia	4.02	50%	4.73	72%	3.70	61%
Poland	4.76	73%	4.33	73%	4.93	73%
Romania	5.90	71%	5.95	79%	6.04	65%
Slovenia	2.29	100%	3.09	92%	2.29	100%
Slovakia	5.94	75%	4.10	75%	5.59	75%
Denmark	2.73	73%	2.62	91%	2.79	76%
Finaldn	2.79	74%	3.60	46%	2.98	72%
Sweden	4.24	51%	5.65	37%	4.22	51%
Average	3.95	73%	4.11	69%	3.85	75%

5.3 Measuring the correction effect of the hybrid models

This section is used to study the improvements in accuracy metrics of the final hybrid forecasts relative to the base linear model forecasts. This way it is possible to conclude whether the neural network yield positive and homogeneous improvements in prediction accuracy across the geographic sample. This is done by comparing the percentage of regional units per country for which average forecasts have improved by applying the corrections. Hence, if a country has the measure of 60%

then for 60% of territorial units in that country the forecasts of the hybrid models are more accurate than the forecasts of the linear model in terms of RMSE.

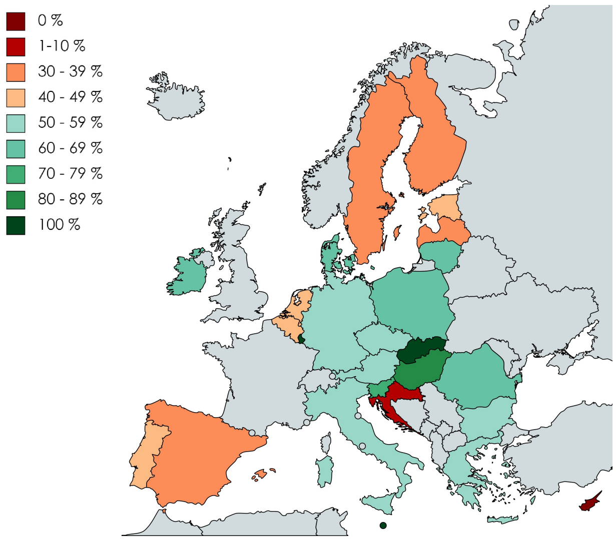

Figure 5.1: One-step ahead: Hybrid-MLP

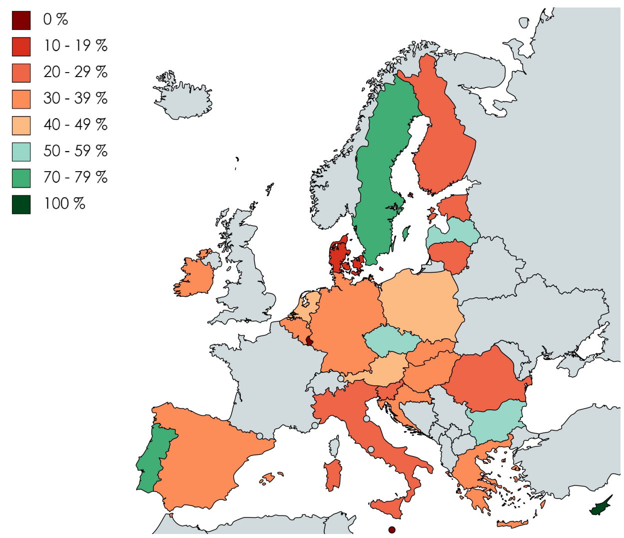

Figure 5.2: One-step ahead: Hybrid-LSTM

For the one-step ahead techniques, the hybrid-MLP model produces moderate improvements over the predictions of the linear panel model. Its corrections improve the accuracy for 553 territorial units out of the 1066 units. This translates to 52% of the units in the dataset. The average improvement in RMSE relative to the base model for the 553 units is 0.89. Nevertheless, there is variation across countries in the level of improvement. Figure 5.1 shows that more than half of the countries in the sample have improvements larger than 50%. Especially in the Central and Eastern Europe, there is a high level of regional units with positive changes in accuracy. Also, in some large countries such as Germany and Italy, the model successfully provides a decrease in error for most regions. This means that the model is capable of capturing patterns of growth that improve the predictions especially in countries with high growth. The hybrid-LSTM provides a worse performance than the hybridization using the MLP. It improves the forecast accuracy in the case of 410 regional units, which constitutes around 38% of the entire sample. This number suggests that the predictions of the linear model are still better for most cross-sectional units, which is confirmed by Figure 5.2. Only a limited number of countries in the dataset have an improvement percentage of more than 50%. Moreover, for units with more accurate forecasts produced by hybrid-LSTM, the average improvement in errors is only 0.23.

For the multi-step ahead forecasts, the hybrid-MLP model produces an improvement in the measured error for 459 territorial units. This hybridization manages to produce positive corrections

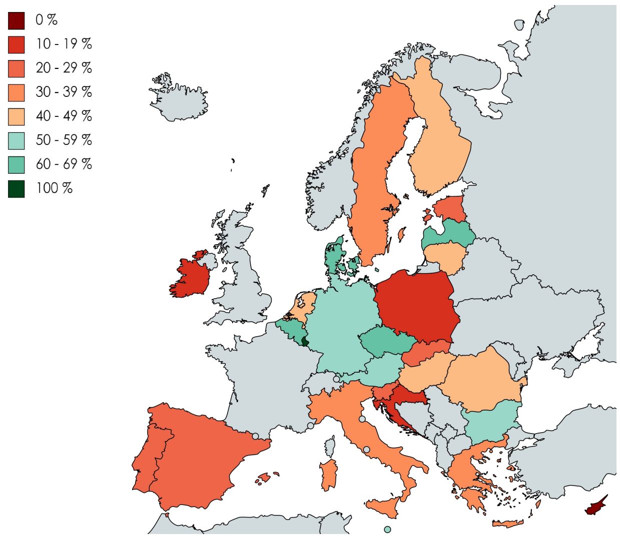

Figure 5.3: Multi-step ahead: Hybrid-MLP

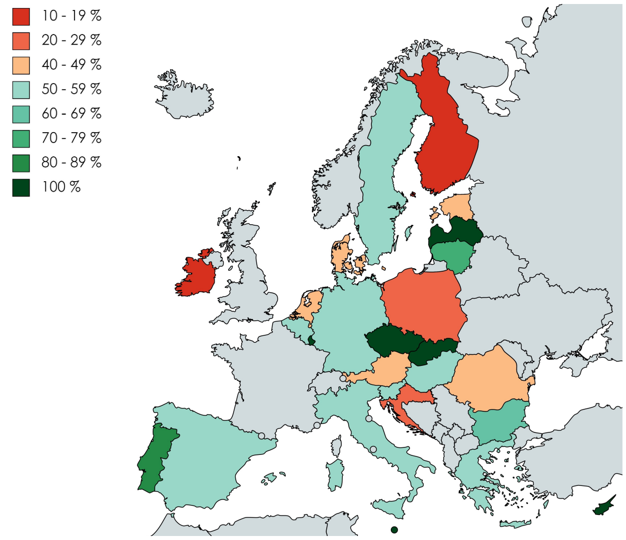

Figure 5.4: Multi-step ahead: Hybrid-LSTM

only in the case of 43% of the units. While the average improvement in RMSE relative to the linear model is 0.88, there are nearly 100 less regional units for the multi-step ahead forecasts than in the case of one-step ahead. The map in Figure 5.3 shows that the model fails to outperform the linear model for a majority of countries. Only for some countries in Western Europe and Central and Eastern Europe, the model produces an improvement in predictions on national levels. This is aligned with the other results based on MLP corrections in this dimension. The hybrid-LSTM model performs better in this dimension. 547 territorial units have an average improvement in RMSE meaning that the corrections provide an improvement for 51% of the units in the sample. The average improvement in the units with a positive change is approximately 0.31. Figure 5.4 shows that there is a large variation across national lines for the positive corrections. For example, all countries in the Southern Europe group have hybrid-LSTM forecasts which outperform the accuracy of the base linear predictions. While in other regions of Europe, there is less homogeneity in the measured improvements. Therefore, it is difficult to understand the patterns which drives the corrections made by the LSTM for the final predictions.

Overall, by studying the correction effects of the hybrid models relative to the base linear model, the initial results have been confirmed. For the one-step ahead predictions, the MLP corrections provide a sizable improvement especially in the subgroup of Central and Eastern Europe and in case of some large countries such as Germany. For the multi-step ahead predictions, the correction is fairly similar in terms of the reduction of the error, but the overall number of units with an improved accuracy is lower. There is also less homogeneity in accuracy across country lines. The hybrid model based on LSTM produces better results both in terms of the reduction in errors and in the percentage of regional units with improved forecast accuracy. In terms of homogeneity across geographic lines, the results don't follow strictly defined regional lines. Hence, it is difficult to assign certain common patterns that lead to better corrections.

6 Discussion and conclusion

The main goal of this research was to study the feasibility of applying linear panel models with an extra non-linear component in the form of neural networks in a practical setting. The chosen situation is the macroeconomic one with regional economic growth data. The scope of the data set covers 1066 regional units given by the NUTS 3 denomination in 26 countries of the European Union. The yearly data spans the time period from 2002 to 2018 with the last three years used for the out-of-sample window. This period includes varying types of economic changes on the European level and on the national and regional levels. Hence, the scope of the research has a high level of difficulty both in terms of the large dataset used, but also in terms of the dynamics and trends seen in these years.

This research considers three different models. First, the linear panel model for which variables include lags of economic growth, dummy variables for economic and financial crises, and a spatial specification based on geographic proximity between the regional units. The linear panel model is built on estimates of a consistent estimator and on heterogeneous effects for cross-sectional units based on Tweedie's formula. Overall, the linear specification provides a suitable panel model which incorporates the time dynamics in economic growth, the spatial dependencies between regions, and the individual effects for the cross-sectional units.

Both hybrid models are built on top of the errors of the linear model. Hence, the goal of the second step is to consider only the dynamics and effects that the linear model fails to capture. Naturally, if the linear model is capable of capturing the linear effects in a satisfactory manner, then only the non-linear effects would be left in the residuals. Given the capacity of neural networks to capture difficult non-linear functions between variables, the hybridization should provide overall better results than the linear specification by itself. The first used neural network model is the Multilayer Perceptron (MLP), which is a widely used method that is used in both cross-sectional and time series problems with satisfactory performance as described beforehand. The second neural network method is the Long Short Term Memory (LSTM) model that is used more for time series questions, given its capacity to incorporate time dependencies. The variables used in the specification for these methods are similar to the ones from the linear model with main difference being that residuals are used instead of the actual economic growth values. After obtaining the forecasts for the corrections in the out-of-sample window, the linear model predictions and the corrections are added up to produce the final results. The final results for the three methods are then measured and compared to identify areas with different performance levels.

Forecasting economic growth on regional level is of high importance for the creation of economic policies at European, national and local levels. Creating accurate yearly predictions for one or more years ahead is a difficult task for economists, which most often use nowcasting for dynamic forecasting for the following time periods. Moreover, most economic forecasting is done on national or regional levels with higher denominations. Therefore, this research provides an interesting exercise about the capacity of panel models and neural networks to produce economic forecasts on both long-term horizons and wider scope of territorial units of small denomination. Accurate forecasts in this case would need to have both a small relative error and predict the correct direction of economic change. Otherwise, when a forecast of growth for a regional unit deviates significantly from the true value, it can have an outsized negative impact on the policymaking process and result in overestimations or underestimations of growth.

6.1 Main findings

The main result from the performed analysis using the three methods is that none is capable of producing highly accurate forecasts in terms of economic growth. The linear panel model with Bayesian corrections fails to produce predictions with a low RMSE across the cross-section. The size of the errors is similar to the average values of economic growth of the regional units. This means that there is a sizeable error in the forecasts for most regions in the data set. Even though the model incorporates the individual effects and is hence more flexible to regional differences, it fails to produce accurate forecasts on the out-of-sample window. This can be explained by the fact that there are either new trends and dynamics in the years from 2016 to 2018 or that the model does not reflect the reality of economic growth accurately.

The forecasts of the hybrid models that incorporate corrections produced by neural network models deliver marginal improvements. In terms of one-step ahead predictions, the hybrid-MLP model provides an improvement both in terms of reduced error and in terms of increased share of regional units with correctly predicted signs of economic change. This is further shown with the country-wise comparisons and the improvements compared to the linear model forecasts. The MLP corrections produce an improvement in accuracy for most countries with a relatively high improvement for those units with improved accuracy. For the one-step ahead predictions, the LSTM based model does not provide an improvement in neither measure. Moreover, the model produces only small corrections that are not accurate for the majority of regional units.

For the multi-step ahead predictions, namely the dynamic forecasts for the years 2016, 2017, and 2018, the forecasts based on the MLP corrections do not produce the same accuracy as in the case of the one-step ahead horizon. Relative to the linear model predictions, the prediction accuracy decreases, especially in terms of the ratio of correctly predicted economic change signs. This is confirmed especially by the country-wise comparisons of the correction effects. The corrections based on LSTM on the other hand produce a marginal improvement in average RMSE and an increase in the ratio of correctly predicted economic change signs. Moreover, as shown by the correction effects compared to the linear model forecasts, the hybrid-LSTM model fails to produce homogeneous predictions for regional units across the national borders. It is also difficult to assign geographic patterns as contributing factors to the quality of the predictions.

To summarize the findings of this research, some points can be made about the forecasting performance of the used models. Even though the linear model technique incorporates both time and spatial dependencies along with individual effects for individual cross-sectional units, its performance is lacking. The high error values point to several difficulties of the technique in incorporating difficult regional patterns in economic growth. Building up on the forecasts of this model, the neural network corrections provide only marginal improvements. One interesting finding is that the MLP provides the best corrections for one-step ahead predictions, meaning that it is perhaps more flexible to new data and incorporates the most recent dynamics in the predictions. The LSTM corrections are better for more long terms predictions. On the multi-step ahead horizon of 3 years, the predictions of the hybrid-LSTM model provide a small improvement in accuracy. This can be down to the capacity of the model to incorporate time series dependencies more efficiently. All in all, the results of the research show that this exercise has more space for improvement and there are several limitations that can be dealt with.

6.2 Limitations and Future Research

The limitations of this research can be categorized in the following manner: the ones related to the data complexity, availability and features, the ones related to the modelling part both in terms of the linear model and in terms of the neural network models used for the corrections. There are some suggestions that can be made for future research.

First, one limitation with a big impact on the final accuracy of the forecasts was the large scope of the research itself. Given that the considered data came from a group of 26 countries in the European Union, it was subject to a big variance in economic numbers over the years. Moreover, by using the data of regional units with the NUTS 3 denomination, even more variance was added to the models. As regional growth can be affected by both local, national and European level variables, there is a lot of complexity that should be captured by the models in order to produce highly accurate forecasts. This was the main reason why additional variables based on time and space dependency between regional units have been introduced in the models. Nevertheless, as could be seen in the results, the results were not as accurate as hoped. Therefore, the complexity of the panel data can be seen as one of the main reasons why there are high errors in the predictions in both one-step and multi-step ahead horizons.

An apparent suggestion for future research in terms of this limitation would be to downsize the problem to a specific group of nations or regional units. For example, one such exercise would be to pick a country with a sufficiently high number of regional units such that there would still be a large enough sample to train complex models. One country with a very large number of regional units with the NUTS 3 denomination is Germany. Hence, it would be possible to do a similar exercise only with German regional units by also including data specific to the German regions and their economic area. Otherwise, it would be also possible to use a sample of smaller countries with similar economic patterns, which would together produce a large enough sample of cross-sectional units. If the goal would be to keep the same large European-wide scope, then other limitations should be tackled accordingly.

Second, the limitations related to the data availability also made it more difficult for the models to create a more accurate reflection of the economic patterns. Even in terms of the economic growth data, the sources are very limited due to the regional data being registered in most countries only after the turn of the century. As mentioned before in the data section, there are some countries that have limited sources for data on regional units such as France, whose regions are completely excluded from this research. Therefore, it was only possible to include the countries in the European Union with a sufficient availability of GDP data on NUTS 3 level. In addition to this, most of the features that could be used to explain the economic growth have limited availability on such low regional denominations. Most of the economic variables and indices are mostly available on national levels without a distinction for smaller parts of these countries. Therefore, the limited time availability of some yearly variables for several countries in the same sample made it difficult to incorporate them even with country-wise values. Also in line with the low availability of data, it can also be mentioned that the prediction accuracy could also suffer from a too short time dimension. Given that models use a relatively high level of lagged values, each cross-sectional unit has a limited number of time series. Hence, accurate individual effects can be more difficult to infer and incorporate in the final output. All in all, for the sample of chosen cross-sectional units, the research was limited by the small number of options in terms of both explanatory variables and data samples over time.

To solve the data availability issues, several options are available for future research. If the same regional denomination is to be kept, then one solution would be downsize the sample of countries to only the ones with the highest data availability. This way, the trade off between a smaller sample and the number of available explanatory variables would be minimized. Another solution would be to create more variables from the available sources or to use feature engineering techniques practiced in the Machine Learning field. This research already used the example of spatial dependencies between regional units to explain growth using the average growth in neighbouring regions. Other similar features could be used to increase the accuracy of predictions for the considered regional units. To tackle the issue of having short time series, using other without many lagged variables can be advised. This should be done in line with a deeper study into the trade-off between having more time dependencies and better data availability.

The limitations that can be attributed to the used methods, meaning the linear panel model and the hybrid methods, also have a big impact on the final quality of the results. The limitations behind the linear approach are mostly related to the requirement of producing consistent model estimates. Given that the used technique requires this condition, this also limits the options of creating a more flexible model. After obtaining the forecasts using the obtained consistent estimates, the individual effects are incorporated using the formula for the effects. Therefore, having a good initial model is very important for the overall research. This is why a limitation of this research could be solved by using additional variables and transformations while not breaking the consistency requirement. Hence, future research could deal more deeply with this issue by exploring more variations of models and perhaps other panel model estimators. One future suggestion regarding the linear panel model would be to test the feasibility of applying a model that incorporates the spatial dimension directly in the panel rather than having a variable for the spatial dependency. A better base linear model with more accurate forecasts would positively impact the rest of the methods as well.

There are several areas that could be tackled to improve the quality of the output of the neural network models. Due to the limited computational capabilities for performing this research, elements of the neural network structure needed to be pre-set in order to have a feasible hyperparameter tuning process. This is why multiple options were not tested for some parameters such as the optimization technique. With a more rigorous hyperparameter selection process, better accuracy metrics could be obtained. Also for the selection process, alternative cross-validation techniques could be used. This research used a limited version of cross-validation whereas a more randomized one could yield better specifications. Moreover, also due to the limited capacity to experiment with both the MLP and LSTM, the feature selection process was also limited. In the end, the neural networks used a similar model to the linear method, but more transformations and feature combinations could be added given that neural networks do not have strict consistency rules and assumptions that need to be respected. Therefore, another idea for future improvement would be to also look at ways to make the feature selection process more complex. Finally, this research made use of two widely used yet basic versions of feedforward and recurrent neural networks. It would be interesting to explore whether more complex architectures of neural networks such as the Bi-LSTM method could produce more accurate results in this panel context.

Overall, there are several avenues that could be explored as a follow-up to this research. Combining more recent advances in classical econometric research such as the used linear panel technique with Machine Learning methods could produce robust and accurate methods for forecasting even in such complex panel settings. Nowadays, there is a high demand in producing accurate forecasts especially in economics and business, mainly thanks to the high value of having accurate forecasts in both short and long term. With this goal in mind, this research has tried to create an innovative hybrid model in a complex panel setting on European level. Regardless of the moderate results in most aspects, there is still space to explore in this field. This research paper managed to create a working model that can be modified and used in other scenarios with potential. In a field with growing interest in both panel econometrics and Machine Learning applications, the contribution of this research hopefully will help other researchers achieve their goals and produce improvements in this quickly moving area of study.

References

Ahmed, N. K., Atiya, A. F., Gayar, N. E., and El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6):594–621.

Amidi, S. and Fagheh Majidi, A. (2020). Geographic proximity, trade and economic growth: a spatial econometrics approach. *Annals of GIS*, 26(1):49–63.

Anderson, T. W. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American statistical Association*, 76(375):598–606.

Anselin, L., Le Gallo, J., and Jayet, H. (2008). Spatial panel econometrics. In *The econometrics of panel data*, pages 625–660. Springer.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1):29–51.

Arkadievich Kholodilin, K., Siliverstovs, B., and Kooths, S. (2008). A dynamic panel data approach to the forecasting of the gdp of german länder. *Spatial Economic Analysis*, 3(2):195–207.

Baltagi, B. H. (2008). Forecasting with panel data. *Journal of forecasting*, 27(2):153–173.

Baltagi, B. H., Bresson, G., and Pirotte, A. (2012). Forecasting with spatial panel data. *Computational Statistics & Data Analysis*, 56(11):3381–3397.

Baltagi, B. H., Fingleton, B., and Pirotte, A. (2014). Estimating and forecasting with a dynamic spatial panel data model. *Oxford Bulletin of Economics and Statistics*, 76(1):112–138.

Baltagi, B. H. and Griffin, J. M. (1997). Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics*, 77(2):303–327.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143.

Choi, H. K. (2018). Stock price correlation coefficient prediction with arima-lstm hybrid model. *arXiv preprint arXiv:1808.01560*.

Cracan, C. (2020). Retail sales forecasting using lstm and arima-lstm: A comparison with traditional econometric models and artificial neural networks.

Galeshchuk, S. (2016). Neural networks performance in exchange rate prediction. *Neurocomputing*, 172:446–452.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2016). Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10):2222–2232.

Hill, T., O'Connor, M., and Remus, W. (1996). Neural network models for time series forecasts. *Management science*, 42(7):1082–1092.

Honoré, B. E. (2002). Nonlinear models with panel data. *Portuguese Economic Journal*, 1(2):163–179.

Hoogstrate, A. J., Palm, F. C., and Pfann, G. A. (2000). Pooling in dynamic panel-data models: An application to forecasting gdp growth rates. *Journal of Business & Economic Statistics*, 18(3):274–283.

Hsiao, C. (2007). Panel data analysis—advantages and challenges. *Test*, 16(1):1–22.

Huang, W., Lai, K. K., Nakamori, Y., Wang, S., and Yu, L. (2007). Neural networks in finance and economics forecasting. *International Journal of Information Technology & Decision Making*, 6(01):113–140.

Jahn, M. (2018). Artificial neural network regression models: Predicting gdp growth. Technical report, HWWI Research Paper.

Jahn, M. (2020). Artificial neural network regression models in a panel setting: Predicting economic growth. *Economic Modelling*, 91:148–154.

Khashei, M. and Bijari, M. (2011). A novel hybridization of artificial neural networks and arima models for time series forecasting. *Applied Soft Computing*, 11(2):2664–2675.

Liu, L., Moon, H. R., and Schorfheide, F. (2020). Forecasting with dynamic panel data models. *Econometrica*, 88(1):171–201.

Longhi, S., Nijkamp, P., Reggianni, A., and Maierhofer, E. (2005). Neural network modeling as a tool for forecasting regional employment patterns. *International Regional Science Review*, 28(3):330–346.

Nuroğlu, E. (2014). Estimating and forecasting trade flows by panel data analysis and neural networks. *İstanbul Üniversitesi İktisat Fakültesi Mecmuası*, 64(1).

Olah, C. (2015). Understanding lstm networks.

Siami-Namini, S. and Namin, A. S. (2018). Forecasting economics and financial time series: Arima vs. lstm. *arXiv preprint arXiv:1803.06386*.

Siami-Namini, S., Tavakoli, N., and Namin, A. S. (2018). A comparison of arima and lstm in forecasting time series. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1394–1401. IEEE.

Siami-Namini, S., Tavakoli, N., and Namin, A. S. (2019). The performance of lstm and bilstm in forecasting time series. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 3285–3292. IEEE.

Tkacz, G. (2001). Neural network forecasting of canadian gdp growth. *International Journal of Forecasting*, 17(1):57–69.

Yunpeng, L., Di, H., Junpeng, B., and Yong, Q. (2017). Multi-step ahead time series forecasting for different data patterns based on lstm recurrent neural network. In *2017 14th web information systems and applications conference (WISA)*, pages 305–310. IEEE.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62.

Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.

Zhang, G. P. and Kline, D. M. (2007). Quarterly time-series forecasting with neural networks. *IEEE transactions on neural networks*, 18(6):1800–1814.

Zhang, G. P. and Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European journal of operational research*, 160(2):501–514.

Appendix

Table 6.1: MLP tuning results

Activation: tanh

Nodes: 10

First set MAPE, RMSE and ratio: 417.72887110710144 ; 4.583242058753967 ; 0.6913225227247222

Second set MAPE, RMSE and ratio: 474.8825967311859 ; 4.14420211315155 ; 0.6411177196723151

Third set MAPE, RMSE and ratio: 436.6786539554596 ; 3.9921947717666626 ; 0.6603986645718775

Activation: tanh

Nodes: 25

First set MAPE, RMSE and ratio: 424.46500062942505 ; 4.603857636451721 ; 0.6881066939737404

Second set MAPE, RMSE and ratio: 501.29249691963196 ; 4.205796539783478 ; 0.6580490405117271

Third set MAPE, RMSE and ratio: 481.27856254577637 ; 4.062228441238403 ; 0.6810192458758837

Activation: tanh

Nodes: 50

First set MAPE, RMSE and ratio: 424.7009217739105 ; 4.584454536437988 ; 0.7044804174615643

Second set MAPE, RMSE and ratio: 478.46630215644836 ; 4.1695942878723145 ; 0.6552470261474582

Third set MAPE, RMSE and ratio: 482.1334421634674 ; 4.071818053722382 ; 0.6819696162046908

Activation: tanh

Nodes: 100

First set MAPE, RMSE and ratio: 406.6805362701416 ; 4.554436087608337 ; 0.6894778924924251

Second set MAPE, RMSE and ratio: 564.5381331443787 ; 4.312972903251648 ; 0.6697796263045673

Third set MAPE, RMSE and ratio: 508.3367109298706 ; 4.107424974441528 ; 0.666034255414656

Activation: sigmoid

Nodes: 10

First set MAPE, RMSE and ratio: 434.0356409549713 ; 4.609079718589783 ; 0.6857044663898553

Second set MAPE, RMSE and ratio: 517.6165282726288 ; 4.189750671386719 ; 0.6149106441476826

Third set MAPE, RMSE and ratio: 413.0335032939911 ; 3.9362969398498535 ; 0.6604582818987768

Activation: sigmoid

Nodes: 25

First set MAPE, RMSE and ratio: 429.7538697719574 ; 4.604688763618469 ; 0.6805668555717652

Second set MAPE, RMSE and ratio: 475.4588782787323 ; 4.131917715072632 ; 0.6176916171024576

Third set MAPE, RMSE and ratio: 392.298948764801 ; 3.922498047351837 ; 0.6711508248232521

Activation: sigmoid

Nodes: 50

First set MAPE, RMSE and ratio: 415.79795479774475 ; 4.559308767318726 ; 0.6899513242060374

Second set MAPE, RMSE and ratio: 461.11482977867126 ; 4.127108097076416 ; 0.6331956289978677

Third set MAPE, RMSE and ratio: 418.02534461021423 ; 3.9519872665405273 ; 0.681503198294243

Table 6.2: MLP tuning results (continued)

Activation: sigmoid

Nodes: 100

First set MAPE, RMSE and ratio: 409.84174609184265 ; 4.554125785827637 ; 0.6857360285040961

Second set MAPE, RMSE and ratio: 462.3896658420563 ; 4.146673262119293 ; 0.6402480080799012

Third set MAPE, RMSE and ratio: 408.430814743042 ; 3.960693061351776 ; 0.6782417798226911

Activation: relu

Nodes: 10

First set MAPE, RMSE and ratio: 426.8590211868286 ; 4.604274868965149 ; 0.6923149758725171

Second set MAPE, RMSE and ratio: 469.9568808078766 ; 4.13179475069046 ; 0.6304497250589159

Third set MAPE, RMSE and ratio: 443.7634587287903 ; 3.992215156555176 ; 0.6571407530019078

Activation: relu

Nodes: 25

First set MAPE, RMSE and ratio: 423.8546907901764 ; 4.5945885181427 ; 0.6913681124452923

Second set MAPE, RMSE and ratio: 501.6677796840668 ; 4.202159106731415 ; 0.6467989002356638

Third set MAPE, RMSE and ratio: 455.15239238739014 ; 4.023081362247467 ; 0.6613841039165077

Activation: relu

Nodes: 50

First set MAPE, RMSE and ratio: 404.7398030757904 ; 4.52893590927124 ; 0.6875981932443047

Second set MAPE, RMSE and ratio: 478.69046330451965 ; 4.184377372264862 ; 0.650063825608798

Third set MAPE, RMSE and ratio: 491.8295383453369 ; 4.085403382778168 ; 0.6707615587476153

Activation: relu

Nodes: 100

First set MAPE, RMSE and ratio: 442.5605237483978 ; 4.6354957818984985 ; 0.6857290147009314

Second set MAPE, RMSE and ratio: 476.0965049266815 ; 4.175436556339264 ; 0.644477331388172

Third set MAPE, RMSE and ratio: 486.1705958843231 ; 4.071379721164703 ; 0.6692921669846257

Table 6.3: LSTM tuning results

Nodes: 25

Activation function: tanh

First set MAPE, RMSE, and ratio: 195.67001407748918 ; 3.822704479886764 ; 0.5654317697228145

Second set MAPE, RMSE, and ratio: 361.51725093842384 ; 4.068791864982419 ; 0.4673051565480866

Third set MAPE, RMSE, and ratio: 302.29752079591657 ; 3.7545938500242584 ; 0.4822410503871619

Nodes: 25

Activation function: sigmoid

First set MAPE, RMSE, and ratio: 183.40808494743396 ; 3.8466947942557175 ; 0.5777094321624958

Second set MAPE, RMSE, and ratio: 611.291605717264 ; 4.1202110387053565 ; 0.5219882729211087

Third set MAPE, RMSE, and ratio: 399.70236366883836 ; 3.9101243729759463 ; 0.48785559982044663

Nodes: 25

Activation function: relu

First set MAPE, RMSE, and ratio: 217.05830217439478 ; 3.9408683022864874 ; 0.500792559757603

Second set MAPE, RMSE, and ratio: nan ; nan ; 0.3884840365839973

Third set MAPE, RMSE, and ratio: nan ; nan ; 0.38438096173269

Nodes: 50

Activation function: tanh

First set MAPE, RMSE, and ratio: 179.4260554713397 ; 3.8289298798039497 ; 0.5952649814835597

Second set MAPE, RMSE, and ratio: 202.78367103743682 ; 3.8088980598896427 ; 0.5182674503422736

Third set MAPE, RMSE, and ratio: 220.88546937913821 ; 3.722690846173327 ; 0.4721797497475031

Nodes: 50

Activation function: sigmoid

First set MAPE, RMSE, and ratio: 281.63859507835303 ; 4.213028593258731 ; 0.47608293120861855

Second set MAPE, RMSE, and ratio: 695.0031488381934 ; 4.227448874369012 ; 0.5215288688138257

Third set MAPE, RMSE, and ratio: 561.4670981364204 ; 4.174685109385146 ; 0.483626276512176

Nodes: 50

Activation function: relu

First set MAPE, RMSE, and ratio: 273.66418602582706 ; 4.06567176486237 ; 0.5253934743575356

Second set MAPE, RMSE, and ratio: nan ; nan ; 0.2505505835484233

Third set MAPE, RMSE, and ratio: nan ; nan ; 0.22979323308270677