TU DORTMUND

CASE STUDY

Project 3: Forecasting the electricity price Comparisons and conclusions

Lecturers:

Prof. Dr. Matei Demetrescu,

Dr. Paul Navas

Author: Mukta Ghosh

Matriculation no: 231720

Group members: Sultan Mahmud Chomon (230668)

July 20, 2024

Contents

1 Introduction

In various sectors, ranging from energy producers and consumers to financial markets, The prediction of energy prices holds importance. Understanding and accurately forecasting energy prices enables stakeholders to manage risks, optimize resource allocation, and also enhance economic efficiency.

The previous report shows one-step-ahead $Return$ series forecasting using machine learning algorithms. In this report, the $Return$ forecasting results for linear models and non-linear models without rolling window and with rolling window are analyzed. Our analysis encloses the period from January 1st, 2015 to March 15th, 2024. The raw data is sourced from (ENTSOE, 2015), the European Network of Transmission System Operators for Electricity, which is the association for the cooperation of the European transmission system operators (TSOs).

This report systematically explores dynamics, descriptive analysis, trends, and stationarity properties of the Day-Ahead return series through the Partial Autocorrelation Function (PACF), histogram, which provides crucial insights into the data. First an AR model, single predictor models, a full model with all external covariates, and a model with forward-selected covariates for the machine learning method are fitted for forecasting $Return$. Again forecasting for all models with rolling window are also shown in this report. Later, a Diebold-Mariano Test is performed for each pair of forecasts of all models.

The second section provides a more detailed overview of the dataset, data quality and description, data preprocessing. The necessary statistical methods are presented and explained in the third section. In the fourth section, the statistical analysis, and interpretation of the results are presented. Finally, in the fifth section, the main findings of this project are summarized.

2 Problem statement

2.1 Description of the dataset

The data used for analysis is provided by TU Dortmund, a Case Study program that is sourced from ENTSO-E, launched in 2015. It's a Central collection of Electricity generation, transportation, and consumption data for the Pan-European market (ENTSOE, 2015).

The provided dataset encompasses the period from 2015 to 2024, with observations collected from different countries across Europe. According to the aim of this project, only data related to Germany is separated for analysis. The dataset contains day-ahead electricity prices at an hourly resolution. This project aim is to analyze the $return$ series, a simple transformation of price data, typically calculated as:

$$\text{Return}_t = \frac{\text{Price}_t - \text{Price}_{t-1}}{\text{Price}_{t-1}}$$

2.2 External predictors

Among other external predictors, certain variables, such as the future price of $carbon$, are recorded on a daily basis. To align with our target variable, $Return$, which is recorded hourly, the data for these other predictors is adjusted to an hourly frequency. $Total.Load$, $Output\ of\ aggregated\ load\ data$ for different generation types (nuclear, renewable, fossil fuels,), $Temp$, $Netexport$, $WindSolarforcast$, $Allocatedtransfercapacity$, $remainingcapacityofload$ and $Gasfutureprice$ are selected as external predictors for predicting day-ahead energy price of the german market. The temperature data is fetched from the API url `https://archive-api.open-meteo.com/v1/archive?` with specific latitude, longitude, and timezone Europe Berlin.

2.2.1 External predictors preprocessing

For consistency, the $Return$ is lagged up to 10 periods, while external covariates are lagged by 1 period. The external predictors are preprocessed to prepare for model training and testing by generating lagged features, clipping outliers, centering, and scaling the data. Removed outliers from the selected columns by clipping values outside the 5th and 95th percentiles to the respective quantile values for each external predictor. Next, standardized the predictors to have zero mean and unit variance. Later, the processed data frame is divided into training (70%) and testing (30%) sets.

2.2.2 DST adjustment

As only data related to Germany have to be accounted for, data should be DST adjusted by adjusting the electricity consumption data for the spring forward and fallback changes. The dataset of day ahead price and other external datasets used for the previous report including load are DST adjusted.

2.3 dataset Quality

Most of the selected dataset's quality to predict one hour ahead return, are not clean enough. All dataset have in total 81096 observations. Several columns have NA or 0 values which are omitted. Missing data for certain days are filled by using the last available value.

2.4 Project objectives

The primary objective of this project is to create precise one-hour-ahead electricity return series predictions for the German market. This will be achieved by employing several models, machine learning methods (decision tree and deep neural network), and rolling window forecasting. The performance of the fitted models will be assessed using the root mean squared forecasting error (RMSFE) will be calculated to find out forecasting accuracy. Additionally, the project aims to make a forecast comparison through the Diebold-Mariano Test for all models and find out the best model.

3 Statistical methods

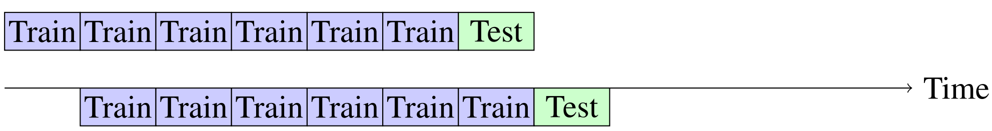
In this section various statistical methods are represented that will be used to analyze the provided dataset according to the objectives of this report. To perform analyses and visualizations the statistical software R, version 4.2.3 (R Development Core Team, 2023) is used.

3.1 Rolling window approach for forecasting

Deterministic trends in a time series follow a fixed, predictable pattern. On the other hand, Stochastic trends, are trends that incorporate random variations. By allowing for changes over time, stochastic trends provide more realistic forecasts and appropriately wide prediction intervals that reflect the inherent uncertainty in the future. The rolling window approach with a forecast horizon of N steps where the training window also moves forward by N steps in each iteration, benefits significantly from the principles of dynamic regression models that incorporate stochastic trends(Hyndman and Athanasopoulos, 2021).

3.1.1 Rolling window with forecast horizon 1

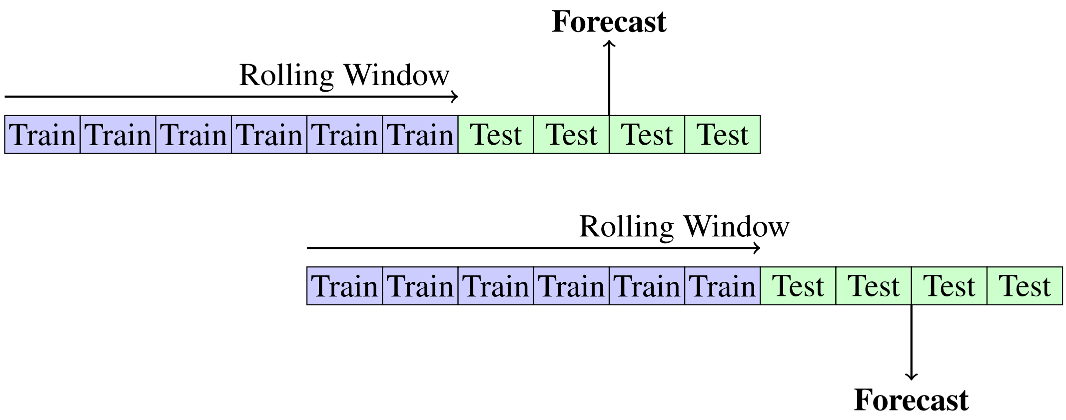
The following figure shows the rolling window movements in each iteration for window size 6 and forecasting horizon 1. In each prediction, one single observation from test data is predicted by using recent 6 observations.



3.1.2 Rolling window with forecast horizon N

The training window moves forward by N steps after each iteration. This means the model is retrained on the most recent data, and older data points are discarded. The model predicts the next N steps instead of just one step ahead. This method captures longer-term dependencies and trends in the data. In the rolling window approach context, stochastic trends can be beneficial because each new training window might exhibit different characteristics. Stochastic models provide wider prediction intervals that account for greater uncertainty, which is especially useful for multi-step forecasting where the uncertainty increases with the forecast horizon.

Model tuning is important in each iteration to capture the new characteristics in train data. The following figure, shows the rolling window approach for window size 6, forecast horizon 4 where the training window moves forward by 4 steps ahead.



3.2 Diebold-Mariano Test

The Diebold-Mariano (DM) test is a statistical test used to compare the accuracy of two competing forecast models. The test assesses if there is a statistically significant difference between the forecast errors of two models. The null hypothesis of the test is that the pair of models have equal predictive accuracy. The alternative hypothesis is that there is a difference in their predictive accuracy. Reject H_0 if the p-value is less than the chosen significance level. First, the forecast errors of each model have to be calculated, and then calculate the loss differential using the loss function. By default, the `dm.test` function in R utilizes the squared error loss function. Then calculate the mean and variance of the loss differential. And calculate DM test statistic, follows a standard normal distribution under the null hypothesis that there is no difference in predictive accuracy between the two models (Diebold and Mariano, 1995) and (Diebold, 2012).

DM Test Statistic:

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2/T}} \tag{1}$$

This statistic follows a standard normal distribution under the null hypothesis.

4 Statistical analysis

This section illustrates all the methods and procedures applied. For the calculation and visualization, the R software (version 4.2.3) is used (R Core Team, 2022). The R packages ggplot2 (Wickham, 2016) and plyr (Wickham, 2022), "tseries", "readr", "urca", "tidyverse", "rstatix", "reticulate", "keras", "caret", "tensorflow" "leaps", "lubridate", "dplyr", "forecast", "stats", "stlplus", "dynlm", "zoo" are used for this task. For visualization neural networks "visNetwork", "graphViz", and "DiagrammeR" packages were used.

4.1 Model with linear form

4.1.1 autoregressive (AR) model selection and forecasting

In the process of selecting autoregressive (AR) models for predicting financial returns, selecting the appropriate model specification is crucial. One effective approach for model selection is to use information criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These criteria help identify the model that best balances goodness-of-fit with model complexity. An algorithm was developed to generate all possible combinations of predictor variables. Each combination represents a potential AR model specification.

AR benchmark Model 1:

$$\text{Return} \sim \text{Return\_lag}_1 + \text{Return\_lag}_2 + \ldots + \text{Return\_lag}_{10} \tag{2}$$

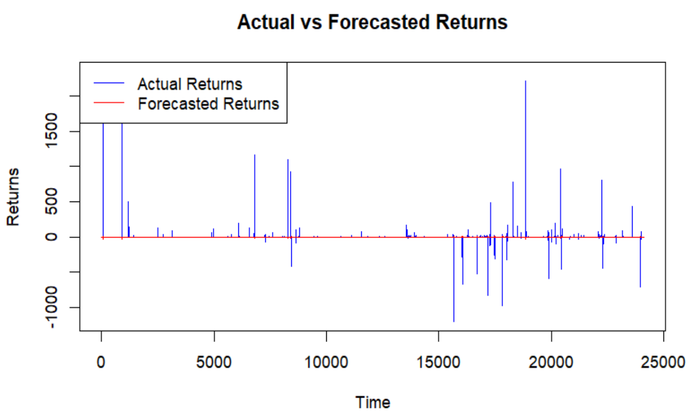


Figure 1: Forcast and actual returns by AR model

For the sake of compatibility lag 1 to lag 10 of the $Return$ series are accounted for. The model with lag 1 to 10 is selected as the benchmark model. Though all combinations of 10 lags have nearly similar AIC values, We found lag 3 and lag 4 combinations have the lowest AIC. Using the AR model, the RMSFE for the test data was found to be 34.0106. Fig 1 shown the actual return and forecasted return values for the test data frame by AR model. Due to the limited selection of lags, the forecasted values are unable to improve significantly.

4.1.2 forecasting by single predictive model

In this analysis, to identify the best predictor for forecasting returns, linear predictive models are fitted on the training dataset. To evaluate the performance of different predictors, each predictor was used to estimate its respective model using Ordinary Least Squares (OLS) regression and RMSFE are calculated for these models. The best predictor is $wind\ solar\ forcast$ with RMSFE 31.204. Table 1 shows the all RMSFE of all single predictive models.

4.1.3 Rolling window forecasting by single predictive model

After fitting the model, forecasts without using a rolling window as well as forecasts using a rolling window for different horizons: a single horizon (1), and a daily horizon (24) are implemented. Here, a daily horizon forecast means making predictions for the next 24 hours. In the context of day-ahead hourly resolution data, a 24-hour horizon forecast involves predicting the hourly $Return$ for the next 24 hours.

Table 1: Summary of benchmark model and other models with linear form

Response	Models	RMSFE	RMSFE (roll-window forecast horizon 24)
Return	AR (lag1+...+lag10)bench**	34.011304	
Return	AR (lag3+lag4)	34.010667	
Return	AR+ all externals	31.288880	
Return	Temp	31.208054	31.128
Return	Load	31.208406	31.1269
Return	Carbon Futures	31.210295	31.1266
Return	Total Fossil Output	31.205004	31.189
Return	Total Renewable Output	31.205217	31.135
Return	Nuclear Output	31.214821	31.209
Return	Net Export	31.208117	31.207
Return	Wind solar Forecast	31.204473	31.203
Return	Allocated Trans Cap	31.208328	31.207
Return	Remaining Capacity	31.209005	31.2067
Return	Gas Futures	31.348994	31.206

In the rolling window forecasting approach, the initial window size was set to include the most recent 5 years of data from the training dataframe. As the rolling process progresses, the training window incrementally advances forward in steps corresponding to the forecast horizon. This allows the model to use a time window that captures recent trends and patterns, ensuring that the forecasts are based on the most current information available within the specified historical period.

Table 1 shows the RMSFE for the AR benchmark model, the best combination for the AR model according to AIC and single predictor models with and without rolling window forecasting. For AR models without external predictors, the RMSFE is 34.01. With external predictors, the RMSFE is reduced to 31.2889.

4.1.4 Model with forward selection

In time series forecasting, selecting the best set of predictors is crucial for improving model performance. The forward selection method is a stepwise regression technique that adds predictors to the model one at a time based on a specified criterion, such as AIC or BIC. This approach helps in identifying the most promising predictors that balance model fit and complexity. In figure,2 shows the forward selection method which effectively identified a subset of predictors with step 4. And the model with the best subset of predictors is-

Full Model 2:

$$\text{Return} \sim \text{Return\_lag}_1 + \ldots + \text{Return\_lag}_{10} + \text{External\_lag}_1 + \ldots + \text{External\_lag}_{11} \quad (3)$$

Forward selected Model 3:

$$\text{Return} \sim \text{Return\_lag}_3 + \text{Return\_lag}_4 + \text{WS\_Forecast\_lag}_1 + \text{Gas\_Futures\_lag}_1 \quad (4)$$

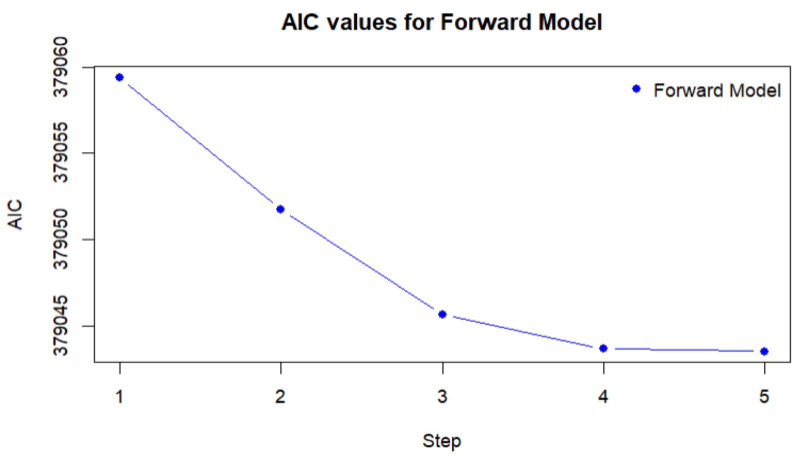


Figure 2: AIC for Forward selection steps

4.2 Model with non-linear form

In this section, 2 machine learning methods, Decision Tree (DT) and Deep Neural Network(DNN) are applied with the full model and forward selected model. After fitting the model, forecasts are performed without using a rolling window as well as forecasts using a rolling window for different horizons: a single horizon (1), a daily horizon (24), and a weekly horizon (168). In the context of day-ahead hourly resolution data, a 24-hour horizon forecast involves predicting the hourly $Return$ for the next 24 hours in each rolling. Comparative to the AR benchmark model, the machine learning model performs better. However, for DT and DNN with the full model in table2, the rolling window forecasting did not outperform the non-rolling window forecasting. In contrast, for the forward model using DT and DNN methods in table3 , the rolling window forecasting approach yielded slightly better results comparatively. So, right model selection is important. in table3 , For the DT with the forward-selected model, the RMSFE without rolling forecast is 31.912, while the RMSFE with a rolling forecast for a 24-hour horizon improves to 31.83.

Table 2: Summary of machine learning methods with model 2(full model)

Response	Models	RMSFE	Rolling Forecast RMSFE	
			Horizon 1	Horizon 24
Return	AR (lag1+...+lag10) bench**	34.011		
Return	Decision tree (model 2)	31.843	32.098	32.135
Return	DNN (model 2)	31.214		31.228

Table 3: Summary of machine learning methods with model 3 (forward selected model)

Response	Models	RMSFE	Rolling Forecast RMSFE	
			Horizon 1	Horizon 24
Return	AR (lag1+...+lag10) bench**	34.011		
Return	Decision tree (model 3)	32.912		31.83
Return	DNN (model 3)	31.210		31.185

4.3 Compare forecasts by Diebold and Mariano test

In this analysis, the DM test was applied to various forecasting models to evaluate their relative performance. A p-value matrix was created to store the results of the DM test for each pair of models. The matrix was then converted into a data frame suitable for plotting and visualize by heatmap. The DM test results reveal that most models do not show statistically significant differences in predictive accuracy when compared directly. figure 3 shown the DM test for full model with DT and DNN method, for the different forcast approach. figure 4 shown the DM test for forward model with DT and DNN method, for the different forcast approach.

DM test was applied to various forecasting of full model with DT and DNN to evaluate their relative performance. AR model forcasting errror is the benchmark here. Here's the list of them performed in the figure 3. In figure 4 shown the forcast comparison for different models included AR, DT with forward selected model, DT with forward selected model for rolling forcast.

- AR (AutoRegressive) model
- DT (Decision Tree) model
- DT_RW_h1 (Decision Tree with rolling window, horizon = 1 hour)
- DT_RW_h24 (Decision Tree with rolling window, horizon = 24 hours)
- DNN (Deep Neural Network) model
- DNN_RW_h24 (Deep Neural Network with rolling window, horizon = 24 hours)

However, the rolling window approach with a 24-hour horizon significantly enhances the performance of the DNN model, as evidenced by the low p-value in the comparison between the forcasting error vectors. This suggests that incorporating a rolling window can improve forecast accuracy, particularly for complex models like deep neural networks. For Decision tree, the hypertuning for each rolling iteration was not performed due to computational cost.

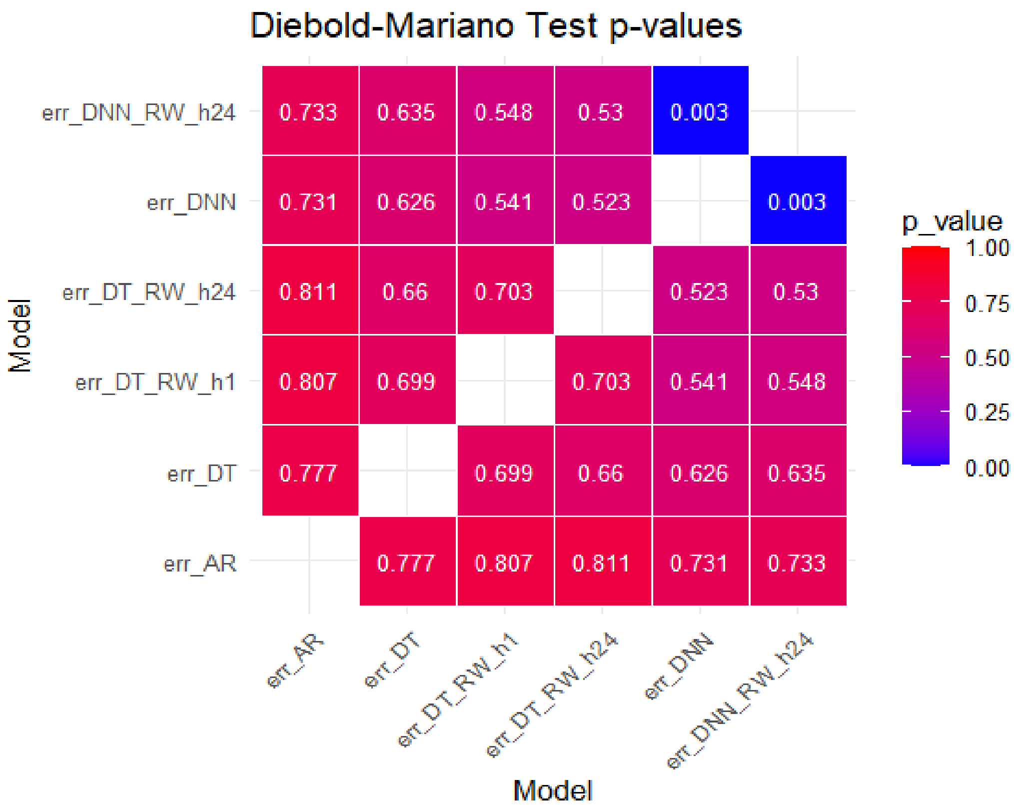


Figure 3: Diebold Mariano test for full model with DT and DNN for different forcasting approach

In figure 4, the forcasting comparison for decision tree with forward selected model without rolling forcast approach and with rolling forcast approach is significatly different as the p value is lower than 0.05.

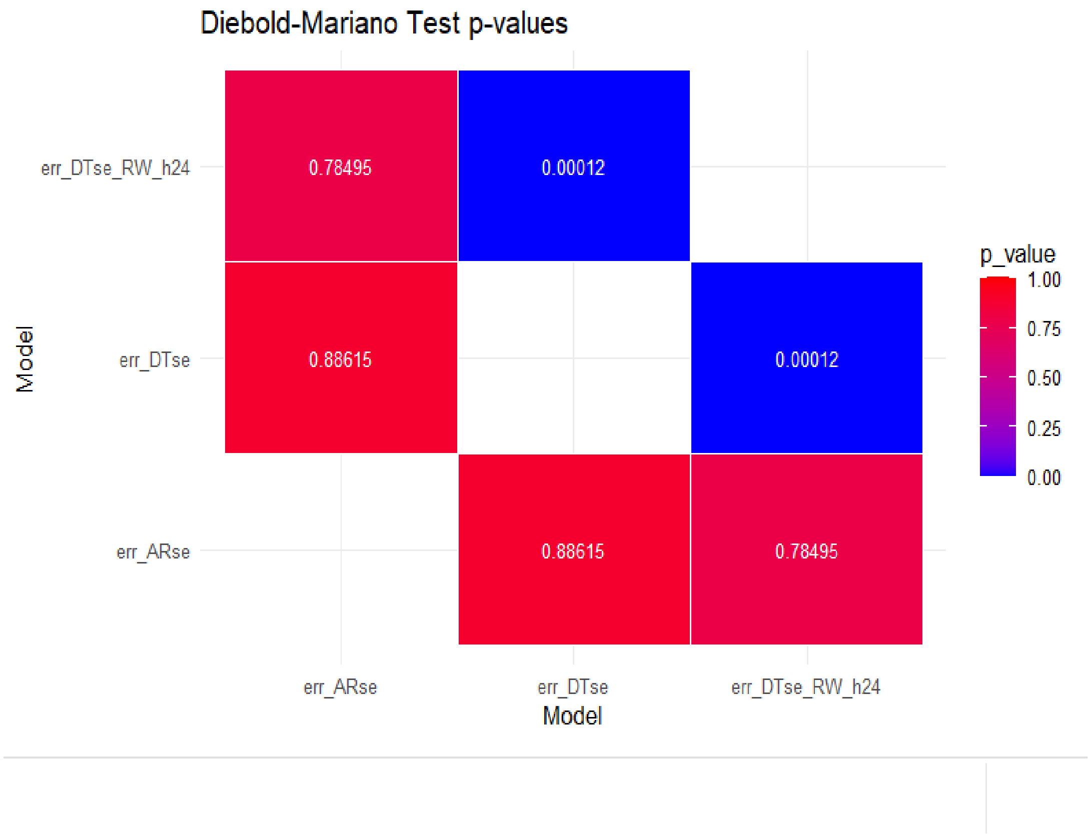


Figure 4: Diebold Mariano test for selected model with DT and DNN for different forcasting approach

5 Summary

Financial and economic data are very complex and unpredictable type because many factors always influence them, and their use in reality often requires consideration of their endogenous variable relationships, which is the core idea of most time series models. As the autocorrelation structure of provided data is not constant over time and includes volatility over data, Return series is not stationary.

By comparing RMSFE of AR model as benchmark, single predictors model, AR model with externals, forward selected model we found the important external predictors for predict day ahead return.BY applying machine learning method on forward slected model and full model, the RMSFE is better for forward selected model for both machine lerning approch. Additionally, model tuning for each rolling window is crucial because the training data changes with each iteration. This approach can lead to more accurate and robust results. However, due to computational complexity, this tuning was not performed.

The deep nural network with rolling forcast approach gives better RMSFE result which is 31.185. By performing Diebold and Mariano test on several models, we found a significant differenece in forcasting with rolling window and without rolling window, where AR model is selected as benchmark model.

Bibliography

Francis X. Diebold. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of the diebold-mariano test. Nber working paper, National Bureau of Economic Research, 2012.

Francis X. Diebold and Roberto S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263, 1995.

ENTSOE. *Central collection of Electricity generation, transportation and consumption data for the Pan-European market*, 2015. URL :`https://transparency.entsoe.eu/`.

Rob J. Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 3 edition, 2021. Chapter 10.4: Stochastic and deterministic trends.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. SpringerVerlag New York, 2016. ISBN 9783319242774. URL `https://ggplot2.tidyverse.org`.

Hadley Wickham. *plyr: Tools for Splitting, Applying and Combining Data*, 2022. URL `https://cran.r-project.org/web/packages/plyr`.

Appendix

A Additional figure

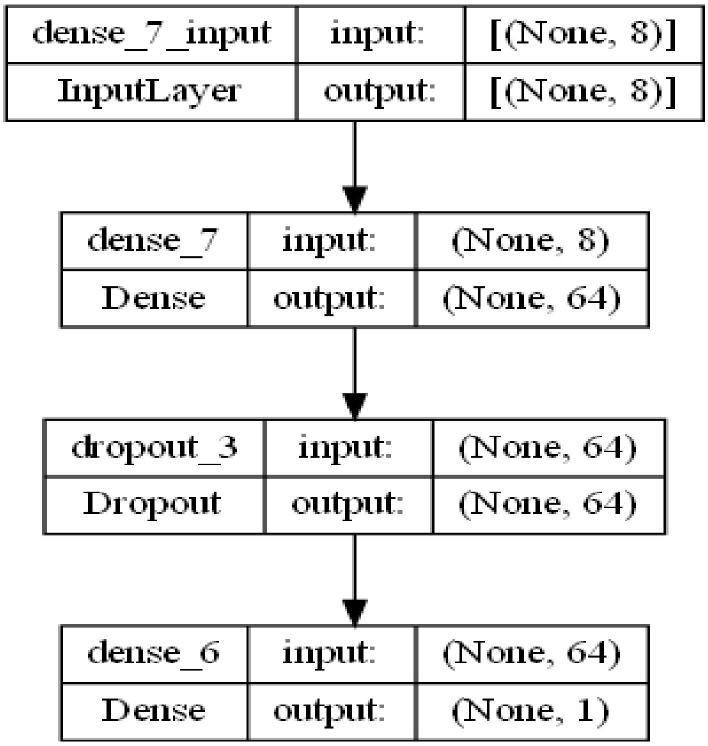


Figure 5: DNN model structure