### Step 2: Stationarity testing and transformations performed

After identifying all the variables to be tested, we applied the stationarity tests. For detecting stationarity, we computed the Augmented Dicky-Fuller (ADF) test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. These two will give us an indication regarding the stationarity of the data.

We performed the tests by using a threshold of 5% for KPSS, 10% for ADF and PP. Then we built a field called adf\_kpss\_pp as a combination between ADF, KPSS and PP tests. It returns TRUE if at least one of the tests demonstrates that the series is stationary. Using this information, we have identified {adfKpssPpTrue} variables to be further used for modeling as being stationary.

Please see reference 4 from Annex 1

We considered those variables and transformations who offer stationary series. For all selected dependent variables, we developed the lagged variables using the following formula:

### Step 3: Rules applied for computing the models

We applied all rules mentioned when defining this step of development. After the previous step, we generated {finalSubsetModelCount} valid models by considering all the combinations could exist between dependent variables and stationary independent variables for {originalIndVars}. The full list of the candidate models, that will be used further in the BLUE tests analysis, is presented in the below file:

Please see reference 8 from Annex 1

Once the possibilities are generated, we proceeded with the next step, represented by the selection.

### Step 4: BLUE tests on selected model

#### 4.1 Model consistency and definition

Table 6 Coefficients of selected model for PDs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** |  | **Pr(>|t|)** |  |
| **{#coefTable}{name}** | {estimate} |  | {pv} | {s}{/} |

The model is validated using the associated probability of {modelPvalue} for the F-statistic test, which is lower than the threshold of 5%. The model is correct defined in this form.

#### 4.2 Standard errors evolution and multicollinearity

In terms of multicollinearity, the results of VIF\_index calculation are presented in the below table:

Table 7 VIF for selected models of PDs

|  |  |  |
| --- | --- | --- |
| **{firstIndependentName}** | **{secondIndependentName}** | **{thirdIndependentName}** |
| {var1VIF} | {var2VIF} | {var3VIF} |

Both being less than 5, we concluded on the stability of the model with limited increases of standard errors compared to if that variable had 0 correlation to other predictor variables in the model.

#### 4.3 Correlation of explanatories variables with Pearson Correlation Coefficient

For testing the correlation using Pearson approach, we combined the number of explanatories variables identified in the model as follows:

* comb1: var1 vs. var2

{#threeVarModel}

* comb2: var1 vs. var3
* comb3: var2 vs. var3

{/threeVarModel}

The order of variables within the model is the one presented in executive summary and tables before this step.

Table 8 Pearson correlation for explanatories

|  |  |  |
| --- | --- | --- |
| **\*.comb1\_Corr\_Pearson** | **\*.comb2\_Corr\_Pearson** | **\*.comb3\_Corr\_Pearson** |
| {comb1} | {comb2} | {comb3} |

#### 4.4 Serial correlation (autocorrelation), heteroskedasticity, and normality of errors

Furthermore, the selected model should be tested for autocorrelation, heteroskedasticity, and normality of errors. The results of tests applied are reflected in the below table.

Table 9 Blue tests applied for selected model of PDs

|  |  |  |
| --- | --- | --- |
| **Statistics or p-value of test** | **Value** | **Decision** |
| Serial correlation using Breusch-Godfrey test 1 | {bg1} | {bg1test} |
| Serial correlation using Breusch-Godfrey test 2 | {bg2} | {bg2test} |
| Serial correlation using Breusch-Godfrey test 3 | {bg3} | {bg3test} |
| Serial correlation using Breusch-Godfrey test 4 | {bg4} | {bg4test} |
| Autocorrelation using Durbin - Watson | {dw} | {dwtest} |
| Normality of errors using Shapiro-Wilk | {sw} | {swtest} |
| Normality of errors using with Jarque - Bera | {jb} | {jbtest} |
| Heteroscedasticity using Breusch-Pagan | {bp} | {bptest} |
| Heteroscedasticity using White test | {wh} | {whtest} |

#### 4.5 Models with inconsistent estimators

{#isInconsistent}

In this case, the confidence levels of estimators are biased due to **autocorrelation** (Serial correlation using Breusch-Godfrey test 1 OR Serial correlation using Breusch-Godfrey test 2 OR Serial correlation using Breusch-Godfrey test 3 OR Serial correlation using Breusch-Godfrey test 4 OR Autocorrelation using Durbin – Watson) and **heteroskedasticity** (Heteroscedasticity using Breusch-Pagan OR Heteroscedasticity using White test). To defend the selected model, we proceed by applying Newey-West approach. The results are stated in the below table.

{/isInconsistent}

{^isInconsistent}

In this case, the confidence levels of estimators are not biased because we did not identify serial correlation or heteroscedasticity presence. However, to better defend the model, we proceed by applying Newey-West approach.

{/isInconsistent}

{#isInconsistent}

Table 10 Newey-West estimators of PDs

|  |  |  |
| --- | --- | --- |
|  | **Pr(>|t|)** |  |
| **{#nwTable}{name}** | {pv} | {s}{/} |

{/isInconsistent}

The confidence levels of estimators in the table are still consistent and we concluded that the significance of each variable in the model is kept.

### Step 5: Rules applied for selecting the proper model to be tested and used

For deciding on a proper model to forecast the probability of defaults for portfolio, we automatically performed (in Python language) the above-mentioned criteria at step 5 in the methodology description.

The result of automatically applying the criteria, we obtained the file 12.

Please see reference 12 from Annex 1

The results of analysis of models from file 12 are presented as follows:

1. Best model by AICc criterion is no. {bestAiccNumber} having the AICc = {bestAicc}
2. Best model by Adjusted R-squared is no. {bestRsqNumber}with adjusted R – squared {bestRsq};

Several other filters were applied based on Bank’s expectations and statistical meaning:

* 1. DE ACTUALIZAT LA FIECARE MODEL IN PARTE

{H1H2Paragraphs}

Going further with the analysis, we analyzed the remaining models with the following conclusions:

* All models are statistically significant, with predictions around aligned with the business expectations.
* Model no. **{finalModelNumber}** had the most appropriate average of predictions with the business expectations for PDs/LGDs and the best combination of Adjusted Rsquared and AIC indicators.

Considering the mentions above, the first model criterion that respects all necessary criteria is no. **{finalModelNumber}**. In this case, the independent variables are **{finalIndependentVars}. {dependent}** is the dependent variable. {independentVariableDescriptions}

As a final step in selection, we compared the champion model with two different alternatives and introduced the results in the below table:

Table 11 Feasible models for PDs

|  |  |  |  |
| --- | --- | --- | --- |
| Variable\_name | Champion Model | Comparison Model 1 | Comparison Model 2 |
| {#modelComparisonTable}{0} | {1} | {2} | {3}{/modelComparisonTable} |

From all the models above presented, model no. **{finalModelNumber}** offers the best solution in terms of AICc, Adjusted R – squared and the p – value after applying the Newey – West approach for robust estimators. In addition, the predictions have economic meaning and offers correlated, with economic evolution, results.

Our conclusion for the model presented for entitys portfolio ({portfolio}) is that is the best option we can have at this time.

### Annex 1

|  |  |
| --- | --- |
| Reference | File name |
| 4 | {stationarityFileName} |
| 8 | {finalSubsetFileName} |
| 12 | {finalFeasibleFileName} |