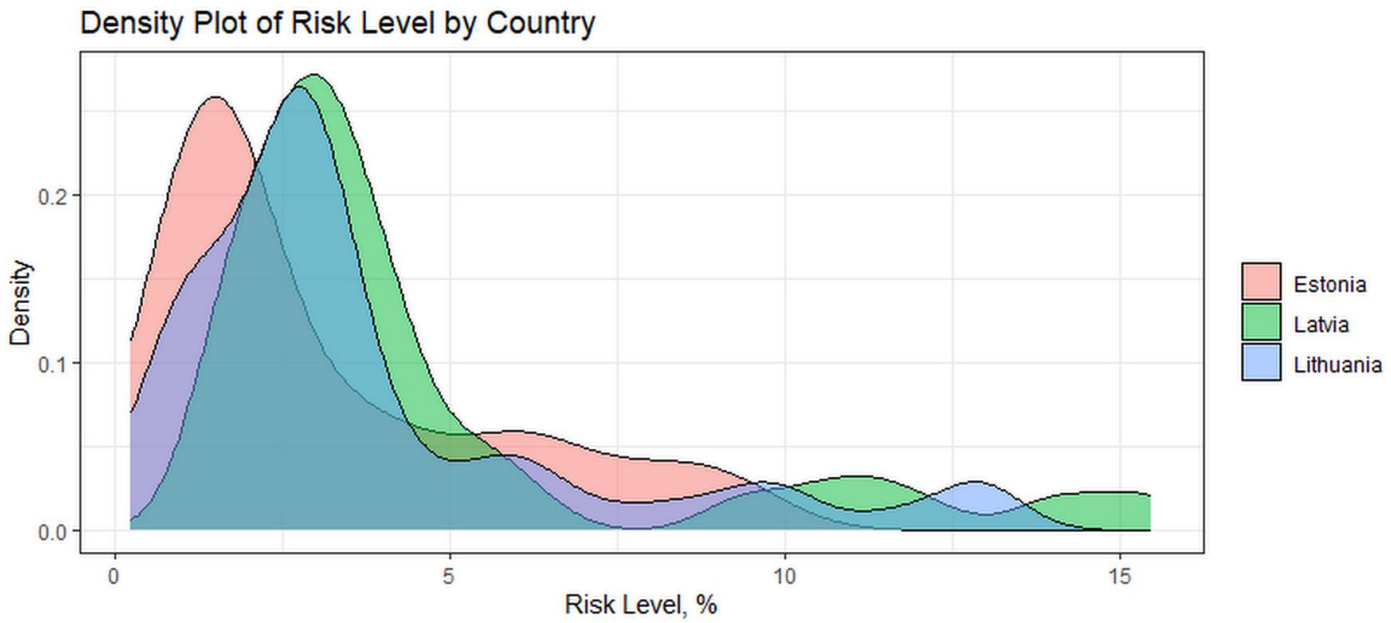
Exploratory Data Analysis

Taking a closer look at the sample data (2006 Q1 – 2022 Q4)

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
GDP	4016	4852	5263	5316	5750	7058
UR	0,04	0,06	0,07	0,08	0,09	0,20
HPI	56	77	100	104	118	191
HICP	213	263	299	297	326	432
EURIBOR (6M)	-0,005	-0,003	0,003	0,010	0,016	0,052
Wage	549	812	998	1069	1326	1762
Risk Level	0,01	0,03	0,02	0,03	0,04	0,09

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
GDP	4595	5678	6149	6221	6766	7590
UR	0,05	0,07	0,09	0,10	0,12	0,21
HPI	79	96	110	119	144	188
HICP	210	281	298	299	318	421
EURIBOR (6M)	-0,005	-0,003	0,003	0,010	0,016	0,052
Wage	383	662	767	839	1013	1428
Risk Level	0,01	0,03	0,03	0,05	0,05	0,16

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
GDP	6985	8335	9263	9408	10340	12637
UR	0,04	0,06	0,08	0,09	0,12	0,18
HPI	83	90	107	115	130	203
HICP	220	277	301	301	322	440
EURIBOR (6M)	-0,005	-0,003	0,003	0,010	0,016	0,052
Wage	405	599	674	862	936	1885
Risk Level	0,003	0,020	0,028	0,038	0,041	0,131



Regression models

Three separate models for forecasting an in-sample Risk Level Index

- $riskLevel_{EE} = 2.549 \cdot EE_{ur} + 0.9277 \cdot EE_{euribor6m} + 0.0001218 \cdot EE_{wage} - 0.006557 \cdot EE_{ur} \cdot EE_{hicp}$
- $riskLevel_{LV} = 0.00008052 \cdot LV_{gdp} + 7.625 \cdot LV_{ur} + 0.003494 \cdot LV_{hicp} + 4.528 \cdot LV_{euribor6m} - 0.0000002370 \cdot LV_{gdp} \cdot LV_{hicp} - 0.02326 \cdot LV_{ur} \cdot LV_{hicp} - 0.01396 \cdot LV_{hicp} \cdot LV_{euribor6m}$
- $riskLevel_{LT} = -0.00003088 \cdot LT_{gdp} + 8.929 \cdot LT_{ur} - 0.00009612 \cdot LT_{gdp} \cdot LT_{ur} + 0.0000001865 \cdot LT_{gdp} \cdot LT_{hicp} + 0.0001769 \cdot LT_{gdp} \cdot LT_{euribor6m} - 0.00000001599 \cdot LT_{gdp} \cdot LT_{wage} - 0.03075 \cdot LT_{ur} \cdot LT_{hicp} + 0.001964 \cdot LT_{ur} \cdot LT_{wage} - 0.002381 \cdot LT_{euribor6m} \cdot LT_{wage}$

where EE – Estonia, LV – Latvia, LT - Lithuania

	EE	***LV***	***LT***
Multiple R^2	0.715	0.9151	0.9573
Adjusted R^2	0.6879	0.8999	0.947
F-statistic	26.35	60.09	92.27
F-statistic (p-value)	5.686×10^{-11}	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$

- In order to use the best practices, it was decided to split the sample dataset into two parts: training set for fitting the model and testing set for making predictions (forecasting values).
- All of the coefficients that were used in each model are statistically significant and were diligently chosen in order to achieve the best result.
- As can be seen, the *Adjusted R^2* is relatively good among all of the models and is the highest for Lithuania. However, it is crucial to note that a high Adjusted R-squared value does not necessarily mean that the model is the best fit. It would be beneficial to do more testing to make sure that the model is as good as it can get.
- As for *F-statistic (p-value),* we see that it is less than significance level (0.05). This means that, in general, all of the three models are statistically significant, suggesting that at least one of the predictors in the model has a significant relationship with the dependent variable.

Conclusions of the models validity

- After having fitted the models for Estonia, Latvia and Lithuania, we have moved on forecasting an in-sample Risk Level Index.

- Previously, we have divided the sample dataset into two parts: training set and testing set. The testing set was used in order to provide the model with unseen data and check how it performs. For this purpose, an R function was used to predict Risk Level Index values.

- Once the forecasting was done, it was time to compare the results to the actual values of Risk Level Index for each country. Surprisingly enough, they were very similar (by just comparing them one to one). Nonetheless, a more accurate test was needed to determine the error between an original value and a predicted one.

- For this reason, a *Root Mean Square Error (RMSE)* was used. It is a frequently used measure of the differences between values predicted by a model or an estimator and the values observed. The results for all three models were relatively similar: $RMSE_{EE}$ = 0.012; $RMSE_{LV}$ = 0.016; $RMSE_{LT}$ = 0.014.

- Let us take a smallest value, that is for Estonia, and interpret it. It means that, on average, the predictions made by the model have an error of approximately 0.012 units when compared to the true values of Risk Level Index in the testing set. Providing that *RMSE* is relatively small, it suggests that the model is performing quite well in predicting Risk Level Index. This goes for the other two models as well.

- All in all, the models showed a great performance on forecasting an in-sample values for Risk Level.