# Conclusions

In this paper, the application of multiple survival analysis techniques in predicting the probability of default to estimate expected credit loss is assessed against the results of logistic regression based on a portfolio of mortgages using real-life data. The assessment of these models takes into account their discriminatory power and goodness of fit, by comparing their probability of default estimates with observed default events over a one-year horizon. The findings of the study reveal that survival models consistently outperform logistic regression in terms of discriminatory power, as evidenced by higher area under the curve values for both out-of-sample and out-of-time observations. These results indicate that the survival models exhibit a greater ability to distinguish between good and bad clients (i.e., those who do not default and those who default). Notably, the DeepHit model stands out as particularly effective, achieving the highest AUC values and demonstrating robust discriminatory power, especially over time and within the subset of clients with no overdue payments, as compared to logistic regression results. Furthermore, the findings reveal that logistic regression consistently provides underestimated PD estimates for out-of-time data, while the DeepHit model offers accurate PD estimates. This results in a consistent underestimation of ECL estimates using logistic regression results, leading to significant deviations from observed losses over a one-year horizon for out-of-time observations. In contrast, ECL estimates derived by the DeepHit model accurately predict losses over a one-year horizon for out-of-time data, accurately reflecting observed losses. These results suggest that the DeepHit model is a reliable choice for assessing credit risk and predicting potential losses, not only improving the discriminatory power and accuracy of ECL estimates compared to logistic regression, but also enabling more dynamic estimates due to its survival approach features, which are consistent with the criteria of IFRS 9 requirements. Other survival models, compared to logistic regression, do not perform particularly better in terms of calibration accuracy than logistic regression, and their PD and ECL estimates also deviate from observed default and loss values.

Starting from these findings, it would be interesting to further explore the application of deep learning-based survival analysis models, such as DeepHit, and evaluate their performance on a wider range of data sets and banking portfolios. This could include assessing their robustness in different economic climates or regions to determine their generalizability. Additionally, investigating the interpretability of these deep survival models and understanding the factors influencing their predictions could provide valuable information for risk management practitioners and regulators. Furthermore, the analysis highlights the potential benefits of using ensemble methods, as the models produce correlated PD estimates of varying strength. Models with strong positive correlations demonstrate consistency in predicting defaults, while those with weaker correlations may benefit from further investigation, possibly incorporating ensemble methods to improve accuracy. Less correlated forecasts, arising from different types of model error, may have the potential to achieve better performance when combined.