Supplementary materials 1

A comparison of hyperparameter tuning procedures for clinical prediction models in low-dimensional data: a simulation study

Zoë S Dunias

1 Estimation error (handling)

Table S1: Summary of simulation estimation errors and their handling.

	Affected runs n (%)
Number of models fitted	333,000 (100%)
Separation	0 (0%)
Complete / near degenerate outcome distributions	
0 events or non-events generated	0 (0%)
less than 8 events or non-events generated	0 (0%)
Degenerate linear predictors*	
Ridge logistic regression	0 (0%)
Lasso logistic regression	104 (0.03%)
Elastic Net	0 (0%)
Random Forest	0 (0%)

* Calibration slopes were replaced by the maximum value of calibration slope for the corresponding scenario.

2 Results predictive performance outcomes for all models

2.1 Calibration slopes for the penalized regression models

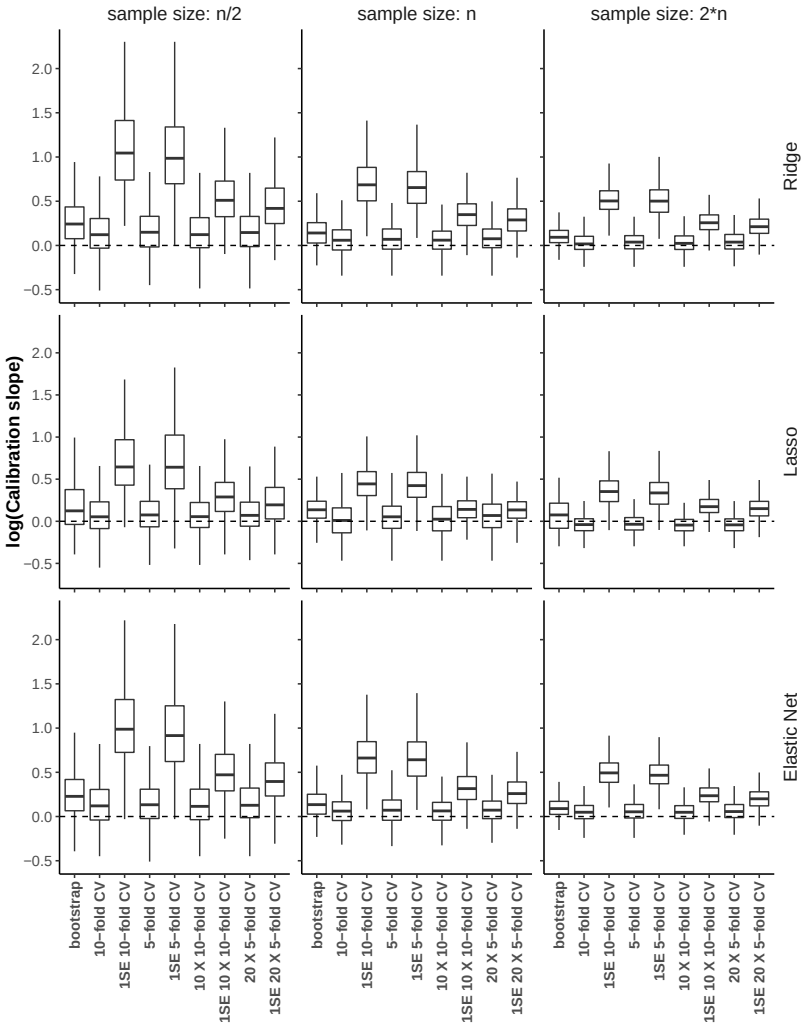

Figure S1: Box plots of the log(calibration slopes) for penalized regression models. Scenario: number of predictors = 8, events fraction = 0.1.

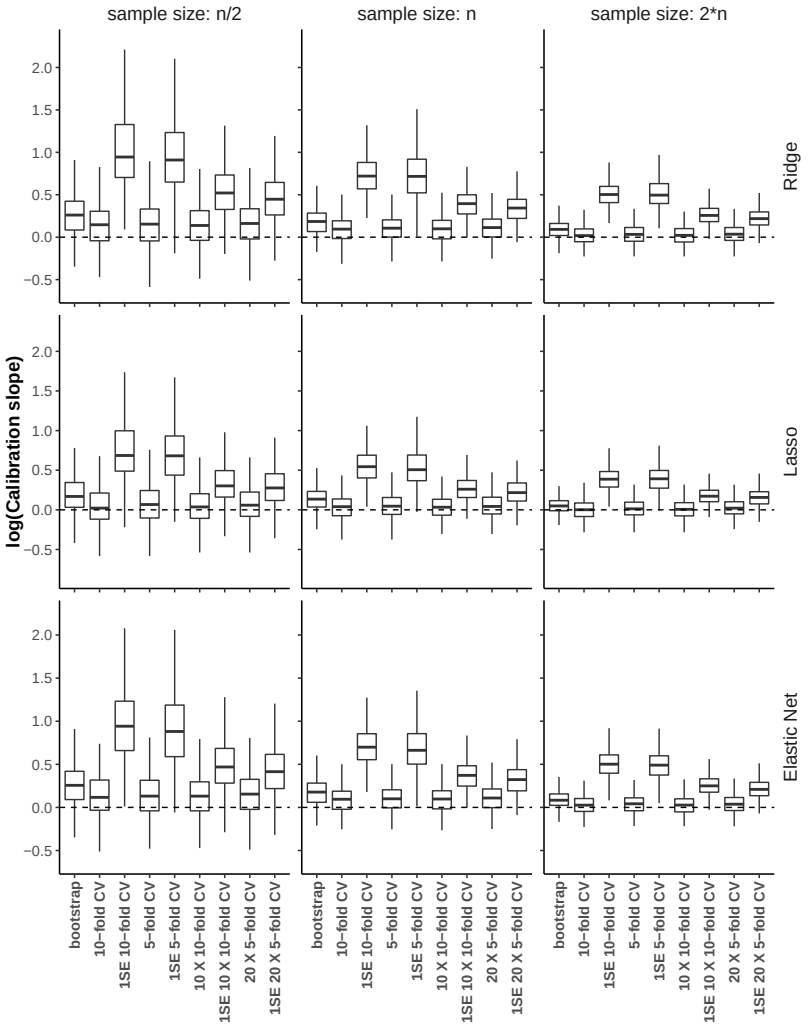

Figure S2: Box plots of the log(calibration slopes) for penalized regression models. Scenario: number of predictors = 8, events fraction = 0.3.

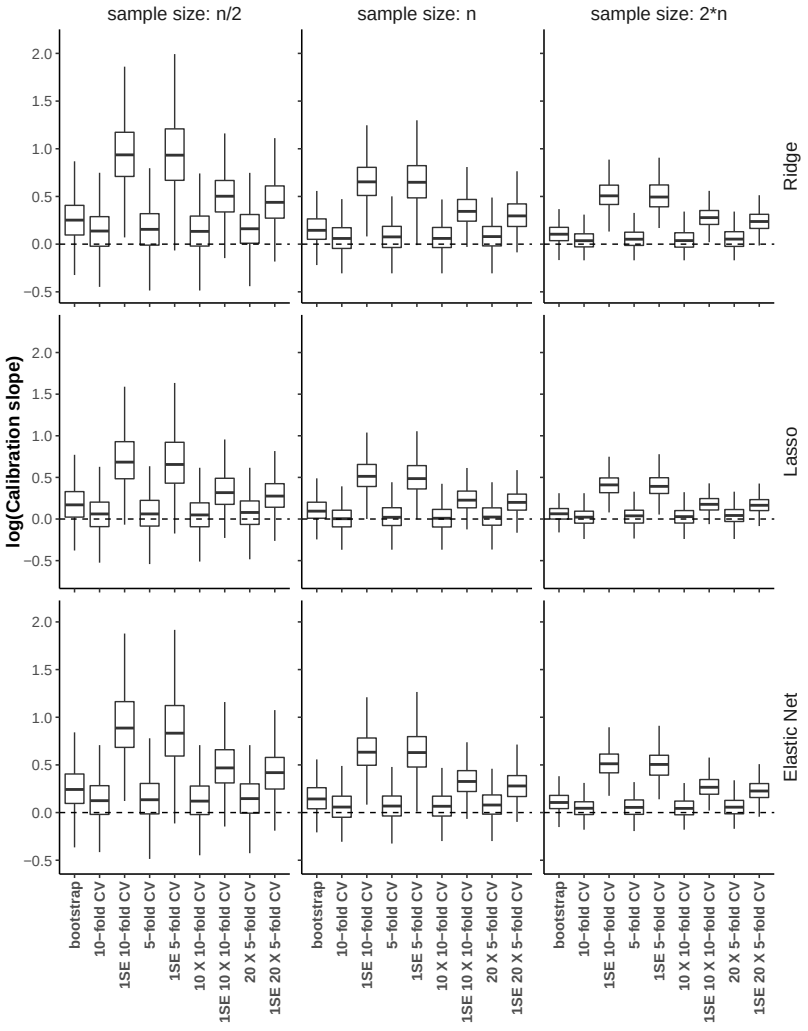

Figure S3: Box plots of the log(calibration slopes) for penalized regression models. Scenario: number of predictors = 8, events fraction = 0.5.

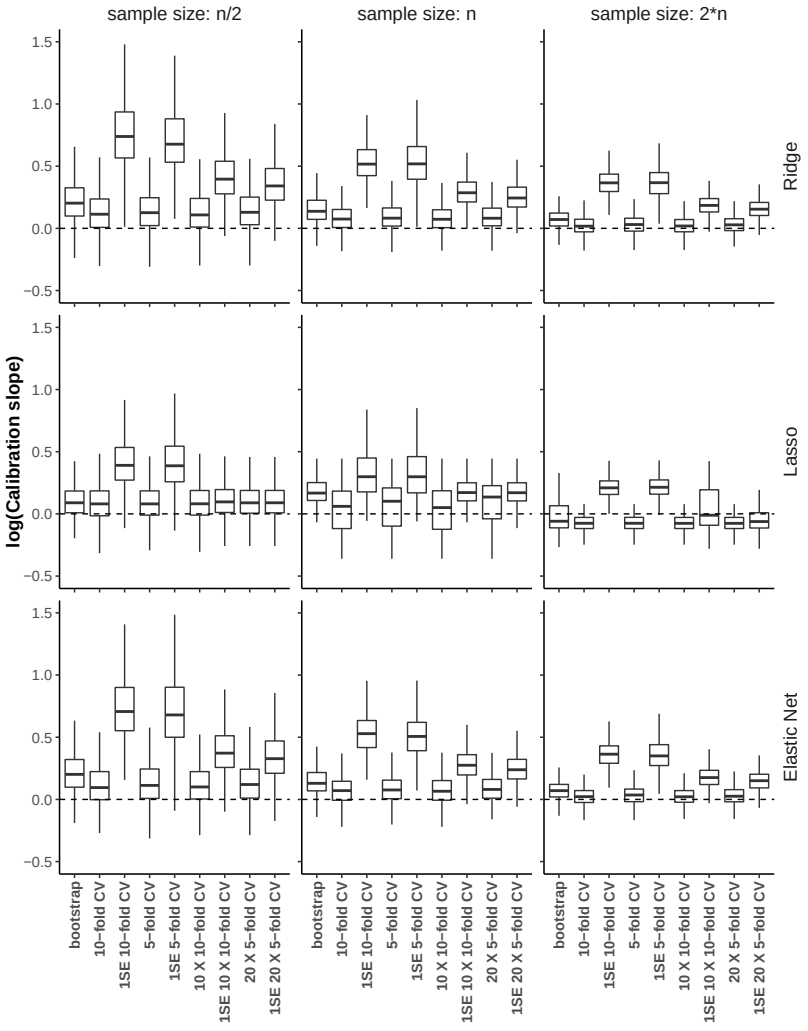

Figure S4: Box plots of the log(calibration slopes) for penalized regression models. Scenario: number of predictors = 16, events fraction = 0.1.

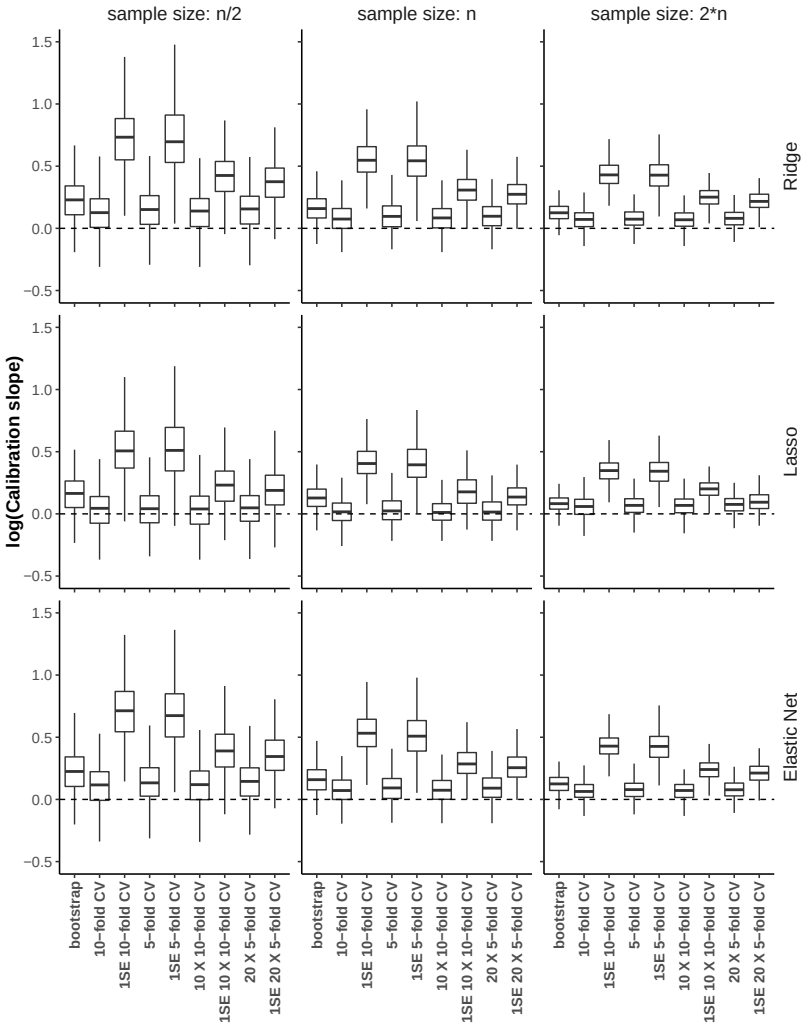

Figure S5: Box plots of the log(calibration slopes) for penalized regression models. Scenario: number of predictors = 16, events fraction = 0.5.

2.2 Calibration slopes for the Random Forest model

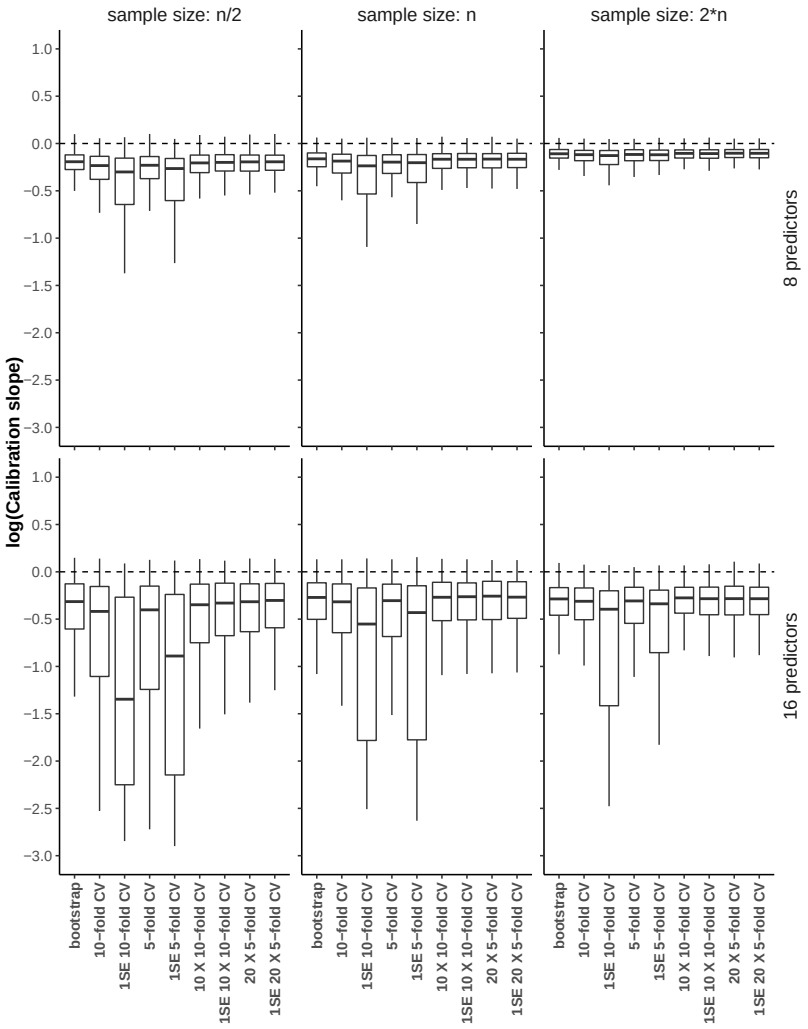

Figure S6: Box plots of the log(calibration slopes) for Random Forest model. Scenario: events fraction = 0.1.

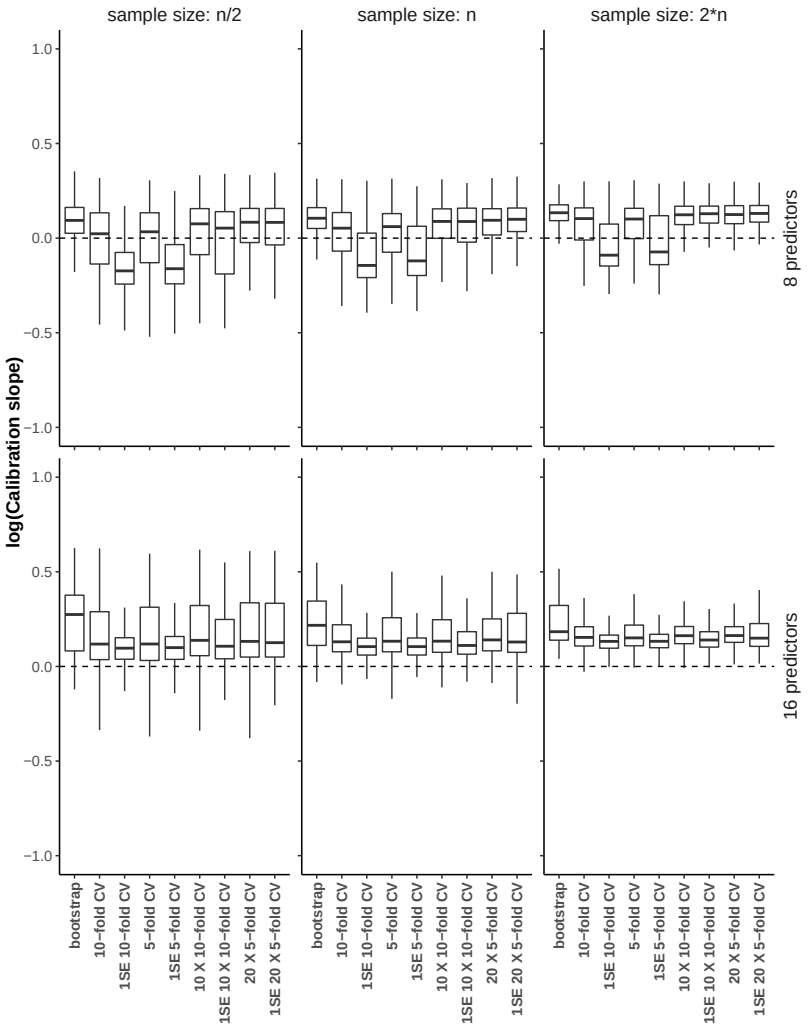

Figure S7: Box plots of the log(calibration slopes) for Random Forest model. Scenario: events fraction = 0.5.

2.3 Root mean squared distance of the target value (RMSD) of the calibration slope for Elastic Net, Ridge, Lasso logistic regression and Random Forest

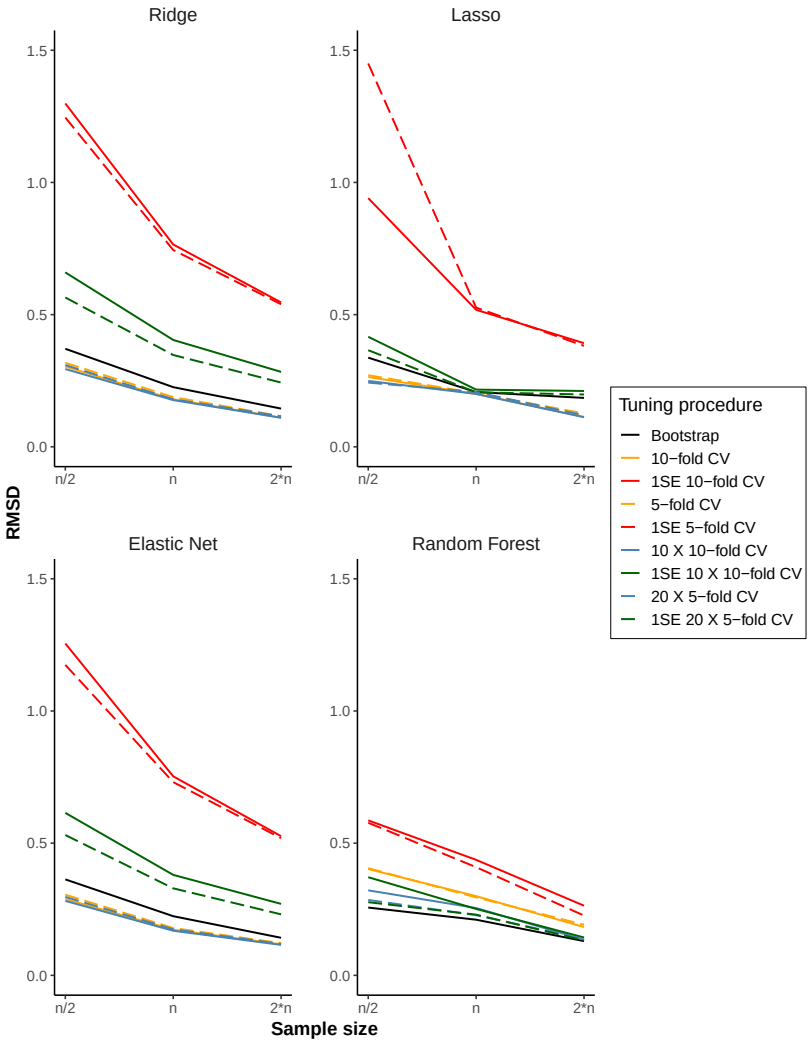

Figure S8: Root mean squared distance (RMSD) of the calibration slope for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.1.

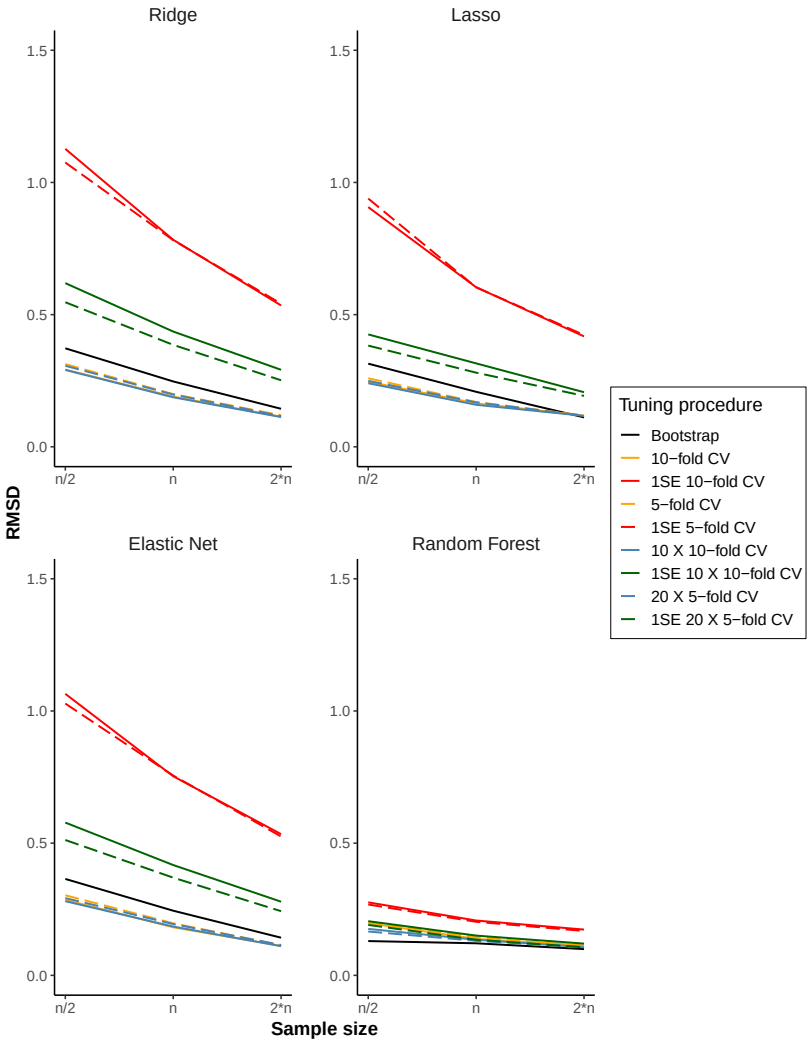

Figure S9: Root mean squared distance (RMSD) of the calibration slope for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.3.

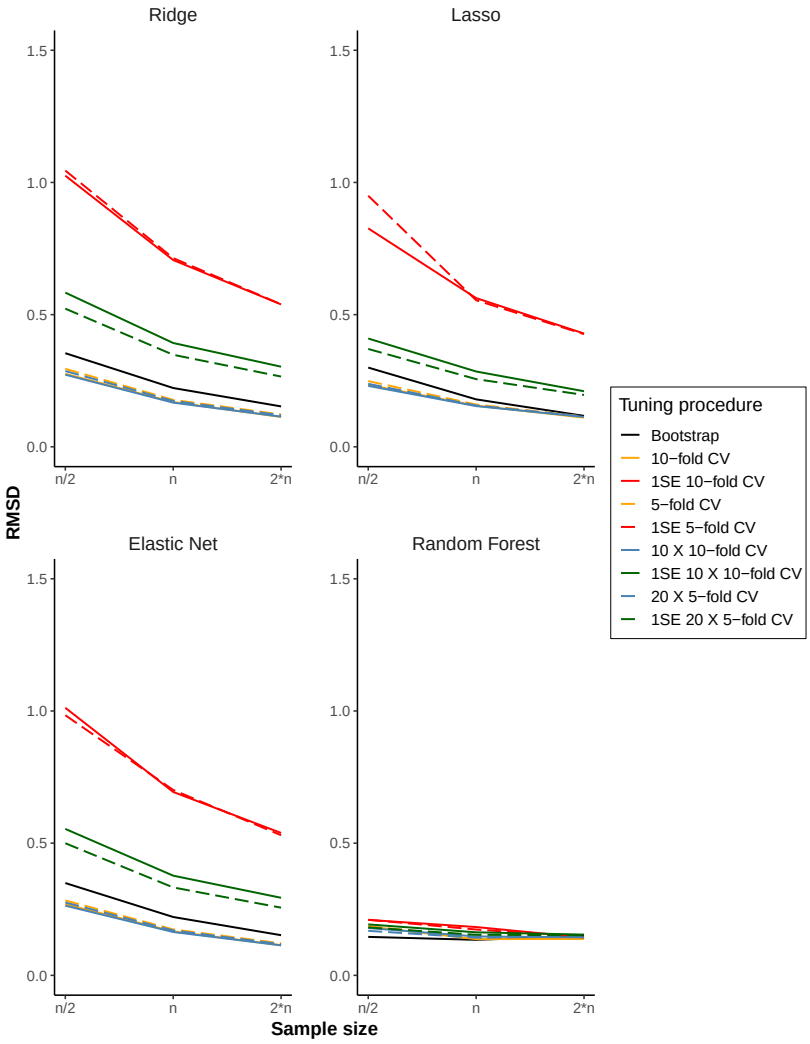

Figure S10: Root mean squared distance (RMSD) of the calibration slope for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.5.

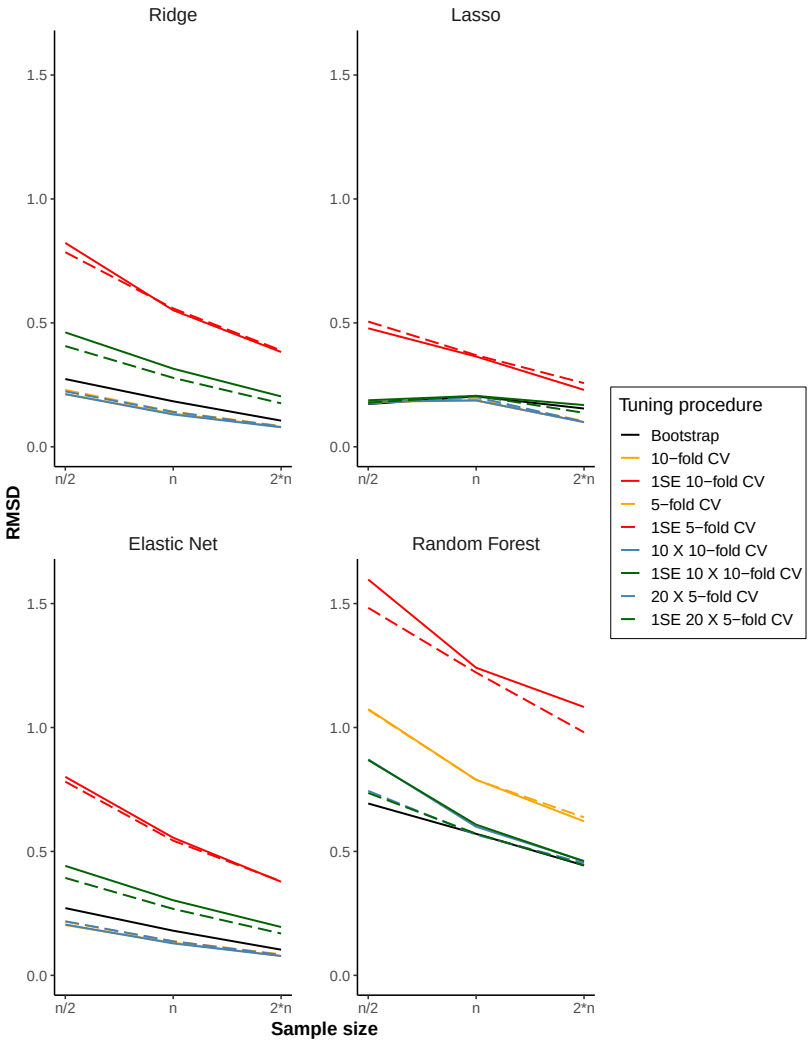

Figure S11: Root mean squared distance (RMSD) of the calibration slope for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 16, events fraction = 0.1.

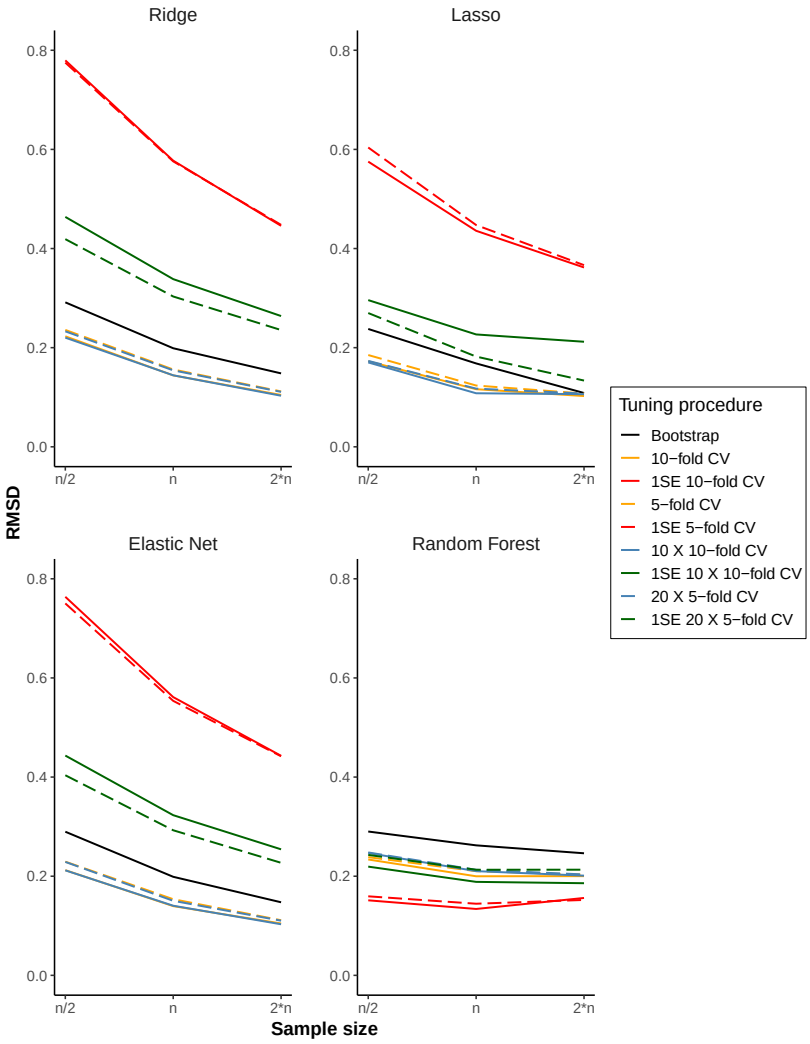

Figure S12: Root mean squared distance (RMSD) of the calibration slope for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 16, events fraction = 0.5.

2.4 C-statistic for Elastic Net, Ridge, Lasso logistic regression and Random Forest

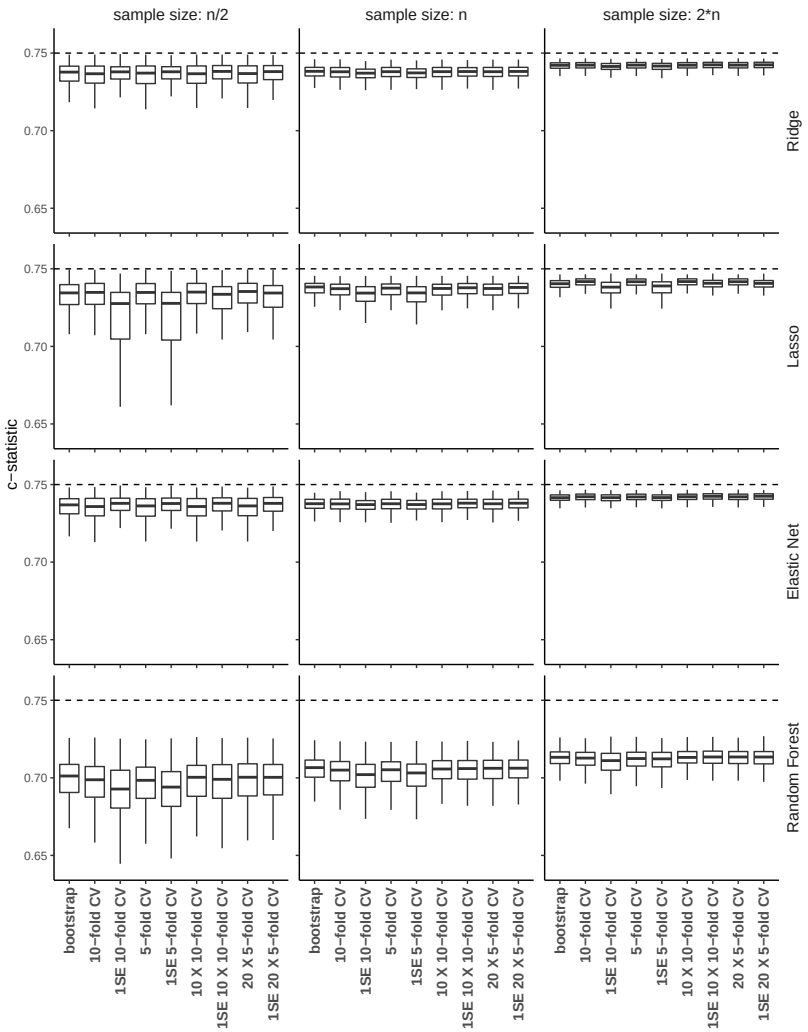

Figure S13: Box plots of the c-statistic for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.1.

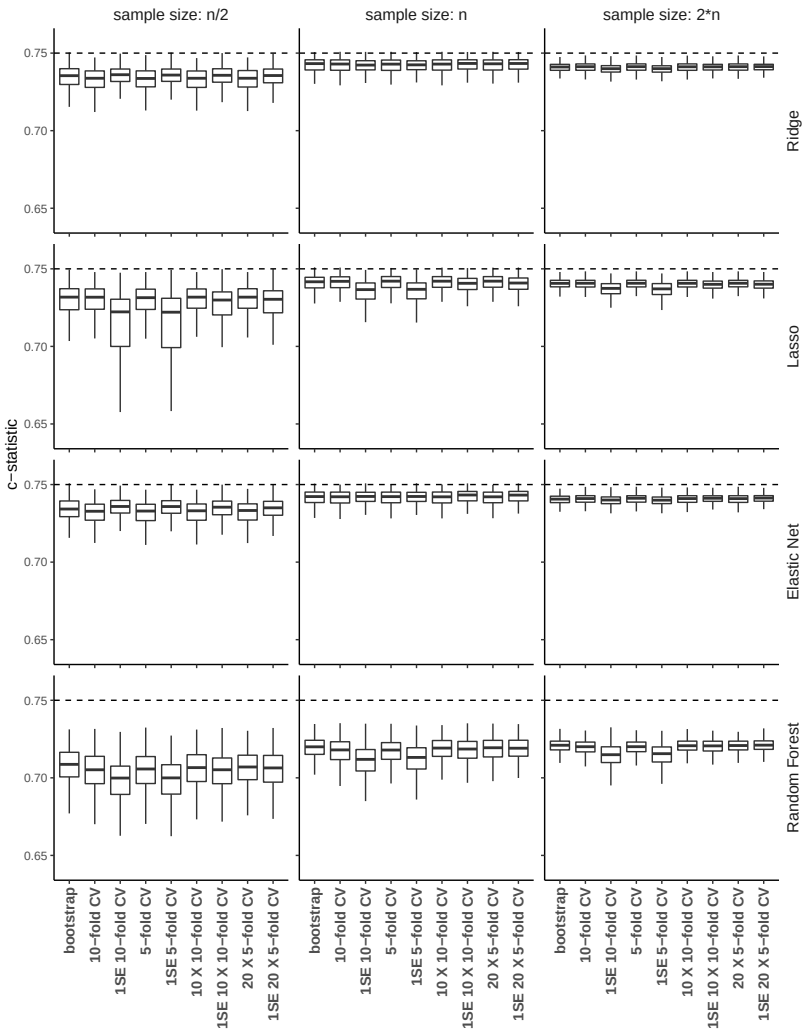

Figure S14: Box plots of the c-statistic for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.3.

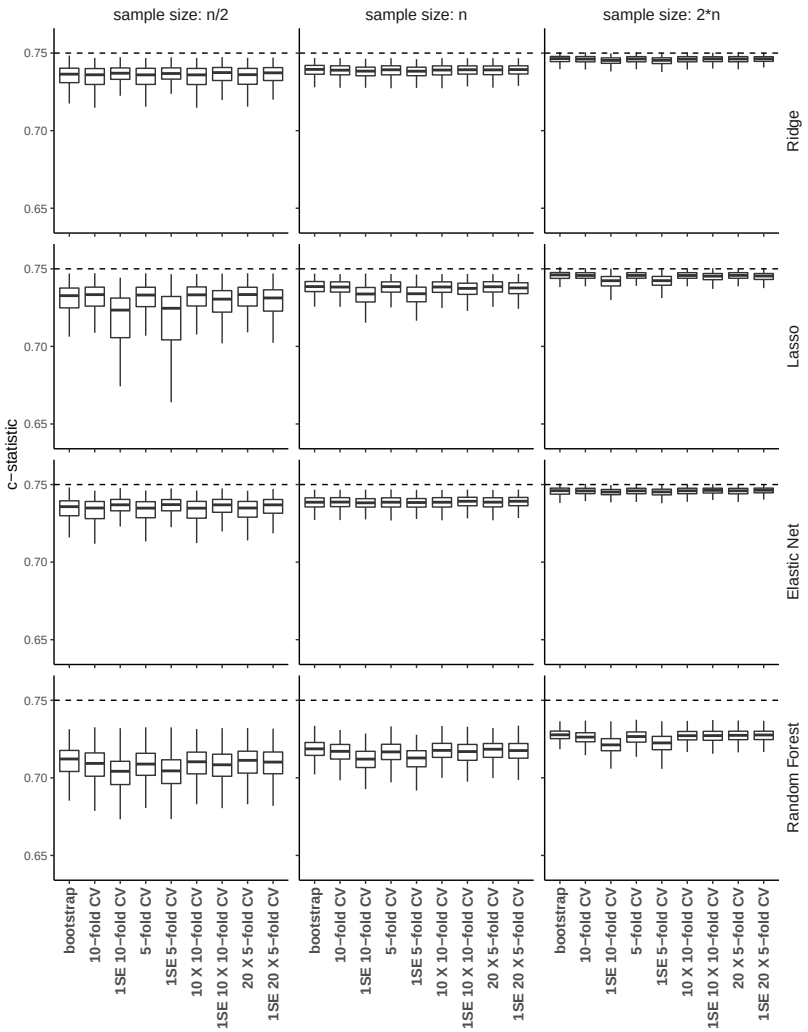

Figure S15: Box plots of the c-statistic for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.5.

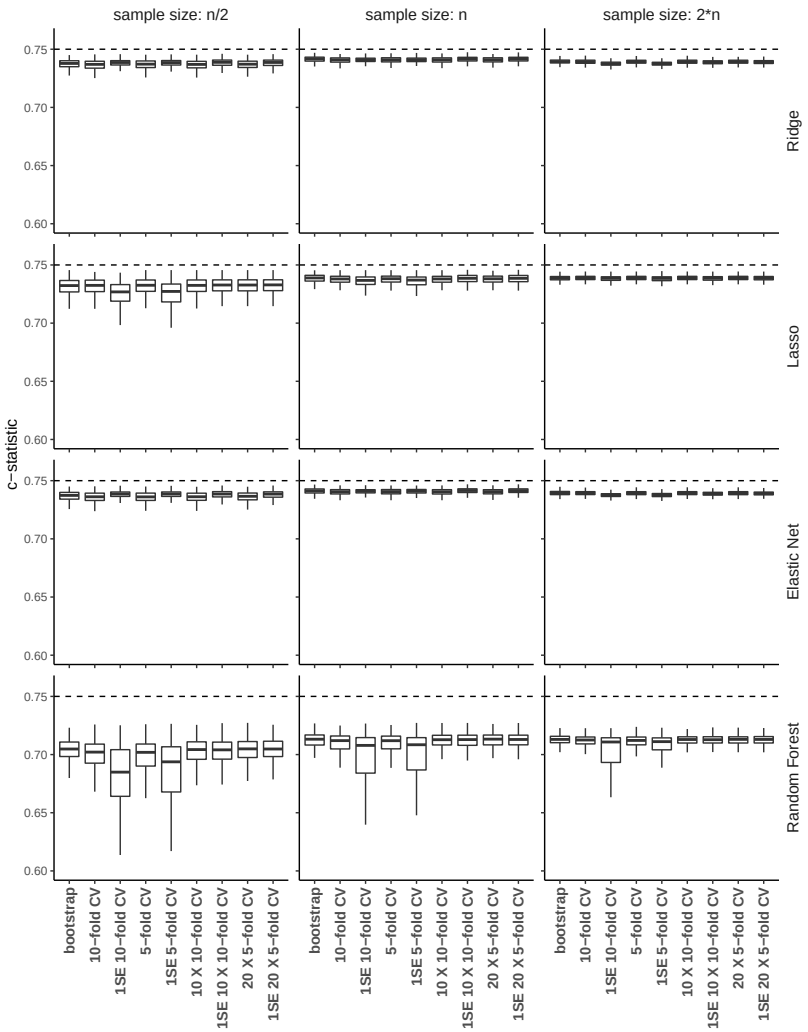

Figure S16: Box plots of the c-statistic for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 16, events fraction = 0.1.

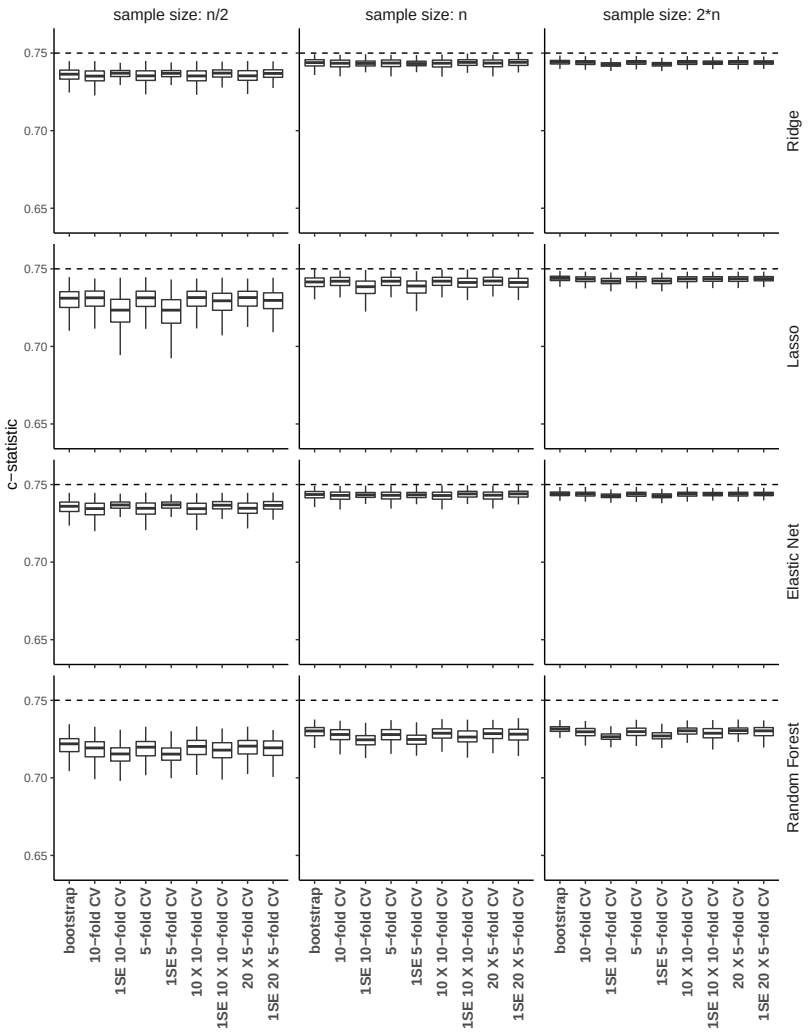

Figure S17: Box plots of the c-statistic for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 16, events fraction = 0.3.

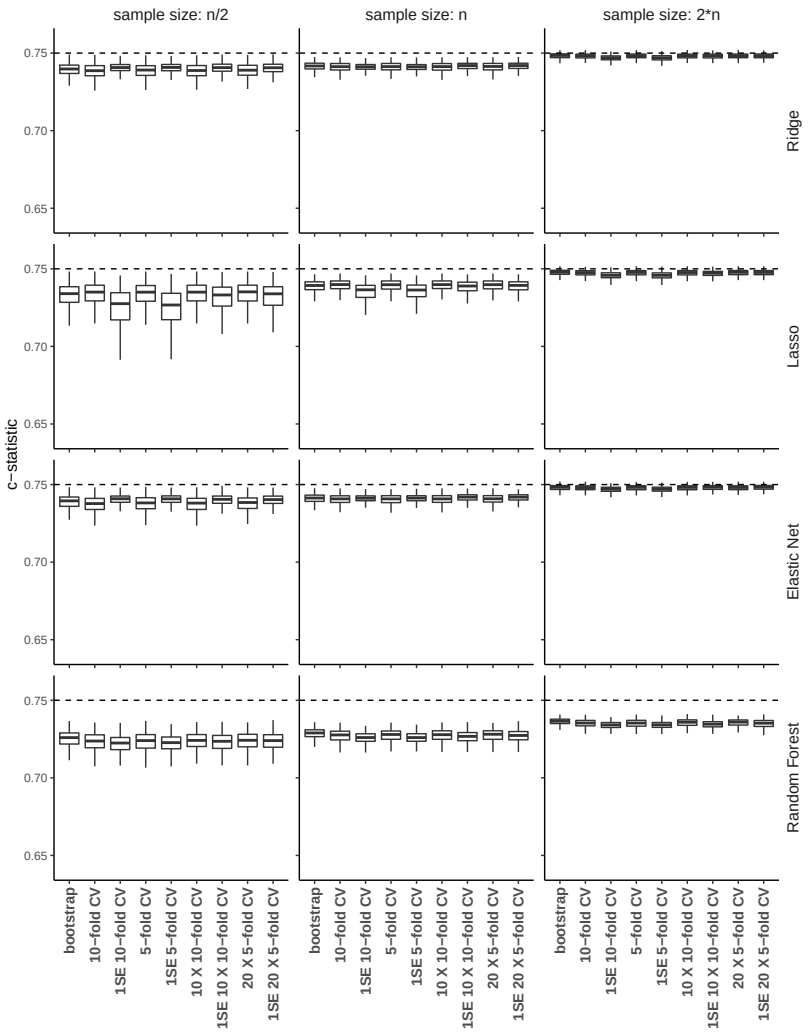

Figure S18: Box plots of the c-statistic for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 16, events fraction = 0.5.

2.5 Mean absolute prediction error (MAPE) for Elastic Net, Ridge, Lasso logistic regression and Random Forest

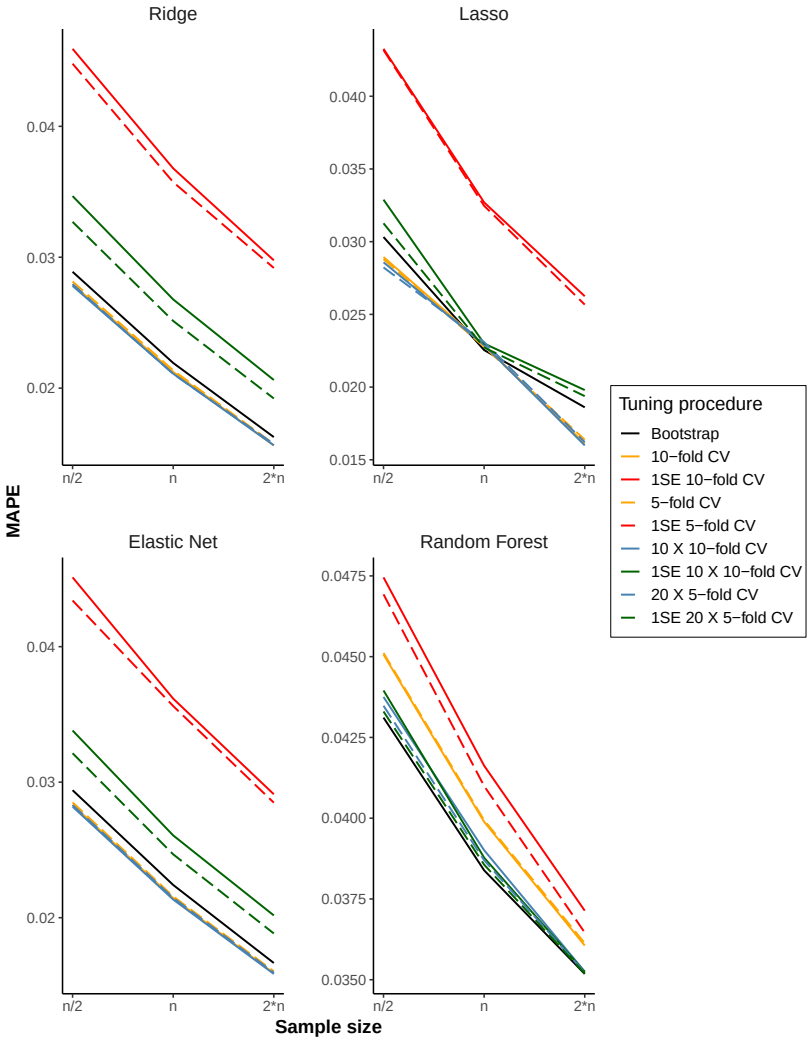

Figure S19: Mean absolute prediction error (MAPE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.1.

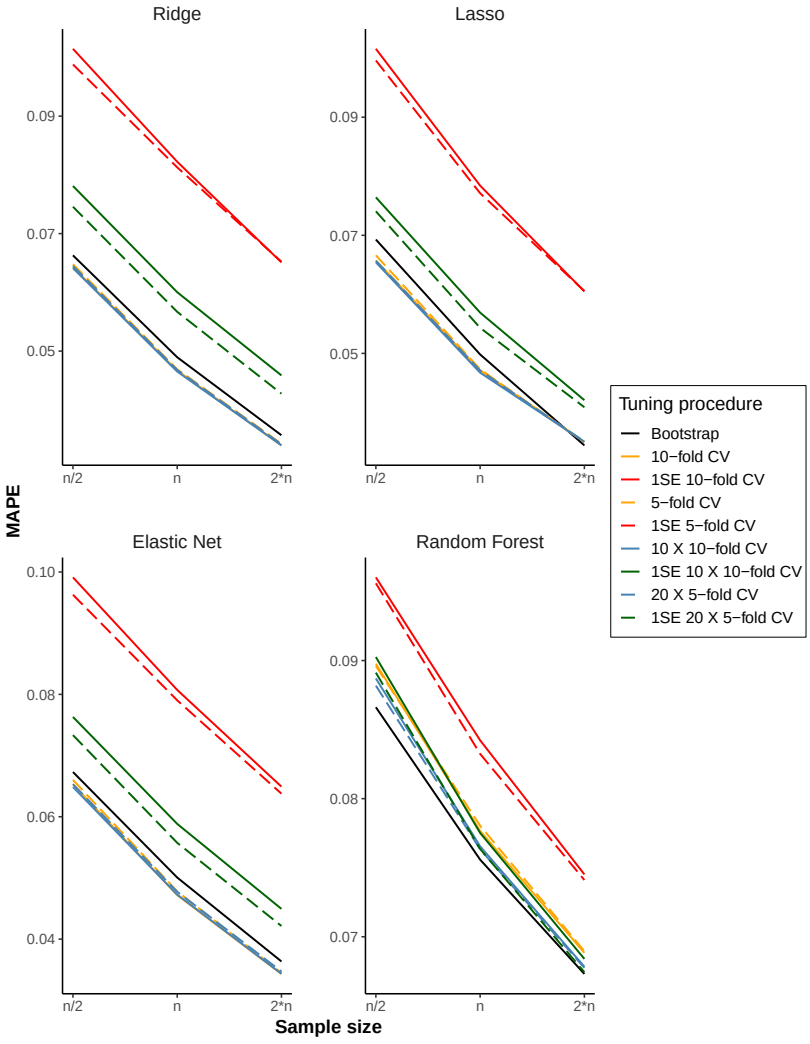

Figure S20: Mean absolute prediction error (MAPE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.3.

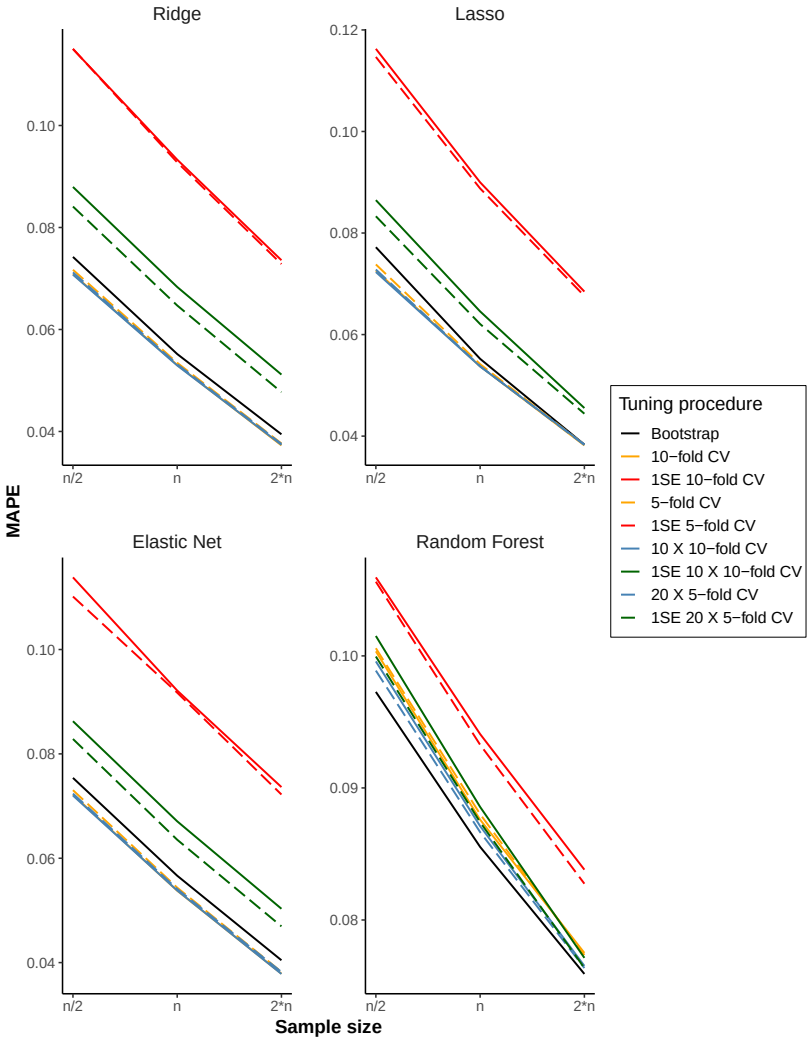

Figure S21: Mean absolute prediction error (MAPE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.5.

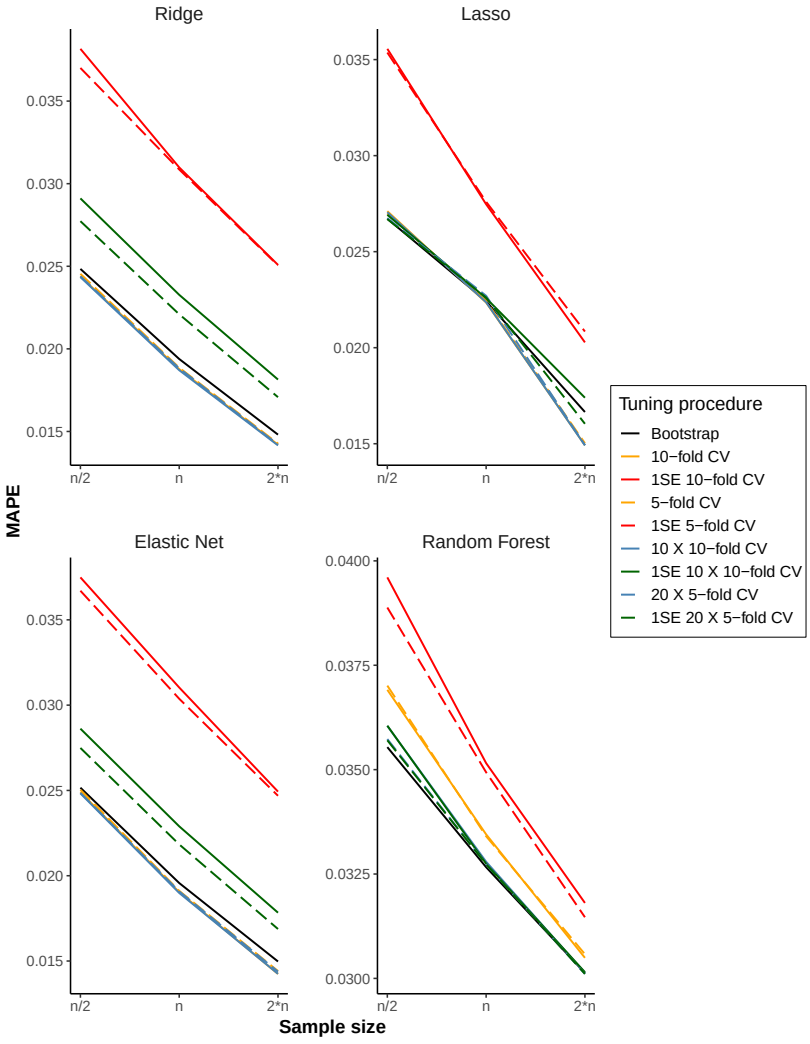

Figure S22: Mean absolute prediction error (MAPE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 16, events fraction = 0.1.

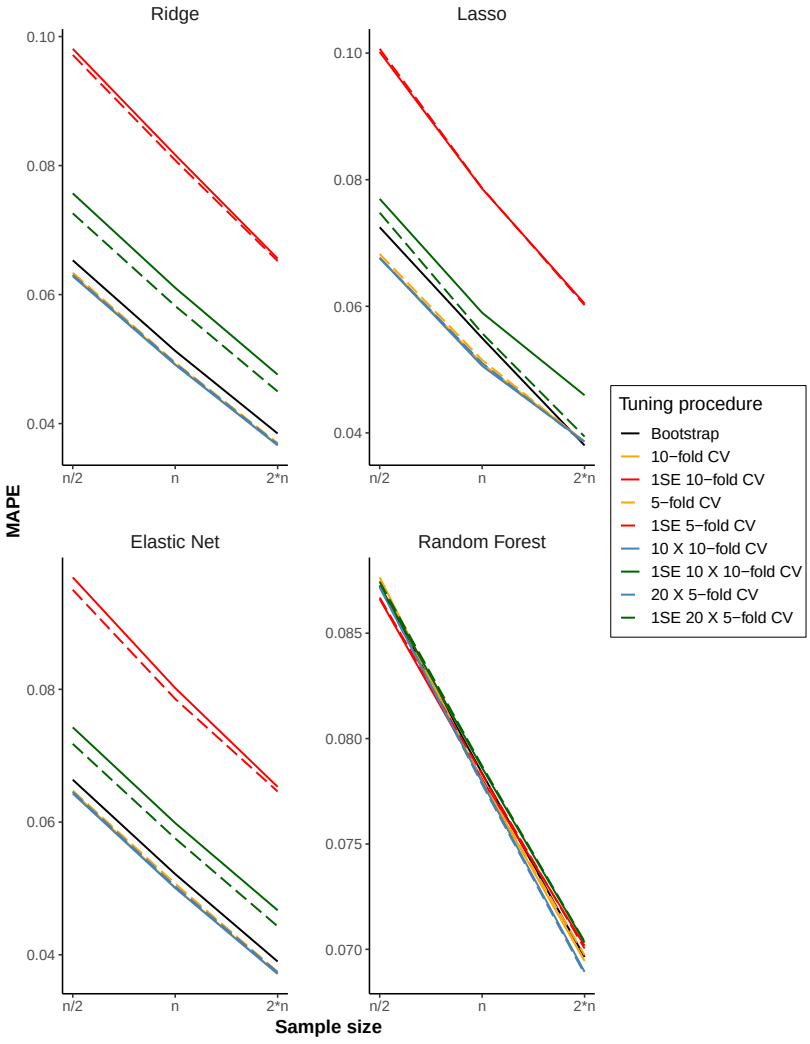

Figure S23: Mean absolute prediction error (MAPE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 16, events fraction = 0.5.

2.6 Square root of the mean squared prediction error (rMPSE) for Elastic Net, Ridge, Lasso logistic regression and Random Forest

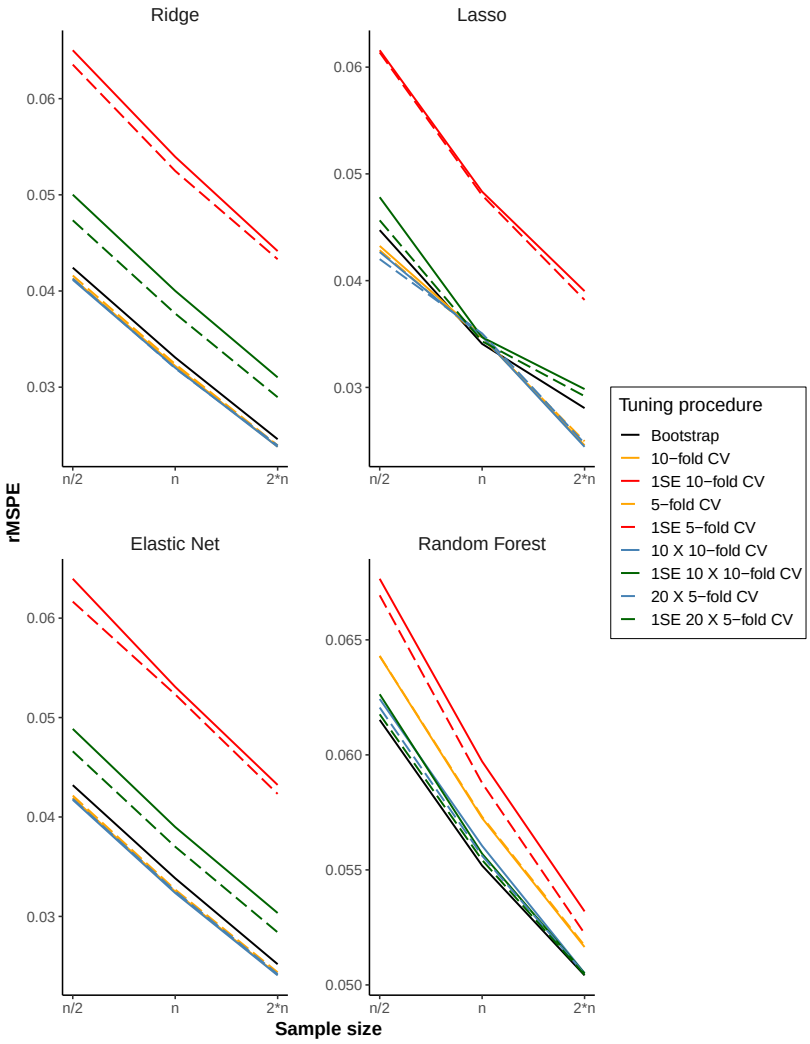

Figure S24: square root of the mean squared prediction error (rMPSE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.1.

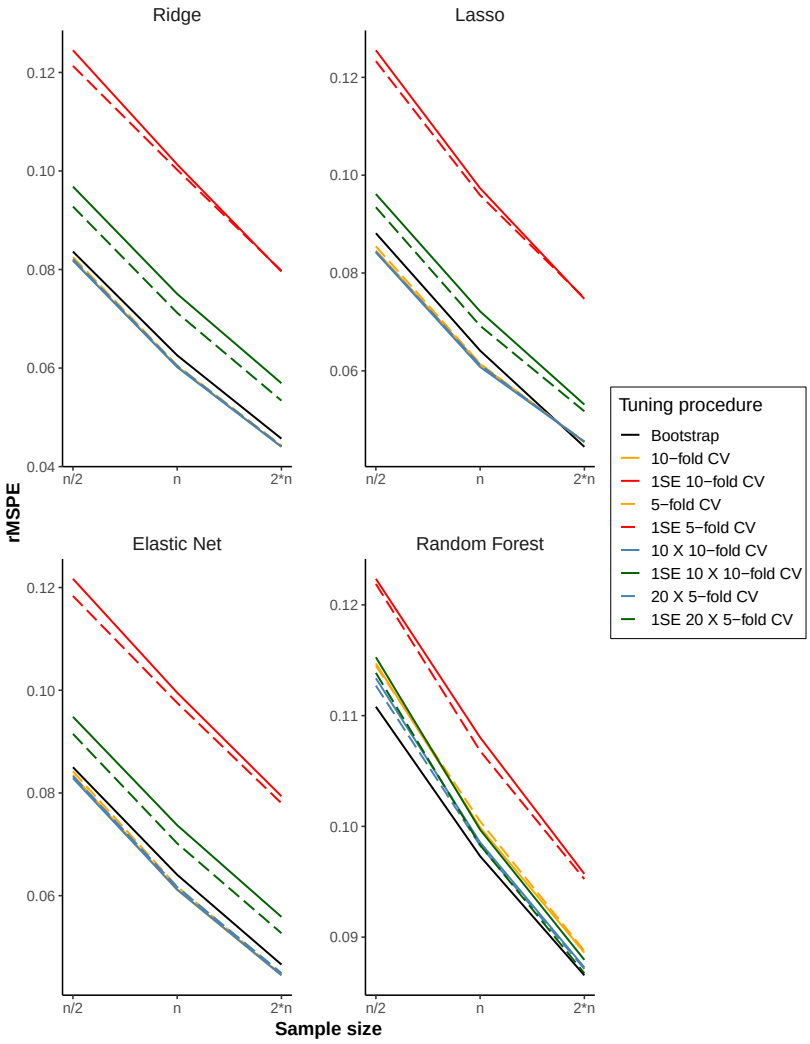

Figure S25: square root of the mean squared prediction error (rMPSE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.3.

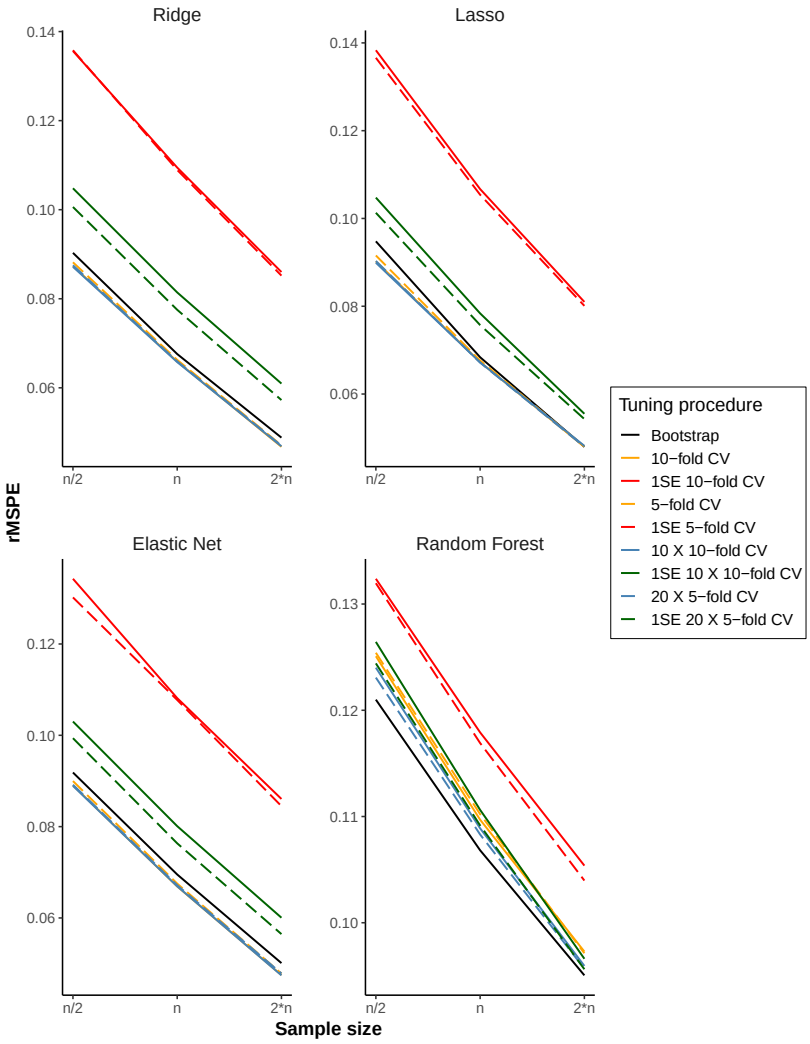

Figure S26: square root of the mean squared prediction error (rMPSE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 8, events fraction = 0.5.

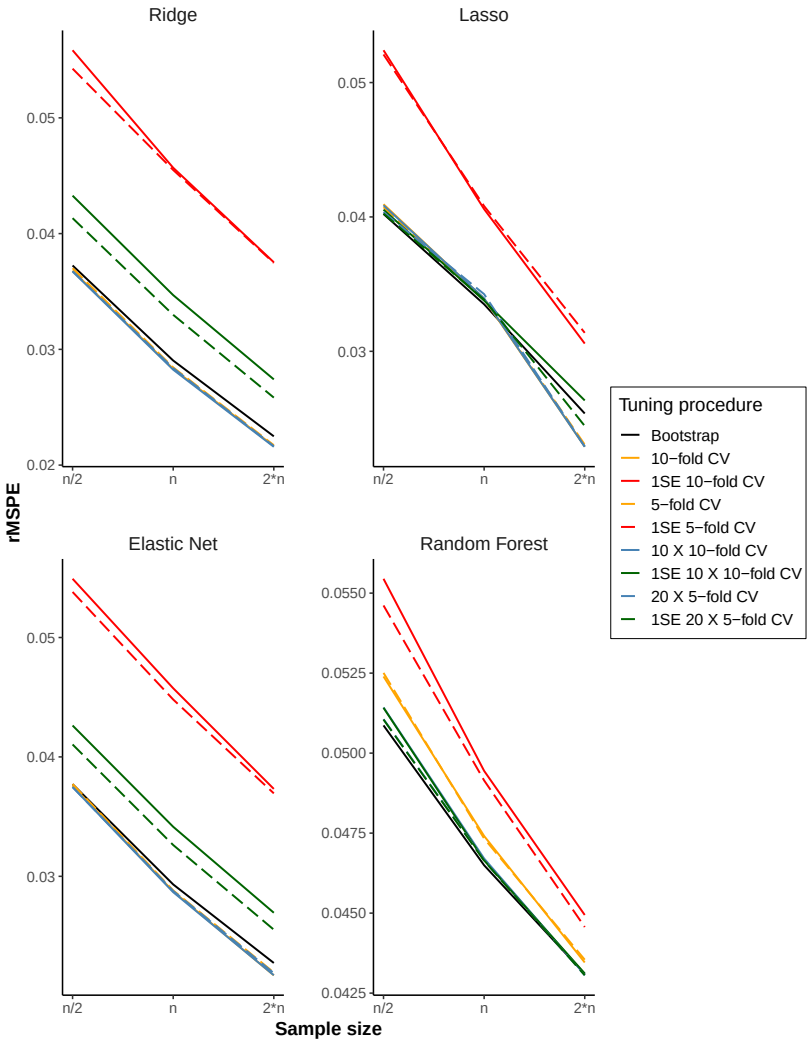

Figure S27: square root of the mean squared prediction error (rMPSE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 16, events fraction = 0.1.

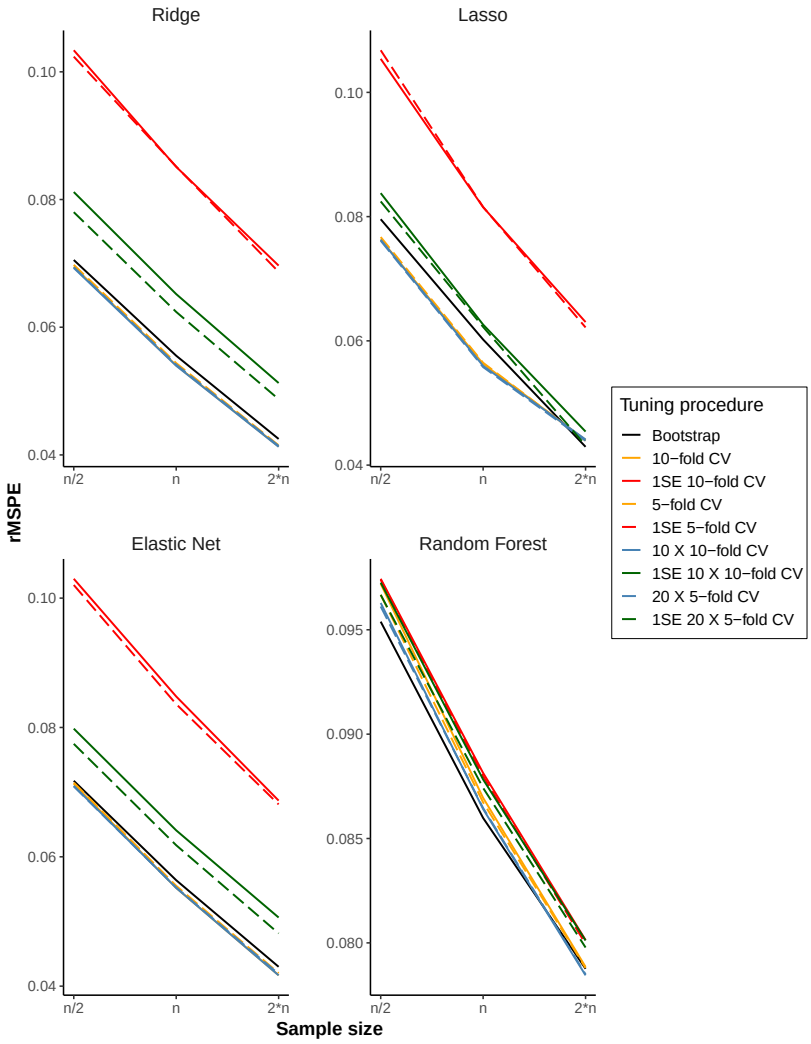

Figure S28: square root of the mean squared prediction error (rMPSE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 16, events fraction = 0.3.

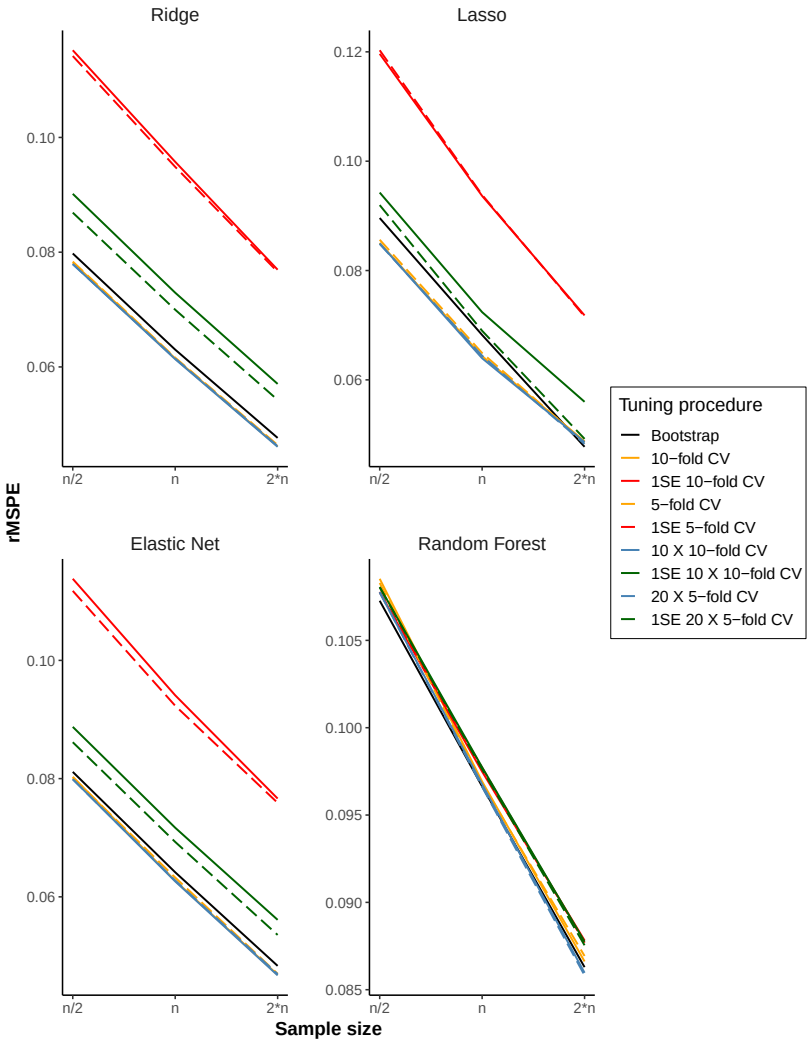

Figure S29: square root of the mean squared prediction error (rMPSE) for all models (Ridge, Lasso, Elastic Net and Random Forest). Scenario: number of predictors = 16, events fraction = 0.5.