Associations and Predictions in Challenging Panel Data Settings

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

Dr. rer. pol.

von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte
Dissertation

von
M.Sc. Lotta Rüter

Tag der mündlichen Prüfung: 06. Februar 2026

Referentin: Prof. Dr. Melanie Schienle
Korreferent: Prof. Dr. Carsten Trenkler

Karlsruhe (2026)

“It is the glory of God to conceal a thing, but the honor of kings is to search out a matter.”

Proverbs 25:2, NIV

Acknowledgements

Soli deo gloria. “Denn durch [Jesus] wurde alles erschaffen ... und alles besteht durch ihn.“ (Kolosser 1,16-17, NGÜ). Ohne Ihn wäre ich nicht, und ohne Ihn wäre auch diese Arbeit nicht entstanden. Seine Liebe bedeutet mir mehr als alles andere auf dieser Welt. Ihm sei alle Ehre, jetzt und für immer.

Mein tief empfundener Dank gilt Melanie. Danke für die vielen wunderbaren Jahre, in denen ich unfassbar viel von Dir lernen durfte, und insbesondere für Deine stets zeitnahe Bereitschaft, Raum zu schaffen, um inhaltliche Fragen, neue Herausforderungen und weiterführende Überlegungen zu besprechen. Ich habe in hohem Maße von Deiner Leitung profitiert, nicht nur fachlich, sondern insbesondere auch menschlich. Besonders danke ich Dir für die Freiheit, Freude und Unterstützung, die Du lebst und mir jederzeit entgegengebracht hast.

Ein herzlicher Dank geht an meine (aktuellen, ehemaligen, neuen und vorübergehenden) Kolleginnen und Kollegen: Andy, Barbora, Daniel, Fabian, Jieyu, Johannes, Konstantin, Lisa, Lora, Marc, Mátyás, Nils, Nina, Osama, Rebekka, Rike, Sam, Sebastian, Sophia, Theo, Tobi Bi., Tobi Bo., Vincent und Zhongwei. Danke für die großartige Zusammenarbeit, jedes hilfreiche Feedback, das Gegenlesen und die vielen fachlichen Diskussionen zu Vorträgen und Zwischenständen dieser Arbeit. Besonders die zahllosen, langen und ausgesprochen lustigen Mittagspausen bleiben mir in bester Erinnerung. Ihr seid mir sehr ans Herz gewachsen und echte Freunde geworden.

Danke an Theda und Mirjam für Eure organisatorische Unterstützung sowie für jedes freundliche Gespräch im Sekretariat und zwischen Tür und Angel.

Danke, Carsten, für die verschiedenen Gutachten und die stets sehr angenehmen Begegnungen in Vorlesungen und Workshops.

Mein Dank gilt außerdem der Mannheim Crew, die mir den Start ins Doktorandenleben erheblich versüßt hat.

Von Herzen danke ich darüber hinaus meiner Familie, die mich besonders in der Abgabephase durch Memes, Mahlzeiten, Kuchen und Care-Pakete seelisch wie körperlich am Laufen gehalten hat.

Zuletzt geht mein besonderer Dank an meine lieben Freunde und Prayer Crew: Anna K., Anna W., Anne, Arved, Biggi, Dani, Doro, Dona, Geli, Gona, Hannah, Jana, Jess, Jules, Labi, Lena A., Lena Wa., Lena Wü., Luisa, Moni, Renate, Rita, Simone, Theresa und Ulrike. Danke für Euer offenes Ohr, gemeinsame Reisen, viele spaßige Stunden, Eure Gebete und den emotionalen Support über die gesamte Zeit dieses Projekts hinweg.

Contents

List of Figures ix

List of Tables xi

List of Abbreviations xiii

1 Introduction 1

2 Model Determination for High-Dimensional Longitudinal Data with Missing Observations: An Application to Microfinance Data 7

2.1 Introduction . . . 7
2.2 Data . . . 12
2.2.1 Choice of the Dependent Variables . . . 15
2.2.2 Included Explanatory Factors . . . 16
2.2.3 Characteristics and Challenges of the Final Data Set . . . 17
2.3 Method . . . 19
2.3.1 The Adapted MIRL Procedure . . . 20
2.3.2 Adaptive Tuning Parameter Choice and Evaluation Criteria . . . 25
2.4 Results . . . 27
2.4.1 Key Financial Success Factors . . . 28
2.4.2 Key Social Success Factors . . . 30
2.4.3 Robustness . . . 34
2.5 Conclusion . . . 38

3 Comparing Forecast Performance on Large Panel Data with Unknown Clustering Structure 41

3.1 Introduction . . . 41
3.2 Theory . . . 44
3.2.1 Test Set-Up . . . 44
3.2.2 Assumptions . . . 45
3.2.3 The Thresholding Variance Estimator . . . 47
3.2.4 The Main Result . . . 50
3.3 Empirical Study: Comparing Sovereign CDS Forecasts . . . 50
3.3.1 Data . . . 51
3.3.2 Model Specifications . . . 53
3.3.3 Results . . . 54
3.4 Conclusion . . . 59

4 Direction Augmentation in the Evaluation of Armed Conflict Predictions 61
4.1 The 2020 VIEWS Prediction Competition and Forecast Evaluation 61
4.2 Definition of the TADDA Score 64
4.3 Scoring Functions and Incentives 65
4.4 Incentives Created by TADDA 67
4.5 Empirical Example 70
4.6 Discussion 73

5 Contribution to the 2023/24 VIEWS Prediction Challenge 75
5.1 Introduction 75
5.2 Data 77
5.3 Modelling Approaches 78
5.3.1 Negative Binomial Distribution 78
5.3.2 Hurdle Model 79
5.3.3 Neural Networks 79
5.4 Results 80
5.4.1 Optimal Window Length for the NB and the Hurdle Model 80
5.4.2 Model Performance 81
5.5 Conclusion 83

6 Challenges in Evaluating Conflict Fatality Forecasts from an Onset Perspective 85
6.1 Introduction 85
6.2 The 2023/24 VIEWS Prediction Challenge 88
6.2.1 Challenge Format and Timeline 88
6.2.2 Targets and Notation 88
6.2.3 Selected Models 90
6.3 A Unified Evaluation Framework 90
6.3.1 Proper Scoring Rules 90
6.3.2 Calibration and Discrimination of Onset Forecasts 93
6.4 Results 94
6.4.1 Global Distribution of Conflict Magnitude and Onset 94
6.4.2 Mean CRPS of Magnitude Forecasts Is Dominated by High-Intensity Conflicts 95
6.4.3 Binary Onset Targets Represent a Distinct Dimension of Performance 98
6.4.4 Calibration of Onset Forecasts Is Challenging 99
6.5 Discussion 100

Appendices **103**

A Appendix to Chapter 2: Model Determination for High-Dimensional Longitudinal Data with Missing Observations: An Application to Microfinance Data **105**
A.1 Details on the Construction of the Panel Data Set 105
A.2 Descriptive Statistics 116
A.3 Detailed aMIRL Results 122
A.4 Detailed MIRL Results 126
A.5 Detailed Two-Step Lasso Results 134

B Appendix to Chapter 3: Comparing Forecast Performance on Large Panel Data with Unknown Clustering Structure **149**
B.1 Proof of Consistency 149
B.2 Descriptive Statistics 154
B.3 Detailed Results 155

C Appendix to Chapter 4: Direction Augmentation in the Evaluation of Armed Conflict Predictions **161**
C.1 Derivation of the Optimal Point Forecast for $\text{TADDA1}_{\epsilon}^{\text{L1}}$ 161
C.2 Derivation of the Optimal Point Forecast for $\text{TADDA1}_{\epsilon}^{\text{L2}}$ 165
C.3 Derivation of the Optimal Point Forecast for $\text{TADDA2}_{\epsilon}^{\text{L1}}$ 166

D Appendix to Chapter 5: Contribution to the 2023/24 VIEWS Prediction Competition **173**
D.1 Modelling Details 173
D.1.1 Negative Binomial Distribution 173
D.1.2 Hurdle Model 174
D.1.3 Neural Networks 174

E Appendix to Chapter 6: Challenges in Evaluating Conflict Fatality Forecasts from an Onset Perspective **177**
E.1 Supplementary Figures 177
E.2 Mathematical Background 178

Bibliography **179**

List of Figures

2.1 Proportion of variables per factor in the final data set 13
2.2 Map of the distribution of MFIs across countries 17
2.3 (Average) correlation in the final, balanced (imputed) panel data set . . 35

3.1 Threshold covariance estimates for average forecast loss differentials of model comparisons with $M > 0$, respective forecasts and CDS spreads . 58

4.1 Illustration of the absolute error and TADDA$_\epsilon$ with $\epsilon = 0.048$ 65
4.2 Probability density function of a skew-normal distribution 68
4.3 Construction of the forecast distribution 72

5.1 Categorisation of countries by conflict level 78
5.2 Average monthly CRPS for retrospective forecasting task, disaggregated by conflict intensity 83
5.3 Optimal models per country by average CRPS per year for retrospective forecasting task 84

6.1 Monthly totals of fatalities and onset events 96
6.2 CRPS results for conflict magnitude 97
6.3 CRPS and Brier Score evaluation under previous peace 98
6.4 Discrimination and calibration of onset forecasts 100

Appendix C

C.1 Illustration of the CDFs F and G for Case 2 of TADDA1$_\epsilon^{L1}$ 163
C.2 Illustration of the CDFs F and G for Case 3 of TADDA1$_\epsilon^{L1}$ 164
C.3 Illustration of the absolute error and TADDA2$_\epsilon$ with $\epsilon = 0.048$ 168

Appendix E

E.1 Classification of monthly conflict situations for countries with at least one conflict between 2018 and 2023 177

List of Tables

2.1 Balanced-panel availability across years in the raw data set 14
2.2 Number of MFIs (countries) per continent and region 18
2.3 Overview of aMIRL coefficients for all target variables 28
2.4 Selection probabilities and aMIRL coefficients for OSS 29
2.5 Selection probabilities and aMIRL coefficients for log(NAB) 31
2.6 Selection probabilities and aMIRL coefficients for –log(ALBG) 33
2.7 Selection probabilities and aMIRL coefficients for –ALBG 37

3.1 Notation and definitions of intraday and daily regressors 52
3.2 Model specifications 55
3.3 Mean squared prediction error of all model specifications 56
3.4 Diebold-Mariano test results for equal predictive accuracy of selected model pairs 57

4.1 Expected scores for different functionals and scoring functions under a skew-normal distribution 70
4.2 Average evaluation scores for different lead times and window lengths . . 73
4.3 Empirical quantiles of different optimal point forecasts 73

5.1 Characteristics of our modelling approaches and their flexibility 76
5.2 Four approaches to determining the optimal window length 79
5.3 CRPS results for the negative binomial distribution and the hurdle model 81
5.4 Evaluation metrics of models and VIEWS benchmarks for retrospective forecasting task 82

6.1 Selected models for our analysis from the 2023/24 VIEWS Prediction Challenge 91

Appendix A

A.1 All variables concerning debt collection practices in the raw data set . . 106
A.2 Details on the construction of the financial and social performance data 107
A.3 Descriptive statistics of the balanced panel data 116
A.4 Descriptive statistics of the unbalanced panel data 119
A.5 Detailed aMIRL results for OSS 122
A.6 Detailed aMIRL results for log(NAB) 123
A.7 Detailed aMIRL results for –log(ALBG) 124
A.8 MIRL results for OSS 126
A.9 MIRL results for log(NAB) 128

A.10 MIRL results for –log(ALBG) 131
A.11 Two-step lasso results for OSS 134
A.12 Two-step lasso results for log(NAB) 138
A.13 Two-step lasso results for –log(ALBG) 143

Appendix B

B.1 Descriptive statistics 154
B.2 Diebold-Mariano test results for equal predictive accuracy of all model pairs 155

Appendix D

D.1 Sampling distributions for neural-network hyperparameters 175

List of Abbreviations

AE	Absolute Error
AIC	Akaike Information Criterion
ALBG	Average Loan Balance per GNI per Capita
aMIRL	Adapted Multiple Imputation Random Lasso
AT	Austria
AUC	Area Under the ROC Curve
BE	Belgium
BIC	Bayesian Information Criterion
BS	Brier Score
CCEW	Center for Crisis Early Warning
CDF	Cumulative Distribution Function
CDS	Credit Default Swap
CET	Central Eastern Time
CHF	Swiss Franc
CRPS	Continuous Ranked Probability Score
CRSP	Center for Research in Security Prices
DE	Germany
DK	Driscoll-Kraay
DM	Diebold-Mariano Test
ES	Spain
EUR	Euro
EVZ	EuroStoxx 50 Volatility Index
FERN	Finance and Econometrics Research and Networking
FI	Finland
FPR	False Positive Rate
FR	France
FS	Funding Spread
FX	Foreign Exchange
GAM	Generalised Additive Model

GARCH	Generalised Autoregressive Conditional Heteroscedasticity
GLMM	Generalised Linear Mixed Model
GLP	Gross Loan Portfolio
GNI	Gross National Income
GR	Greece
HAC	Heteroscedasticity and Autocorrelation Consistent
HKMetrics	Joint initiative of econometricians and statisticians from Heidelberg University, Karlsruhe Institute of Technology and University of Mannheim
HR	Human Resources
IE	Ireland
IGN	Ignorance Score
IMS	Institute of Mathematical Statistics
IT	Italy
LASSO	Least Absolute Shrinkage and Selection Operator
MathSEE	Mathematics in Sciences, Engineering, and Economics
MFI	Microfinance Institution
MICE	Multivariate Imputation by Chained Equations
MIRL	Multiple Imputation Random Lasso
MIS	Mean Interval Score
MIX	Microfinance Information eXchange, Inc.
MM	Markov Model
MSE	Mean Squared Error
MSPE	Mean Squared Prediction Error
MV	Multivariate
NA	Not Available
NAB	Number of Active Borrowers
NB	Negative Binomial Distribution
NL	Netherlands
NLO	Number of Loans Outstanding
NN	Neural Network
NOK	Norwegian Krone
NW	Newey-West
OC	Optimality Criterion
OLS	Ordinary Least Squares
OPF	Optimal Point Forecast

OSS	Operational Self-Sufficiency
PAR	Portfolio at Risk
PMF	Probability Mass Function
PRIO	Peace Research Institute Oslo
PRIO-GRID	Vector grid with a resolution of 0.5 × 0.5 decimal degrees, covering all terrestrial areas of the world
PT	Portugal
PVAR-X	Panel Vector Autoregression with Exogenous Variables
R	R Statistical Computing Environment
RE-EM	Random-Effects Expectation-Maximisation
ReLU	Rectified Linear Unit
RF	Random Forest
ROC	Receiver Operating Characteristic
RP	Reporting Period
RPS	Ranked Probability Score
RSS	Royal Statistical Society
RV	Realised Volatility
SC	Scoring Committee
SE	Squared Error
SHG	Self-Help Group
SPM	Social Performance Management
SUR	Seemingly Unrelated Regression
TADDA	Targeted Absolute Distance with Direction Augmentation
TNB	Truncated Negative Binomial Distribution
TPR	True Positive Rate
UCDP	Uppsala Conflict Data Program
UK	United Kingdom
USD	United States Dollar
UV	Univariate
VIEWS	Violence & Impacts Early Warning System
VIX	Chicago Board Options Exchange's Volatility Index
XGBoost	eXtreme Gradient Boosting
ZL	Zero-Lag

1 Introduction

Economic and political decisions of considerable consequence are increasingly guided by quantitative analyses and forecasts (Manski, 2013; Athey and Imbens, 2017). Governments rely on projections of economic activity to design fiscal and monetary policy, development agencies depend on predictions to allocate scarce resources effectively, and humanitarian actors monitor early warning systems to anticipate outbreaks of violent conflict (Rød et al., 2024; Banerjee and Duflo, 2025; Becker et al., 2025). In all these settings, both the discovery of associations within data and the task of forecasting are not merely academic exercises: they shape real interventions, affect lives, and influence how societies respond to uncertainty. The validity of the resulting insights hinges on whether the underlying statistical methods are appropriate for the complex data environments in which they are applied.

Modern empirical research in economics and related fields is characterised by datasets that are simultaneously rich and challenging (Varian, 2014). Panels often combine many cross-sectional units with time dimensions that can range from very short to very long, substantial and non-monotone missingness, strong temporal persistence, and complex, often unknown cross-sectional dependence driven by shared exposures or latent group structures (Baltagi, 2008; Little and Rubin, 2019). At the same time, the number of potential predictors frequently rivals or exceeds the number of observations, often alongside multicollinearity, each complicating model selection (Meinshausen and Bühlmann, 2010; Shah and Samworth, 2013). While advances in data availability have greatly expanded the scope of empirical research, the assumptions underlying many standard econometric models have become increasingly difficult to justify.

These challenges are particularly acute in forecasting and forecast evaluation. Comparing predictive models requires evaluation procedures that remain valid under dependence, heterogeneity, and structural uncertainty—conditions that are the rule rather than the exception in applied work (Hewamalage et al., 2023). Moreover, different forecasting tasks demand different notions of predictive success. Accurately predicting magnitudes may be crucial in some contexts, while correctly anticipating directional changes or rare onsets may be far more relevant in others. Scoring rules and test procedures that are misaligned with these substantive objectives risk rewarding uninformative or overly conservative forecasts, thereby undermining the very purpose of prediction exercises (Gneiting, 2011).

This dissertation is motivated by the need for statistical methods that take these realities seriously. It contributes to applied econometrics and statistical forecasting by developing tools explicitly designed for high-dimensional, dependent, and non-standard panel data, and by studying how forecasts should be evaluated when predictive goals are nuanced and policy-relevant. Across its chapters, the thesis advances a coherent agenda:

(i) to design model determination and inference procedures that respect complex data-generating processes;

(ii) to evaluate forecasts using methods that are robust to unknown dependence structures and aligned with the forecasting task at hand; and

(iii) to embed empirical analysis within transparent and reproducible workflows.

The relevance of this agenda is illustrated through applications to development economics, financial forecasting, and conflict prediction. In development economics, understanding which institutional characteristics drive the social and financial performance of microfinance institutions is essential for designing policies that effectively support the world's poorest populations (Yunus, 2009; Banerjee et al., 2015; Hermes and Hudon, 2019). In finance, properly accounting for unknown cross-sectional dependence is crucial for conducting valid tests when comparing competing model forecasts (Petersen, 2009). In conflict forecasting, the stakes are particularly high: Russia's sudden invasion of Ukraine underscored how difficult—and how important—it is to anticipate escalations and onsets of violence (Mueller and Rauh, 2022a). Initiatives such as the Violence & Impacts Early Warning System (VIEWS) seek to advance this goal by promoting systematic forecasting through international prediction competitions, but their success crucially depends on the properties of the scoring rules and evaluation frameworks employed (Gneiting and Raftery, 2007; Gneiting, 2011; Hegre et al., 2019, 2022).

This thesis engages with these challenges both theoretically and empirically. It develops new methodologies for model determination in high-dimensional panels with missing data, proposes a novel test for equal predictive accuracy under unknown clustering structures, and provides a detailed analysis of scoring rules and forecast evaluation strategies in the context of conflict prediction. Throughout, particular emphasis is placed on principled statistical reasoning, practical relevance, and reproducibility. All results presented in this work are supported by openly available code and carefully documented data-processing pipelines. The data employed are likewise openly available, with the exception of the application in Chapter 3, for which the code and data-processing pipelines are public, but the data themselves are proprietary. The chapters of this thesis are outlined in more detail below.

Chapter 2 proposes a methodology for model determination in high-dimensional longitudinal data with complex missingness, characterised by a short time-series and a large number of potential regressors. The chapter extends the multiple imputation random lasso framework of Liu et al. (2016) to a fixed-effects panel setting, providing a unified treatment of severe multicollinearity, extensive non-monotone missing data, and unobserved time-invariant heterogeneity. By integrating multiple imputation, stability selection, and panel techniques, the approach delivers principled and data-driven variable selection and estimation that remain robust in environments where standard methods often fail (Dernoncourt et al., 2014). The methodological contribution is illustrated through an extensive application to a global balanced panel of 213 microfinance institutions comprising observations of 136 characteristics over six years, providing novel insights into the drivers of their success. The results identify staff composition as a key determinant of social outreach measures and profitability as the dominant driver of financial sustainability. They further indicate that financial sustainability and breadth of outreach are not inherently in conflict, while the relationship with depth of outreach is more nuanced. This chapter is joint work with Melanie Schienle and was published in the *Journal of the Royal Statistical Society: Series A* (Rüter and Schienle, 2025).

Chapter 3 develops a novel Diebold-Mariano type test for evaluating the equal predictive accuracy of forecast models in panels with a large time series and a possibly large cross section. The framework accommodates forecast loss differentials with substantial heterogeneity, unknown clustered cross-sectional dependence, and serial correlation over time. A key feature of the approach is a thresholded variance estimator, which does not require prior knowledge of the number or composition of clusters and allows for overlapping or non-independent clusters, making the test particularly well suited for complex data environments such as financial forecasting. An empirical application to intraday sovereign CDS spread forecasts illustrates that thresholding is especially important when average forecast score differences are small.

Chapter 4 analyses the targeted absolute deviation with direction augmentation (TADDA) score, which was introduced in the 2020 VIEWS Prediction Competition to account for both the sign and magnitude of log-changes in fatalities (Hegre et al., 2022). Although the score has an intuitive motivation, the empirical results revealed a striking dominance of a no-change benchmark model. This chapter provides a statistical explanation for this outcome, showing that TADDA creates incentives for forecasters to issue overly conservative predictions concentrated near zero. It further demonstrates how performance can be improved by tailoring forecasts to the score's properties and concludes by pointing to alternative scoring approaches that more appropriately reflect directional forecasting goals. This contribution is joint work with Johannes Bracher, Fabian Krüger, Sebastian Lerch, and Melanie Schienle and was published in *International Interactions* (Bracher et al., 2023).

Chapter 5 presents a contribution to the 2023/24 VIEWS Prediction Challenge which adopts a transparent and data-driven modelling strategy that targets optimal performance under the Continuous Ranked Probability Score (CRPS), the competition's primary evaluation metric. Comparing three approaches ranging from simple parametric specifications to neural networks, the results show that a simple negative binomial distribution fitted on past data performs best on the initially available test data, outperforming more complex alternatives while remaining interpretable and computationally efficient. The model captures the persistence characteristic of conflict dynamics but is inherently limited in predicting first-onset events. Additional test data released closer to the submission deadline indicate improved performance of neural networks, suggesting that more flexible models may become advantageous as conflict dynamics shift over time. This chapter represents joint work with Tobias Bodentien, and the corresponding submission was featured in the competition's summary article published in the *Journal of Peace Research* (Hegre et al., 2025).

Chapter 6 evaluates the ability of probabilistic fatality forecasts submitted by different teams in the 2023/24 VIEWS Prediction Competition to capture conflict onset risk (Hegre et al., 2025). Onset probabilities are derived from predicted fatality distributions and evaluated using the Brier score, complemented by calibration and discrimination diagnostics. The results show that while most models distinguish reasonably well between high- and low-risk contexts, they are calibrated with mixed accuracy. Differences in performance are driven primarily by calibration rather than discrimination, and increased model complexity does not necessarily lead to superior onset prediction. These findings highlight both the difficulty of forecasting rare political events and the need for models explicitly designed to target onset rather than magnitude. The chapter further presents an evaluation framework in which probabilistic forecasts of relative changes in fatalities, as considered in the 2020 VIEWS Prediction Challenge, are derived from the original magnitude forecasts and assessed using the threshold-weighted CRPS, which places greater weight on large or policy-relevant changes. This contribution represents joint work with Tobias Bodentien, Johannes Bracher, and Melanie Schienle.

2 Model Determination for High-Dimensional Longitudinal Data with Missing Observations: An Application to Microfinance Data

This chapter is based on joint work with Melanie Schienle, published in the *Journal of the Royal Statistical Society: Series A* (Rüter and Schienle, 2025)[1], and presented at the *HKMetrics Workshop*, University of Mannheim (05/2022), the *FERN Seminar*, KIT (06/2022), the *8th Annual Conference of the International Association for Applied Econometrics*, King's College London (06/2022), the *Africa Meeting of the Econometric Society*, Nairobi (06/2023), and the *28th International Panel Data Conference*, University of Amsterdam (07/2023). It introduces an adaptation of the multiple-imputation random lasso (aMIRL) for longitudinal data with unobserved fixed effects, enabling robust variable selection under complex missingness and high dimensionality, and applies it to uncover key social and financial success drivers of microfinance institutions (MFIs) worldwide.

The code for the proposed aMIRL technique and all reference models is available at `https://github.com/lottarueter/aMIRL`. The repository also includes the raw data and the code for all preprocessing steps, ensuring full reproducibility of the empirical study.

2.1 Introduction

Longitudinal data naturally emerge in many areas of research such as biostatistics, sociology, health, labour and development economics. Such data are often incomplete,

[1]Rüter, L. and M. Schienle (2025): "Model Determination for High-Dimensional Longitudinal Data with Missing Observations: An Application to Microfinance Data," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, qnae144. Reprinted by permission of Oxford University Press.

where a moderate share of overall missing values is distributed across observations so unfavourably that the amount of complete cases becomes negligible. Often, the underlying problem is also high-dimensional, with a large cross-sectional dimension and few available observations, where covariates additionally might be highly correlated variables. Moreover, despite many included covariates in general, substantial amounts of subject-specific unobserved heterogeneity remain that need to be captured by appropriate panel data methods. Aiming at model selection and inference, these challenges are usually addressed separately, with the main approaches being the following. In practice, missing data are mostly either list-wise deleted, which can (obviously) lead to a substantial loss of valuable information, or imputed via multiple imputation. The latter method replaces the missing values with draws from probability distributions, commonly using either the joint posterior distribution of all variables with missing observations (Little and Rubin, 2019) or the conditional distribution of each variable conditioned on other variables in the data (van Buuren, 2007). Many extensions have been proposed, e.g., to include interaction effects (Goldstein et al., 2014), general non-linear effects (Bartlett et al., 2015) or to account for sampling weights (Zhou et al., 2016). To deal with the high dimensionality of the data, variable selection methods such as the lasso and the elastic net can be used. These methods, however, have undesirable properties under multicollinearity of the data in that they tend to select only one of the highly correlated variables and shrink the impact of all others to zero (Wang et al., 2011). Lastly, in the specific context of individual-level time-invariant heterogeneity, these methods may have "poor estimation and inference properties" (e.g., Belloni et al., 2016).

In this chapter, we jointly address the above points by building on the multiple imputation random lasso method introduced by Liu et al. (2016) and by adapting it to account for unobserved idiosyncratic effects in the data. We thereby obtain a methodology that yields robust results in the presence of high dimensionality and multicollinearity for rather complex structures of missingness in longitudinal data as indicated above. The robustness we confirm by selection consistency across different base optimisation criteria and sign consistency of effects for different ways of assessing post-selection effects. Our suggested aMIRL procedure generates values for missing entries via multiple imputation, where the imputation step uses the panel structure of the observations and is as general as possible in imposed functional forms. In particular, we propose the use of a combination of mixed effects models and regression trees based on the random effects expectation maximisation (RE-EM) tree technique of Sela and Simonoff (2012) for continuous variables and classification trees for binary covariates. The imputation step is followed by random lasso and stability selection, which are both performed on the within-transformed imputed data for the inclusion of fixed effects in the final linear model. The combination of imputation and stability-selection-enhanced model determination yields robust feature selection and model estimation in the presence of high dimensionality and multicollinearity; see Meinshausen and Bühlmann (2010) and Wang et al. (2011). An overview of further methods that perform variable selection on imputed data is provided by Zhao and Long (2017).

We employ the aMIRL method for identifying success factors of MFIs in a purely data-driven manner using the MIX Market data set from the World Bank Data Catalog that is characterised by many potentially relevant covariates with incomplete observations and therefore pronounced missingness.[2] The problem is of substantial interest since, over the last decades, microfinance institutions have been established to counter the problem that the poor have little access to financial help, since they are not considered creditworthy by most banks due to their (obvious) lack of financial securities. MFIs hand out small credits (usually a few hundred USD) on terms and conditions different from those of common banks (e.g., Morduch, 1999; Brau and Woller, 2004; van Rooyen et al., 2012; Quayes, 2015). Instead of demanding financial securities, they rather come with obligations such as regular meetings with a liability group or participation in special training. Given that loans are accompanied by certain safety measures, such as flexible repayment horizons and repayment limits, microcredits provide an important and successful tool to fight poverty not only on an individual but also on a macroeconomic level, as shown by Yunus (2009) and Imai et al. (2012). Though the impact of microcredits is highly context-sensitive (Brau and Woller, 2004; Hulme, 2000), multiple studies have determined significant individual-level effects that go beyond financial aid. These include the empowerment of women (Cheston and Kuhn, 2002; Brau and Woller, 2004; Yunus, 2009), the generation of businesses and new jobs (Brau and Woller, 2004), and positive changes in work ethics (Banerjee et al., 2015). Significant long-term effects on the welfare of villages and economies through higher wage and employment levels have been documented by Brau and Woller (2004), Imai et al. (2012) and Buera et al. (2021). Given these positive effects of microcredits, it is argued that MFIs are most successful when

[2] https://datacatalog.worldbank.org/dataset/mix-market, retrieved on 4 September 2023.

they reach a large number (breadth of outreach) of especially poor (depth of outreach) borrowers while being financially sustainable and hence independent of external funding (e.g., Hartarska and Nadolnyak, 2007; Hermes and Lensink, 2007; Bogan, 2012; Quayes, 2015).

In this work, we identify the determinants of such social and financial success in a data-driven way. To account for individual-level time-invariant heterogeneity via fixed effects, we construct a balanced panel comprising 1278 observations of 136 variables from 213 MFIs operating in 55 different countries that covers a span of six years (2009 to 2014). The final data set is thoroughly built, ensuring that all potentially meaningful and important variables, as specified in the literature (e.g., Basharat et al., 2015; Quayes, 2015; Hermes and Hudon, 2019), are contained, but redundant or uninformative measures are omitted. We apply the aMIRL method to account for the challenging structure and degree of missingness (95.5% of the observations are incomplete with 13.7% missing values in total), high dimensionality and multicollinearity while including MFI-specific fixed effects.

To demonstrate the importance of accounting for the unobserved heterogeneity in the data as well as the additional robustness in the variable selection step, we compare our aMIRL results with fixed effects and pooled regression results from the original MIRL method as well as conventional lasso estimates with column-mean imputations. We further supplement the pooled regression results of the balanced panel data set with those of a large unbalanced data set with 3846 observations of 1026 MFIs located in 100 different countries from 2007 to 2018.

To our knowledge, this work is the first to quantify and determine the importance of the personnel structure (rather than focusing on the role of management only, as analysed in Kyereboah-Coleman and Osei, 2008 or described in Hermes and Hudon, 2019) as a key driver for the social success of MFIs. Amongst others, a greater borrower-staff ratio and an increased number of employees further an MFI's outreach. Both financial and social success benefit on average from reduced roles of management or board. Other drivers of an MFI's overall success are greater financial performance (main determinant of financial success) and lower costs. Breadth of outreach can be increased by setting certain new staff incentives and targeting specific (new) borrower groups. Depth of outreach is associated with higher charged interest rates and risk of default, as also noted by Yunus (2009). We find that financial sustainability and breadth of outreach can go hand in hand, while the relationship between financial success and depth of outreach is less pronounced. Our results confirm the presence of mission drift, meaning that reaching more borrowers can lead to targeting wealthier borrowers and thus deviating from the original mission of serving the poorest (Armendáriz and Szafarz, 2011). On the other hand, issuing smaller loans can help to increase an MFI's breadth of outreach. Our aMIRL models significantly outperform the results of existing studies that also analyse MFI success based on the MIX Market data set in terms of goodness of fit measured by R^2. For example, our models for financial sustainability yield an R^2 that is 0.54 higher (0.88 vs. 0.34) than that of Quayes (2015), who also uses balanced panel data and employs a fixed effects model to estimate the effects of potential drivers on operational sustainability.

Most studies on microfinance success that use non-experimental data employ linear panel models with pre-selected predictors, where the choice of regressors often depends on data availability and observations with missing values in selected regressors are simply dropped (Ayayi and Sene, 2010; Quayes, 2015). They analyse the sustainability and/or outreach of MFIs with regard to certain aspects such as competition (Assefa et al., 2013), poverty reduction (Khandker, 2005), profit orientation (Roberts, 2013), governance (Hartarska and Nadolnyak, 2007; Kyereboah-Coleman and Osei, 2008), and capital structure (Bogan, 2012). Hermes and Hudon (2019) provide a systematic review of literature on the determinants of social and financial performance of MFIs and state that "research on MFI performance is still in its infancy". The objective of this study is to contribute to the existing literature by examining both financial and social success drivers, thereby allowing for a direct comparison of the determinants of these dimensions of MFI success, while simultaneously accounting for the challenging missingness structure in the data, performing data-driven variable selection and addressing the high dimensionality in the data.

Our empirical findings complement theoretical economic model-driven studies that analyse specific aspects such as the performance of MFIs in the presence of competition (McIntosh and Wydick, 2005), potential deviation from their mission to reach the poor (Armendáriz and Szafarz, 2011), the diffusion of microcredits (Banerjee et al., 2013) and their impact on whole economies (Buera et al., 2021).

Lastly, a large portion of the microfinance literature stems from randomised experiments where, mostly, ordinary linear models with control variables and/or treatment dummies are used (Field and Pande, 2008; McIntosh, 2008; Swain and Wallentin, 2009; Field et al., 2013; Berge et al., 2015; Banerjee et al., 2015). Due to the specificity of these setups in spatial design and research question, however, strong assumptions are required to derive general implications from such data.

This work is organised as follows. The next section provides details on the construction of the balanced panel data used. Section 2.3 describes the method under investigation and Section 2.4 presents our empirical results. Final conclusions and an outlook are given in Section 2.5.

2.2 Data

The data used in this work originate from the MIX Market data set, which was made available in the World Bank Data Catalog on 28 October 2019.[2] MIX is an acronym for Microfinance Information eXchange, Inc., a provider of global self-reported financial and social performance data of microfinance institutions launched in 2002 (Imai et al., 2012). The MIX Market data set is the largest, most reliable data source on MFIs and comprises both financial and social performance data. In particular, our raw data are obtained from an inner join of the *Financial Performance Data Set in USD* and the *Social Performance Data Set* of the MIX Market data, i.e., it comprises all MFIs that appear in both source data sets. Concerning the reliability of the data source, according to the World Bank's Data Catalog, data collection and reporting for MIX took place "in line with broadly recognised reporting standards within microfinance and inclusive finance".[2] That means predetermined reporting formats were used and internal as well as external cross-checks were performed for validation (Quayes, 2015). For reproducibility, all following data pre-processing steps described below and in the appendix are available as software code in the GitHub repository `https://github.com/lottarueter/aMIRL`. Additionally, we kept the (sub)categories and variable names of the original data set while constructing the panel used in this work; see Table A.2 in the appendix.[3] However, for reasons of readability and interpretation, we introduce more compact names for the resulting variables and group them into different new factors, which are represented by different colours as shown in Figure 2.1.

[3]Note that the character ">" functions as a subcategory indicator in the variable names of the original data set. Hence, the variable *Av. Loan Size > Gender > Female* denotes the variable *Av. Loan Size of Female Borrowers*. We employ a similar notation in this work. Instead of ">" we use "▷".

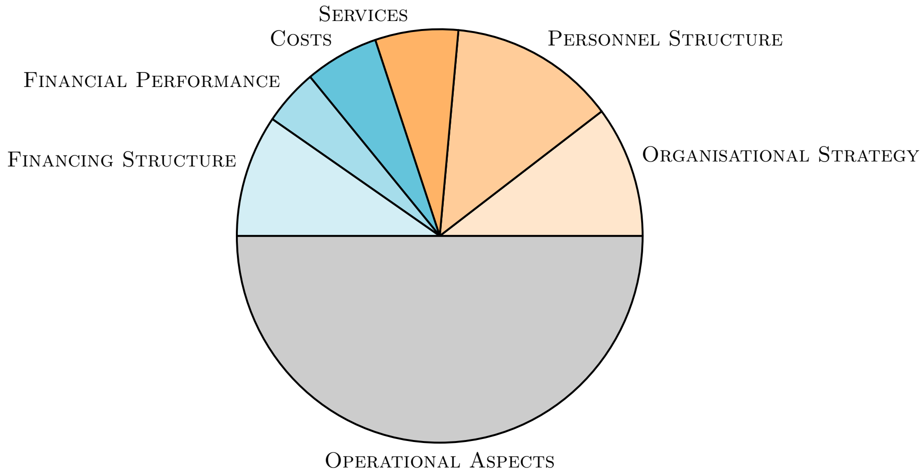

Figure 2.1: Proportion of variables per social (orange), operational (gray) and financial (blue) factors in the final data set

Note that the raw data consist of a largely unbalanced panel comprising the years 2007 to 2018, where from year to year many new MFIs appear, existing ones disappear, and sometimes previously contained ones reappear (see Table 2.1). Conditional on even the largest set of observable characteristics for MFIs, however, it is well known that MFIs across the globe are quite heterogeneous and have their own peculiar specificity (see, e.g., Fall et al., 2023). In order to capture such MFI-specific heterogeneity in a practically feasible fixed effects approach, we construct a balanced panel from the unbalanced raw data. This is the main basis for our empirical analysis. If we used unbalanced data instead, different sample sizes in the time dimension $T = T_i$ would affect the within-transformation differently for different MFIs, resulting in non-standard statistical properties of the final estimates that are beyond the scope of the present project. We would further still have to exclude the many cases with $T_i < 2$ that do not allow for the identification of a fixed effect.

In the construction of the balanced panel, we select the timing and time span with consecutively available MFIs such that it yields the maximum number of available balanced panel elements. An MFI is classified as "available" in the raw data if for all time points in the respective period it has at least one non-zero entry per target variable and across all factors specified below in detail. Please see Table 2.1, which displays all possible allocation options of a balanced panel within the unbalanced raw one with corresponding sample sizes. As the optimal window for our analysis, we chose $w^* = [2009, 2014]$ with

Table 2.1: Data availability for balanced panel allocations within the raw unbalanced data

	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
2008	199 (199)	280 (140)	417 (139)	540 (135)	625 (125)	666 (111)	700 (100)	56 (7)	18 (2)	0 (0)	0 (0)
2009		330 (330)	636 (318)	924 (308)	1,084 (271)	1,160 (232)	**1,278 (213)**	63 (9)	24 (3)	9 (1)	10 (1)
2010			375 (375)	728 (364)	912 (304)	1,012 (253)	1,140 (228)	60 (10)	28 (4)	16 (2)	18 (2)
2011				552 (552)	918 (459)	1,101 (367)	1,284 (321)	55 (11)	30 (5)	21 (3)	16 (2)
2012					612 (612)	922 (461)	1,188 (396)	76 (19)	35 (7)	18 (3)	14 (2)
2013						617 (617)	1,042 (521)	84 (28)	44 (11)	35 (7)	12 (2)
2014							671 (671)	82 (41)	51 (17)	32 (8)	15 (3)
2015								46 (46)	34 (17)	24 (8)	12 (3)
2016									155 (155)	146 (73)	123 (41)
2017										111 (111)	110 (55)
2018											163 (163)

Note: Rows indicate the start year of the window, columns the end year. In each cell, the top number shows the balanced panel size $N_w \times T_w$, with the number of MFIs N_w in parentheses. Start-year 2007 windows are omitted because only 15 MFIs existed in that year.

$N_{w^*} \times T_{w^*} = 213 \times 6 = 1,278$ available panel observations. We prefer this to the case that comprises only four years instead of six with an only slightly larger sample size of 1284. For our goal of determining the success factors of MFIs, we consider a larger time horizon crucial for capturing the evolution of MFIs but also for improving the precision of fixed effect estimates and the respective within-transformations in a panel model.

This results in a balanced panel data set that contains observations of 213 MFIs operating in 55 different countries for six consecutive years (2009 to 2014). For all MFIs in the balanced panel, we have observations on the considered three different target variables and 136 explanatory factors as detailed in Sections 2.2.1 and 2.2.2 below. Detailed summary statistics are provided in Table A.3 in the appendix. The data structure in the final balanced panel contains major challenges that comprise in particular the degree and structure of missingness in covariates and that require specialised statistical techniques to address them. Please see Section 2.2.3 below.

Due to the limited sample size of the constructed balanced panel, we use an unbalanced panel that comprises all years 2007 to 2018 for robustness checks (see Section 2.4.3). It comprises 1026 MFIs and 3846 observations for the same set of target variables and covariates as in the balanced panel. Its summary statistics are given in Table A.4. The table illustrates that not only the non-consecutive years in the panel but also the structure and degree of missingness of covariates pose an additional challenge in this case.

2.2.1 Choice of the Dependent Variables

Since poverty reduction should be the main objective of MFIs, they are considered successful if they maximise their social impact rather than, e.g., their profits.[4] Given the various positive effects of microcredits mentioned in the introduction, MFIs are defined as socially successful if they reach (i) many and (ii) particularly poor borrowers (Cull et al., 2007, e.g.,). The first dimension, breadth of outreach, is measured by the logarithm of the *Number of Active Borrowers*, *log(NAB)* (Hartarska and Nadolnyak, 2007; Quayes, 2015). However, direct measures of the second dimension, depth of outreach, such as the income level of borrowers, are not available. Under the assumption that poorer borrowers generally receive smaller loans, we use the negative of the logarithm of *Average Loan Balance per Borrower / GNI per Capita*, *–log(ALBG)*, as a standardised proxy, as done by Hartarska and Nadolnyak (2007) and Armendáriz and Szafarz (2011), amongst others. To maximise long-term social success, MFIs are required to be operationally sustainable and hence independent of external funding. The money originally used for financing can then be used to set up new MFIs, which increases the overall social impact. An MFI is operationally sustainable when its revenues exceed its costs. The most common indicator of financial MFI performance is therefore *Operational Self-Sufficiency*, *OSS* (Hartarska and Nadolnyak, 2007; Assefa et al., 2013; Quayes, 2015). It relates the total income of the MFI to the expenditure required for its operation; see Table A.2 for the exact definition. To summarise, we study success determinants of MFIs with regard to *log(NAB)*, *–log(ALBG)* and *OSS*.

[4]Profit maximisation is argued to be a dangerous goal by Yunus (2009), since it likely leads to the exploitation of the poor rather than their benefit.

2.2.2 Included Explanatory Factors

We only include variables with fewer than 50% missing entries and perform further transformations, which are described in detail in Section A.1 and Table A.2 in the appendix. For the resulting full list of the 136 considered variables and details of their definitions as well as their respective transformations, we also refer to Section A.1 and Table A.2 in the appendix. The final set of characteristics in the balanced panel covers a wide span of potential success factors, ranging from performance indicators and measures of the financial structure of the MFI to its borrower structure, aggregated personnel data, information on staff incentives and development targets.

The 105 included variables from the *Financial Performance Data Set in USD* provide information on the distribution of the loan portfolio, the client as well as the personnel structure, and the financial situation of each MFI. Here, we focus on an overview of the considered different categories in the original data set. Data on the shares of the loan portfolio and loan sizes per borrower type, lending methodology and credit delay are given in the categories *Clients*, *Credit Products*, *Delinquency* and *Outreach*. Additionally, information on the number of (new) borrowers and the number of loans outstanding is included. Further, balance sheet positions and ratios describing the financing, risk and liquidity structures as well as the income and expenses are presented in the corresponding categories *Balance Sheet*, *Financing Structure*, *Risk & Liquidity* and *Income*. Financial performance, productivity and efficiency measures and further revenue and expenses sizes are given by the categories *Financial Performance*, *Productivity & Efficiency* and *Revenue & Expenses*. The personnel structure is portrayed in the categories *Infrastructure* and *Social Performance*. We implicitly account for macroeconomic characteristics, since the variables *GNI per Capita* and *Inflation Rate* are contained in quantities such as *Av. Loan Bal. / GNI p. c.*, *Av. Salary / GNI p. c.* and *Yield on GLP (Real)*.[5]

The 31 considered variables from the *Social Performance Data Set* contain information on client protection measures, staff incentives and additional services provided by the MFI; see the categories *Client Protection*, *Governance & HR* and *Products & Services*. The category *Social Goals* further comprises development goals, a quantification of the MFI's focus on poverty reduction, and its target markets.

[5]The explicit consideration of these variables only increases the multicollinearity in the data but does not lead to significantly different results. Corresponding results can be made available on request.

2.2.3 Characteristics and Challenges of the Final Data Set

53% of the included MFIs operate in South America and 35% in Asia, while Europe and Africa each provide 6% of the data, as shown on the map in Figure 2.2 and Table 2.2. More than 75% of the MFIs are operationally sustainable (i.e., *Operational Self-Sufficiency* ≥ 1). 50% have assets of more than USD 25 million and a gross loan portfolio of more than USD 20 million while serving at least 25,813 borrowers; see Table A.3 in the appendix. In addition, clearly only MFIs that have been operating successfully for at least six years are included. These aspects should be taken into account when interpreting our results. While our findings may contain practical insights for the establishment of new MFIs globally, statistically we can only identify variables that are associated with increased success of already established MFIs operating primarily in South America and Asia.

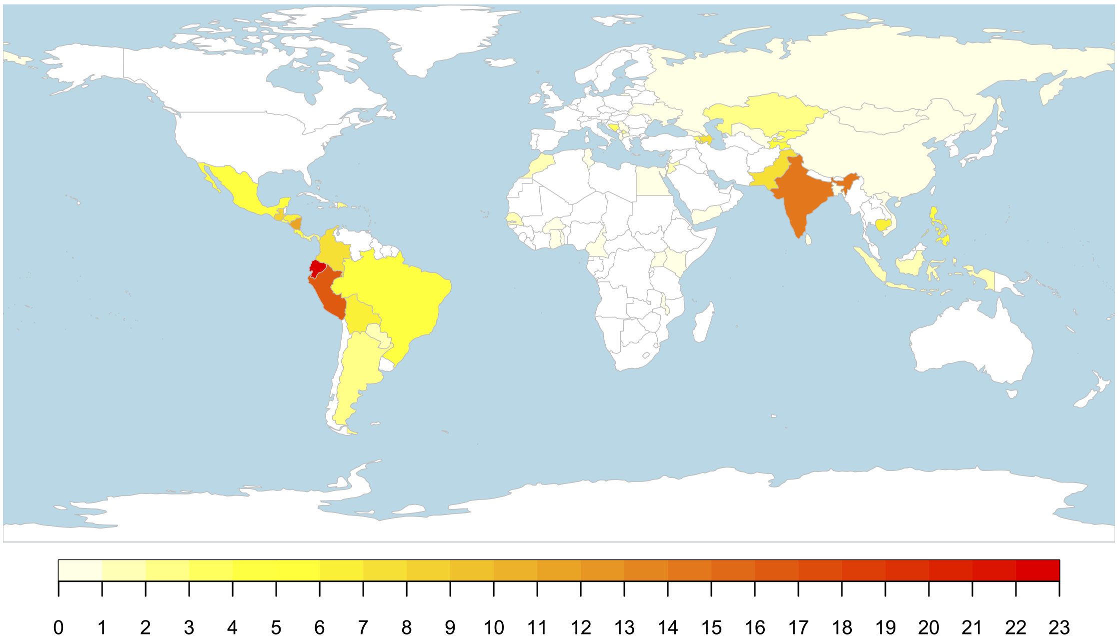

Figure 2.2: Map of the the number of microfinance institutions per country in the final data set

Many of the potential regressor variables are moderately or highly correlated; see Figure 2.3 in Section 2.4.3. Depending on the target variable, we remove those regressors

Table 2.2: Number of different microfinance institutions (countries) per continent and region in the final data set

	Africa	America	Asia	Europe	
East Asia & Pacific	0	0	18 (7)	0	18 (7)
Europe & Central Asia	0	0	27 (7)	12 (7)	39 (14)
Latin America & Caribbean	0	113 (15)	0	0	113 (15)
Middle East & North Africa	4 (3)	0	5 (4)	0	9 (7)
South Asia	0	0	25 (4)	0	25 (4)
Sub-Saharan Africa	9 (8)	0	0	0	9 (8)
	13 (11)	113 (15)	75 (22)	12 (7)	213 (55)

that are by definition almost identical to the target in the respective analysis.[6] The resulting maximum correlation between an outcome variable and a potential regressor is 0.73.

The main statistical challenge of the final balanced panel data set, however, consists in the substantial amount of missingness in covariates and target variables, as documented in Table A.3 in the appendix. In particular, only for 58 of all considered 1,278 instances it ($i = 1, \ldots, 213$ and $t = 1, \ldots, 6$ with $213 \times 6 = 1,278$) in the balanced panel are all 136 factors completely observed.[7] Note that it is the allocation of the missing values that causes 95.5% of instances in the panel to suffer from incompletely observed characteristics. When counting the aggregate number of unobserved factors and target variables across all MFIs and time points relative to the total possible number of $1,278 \times 136$ in the balanced panel, there are only 13.7% missing values, causing the degree of overall missingness to appear rather moderate. Concerning the allocation of missing observations across covariates, note that there is no inherent ordering in the factors and no further structure in missing observations across covariates. This implies that the missingness is non-monotone in covariates, i.e., variables of the predictor set cannot be sorted in such a way that the missingness of one variable implies the missingness of all subsequent variables.

[6] In particular, for target variable *log(NAB)* we exclude *# Active Borrowers* and *# Loans Outstanding* from the set of potential regressors. For *–log(ALBG)* we remove *Av. Loan Bal. / GNI p. c.* and *Av. Outst. Bal. / GNI p. c.*. *log(NAB)* and *–log(ALBG)* are only used as target variables. In the set of potential regressors, their untransformed versions *# Active Borrowers* and *Av. Loan Bal. / GNI p. c.* are included.

[7] In the unbalanced panel from 2007 to 2018 in the comparison study, only 177 of 3,846 instances are complete. See Table A.4 for details in this case.

The multicollinearity and missingness in the data hence pose non-trivial challenges for the model determination and estimation in this work, which requires specialised statistical techniques.

2.3 Method

Our goal to identify MFI success factors in a purely data-driven manner using the panel data presented calls for a method with corresponding characteristics. First, the method must be able to identify the most important variables and filter out their singular effects in the presence of high dimensionality and correlated variables. Second, it must be able to handle the considerable amount of complex missingness. Third, it should incorporate the longitudinal structure of the data.

We build on the Multiple Imputation Random Lasso procedure (MIRL) as introduced by Liu et al. (2016), which combines random lasso for variable determination with multiple imputation and stability selection, specifically to handle the high degree of missingness in the data. Random lasso is a two-step procedure that provides variable selection and parameter estimation of linear models in high-dimensional settings with multicollinearity. It has been shown to outperform similar existing methods in such cases in terms of prediction accuracy (Wang et al., 2011). Combined with multiple imputation, it systematically accounts for missingness in the data (Azur et al., 2011). The incorporation of stability selection further yields an importance ranking of all variables, where a data-adapted threshold determines the final set of predictors. In particular, the procedure with stability selection requires much weaker conditions than the original lasso in order to yield consistent variable selection for dependent data (Meinshausen and Bühlmann, 2010).

We propose an adaptation of the original MIRL approach that accounts for the longitudinal structure in our data. For the estimation and variable selection step, we use a linear fixed effects panel setting, where MFI-specific unobserved characteristics are captured by time-invariant individual effects α_i, so that

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \epsilon_{it}, \tag{2.1}$$

with y_{it} denoting the value of the target variable for MFI i at time t for $i = 1, \dots, N$ and $t = 1, \dots, T$, where $N = 213$ and $T = 6$ in our case. Moreover, $\mathbf{x}_{it} = (x_{1,it}, \dots, x_{p,it})'$

comprises the observations of p explanatory variables for i at time t, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ contains the linear effects of $x_1, \ldots, x_p$ on y, and ϵ_{it} is the random error term of the model for individual i at time t, satisfying the standard panel fixed effects exogeneity assumptions. Using the usual within-transformation as in Belloni et al. (2016), a pilot estimate $\hat{\boldsymbol{\beta}}$ of the full model can be obtained via OLS on the time-demeaned model, i.e.,

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}'\boldsymbol{\beta} + \ddot{\epsilon}_{it}, \tag{2.2}$$

with $\ddot{y}_{it} = y_{it} - \bar{y}_i$ where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, and $\ddot{\mathbf{x}}_{it}$ and $\ddot{\epsilon}_{it}$ are transformed accordingly; see Wooldridge (2012). We take this time-demeaned form (2.2) as the starting point for Steps 2–4 of our adapted MIRL procedure, which addresses the complex challenges in the data such as high dimensionality, missingness and multicollinearity.

2.3.1 The Adapted MIRL Procedure

The adapted MIRL (aMIRL) technique comprises four steps, which are presented in detail in the following. In essence, the aMIRL simultaneously selects variables and estimates parameters across panel bootstrap samples of a sequence of imputed data sets, where the final parameter estimates are robust aggregates across samples and relevant components are ranked and selected according to stability selection (Meinshausen and Bühlmann, 2010). Thus, the first step of aMIRL generates multiple imputations and carries out the standardisation and the within-transformation of the data. Steps 2 and 3 represent an adaptation of the random lasso algorithm (Wang et al., 2011; Liu et al., 2016) in the panel setting, and Step 4 comprises the selection and ranking of variables performed by stability selection.

Step 1: Imputation and Standardisation

All missing values in the data are imputed using the Multivariate Imputation by Chained Equations (MICE) technique (Raghunathan et al., 2001; van Buuren, 2007; Azur et al., 2011). It is based on the Missing At Random assumption, whereby the probability that a value is missing depends only on the observed data but not on unobserved components (Schafer and Graham, 2002). The MICE procedure yields M completed data sets that differ in their final imputed values. We choose $M = 10$ in line with the literature, which suggests a range from $M = 5$ to $M = 40$ but offers no further criterion for determining

M (Azur et al., 2011). For each $m \in \{1, \ldots, M\}$, we obtain the final imputed values according to the following procedure.

We start by replacing each missing value in the data with the mean of the available observations of the corresponding variable (placeholder imputation). We then fix a certain variable x_j and set its imputed values back to missing. The observed values of x_j are then regressed on the imputed and observed values of all other variables in the data set. We use nonlinear regression and classification trees to counteract the complexity and correlation challenges in the data. For continuous x_j, we employ RE-EM trees by Sela and Simonoff (2012), which are regression trees with random effects that accommodate nonlinearities in the data while taking the panel structure into account. The resulting imputation procedure for continuous outcomes is described in Algorithm 1.

The use of fixed effects instead of random effects is impracticable here, as the iterative nature of the estimation procedure would lead to an overestimation of the individual effects and an underestimation of the component f. This would particularly affect the imputation of missing values of MFIs i' that are not included in the estimation data set (Case 2 in Part B of Algorithm 1). In this case, $u_{i'}$ cannot be estimated and is therefore set to zero, which would lead to imputed values close to zero. In the binary setting, the inclusion of random effects is not straightforward, and we therefore rely on conventional classification trees for imputing the few dummy variables contained in the data. Subsequently, the missing values of x_j are replaced with predictions from the fitted regression model, i.e., using Algorithm 1 for continuous outcomes or estimated classification trees for binary outcomes.[8] These imputed values then replace the corresponding placeholders or previously imputed values. Iterating through all p variables and repeating the aforementioned steps forms a *cycle*. At the end of the first cycle, all missing values have been imputed once. In total, we perform C cycles, updating the imputations in each cycle and permuting the order of the components j in the updating steps. We set $C = 20$, which is the maximum in the suggested range of 10 to 20 (see van Buuren and Groothuis-Oudshoorn, 2011) to cope with the strong correlations in our data (White et al., 2011). For each completed data set m, the ensuing lasso procedure requires coefficients of comparable size. The resulting data are therefore standardised to have zero mean and unit variance. In this fixed effects panel adaptation of the algorithm,

[8]For very few (0.65%) of the missing observations, the RE-EM trees predicted values < 0 or > 1, although only values from 0 to 1 had been observed in the respective variable. We manually replaced these values with 0 and 1.

Algorithm 1 Imputation of continuous covariate x_j using the panel structure.

Variable x_j is modelled as $x_{j,it} = u_i + f(\mathbf{x}_{-j,it}) + e_{it}$, where u_i is an MFI-specific random effect, $\mathbf{x}_{-j,it}$ contains the values of all p regressors except x_j for MFI i at time t, f is an unknown function approximated via regression trees, and e_{it} denotes the error term. Let $\mathcal{I}_j^{\text{obs}}$ denote the set of indices $(i,t) \in \{(1,1),\ldots,(213,6)\}$ for which x_j is observed, and $\mathcal{I}_j^{\text{miss}}$ the set of indices (i',t') for which the values of x_j are missing.

Part A (Estimation): Estimate the RE-EM trees using all observations $(i,t) \in \mathcal{I}_j^{\text{obs}}$:

1. **Initialisation**: Set the initial estimated random effects $\hat{u}_i$ to zero.
2. **Iterative Estimation**: Repeat until the estimated random effects $\hat{u}_i$ converge (i.e., until the change in the restricted likelihood function falls below a pre-defined tolerance level):
   a) **Tree Estimation**: Estimate a regression tree approximating f, using the adjusted target variable $x_{j,it} - \hat{u}_i$ and the attributes $\mathbf{x}_{-j,it}$. From this regression tree, generate a set of indicator variables $I(\mathbf{x}_{-j,it} \in g_p)$, where g_p represents the terminal nodes of the tree.
   b) **Mixed Effects Model Fit**: Fit a linear mixed effects model of the form

   $$x_{j,it} = u_i + I(\mathbf{x}_{-j,it} \in g_p)\mu_p + e_{it},$$

   where μ_p is the prediction for leaf p. Extract the updated estimates $\hat{u}_i$.
3. **Response Adjustment**: After convergence, replace the predicted response at each terminal node with the estimated $\hat{\mu}_p$ derived from the mixed effects model in 2(b).

Part B (Imputation): Replace the missing values of x_j for $(i',t') \in \mathcal{I}_j^{\text{miss}}$ as follows:

- **Case 1**: MFI i' appears in the estimation sample. Impute

$$x_{j,i't'} := \hat{u}_{i'} + I(\mathbf{x}_{-j,i't'} \in g_p)\,\hat{\mu}_p.$$

- **Case 2**: MFI i' does *not* appear in the estimation sample. Set $u_{i'} = 0$ and impute

$$x_{j,i't'} := I(\mathbf{x}_{-j,i't'} \in g_p)\,\hat{\mu}_p.$$

the data are further time-demeaned, since we estimate $\boldsymbol{\beta}$ using the within-transformed model (2.2).

Step 2: Bootstrap Samples and Importance Measures

We obtain the parameter estimates by applying a slightly modified version of the random lasso introduced by Wang et al. (2011), which re-estimates a lasso model on multiple bootstrapped data sets. It has been shown to outperform the elastic-net method in terms of efficiency in selecting or removing highly correlated variables and flexibility in coefficient estimation.

From each of the M imputed data sets, B bootstrap samples are drawn, resulting in $M \times B$ bootstrapped data sets. To maintain the panel structure in each sample, we draw with replacement from vectors of $T = 6$ observations per $N = 213$ different MFIs. Based on the recommendations of Wang et al. (2011), we choose $B = 100$ and obtain $M \times B = 1,000$ bootstrapped data sets that are of the same size as the original $N \times T$ panel data set.

The subsequently computed importance measures are required for the subset selection of the random lasso procedure in Step 3. Intuitively, important (unimportant) variables j are likely to have consistently large (small) lasso estimates $\hat{\beta}_j$ in different bootstrap samples, where the fixed effects lasso estimator $\hat{\boldsymbol{\beta}}$ for model (2.1) solves the penalised regression

$$\min_{\theta \in \mathbb{R}^p} \|\ddot{\mathbf{y}} - \ddot{\mathbf{X}}\theta\|_{NT}^2 + \lambda^{\mathrm{OC}} \|\theta\|_1, \tag{2.3}$$

where $\|z\|_{NT}^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}^2$ is the empirical norm, $\|\theta\|_1 = \sum_{j=1}^{p} |\theta_j|$ denotes the ℓ^1-norm, and $\ddot{\mathbf{X}} = (\ddot{\mathbf{x}}_{11}', \ldots, \ddot{\mathbf{x}}_{1T}', \ddot{\mathbf{x}}_{21}', \ldots, \ddot{\mathbf{x}}_{NT}')'$ contains all $N \times T$ observations of the p regressors. The tuning parameter λ^{OC} is chosen in a data-driven way as outlined in Section 2.3.2, minimising one of the three optimality criteria (OC) BIC, AIC and Mallows' C_p. We obtain lasso pre-estimates $\hat{\boldsymbol{\beta}}_m^{(b),\mathrm{OC}}$ for every pair of imputation and bootstrap samples (m, b) and consider a variable j as relevant if $\hat{\beta}_{mj}^{(b),\mathrm{OC}} \neq 0$.

In the following, we work with the bias-reduced two-step estimate $\tilde{\boldsymbol{\beta}}_m^{(b),\mathrm{OC}}$ (e.g., Belloni and Chernozhukov 2013), which sets $\tilde{\beta}_{m,j}^{(b),\mathrm{OC}} = 0$ if component j was not lasso-selected, and replaces all other components by OLS estimates using only the lasso pre-selected components as regressors.

Hence, a straightforward, non-negative measure of variable importance for $x_j \in \{x_1, \dots, x_p\}$ is the absolute average of the estimates across all data sets (m, b), computed as

$$I_j^{\mathrm{OC}} = \frac{1}{MB}\left|\sum_{m=1}^{M}\sum_{b=1}^{B}\tilde{\beta}_{mj}^{(b),\mathrm{OC}}\right|. \tag{2.4}$$

Step 3: Initial aMIRL Estimates

To counteract high correlations among the considered variables, we use only one third of the variables per bootstrap sample in the second step of the modified random lasso procedure. Thus, for each standardised and time-demeaned imputation–bootstrap sample (m, b), we randomly select $\lfloor p/3 \rfloor$ candidate variables, where the selection probability for component j is proportional to its importance measure I_j^{OC} from Equation 2.4.[9] We perform lasso–OLS on the time-demeaned bootstrapped and imputed data including only the chosen $\lfloor p/3 \rfloor$ candidate variables in the potential set of regressors. For each variable x_j we thereby obtain $\tilde{\beta}_{mj}^{(b),\mathrm{OC},\lfloor p/3 \rfloor}$, which equals zero if variable j was either not included in the set of potential regressors or not selected by the lasso–OLS procedure.

The vector of initial aMIRL estimates $\hat{\mathbf{b}}^{\mathrm{init,OC}} = (\hat{b}_1^{\mathrm{init,OC}}, \dots, \hat{b}_p^{\mathrm{init,OC}})'$ is then computed by averaging the resulting estimates:

$$\hat{b}_j^{\mathrm{init,OC}} = \frac{1}{MB}\sum_{m=1}^{M}\sum_{b=1}^{B}\tilde{\beta}_{mj}^{(b),\mathrm{OC},\lfloor p/3 \rfloor}. \tag{2.5}$$

Step 4: Stability Selection and aMIRL Estimates

Stability selection, which produces a variable ranking and ultimately the set of variables to be included in the final model, is similar to the random lasso in the previous step in that we consider lasso estimates based on a subset of the data. Here, too, $\lfloor p/3 \rfloor$ randomly selected candidate variables per imputed–bootstrapped data set are used as potential regressors. Instead of employing the OC-optimal tuning parameter λ^{OC}, however, we perform the lasso step over a grid Λ.

We then obtain the empirical selection probabilities $\hat{\pi}_j^{\mathrm{OC}}$ as

$$\hat{\pi}_j^{\mathrm{OC}} = \max_{\lambda \in \Lambda} \frac{1}{MB}\sum_{m=1}^{M}\sum_{b=1}^{B}\mathbb{1}\left\{\hat{\beta}_{mj}^{(b),\lfloor p/3 \rfloor,\lambda} \neq 0\right\},$$

[9] Liu et al. (2016) select $p/2$ instead of $p/3$, but state that both variants yield similar results.

where the OC-dependence stems from the fact that the sets of $\lfloor p/3 \rfloor$ potential regressors differ for each optimality criterion and pair (m, b) due to the differing information criteria I_j^{OC}. These selection probabilities introduce a natural importance ranking of all considered variables, and the stable set is determined as

$$\hat{S}^{\text{stable,OC}} = \{j : \hat{\pi}_j^{\text{OC}} \geq \pi^{*,\text{OC}}\},$$

where $\pi^{*,\text{OC}}$ denotes a data-driven threshold (Section 2.3.2).

The aMIRL estimates $b_{\text{aMIRL},j}^{\text{OC}}$ are therefore non-zero only for the components in the stable set and are computed as

$$b_{\text{aMIRL},j}^{\text{OC}} = \hat{b}_j^{\text{init,OC}} \times \mathbb{1}\{j \in \hat{S}^{\text{stable,OC}}\}. \tag{2.6}$$

2.3.2 Adaptive Tuning Parameter Choice and Evaluation Criteria

In this work, we aim to maximise the explanatory power of our models and therefore optimise their in-sample performance with respect to three optimality criteria: the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC) and Mallows' C_p. Since we can retrieve estimates $\hat{\alpha}_i$ for α_i as $\hat{\alpha}_i = \bar{y}_i - \hat{\beta}_1 \bar{x}_{i1} - \cdots - \hat{\beta}_K \bar{x}_{iK}$ for $i = 1, \ldots, N$ (see Wooldridge, 2012), we have

$$y_{it} - \hat{\alpha}_i - \mathbf{x}_{it}' \hat{\boldsymbol{\beta}} = \ddot{y}_{it} - \ddot{\mathbf{x}}_{it}' \hat{\boldsymbol{\beta}}.$$

This leads to minimising

$$\text{BIC} = NT \log \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\ddot{y}_{it} - \ddot{\mathbf{x}}_{it}' \hat{\boldsymbol{\beta}})^2 \right] + \log(NT)\,(N + K),$$

$$\text{AIC} = NT \log \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\ddot{y}_{it} - \ddot{\mathbf{x}}_{it}' \hat{\boldsymbol{\beta}})^2 \right] + 2\,(N + K),$$

$$C_p = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^{N} \sum_{t=1}^{T} (\ddot{y}_{it} - \ddot{\mathbf{x}}_{it}' \hat{\boldsymbol{\beta}})^2 - NT + 2\,(N + K),$$

with K denoting the number of included regressors (non-zero components of $\hat{\boldsymbol{\beta}}$), and

$$\hat{\sigma}^2 = \frac{1}{NT - N - p} \sum_{i=1}^{N} \sum_{t=1}^{T} (\ddot{y}_{it} - \ddot{\mathbf{x}}'_{it} \hat{\boldsymbol{\beta}}_{\text{full}})^2,$$

where $\hat{\boldsymbol{\beta}}_{\text{full}}$ denotes the within-estimate for $\boldsymbol{\beta}$ of the full model (Mallows, 2000).[10]

Following Zheng and Loh (1995), we determine the optimal threshold $\pi^{*,\text{OC}}$ via minimising the BIC—specifically, the average BIC across all imputed data sets $m = 1, \ldots, M$—over all levels of π, i.e.,

$$\pi^{*,\text{OC}} = \arg \min_{\pi \in \{\hat{\pi}_1, \ldots, \hat{\pi}_p\}} \frac{1}{M} \sum_{m=1}^{M} \text{BIC}(m, \pi),$$

where

$$\text{BIC}(m, \pi) = NT \log \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\ddot{y}_{m,it} - \ddot{\mathbf{x}}'_{m,it} \hat{\mathbf{b}}^{\text{init,OC},\pi})^2 \right] + \log(NT)\,(N + K),$$

with $\hat{\mathbf{b}}^{\text{init,OC},\pi} = \hat{\mathbf{b}}^{\text{init,OC}} \times \mathbb{1}\{j : \hat{\pi}_j \geq \pi\}$.

Regarding the initial tuning parameters $\Lambda = \{\lambda_k\}_{k=1}^{K}$, Friedman et al. (2010) suggest using a logarithmically spaced sequence from $\lambda_{\max}$ down to $\lambda_{\min} = \delta\lambda_{\max}$, where $\lambda_{\max}$ is the smallest tuning parameter for which the lasso estimate is identically zero. Thus,

$$\lambda_k = \exp(s_k),$$

where $\{s_k\}_{k=1}^{K}$ is an equally spaced sequence from $\log \lambda_{\max}$ to $\log \lambda_{\min}$. They recommend $K = 100$ and $\delta = 0.001$, which we also adopt. We determine

$$\lambda_{\max} = \max_{m,b} \lambda_0^{m,b},$$

where $\lambda_0^{m,b}$ is the smallest tuning parameter for which the fixed effects lasso estimator in bootstrap sample b of imputation m equals zero.

For evaluation of our results in Section 2.4, we compute significance of the obtained aMIRL estimates via nonparametric, bias-corrected and accelerated (BCa) bootstrap

[10] The definitions of BIC and AIC differ from the actual information criteria since the first term does not equal $-2\log(\hat{L})$, $\hat{L}$ being the estimated likelihood of the model, but under normality they have the same minimiser.

confidence intervals, which do not rely on potentially misspecified parametric assumptions and are robust to the quality of asymptotic approximations in finite samples (Efron and Narasimhan, 2020). Unlike conventional bootstrap intervals, BCa intervals accommodate (i) non-normality of the b_{aMIRL} estimate, (ii) bias in the estimate and (iii) potentially non-constant standard errors. Further details are provided in Efron and Narasimhan (2020). We use jackknife resampling as implemented in the function `bcajack` in the R package `bcaboot`. The effect of variable j is significant at level α if the respective bootstrap confidence interval does not contain zero.

As goodness-of-fit measures, we use the ordinary (adjusted) R^2 as well as the (adjusted) within-R^2, which is defined as the ordinary (adjusted) R^2 of the time-demeaned model. The latter can be interpreted as the amount of time variation in the target variable y_{it} explained by the time variation in the predictors $\mathbf{x}_{it}$.

2.4 Results

We apply the aMIRL procedure from Section 2.3 to the balanced panel data described in Section 2.2. For the key financial indicator *OSS* as well as for each of the two social impact target variables *log(NAB)* and *–log(ALBG)*, see Section 2.2.1 for more details, we run the procedure three times, i.e., once per optimality criterion: BIC, AIC and Mallows' C_p. While optimising for BIC yields the smallest models, minimising AIC and C_p results in larger models that are very similar to each other.[11] The direction and size of the respective effects per target variable are consistent for all three optimality criteria. Hence, the essence of the models is the same. For simplicity, in this section we report and refer to the estimates for AIC for those variables selected in at least two of the final models. A qualitative overview of the estimated selected effects per factor is given in Table 2.3, while quantitative results are described in the following two Sections 2.4.1 and 2.4.2 and shown in Tables 2.4–2.6. Detailed results are provided in Tables A.5–A.7 in the appendix.

We find that financial success is mostly driven by financial aspects of the MFI, while social success depends largely on staff characteristics and operational features. By definition, the operational self-sufficiency of MFIs can be promoted by increasing profitability and reducing costs. The same applies, to a lesser extent, to both dimensions of MFI

[11] In fact, AIC and C_p have been shown to be equivalent in the case of Gaussian linear regression, which may explain the similarity in the results here (Boisbunon et al., 2013).

Table 2.3: Overview per target variable of the direction and approximate size of the aMIRL coefficients of factors for the AIC optimality criterion

Factor groups	***OSS***	***log(NAB)***	***–log(ALBG)***
Costs	$\Large\oplus\Large\ominus\Large\ominus\ominus$	$\oplus\ominus\ominus$	$\Large\oplus\oplus\Large\ominus\ominus\ominus$
Financial Performance	$\Large\oplus\Large\oplus\oplus$	$\oplus$	$\oplus\ominus$
Financing Structure		$\ominus$	$\ominus$
Operational Aspects	$\oplus\oplus\Large\ominus$	$\oplus\oplus\oplus\oplus\oplus\Large\ominus\ominus\ominus$	$\oplus\oplus\oplus\oplus\oplus\oplus\ominus\ominus\ominus\ominus\ominus\ominus$
Organisational Strategy		$\oplus\oplus\oplus\ominus$	
Personnel Structure	$\ominus$	$\Large\oplus\Large\oplus\oplus\Large\ominus\ominus\ominus$	$\Large\oplus\Large\oplus\Large\oplus\oplus\Large\ominus\ominus$
Services		$\oplus$	

Note: We only focus on variables that are selected by the aMIRL procedure for at least two optimality criteria from the group AIC, BIC and C_p. All respective variables have effects of size ≥ 0.01 in absolute value and are significant at the 1% level based on bootstrap confidence intervals. We roughly distinguish the cases $\Large\oplus$: $b^{AIC}_{\text{aMIRL}} \geq 0.1$, $\oplus$: $0.1 > b^{AIC}_{\text{aMIRL}} \geq 0.01$, negative analogously.

outreach. The personnel structure plays an important role in all three dimensions of MFI success. While a larger management has a significant negative impact on all three measures of success, changes in the size, efficiency and salary structure of the staff are crucial determinants of greater breadth and depth of outreach. Operational aspects are important for social success, but less so for financial success. In general, financial and social success of MFIs are not in opposition to each other, which is in line with the findings of Quayes (2015). The relationship between the two measures of social success is more ambiguous. While issuing smaller loans can contribute to increased breadth of outreach, reaching more borrowers usually leads to targeting wealthier customers, a phenomenon referred to in the literature as *mission drift* (Armendáriz and Szafarz, 2011).

2.4.1 Key Financial Success Factors

By analysing *OSS*, we find that the key characteristics of operational sustainability correspond to financial measures, while operational and personnel aspects are of secondary importance; see Table 2.4 for an overview of the results and Table A.5 in the appendix for detailed results of the aMIRL models for all three optimality criteria. Our models yield an R^2 of 0.88, which is 0.54 (2.6 times) higher than that of Quayes (2015), who also studies *OSS* with a fixed-effects model on a balanced panel of similar size and obtains an R^2 of 0.34. Other empirical studies also obtain models with smaller R^2 values than ours, see Kinde (2012) (0.58), Bogan (2012) (0.37), Strøm et al. (2014) (0.25) and Bassem (2012) (0.12), among others. The variables contained in their models are also included in our potential set of regressors, except for age (which is accounted for indirectly in our setting by the fact that we only consider MFIs that have reported and therefore survived for at least six years), type of organisation and country-specific controls (both of which are accounted for by a fixed effect, as they are (likely) constant over time).

Table 2.4: Empirical selection probabilities and aMIRL coefficients for target variable OSS

Variable	$\hat{\pi}^{AIC}_{\text{aMIRL}}$	b^{AIC}_{aMIRL}	**Variable**	$\hat{\pi}^{AIC}_{\text{aMIRL}}$	b^{AIC}_{aMIRL}
Operating Expense / Assets	**0.866**	**−0.229**	Return on Assets	0.913	0.070
Financial Expense / Assets	0.982	−0.183	Av. Loan Size: Urban	0.907	−0.106•••
Operating Expense / GLP	0.826	0.111	Tax Expense / Assets	0.872	0.059
Cost per Borrower	0.780	−0.044	Write-Off Ratio	0.776	0.053•••
Profit Margin	**1.000**	**0.484**	% Staff: Managers	0.863	−0.075•••
Financial Revenue / Assets	**0.972**	**0.263**			

Note: We only focus on variables that are selected by the aMIRL procedure for at least two optimality criteria from the group AIC, BIC and C_p and report the corresponding AIC estimates. The colours represent the factors from Figure 2.1. All effects are significant at the 1% level based on bootstrap confidence intervals. The coloured bullets ••• indicate the signs (• positive, • negative) of the post-imputation and post-selection quantile regression estimates, see Section 2.4.3, for $\tau = 0.25, 0.5, 0.75$ where at least one differs from the aMIRL estimate. Bold values indicate the variables with the three largest effects.

Given the definition of *OSS*, it is unsurprising that financial performance characteristics and costs are the main determinants of an MFI's operational self-sufficiency, as reflected by the five largest estimated effects: *Profit Margin*, *Financial Revenue / Assets*, *Operating Expense / Assets*, *Financial Expense / Assets* and *Operating Expense / GLP*. Similarly, higher relative tax expenses, an indicator of increased taxable capital, are positively associated with operational sustainability. More surprising is the negative role of the percentage of managers in the MFI personnel, which is mainly driven by less successful MFIs, since the effect is positive for the median and the 75% quantile. Operationally less sustainable MFIs[12] are most likely run by a less effective management, so increasing the size of management in that case has a negative impact. The opposite holds true for financially successful MFIs. We further find that smaller loan sizes for urban borrowers (and hence greater depth of outreach) can improve financial performance. However, this effect is not confirmed by the respective 50% and 75% quantile estimates.

[12] Reminder: The 25% quantile of *OSS* in our data set is 1.032, see Table A.3, where 1 is the critical point for operational sustainability. Hence, the effects of the 25% quantile describe MFIs on the verge of being operationally sustainable.

In summary, the operational sustainability of an MFI can be increased by focusing on higher profitability and lower costs and ensuring a staff structure with an effective management of moderate size.

These findings complement and significantly extend previous analyses. We find that most of the variables selected in our final models are not included in models used in the literature (Bassem, 2012; Bogan, 2012; Kinde, 2012; Strøm et al., 2014; Quayes, 2015), highlighting the importance of data-driven variable selection. While Quayes (2015) also recognises a negative influence of cost and loan size on *OSS*, many previously reported significant relationships are not confirmed by our results, possibly due to omitted variables in those studies. For example, Kinde (2012) finds a significant positive correlation with outreach and a negative one with the proportion of donated capital; Bogan (2012) analyses the role of capital structure and Strøm et al. (2014) that of female leadership—none of which are supported by our models.

2.4.2 Key Social Success Factors

Changes in the personnel structure of an MFI are *the* most relevant driver of both dimensions of social success: the breadth of outreach (reaching as many borrowers as possible) and the depth of outreach (reaching particularly poor borrowers). Corresponding variables are selected with probability greater than 0.98 and have the largest marginal post-lasso effects, see Tables 2.5 and 2.6 for an overview and Tables A.6 and A.7 in the appendix for detailed results of both aMIRL analyses. More specifically, a larger borrower–staff ratio and a greater number of employees, combined with flat hierarchies, are crucial for increased social performance. Costs are overall negatively associated with social success, while profitability plays a minor but generally positive role. An increase in liabilities in the capital structure can lead to higher social success, reflected by the positive effect of *Fin. Exp. Fund. Liab. / Assets* in both cases and the negative effect of *Capital / Assets* for *log(NAB).*

Regarding the relationship between the two measures of social success, we find that increased depth of outreach can lead to serving a greater number of borrowers. However, reaching more borrowers typically results in serving wealthier borrowers. This is shown by the combined negative effect of *Av. Loan Bal. / GNI p. c.* and *Av. Outst. Bal. / GNI p. c.* on *log(NAB)* and the negative impact of *# Active Borrowers* on *–log(ALBG).* MFIs should therefore carefully choose their primary social objective and tailor their activities accordingly. The two types of social success have additional specific and unique characteristics, discussed in the following Sections 2.4.2 and 2.4.2.

To our knowledge, the prominent role of staff structure in the social success of MFIs is new to the literature. We find that our aMIRL results yield a significantly better fit than related models from authors who also use the MIX Market data set. Bogan (2012) analyses *log(NAB)* using a linear model with an R^2 of 0.72, which is 0.26 smaller than that of our aMIRL models, and Quayes (2012) models *–log(ALBG)* with an R^2 of 0.65, which is 0.32 lower than that of our models. Quayes and Joseph (2022) use NAB and $ALBG$ as target variables and reach R^2 values of 0.73 and 0.27, again far below our values for the logarithm. Other popular MFI outreach studies often do not report R^2 values, making comparisons difficult, see Hartarska and Nadolnyak (2007), Hermes et al. (2011) and Awaworyi Churchill (2020). Moreover, many such models contain statistically insignificant regressors, again underlining the need for data-driven variable selection.

Breadth of Outreach

Table 2.5: Empirical selection probabilities and aMIRL coefficients for target variable log(NAB)

Variable	$\hat{\pi}^{AIC}_{\text{aMIRL}}$	b^{AIC}_{aMIRL}	Variable	$\hat{\pi}^{AIC}_{\text{aMIRL}}$	b^{AIC}_{aMIRL}
Cost per Borrower	0.630	−0.024	% GLP: Solidarity Group	0.528	0.013
Financial Expense / Assets	0.682	−0.015	Incentives: Portf. Qual.	0.483	0.016
Fin. Exp. Fund. Liab. / Assets	0.627	0.011	Target Market: Youth	0.485	0.014
Return on Assets	0.742	0.024	Goals: Health Infrastructure	0.524	0.012
Capital / Assets	0.755	−0.053	Goals: Econ. Improvement	0.508	−0.011
Av. Loan Bal. / GNI p. c.	**0.992**	**−0.201**	**# Personnel**	**0.998**	**0.270**
Av. Outst. Bal. / GNI p. c.	0.917	0.097	**Borrowers / Staff Member**	**0.997**	**0.229**
% GLP: Microenterprise	0.907	−0.043	% Staff: Board Members	0.992	−0.179
Deposit Accounts / Staff	0.661	0.032	Av. Salary / GNI p. c.	0.757	0.032
% Borrowers: Urban	0.704	−0.026	Personnel Expense / GLP	0.710	−0.028
% GLP: Enterprise Finance	0.823	0.023	% Staff: Managers	0.505	−0.011
Loan Impairm. Prov. / Assets	0.609	0.015	Offers Other Fin. Services	0.523	0.016

Note: We only focus on variables that are selected by the aMIRL procedure for at least two optimality criteria from the group AIC, BIC and C_p and report the corresponding AIC estimates. The colours represent the factors from Figure 2.1. All effects are significant at the 1% level based on bootstrap confidence intervals. The signs of the aMIRL estimates coincide with those of the postimputation and postselection quantile regression estimates, see Section 2.4.3, for $\tau = 0.25, 0.5, 0.75$. Bold values indicate the variables with the three largest effects.

In contrast to services provided by traditional banks, the success of microfinance loans depends heavily on the support that borrowers receive from the MFI, particularly from its staff (Yunus, 2009). We find that a larger number of borrowers is reached primarily through the expansion of microfinance activities. The key components of this are a greater number of employees and their efficient deployment. It is not beneficial to focus on a larger board. The emphasis should be on additional loan officers who run the microfinance business by directly serving borrowers. Each staff member should be trained to serve more borrowers (*Borrowers / Staff*) and be given the right incentives and goals (*Incentives: Portfolio Quality*, *Goals: Health Infrastructure*). More borrowers are also attained by targeting new borrower groups, especially young people, and by offering additional services (*Target Market: Youth*, *Deposit Accounts / Staff*, *Offers Other Fin. Services*).

The adjustment of the staff structure should be accompanied by changes in the gross loan portfolio (GLP). It should either be redistributed or increased. Redistribution can be achieved by reducing the average loan size and lending to groups instead of individuals. Consequently, the depth of outreach can contribute not only to financial success but also to the breadth of outreach. Additional GLP may also be financed through liabilities, which increases relative financial expenses on funding liabilities. These offer a convenient way to acquire new resources quickly. As before, costs are negatively related to the breadth of outreach. Profitability, on the other hand, promotes it.

Depth of Outreach

The impact of staff structure on the depth of outreach is determined largely by the same variables as in the analysis of the breadth of outreach. The ratio of borrowers to employees remains decisive, while the size of the workforce is reflected to a lesser extent in the number of staff and more clearly in the combination of the positive effect of personnel expenses and the negative effect of an increase in average wages. Additional personnel expenses should therefore be directed towards new staff, especially new loan officers, rather than management, who are assumed to receive higher salaries.

We do not find that greater depth of outreach is necessarily associated with higher costs. The results are mixed here: while *Financial Expense / Assets*, *Cost per Borrower* and *Cost per Loan* have a large negative impact, greater funding costs for liabilities and greater *Operating Expense / GLP* are positively associated with depth of outreach.

Table 2.6: Empirical selection probabilities and aMIRL coefficients for target variable –log(ALBG)

Variable	$\hat{\pi}^{AIC}_{\text{aMIRL}}$	b^{AIC}_{aMIRL}	Variable	$\hat{\pi}^{AIC}_{\text{aMIRL}}$	b^{AIC}_{aMIRL}
Financial Expense / Assets	**0.987**	**−0.189**	Portfolio at Risk: > 90 Days	0.849	0.029
Fin. Exp. Fund. Liab. / Assets	0.982	0.180	Av. Loan Size: Microenterpr.	0.575	−0.029
Cost per Borrower	0.830	−0.069	% GLP: Individual	0.717	0.028
Cost per Loan	0.680	−0.045	Av. Loan Size: Urban	0.544	−0.027
Operating Expense / GLP	0.640	0.037	% Borrowers: Male	0.510	−0.022
Return on Assets	0.736	0.034	% GLP: Renegotiated Loans	0.590	−0.015
Profit Margin	0.682	−0.031	Borrower Retention Rate	0.533	0.013
Assets	0.865	−0.071	**Borrowers / Staff Member**	**0.995**	**0.247**
# Active Borrowers	0.823	−0.091	**Av. Salary / GNI p. c.**	**0.982**	**−0.194**
Interest Income on GLP / GLP	0.776	0.048	Personnel Expense / GLP	0.868	0.130
% GLP: Solidarity Group	0.707	0.041	Personnel Expense / Assets	0.887	0.102
% GLP: Village Banking (SHG)	0.685	0.039	# Personnel	0.836	0.058
GLP / Total Assets	0.636	−0.033	% Staff: Managers	0.582	−0.021

Note: We only focus on variables that are selected by the aMIRL procedure for at least two optimality criteria from the group AIC, BIC and C_p and report the corresponding AIC estimates. The colours represent the factors from Figure 2.1. All effects are significant at the 1% level based on bootstrap confidence intervals. The signs of the aMIRL estimates coincide with those of the postimputation and postselection quantile regression estimates, see Section 2.4.3, for $\tau = 0.25, 0.5, 0.75$. Grey font indicates variables not included in the final AIC model, but in both the BIC and C_p models. Bold values indicate the variables with the three largest effects.

As noted by Yunus (2009), small loans are usually charged higher interest rates due to higher default risk. This is confirmed here by the effects of *Interest Income on GLP / GLP*, *Portfolio at Risk: >90 Days* and *% GLP: Renegotiated Loans*. We further find that MFIs expanding their microfinance activity—here reflected by serving more borrowers, increasing the *GLP / Assets* ratio and holding larger *Assets*—may deviate from their primary goal of serving poorer borrowers by extending larger loan sizes. This phenomenon, *mission drift*, aligns with findings by Armendáriz and Szafarz (2011). Focusing on lending to solidarity groups, villages (self-help groups) and individuals (women more than men) rather than small and medium enterprises or large corporations can also increase the depth of outreach. We identify a mixed relationship with profitability.

Both dimensions of an MFI's social success are achieved mainly through adjustments in the personnel structure: increasing the number of staff, ensuring that employees can serve more borrowers and maintaining an efficient, moderately sized management. To a lesser extent, they are driven by aspects of the size and distribution of the loan portfolio. While increasing the GLP and targeting new borrower groups furthers the breadth of outreach, depth is enhanced by focusing on specific groups (solidarity groups, self-help groups and women). Although greater depth of outreach may help an MFI reach more borrowers, MFIs with a greater breadth of outreach tend to neglect their mission to serve the poorest. Pursuing both breadth and depth simultaneously can therefore be empirically conflicting goals.

2.4.3 Robustness

We compare our aMIRL results with those obtained from applying conventional lasso-OLS, with and without fixed effects, as well as the original MIRL algorithm, which does not take into account the panel structure in the data. We also examine the quality of the imputed data, contrast aMIRL estimates for similar outcome variables and compute additional post-selection quantile estimates. We find that all method comparisons confirm our findings. In contrast, the lasso method with fixed effects is unable to perform robust variable selection in our setting, and the remaining methods do not adequately account for the heterogeneity present in the data. The means of two comparable outcome variables are driven by similar factors, and quantile estimates confirm that the driving effects are similar across the distribution of the target variables.

Comparison to Results from Different Model Selection and Imputation Techniques

We compare results from our aMIRL approach with those derived from conventional lasso-OLS estimation. The latter is not robust to multicollinearity or high-dimensionality and does not specifically address the missingness problem. We apply lasso-OLS to the mean-imputed and within-transformed data. This approach accounts for unobserved fixed effects in variable selection and estimation, but not in the imputation step. We find that lasso-OLS yields much larger models than aMIRL while producing comparable, or up to 13% (16%), smaller within (adjusted) R^2, see the left columns in Tables A.11–A.13 in the appendix. The majority of variables selected by aMIRL are also included in the corresponding lasso model with comparable effect sizes, supporting our findings. However, the majority of the additional variables in the lasso models are insignificant. Furthermore,

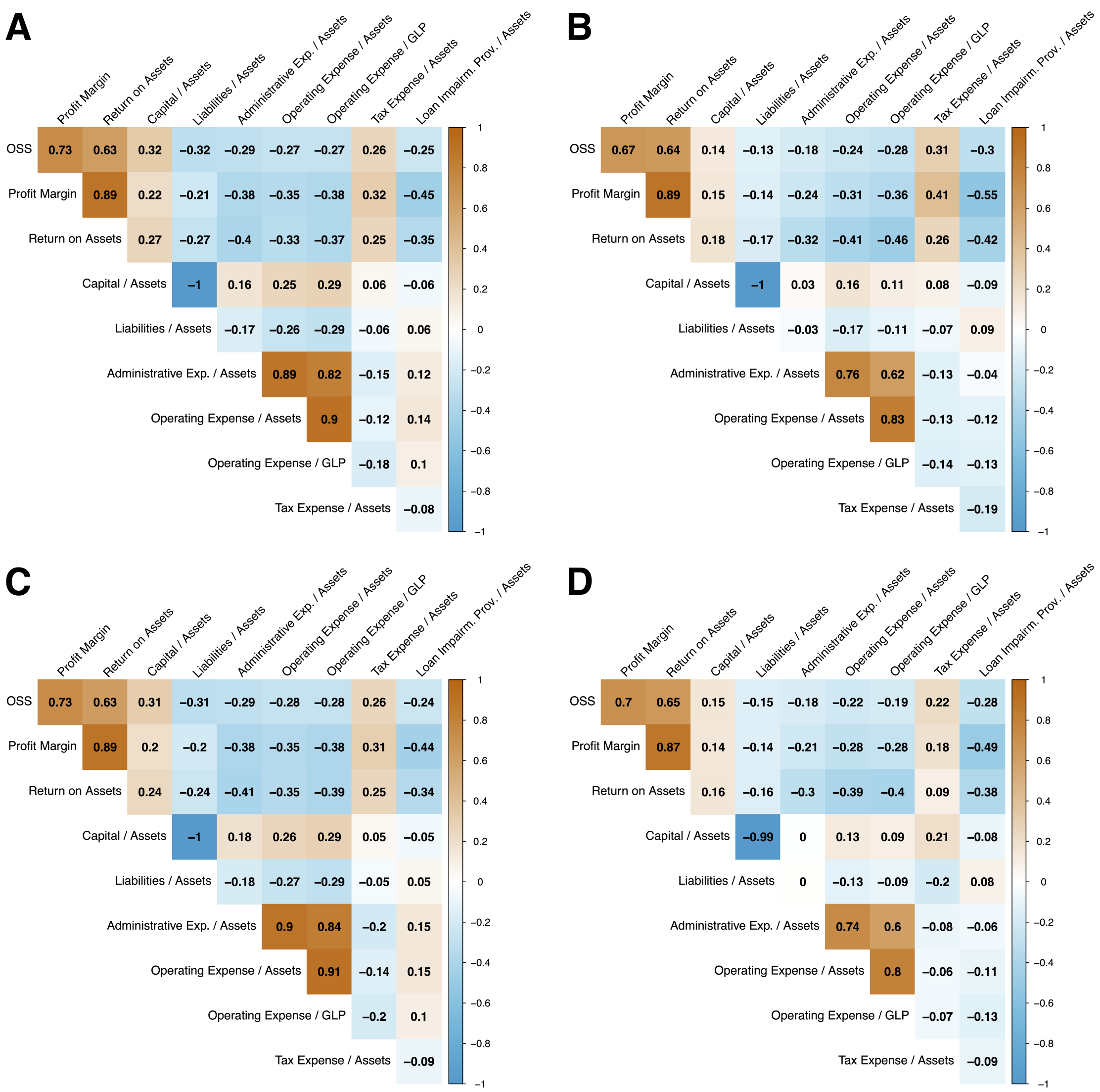

Figure 2.3: (Average) correlation in the final, balanced panel data set for variables most highly correlated with Operational Self-Sufficiency (OSS)

Note: A: pairwise-complete correlation of the cleaned, unimputed data set; B: pairwise-complete correlation of the cleaned, unimputed, standardised and time-demeaned data set; C: average correlation of the 10 imputed data sets; D: average correlation of the 10 imputed, standardised, time-demeaned data sets. All standard deviations of the correlations of the imputed data sets lie between 0 and 0.02.

model sizes differ substantially across optimality criteria, making results highly sensitive to the choice of OC.

We also estimate the original MIRL procedure, which uses a common intercept for all observations instead of MFI-specific fixed effects. It is robust to high-dimensionality and multicollinearity, but neither the imputation step—using classification and regression trees—nor the estimation and variable selection steps specifically account for the heterogeneity introduced by the panel structure. The results (middle columns of Tables A.8–A.10) confirm our main findings. However, while MIRL models contain slightly more time-varying regressors, their R^2 is much lower (up to 25%).

Comparison to Results from the Unbalanced Panel

Since the construction of the balanced panel reduces the sample size relative to the unbalanced raw data, we benchmark our empirical results against results from an unbalanced panel as detailed in Section 2.2. In this case, aMIRL cannot be used, as the within-transformation would eliminate MFIs with only one observed year. For MFIs with few observations, time-demeaning is also problematic.

The standard MIRL method does not require fixed effects imputation and can therefore be applied to the unbalanced panel to handle substantial missingness. Corresponding MIRL estimates are shown in the right column of Tables A.8–A.10. The overall R^2 for *OSS* and *–log(ALBG)* are approximately 0.07 and 0.5, respectively—much smaller than in the balanced panel—indicating that there is even greater heterogeneity in the unbalanced data, which MIRL does not capture well. The same linear model with a single intercept is also estimated using lasso-OLS, which ignores high-dimensionality, multicollinearity, panel structure and missingness. The number of variables included varies widely across target variables, OCs and data sets, ranging from six to 123, as illustrated in the middle and right columns of Tables A.11–A.13.

Since only 58 (177) of all 1,278 (3,846) observations are fully observed in the balanced (unbalanced) panel, comparison with complete-case results is infeasible.

Overall, we conclude that there is considerable heterogeneity in the balanced panel—and even more in the unbalanced one. Combined with high-dimensionality, multicollinearity and pronounced missingness that cannot be ignored, this renders both MIRL and lasso-OLS incapable of providing robust results. Conversely, the aMIRL results are corroborated by all alternative methods.

Imputation Quality

As mentioned in Section 2.3.1, the first step of the aMIRL algorithm—the MICE procedure—is carried out $M = 10$ times on the balanced incomplete panel data, resulting in ten imputed data sets. To assess MICE's ability to maintain linear effects, the pairwise-complete correlations between selected variables in the original data are plotted against the respective average correlations in the imputed data (Figure 2.3). We depict variables most highly correlated with *OSS* in the imputed, time-demeaned data. The plots show that the imputed data maintain the same linear structure as the original data, as the pairwise-complete and average correlations are very similar. Standard deviations across imputed data sets are close to zero, suggesting that each imputed data set captures the linear dependencies similarly well. The same holds for time-demeaned data.

Since RE-EM and classification trees—nonlinear by design (Sela and Simonoff, 2012; De'Ath and Fabricius, 2000)—are used for imputation, we assume that relevant nonlinearities are also adequately represented in the imputed data.

Table 2.7: Empirical selection probabilities and aMIRL coefficients for target variable –ALBG

Variable	$\hat{\pi}^{AIC}_{\text{aMIRL}}$	b^{AIC}_{aMIRL}	**Variable**	$\hat{\pi}^{AIC}_{\text{aMIRL}}$	b^{AIC}_{aMIRL}
Cost per Loan	0.965	−0.210	**Gross Loan Portf. (GLP)**	**0.923**	**0.664**
Financial Expense / Assets	0.846	−0.085	Av. Loan Size: Urban	0.789	−0.071
Operating Expense / Assets	0.746	0.082	Av. Loan Size: Microenterpr.	0.710	−0.050
Fin. Exp. Fund. Liab. / Assets	0.823	0.073	Borrowers / Staff Member	0.942	0.193
Assets	**0.761**	**−0.361**	Personnel Expense / GLP	0.901	0.155
Liabilities	**0.763**	**−0.342**•••	# Personnel	0.873	0.107
Equity	0.897	−0.137•••	Av. Salary / GNI p. c.	0.853	−0.094

Note: We only focus on variables that are selected by the aMIRL procedure for at least two optimality criteria from the group AIC, BIC and C_p and report the corresponding AIC estimates. The colours represent the factors from Figure 2.1. All effects are significant at the 1% level based on bootstrap confidence intervals. The coloured bullets ••• indicate the signs (• positive, • negative) of the post-imputation and post-selection quantile regression estimates, see Section 2.4.3, for $\tau = 0.25, 0.5, 0.75$ where at least one differs from the aMIRL estimate. Grey font indicates variables not included in the final AIC model, but in both the BIC and C_p models. Bold values indicate the variables with the three largest effects.

Comparison of Similar Target Variables

We compare results for the negative *Average Loan Balance per GNI per Capita (–ALBG)*, see Table 2.7, with those for the depth of outreach, assuming percentage changes in *ALBG* and corresponding level changes are governed by similar factors. Our results support this assumption: key findings and takeaways are highly consistent. First, the role of personnel structure is confirmed by almost identical variables. Relative differences in effect sizes vary somewhat but not in a way that changes qualitative conclusions. Second, costs play a similar role. Third, the same average loan sizes impacting *–log(ALBG)* also negatively affect *–ALBG*. Fourth, the results for *–ALBG* strongly confirm mission drift: negative effects of increasing MFI size (*Assets*, and here also *Liabilities* and *Equity*) on depth of outreach. Regarding operational aspects, changes in GLP distribution are less important here than changes in GLP size. The ambiguous role of profitability found in the depth-of-outreach analysis is not substantiated by the *–ALBG* analysis.

Post-Selection Quantile Estimates

For each of the three final models per target variable, we conduct post-selection fixed effects quantile regressions following Koenker (2004), estimating the 25%, 50% and 75% quantiles. Differences in the signs of estimates compared with the aMIRL estimates are indicated in Tables 2.4–2.6 and A.5–A.7. In almost all cases, the direction of marginal quantile effects coincides with that of the mean coefficients. In rare cases where signs differ, the median still matches the aMIRL direction, with only one exception for one factor in the case of Average Loan Balance per GNI per Capita. This indicates robustness across the distribution of impacts.

2.5 Conclusion

We propose the aMIRL method, an adaptation of the MIRL procedure introduced by Liu et al. (2016), that combines random lasso with multiple imputation and stability selection. It is tailored to perform robust variable selection for incomplete, high-dimensional longitudinal data with highly correlated variables and unobserved fixed effects. The longitudinal structure in the data is accounted for by using RE-EM trees (Sela and Simonoff, 2012) in the imputation step and including individual-level fixed effects in the model estimation and variable selection steps.

We utilise this method to determine the success factors of microfinance institutions. To the best of our knowledge, we are the first to do so in a data-driven manner. We do not pre-specify the included regressors, nor do we make the choice of variables depend on their level of missingness. Instead, we apply the aMIRL approach to a transparently constructed balanced panel comprising 1,278 observations from 213 MFIs across 136 variables over a six-year period. In accordance with the existing literature, we measure financial success by *Operational Self-Sustainability*. The breadth of outreach (logarithm of the *Number of Active Borrowers*) and the depth of outreach (negative of the logarithm of the *Average Loan Balance per GNI per Capita*) are employed as measures for the different aspects of social success. Our results demonstrate superior explanatory power compared to existing models with the same target variables. For instance, our models for OSS exhibit an overall R^2 that is 2.6 times larger (0.88 vs. 0.34) than those of Quayes (2015), who also estimates a fixed-effects model on balanced panel data.

Important practical implications for key decision-makers of MFIs can be derived from this work. We show that increased financial sustainability can go hand in hand with greater breadth of outreach, while the relationship between financial success and depth of outreach is not as pronounced. Our results further confirm the existence of *mission drift* (Armendáriz and Szafarz, 2011). That is, in the process of expanding their lending activities, MFIs tend to deviate from their mission to serve the poorest by extending loans to wealthier borrowers. Optimisation for both dimensions of social success therefore proves to be empirically problematic. Still, both breadth and depth of outreach are furthered by similar factors. The employee structure is found to be the strongest determinant of both dimensions of social success. In particular, a larger number of loan officers trained to serve more borrowers is beneficial, as well as the provision of appropriate incentives for the loan officers and targeting specific (new) borrower groups. Moreover, in line with the literature, we find that greater depth of outreach is associated with higher interest rates and default risk (Yunus, 2009). Financial success, on the other hand, is furthered mainly through greater financial performance and efficient management. As far as we know, the role of staff as a key factor in MFI success is new to the literature.

A series of robustness checks was conducted to assess the performance of the method and to evaluate the validity of this study's results. A comparison of the aMIRL results with those of the conventional lasso, with and without fixed effects, as well as MIRL estimates on a large, unbalanced panel comprising 3,846 observations of 1,026 MFIs, reveals that the results of all of these approaches are significantly inferior to those obtained by the aMIRL method. The corresponding estimates are more unstable, more variables are selected (many of which are insignificant), and the goodness of fit is similar or worse than that obtained by aMIRL. Moreover, we show that the correlation structure within the data is preserved by the imputation and that related outcome variables yield similar results. Lastly, post-selection quantile estimates corroborate our findings for the mean.

Our results show that it is crucial to account for MFI-specific heterogeneity through fixed effects, to use data-driven variable selection rather than pre-specified regressors, and to account for missing data in the analysis instead of excluding incomplete observations. They also empirically demonstrate that, under the given conditions of unobserved individual heterogeneity, pronounced missingness, high-dimensionality and multicollinearity, the aMIRL approach is more flexible and robust than the methods we compare it against.

Further research could facilitate a deeper understanding of the factors that contribute to MFI success. The incorporation of nonlinear effects and the study of simultaneous interactions between different outcome variables may yield interesting results. In addition, variable selection methods for fixed-effects quantile regressions could provide more comprehensive insights into the distribution of MFI success. However, extending the aMIRL algorithm for quantiles in panels with fixed effects is not straightforward. This is due to the inherent non-linearity of the quantile, which does not allow for the pre-elimination of fixed effects. This can lead to an excessive number of variables being removed, resulting in an overestimation of individual effects and a failure to account for the inherent variability in the data. We leave the development of a suitable technique for quantiles to a separate statistical paper.

3 Comparing Forecast Performance on Large Panel Data with Unknown Clustering Structure

This chapter draws on joint work with Melanie Schienle, presented at the *HKMetrics Workshop*, University of Mannheim (06/2025), the *RSS International Conference*, Edinburgh (09/2025), and the *IMS International Conference on Statistics and Data Science*, Seville (12/2025). It develops a Diebold-Mariano type test for equal predictive accuracy in large forecast panels, using a threshold variance estimator that accommodates heterogeneous, overlapping, and unknown cross-sectional dependence, with standard HAC treatment in the time-series dimension, and applies it to intraday sovereign CDS forecasts to demonstrate its robustness and practical relevance.

All code for the Diebold-Mariano test, model estimation, and data preprocessing is openly available at `https://github.com/lottarueter/threshold-dm-test`, providing full transparency of the empirical analysis.

3.1 Introduction

Advances in computational power, expanding data availability, and the development of flexible machine-learning methods have made it straightforward to generate forecasts for large panels of economic and financial variables. As a result, researchers and practitioners now routinely work with forecast panels that are characterised by long time series and heterogeneous across a potentially large cross-section. In many applications the primary objective is to evaluate comparative predictive performance and identify the model with superior accuracy. Typically, this is done using averages of loss function differences such as squared or absolute forecast errors. Yet these differences do not reveal whether one model significantly outperforms another. Establishing statistical significance is challenging in this setting, as valid inference requires a variance estimator for forecast-loss differentials that remains robust to complex and in particular mostly unknown forms of cross-sectional and serial dependence.

Economic forecast loss differentials, for example, for individual stocks or sovereigns, often reflect underlying group structures that are typically unknown. Moreover, these structures may vary across model comparisons, making global assumptions particularly restrictive and requiring an approach that flexibly accommodates model-pair–specific cross-sectional dependencies. These features are at odds with conventional assumptions invoked in large-sample panel inference. Standard asymptotic arguments based on independence across cross-sectional units are rarely plausible (Phillips and Moon, 1999; Pesaran and Yamagata, 2008), and ignoring cross-sectional dependence can lead to substantially distorted inference (Petersen, 2009). Likewise, cluster-robust methods require that the researcher knows the cluster assignments and/or that clusters are asymptotically independent, conditions that do not reflect the overlapping and interconnected nature of modern financial markets (Ibragimov and Müller, 2010, 2016; Zhou et al., 2021; Yap, 2025). Spatial heteroscedasticity and autocorrelation consistent (HAC) approaches, which weight units by physical distance, are also unsuitable when no meaningful notion of distance exists (Bester et al., 2011; Vogelsang, 2012; Kim and Sun, 2013; Hidalgo and Schafgans, 2021; Akgun et al., 2024). Recent two-way clustering and factor-based covariance estimators relax some restrictions but still impose a block-exchangeability or low rank latent-factor structure that may not describe the dependence pattern of forecast loss differentials (Cameron et al., 2011; Chen and Vogelsang, 2024; Chiang et al., 2024; Davezies et al., 2025). Bootstrap methods also rely on independent-cluster and therefore resulting exchangeability assumptions (Gonçalves, 2011; Menzel, 2021). Other recent approaches assume an underlying factor or restrictive block-covariance structure (Creal and Kim, 2024; Gao et al., 2024).

Among existing contributions, our paper aligns most closely with Qu et al. (2024). For their overall test of equal predictive accuracy in forecast panels, they state that they do not impose restrictions on the degree of cross-sectional dependence and can accommodate arbitrary dependence patterns. Their results, however, rely on consistency of a variance estimator that does not in fact adjust for such general forms of cross-sectional dependence.

This work develops a new test for equal predictive accuracy in large forecast panels that circumvents these limitations. We introduce a HAC-type variance estimator for the panel Diebold and Mariano (1995)-type statistic that regularises the cross-sectional component through data-driven covariance thresholding. The method builds on modern high-dimensional covariance estimation in the regression context (Bickel and Levina, 2008; Fan et al., 2013; Bai et al., 2024) and exploits the empirical observation that, in many forecasting applications, most cross-sectional covariance terms are small, though not necessarily zero, while a minority are economically meaningful. Our estimator automatically identifies and retains the latter while shrinking the former towards zero. By doing so, it accommodates arbitrary and overlapping cross-sectional dependence structures without requiring the researcher to specify groupings, factors, or distance metrics in advance. At the same time, it nests Newey and West (1987) (no cross-sectional dependence) and Driscoll and Kraay (1998) standard errors (full cross-sectional dependence) as limiting cases, providing a transparent continuum between the two.

We derive the asymptotic distribution of the resulting test statistic under weak conditions on the dependence structure of the forecast loss differentials, which may be unknown but must remain constant over time. The procedure permits general forms of heterogeneity across units and HAC-type serial correlation, making it well suited to financial applications. The method remains applicable when the relevant cross-sectional dependence structure varies across model pairs.

We illustrate our test through an empirical application to forecasts of intraday CDS spread returns for eleven European sovereigns during the euro-area crisis. This setting exhibits pronounced cross-sectional dependence. We find that most model comparisons yield highly significant differences even after accounting for complex dependence patterns. Thresholding proves to be particularly important in cases with small average score differences. However, given the relatively modest cross-sectional dimension of this application, both the Newey-West and Driscoll-Kraay estimators deliver results similar to ours, and the thresholding estimator does not alter the inference relative to these benchmarks. Because the framework is designed for larger panels, we expect the importance of thresholding to increase with N. We will explore this in future applications comparing large panels of forecasts for daily CRSP stock returns.

The remainder of this chapter proceeds as follows. Section 3.2 introduces the pooled test for equal predictive accuracy, presents the thresholded HAC variance estimator, and establishes its asymptotic properties. Section 3.3 applies the procedure to intraday CDS-spread-return forecasts. Section 3.4 concludes.

3.2 Theory

This section develops a Diebold and Mariano (1995)-type predictive accuracy test for panel data settings with large T and potentially large N, allowing for serial correlation and unknown cross-sectional dependence. Such characteristics are common in many empirical applications. For example, in financial forecasting panels, errors for returns, volatilities, or tail risk measures often display serial correlation, while cross-sectional dependence may arise from sectoral linkages, exposure to common risk factors, market-wide shocks, or latent networks.

A key feature of our approach is that we allow for a highly flexible form of cross-sectional dependence. In particular, while some pairs of units, such as equities within the same industry or banks exposed to similar portfolio risks, may exhibit strong dependence, the majority of cross-sectional covariances may be small but non-zero. Crucially, our framework requires neither knowledge of which dependencies are strong, nor pre-specified clusters, sectors, or factor structures. This flexibility is particularly valuable in high-dimensional financial settings where dependence can arise through channels that are difficult to observe.

3.2.1 Test Set-Up

We compare the predictive performance of two forecasting methods, m_1 and m_2, each producing a balanced panel of forecasts $\{\hat{y}_{it,m_1}\}$ and $\{\hat{y}_{it,m_2}\}$ for the realisation y_{it} across units $i = 1, \ldots, N$ and times $t = 1, \ldots, T$. The forecasts are treated as given, and no assumptions are made about how they were generated. This accommodates comparisons of nested forecasts, such as a factor-based return forecast versus an augmented specification, or a raw GARCH volatility forecast versus a bias-corrected or shrinkage-adjusted alternative.

The time dimension T is assumed large, while N may grow with T, subject to $\log N = O(\log T)$. Let $g_{it,m} = g(y_{it}, \hat{y}_{it,m})$ denote the loss for model m. Following Gneiting and Raftery (2007), competing forecasts should target the same functional and be evaluated using a loss function that elicits that functional. Mean-squared error is standard for mean forecasts, while quantile or asymmetric losses are natural for Value-at-Risk forecasts. Beyond this principle, the loss function $g : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ may be chosen freely, provided the resulting loss differentials $d_{it} = g_{it,m_1} - g_{it,m_2}$ satisfy Assumptions 1–2, the moment and dependence conditions introduced in Section 3.2.2.

We test the null hypothesis

$$H_0 : \mathbb{E}[\bar{g}_{m_1}] = \mathbb{E}[\bar{g}_{m_2}] \qquad \Longleftrightarrow \qquad H_0 : \mathbb{E}[\bar{d}] = 0, \tag{3.1}$$

where $\bar{g}_m = (NT)^{-1} \sum_{i,t} g_{it,m}$ and $\bar{d} = (NT)^{-1} \sum_{i,t} d_{it}$. The test statistic is

$$J_{NT} = \sqrt{NT}\,\frac{\bar{d}}{\hat{\sigma}_{\bar{d}}},$$

where $\hat{\sigma}_{\bar{d}}$ is a consistent estimator of the long-run variance of $\sqrt{NT}\,\bar{d}$. We adapt the thresholding variance estimator of Bai et al. (2024), originally developed for regression settings, to the forecasting context. This estimator remains consistent under serial correlation and flexible cross-sectional dependence, including unknown cluster structures, latent factor dependence, and network spillovers. It is developed in detail in Section 3.2.3.

We proceed as follows. We begin by stating the assumptions that ensure consistency of the variance estimator and asymptotic normality of the test statistic in Section 3.2.2. We then introduce the variance estimator in Section 3.2.3, and conclude with the main theoretical result, Theorem 3.2.1 in Section 3.2.4.

3.2.2 Assumptions

Let $d_t = (d_{1t}, \dots, d_{Nt})'$ denote the cross-section of loss differentials at time t. We now state the moment and mixing conditions under which $\sqrt{NT}\,\bar{d}$ is asymptotically normal.

Assumption 1 (*Expectation: Zero Mean and Finite rth Moment*)**.**

(i) (Zero Mean) $\mathbb{E}[d_{it}] = 0$.

(ii) (Finite rth Moment) $\mathbb{E}|d_{it}|^r < \Delta < \infty$ *for some* $r > 2$ *and all* i, t.

(iii) (Strong Mixing) For some $C_1 > 0$ *and all* $T > 0$,

$$\sup_{p \in \mathbb{Z}} \sup_{A \in \mathcal{F}_{-\infty}^{p},\, B \in \mathcal{F}_{T+p}^{\infty}} \left|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)\right| < C_1 T^{-r/(r-2)},$$

where $\mathcal{F}_{-\infty}^{p}$ *and* $\mathcal{F}_{T+p}^{\infty}$ *are the* σ*-algebras generated by* $\{d_t : t \leq p\}$ *and* $\{d_t : t \geq T + p\}$.

Assumption 1 imposes finite moments of order $r > 2$ and a standard form of weak temporal dependence. These conditions ensure that $\sqrt{NT}\,\bar{d}$ satisfies a central limit theorem even when the loss differentials exhibit serial correlation.

To obtain a consistent estimator of the long-run variance of $\sqrt{NT}\,\bar{d}$, we impose two further sets of assumptions. Assumption 2 governs the temporal and cross-sectional dependence in $\{d_{it}\}$, while Assumption 3 imposes regularity on the kernel weights used to approximate the long-run covariance matrix.

Before stating Assumption 2, we introduce several dependence measures. Let $\|A\| = \sqrt{\nu_{\max}(A'A)}$, where $\nu_{\max}(A)$ denotes the maximum eigenvalue of the matrix A. For any lag $h \geq 0$,

$$\gamma_{NT}(h) = \max_{t \leq T} \left(\|\mathbb{E}[d_t d_{t-h}']\| + \|\mathbb{E}[d_{t-h} d_t']\| \right),$$

$$\rho_{ij,h} = \max_{t \leq T} \left(|\mathbb{E}[d_{it} d_{j,t-h}]| + |\mathbb{E}[d_{i,t-h} d_{jt}]| \right),$$

which capture aggregate serial dependence and pairwise temporal dependence across units. For the long-run variance estimator, define

$$z_{h,ij,t} = \left(d_{it} d_{j,t-h} - \mathbb{E}[d_{it} d_{j,t-h}] \right) \omega(h, L)\, \mathbb{1}\{t \geq h\},$$

where $\omega(h, L)$ is a kernel weight and L the truncation lag.

Assumption 2 (*Variance: Weak Serial and Cross-Sectional Dependence*)**.**

(i) (Summability) $\sum_{h=0}^{\infty} \gamma_{NT}(h) \leq C_2$ *for some* $C_2 > 0$.

(ii) (Strong Mixing) For some $C_3 > 0$,

$$\sup_{p \in \mathbb{Z}} \sup_{A \in \mathcal{G}_{-\infty}^{p},\, B \in \mathcal{G}_{T+p}^{\infty}} \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right| < e^{-2C_3 T},$$

where $\mathcal{G}_{-\infty}^{p}$ *and* $\mathcal{G}_{T+p}^{\infty}$ *are the* σ*-algebras generated by* $\{z_{h,ij,t} : t \leq p\}$ *and* $\{z_{h,ij,t} : t \geq T + p\}$.

(iii) (Boundedness) $\max_{h,i,j,t} |z_{h,ij,t}| \leq C_4$ *for some* $C_4 > 0$.

(iv) (Sparse Cross-Sectional Dependence) For some $q \in [0, 1)$,

$$\omega_{NT}^{1-q} \max_{i \le N} \sum_{j=1}^{N} \left(\sum_{h=0}^{L} \rho_{ij,h} \right)^{q} = o(1), \qquad \omega_{NT} = L\sqrt{\frac{\log[(L+1)N^2T]}{T}}.$$

Assumption 2 restricts the strength of temporal and cross-sectional dependence. (i) rules out long memory; (ii) ensures that distant observations become approximately independent; (iii) prevents extreme realisations from dominating the variance estimator;[1] and (iv) allows for flexible cross-sectional dependence, including unknown clusters and factor structures, provided this dependence remains sufficiently sparse in the sense that it does not accumulate too quickly as N grows.

Assumption 3 (Weighting Scheme and Identification)**.**

(i) (Weight Convergence) For each fixed h *and some* $C_5 > 0$*, the kernel function* $\omega(h, L)$ *is such that* $\omega(h, L) \to 1$ *as* $L \to \infty$*, with* $\max_{h \le L} |\omega(h, L)| \le C_5$ *and* $\omega(0, L) = 1$.

(ii) (Identification and Bounded Variance) There is a $C_6 > 0$ *such that* $|s_{ii}| > C_6$ *for all* i *(see Equation 3.2), and the long-run variance satisfies* $0 < \delta < \sigma_{\bar{d}}^2 < \infty$.

Assumption 3 ensures that the kernel weighting scheme behaves well asymptotically and that the long-run variance is well defined and non-degenerate.

3.2.3 The Thresholding Variance Estimator

Under H_0, $\mathbb{E}[\bar{d}] = 0$, and hence

$$\begin{aligned} \sigma_{\bar{d}}^2 &\equiv \operatorname{Var}\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} \right) \\ &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\{ \sum_{t=1}^{T} \mathbb{E}(d_{it} d_{jt}) + \sum_{h=1}^{T-1} \sum_{t=h+1}^{T} \left[\mathbb{E}(d_{it} d_{j,t-h}) + \mathbb{E}(d_{i,t-h} d_{jt}) \right] \right\}. \end{aligned}$$

Define

$$v_{ij} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(d_{it} d_{jt}) + \frac{1}{T} \sum_{h=1}^{T-1} \sum_{t=h+1}^{T} \left[\mathbb{E}(d_{it} d_{j,t-h}) + \mathbb{E}(d_{i,t-h} d_{jt}) \right],$$

[1] This condition is slightly stronger than in Bai et al. (2024), who assume exponential tails but no serial correlation. We instead use a Bernstein-type inequality that accommodates weak serial dependence; see Appendix B.1.

so that

$$\sigma_{\bar{d}}^2 = \frac{1}{N}\sum_{i,j} v_{ij}.$$

Assumption 2 (iv) implies that $\{d_{it}\}$ is weakly correlated both over time and across units, so v_{ij} is nearly zero for most pairs (i,j). For a small $\delta > 0$, partition the index set as

$$\{(i,j) : i,j \leq N\} = S_s \cup S_l,$$

where

$$\begin{aligned} S_s &= \{(i,j) : |\mathbb{E}(d_{it}d_{j,t+h})| \leq \delta \text{ for all } h\}, \\ S_l &= \{(i,j) : |\mathbb{E}(d_{it}d_{j,t+h})| > \delta \text{ for some } h\}. \end{aligned}$$

The sets S_s and S_l collect pairs with small and large dependence, respectively. We assume that most index pairs lie in S_s, while all diagonals (i,i) lie in S_l. Importantly, we do not require prior knowledge of which specific pairs lie in S_s or S_l. Under this sparsity structure,

$$\sigma_{\bar{d}}^2 \approx \frac{1}{N}\sum_{(i,j)\in S_l} v_{ij}.$$

Following Bai et al. (2024), building on the HAC framework of Newey and West (1987), we approximate v_{ij} by a weighted sum of truncated autocovariances:

$$s_{ij} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(d_{it}d_{jt}) + \frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}\left[\mathbb{E}(d_{it}d_{j,t-h}) + \mathbb{E}(d_{i,t-h}d_{jt})\right]. \tag{3.2}$$

We use the Bartlett kernel $\omega(h,L) = 1 - h/(L+1)$ and choose $L = 4(T/100)^{2/9}$ following Newey and West (1994), see Andrews (1991) for an overview of other kernel functions. Then

$$\sigma_{\bar{d}}^2 \approx \frac{1}{N}\sum_{(i,j)\in S_l} s_{ij}.$$

In practice, s_{ij} is replaced by its sample analogue

$$\hat{s}_{ij} = \frac{1}{T}\sum_{t=1}^{T} d_{it}d_{jt} + \frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}\left[d_{it}d_{j,t-h} + d_{i,t-h}d_{jt}\right].$$

To identify S_l, let $M > 0$ be a tuning parameter and define

$$\omega_{NT} = L\sqrt{\frac{\log[(L+1)N^2T]}{T}}, \qquad \lambda_{ij} = M\,\omega_{NT}\sqrt{\hat{s}_{ii}\hat{s}_{jj}}.$$

The set of estimated "large" elements is

$$\hat{S}_l = \{(i,j) : i = j \text{ or } |\hat{s}_{ij}| > \lambda_{ij}\}.$$

The choice of ω_{NT} ensures

$$\max_{i,j\le N} |\hat{s}_{ij} - s_{ij}| = O_P(\omega_{NT}),$$

so that $\hat{S}_l$ consistently identifies non-negligible dependence.

The hard-thresholding estimator of σ_d^2 is

$$\hat{\sigma}_d^2 = \frac{1}{N}\sum_{(i,j)\in\hat{S}_l} \hat{s}_{ij}. \tag{3.3}$$

This construction nests several familiar estimators. Setting $M = 0$ recovers the estimator of Driscoll and Kraay (1998), which retains all cross-sectional dependencies and can therefore be relatively noisy; letting $M \to \infty$ recovers the Newey and West (1987) estimator, which ignores them entirely. Section 3.2.3 describes a data-driven rule for choosing M. Known cluster structures can be incorporated by forcing all pairs within a cluster into $\hat{S}_l$.

Choice of the Tuning Parameter M

The tuning parameter $M > 0$ controls the degree of thresholding applied to $\hat{s}_{ij}$. Small values retain many cross-sectional dependencies, whereas large values retain only strong ones. Since the optimal degree of thresholding is unknown, we select M by cross-validation.

Define

$$\lambda_{ij}(M) = M\,\omega_{NT}\sqrt{\hat{s}_{ii}\hat{s}_{jj}}, \qquad \omega_{NT} = L\sqrt{\frac{\log[(L+1)N^2T]}{T}}.$$

Partition the time dimension into $P = \lfloor \log T \rfloor$ consecutive validation blocks. For each candidate M from a grid $\mathcal{M}$, compute:

1. A *training* long-run variance estimate,

$$\hat{\sigma}^{(M)}_{\bar{d},\text{train}} = \frac{1}{N}\sum_{i,j} \hat{s}^{(M)}_{ij},$$

2. and, for each block $p = 1, \ldots, P$, a *validation* estimate $\hat{\sigma}^{(M)}_{\bar{d},\text{val},p}$ computed analogously using only observations in block p.

Select

$$M_{\text{CV}} = \arg\min_{M\in\mathcal{M}} \frac{1}{P}\sum_{p=1}^{P}\left(\hat{\sigma}^{(M)}_{\bar{d},\text{train}} - \hat{\sigma}^{(M)}_{\bar{d},\text{val},p}\right)^2.$$

This procedure chooses the value of M that yields the most stable variance estimates across validation blocks.

3.2.4 The Main Result

Theorem 3.2.1. *Under Assumptions 1–3,*

$$\sqrt{NT}\,\frac{\bar{d}}{\hat{\sigma}_{\bar{d}}} \xrightarrow{d} \mathcal{N}(0,1) \qquad \textit{as } T \textit{ (and possibly } N\textit{)} \to \infty.$$

Proof. Assumption 1 yields a central limit theorem for $\sqrt{NT}\,\bar{d}$ (Theorem 5.20 in White, 2014), so $\sqrt{NT}\,\bar{d}/\sigma_{\bar{d}} \xrightarrow{d} \mathcal{N}(0,1)$. Assumptions 2–3 imply consistency of the thresholding estimator. Slutsky's theorem gives the result. □

Remark 3.2.2. *Theorem 3.2.1 also holds for the corresponding soft-thresholding estimator:*

$$\hat{\sigma}^2_{\bar{d},Soft} = \frac{1}{N}\sum_{i,j}\hat{s}^{Soft}_{ij}, \qquad \hat{s}^{Soft}_{ij} = \begin{cases} \hat{s}_{ij}, & i = j, \\ \operatorname{sgn}(\hat{s}_{ij})(|\hat{s}_{ij}| - \lambda_{ij}), & |\hat{s}_{ij}| \geq \lambda_{ij},\ i \neq j, \\ 0, & |\hat{s}_{ij}| < \lambda_{ij},\ i \neq j. \end{cases}$$

3.3 Empirical Study: Comparing Sovereign CDS Forecasts

We compare out-of-sample forecasts of 30-minute sovereign CDS changes for eleven European countries over January 2011 to December 2014, using a range of model classes that combine lagged CDS spreads with high-frequency EuroStoxx 50 indicators and daily macro-financial variables. This period covers the core phase of the Euro-area sovereign debt crisis and is characterised by pronounced volatility and substantial cross-country interconnectedness (Buse and Schienle, 2019), creating a demanding environment for short-horizon forecasting. While sovereign CDS series typically co-move strongly in periods of stress, structural differences across economies can imply that some cross-sectional dependencies remain weak or negligible. These patterns are expected to carry over, potentially in amplified form, to the squared forecast-error differentials used in the model comparisons. This combination of strong and weak linkages makes the setting particularly well suited for assessing forecasting methods and for illustrating how our testing procedure accommodates heterogeneous cross-sectional dependence.

3.3.1 Data

Our data set comprises intraday CDS observations for eleven European sovereigns and EuroStoxx 50 prices recorded from 09:30 to 17:00 over 1,422 trading days between January 2009 and December 2014, yielding 22,752 intraday observations. In addition, we include daily data for eight macro-financial variables. All intraday and daily regressors are listed in Table 3.1. Descriptive statistics for all variables are provided in Table B.1 in the appendix.

CDS Data

We use 30-minute mid-quotes of five-year USD-denominated sovereign CDS contracts, following Gyntelberg et al. (2013), as this frequency provides a useful compromise between granularity and data completeness. The sample includes Austria, Belgium, Germany, Spain, Finland, France, Greece, Ireland, Italy, the Netherlands, Portugal, and the United Kingdom (AT, BE, DE, ES, FI, FR, GR, IE, IT, NL, PT, UK). As in Buse et al. (2022), Greece is excluded due to persistent illiquidity and extensive missingness after 2011. Quotes are observed from 08:30 to 17:00 CET, constructed from cleaned bid and offer prices, and merged across countries by timestamp. To limit noise from imputing returns over extensive data gaps, we drop any day on which any country has more than 50% missing observations. Remaining gaps are imputed within each day using a simple local rule: forward and backward fill for boundary missingness, and the average of adjacent values for interior gaps. As a robustness check, we also implement Kalman-filter imputation, which produces nearly identical CDS return series

Table 3.1: Notation and definitions of intraday and daily regressors

Symbol	Definition	Economic Interpretation
	Intraday Variables	
$r_{d,t}^{\text{EUSTX50}}$	$\ln P_{d,t} - \ln P_{d,t-1}$, constructed from the first and last price within interval $(t-1,t]$	30-minute log return of the EuroStoxx 50 index; captures high-frequency equity movements
$\text{RV}_{d,t}^{\text{EUSTX50}}$	$\sqrt{\sum_{k\in(t-1,t]}(\Delta \ln P_{d,k})^2}$, using squared 15-second log returns inside the interval	30-minute realised volatility; measures short-term market stress
	Daily Variables	
FS_d	$\text{Euribor3m}_d - \text{Eonia3m}_d$	Funding spread; indicator of euro-area funding and liquidity stress affecting sovereign CDS
Slope_d	$\text{EuroSwap15y}_d - \text{Euribor3m}_d$	Yield-curve slope; reflects macro outlook and monetary stance driving sovereign risk premia
ΔVIX_d	$\text{VIX}_d - \text{VIX}_{d-1}$	Daily change in the VIX; captures global risk aversion relevant for CDS spreads
ΔEVZ_d	$\text{EVZ}_d - \text{EVZ}_{d-1}$	Daily change in EuroStoxx 50 implied volatility; measures euro-area specific uncertainty
$\Delta\text{iTraxx}_d^{\text{Corp}}$	$\text{iTraxx}_d^{\text{Corp}} - \text{iTraxx}_{d-1}^{\text{Corp}}$	Daily change in iTraxx Europe Non-Financials 5y; proxy for broad credit-market conditions
$\Delta \ln FX_d^{\text{GBP/EUR}}$	$\ln FX_d^{\text{GBP/EUR}} - \ln FX_{d-1}^{\text{GBP/EUR}}$	Log return of the GBP/EUR exchange rate; reflects shifts in relative U.K.-euro-area sentiment
$\Delta \ln FX_d^{\text{CHF/EUR}}$	$\ln FX_d^{\text{CHF/EUR}} - \ln FX_{d-1}^{\text{CHF/EUR}}$	Log return of CHF/EUR; proxy for safe-haven flows
$\Delta \ln FX_d^{\text{NOK/EUR}}$	$\ln FX_d^{\text{NOK/EUR}} - \ln FX_{d-1}^{\text{NOK/EUR}}$	Log return of NOK/EUR; captures regional and commodity-related risk sentiment

and leaves all forecasting and testing results unchanged (available upon request). Finally, preventing overnight-effects to distort our results, observations at 08:30 are removed before computing 30-minute CDS changes, leaving a regular intraday panel spanning 09:30 to 17:00.

Intraday High-Frequency Regressors

EuroStoxx 50 index prices are observed at 15-second frequency between 09:00 and 17:30 CET. To mitigate overnight and opening-auction effects, observations from 09:00 to 09:15 are discarded, consistent with Hautsch (2004). The first intraday interval therefore spans 09:15 to 09:30 (15 minutes), and all subsequent intervals have a length of 30 minutes up to 17:30. From these prices, we construct (i) log returns using the first and last price within each interval, and (ii) realised volatility based on squared 15-second log returns recorded inside the interval. Consequently, observations for the return and volatility measures are available at 30-minute timestamps from 09:30 to 17:30. For the forecasting exercise, only information up to 16:30 is relevant, as these lagged values are used to predict CDS changes up to 17:00. On 1 May in 2012–2014, EuroStoxx prices are missing due to exchange holidays while CDS markets remain open. For these dates, we replace the missing intraday sequence with the previous trading day's EuroStoxx data to preserve alignment across variables.

Daily Macro-Financial Variables

The daily variables capture broader macro-financial conditions relevant for short-term sovereign risk, including funding pressures, interest-rate dynamics, market uncertainty, credit-market stress, and exchange-rate movements. Specifically, we use the interbank funding spread (Euribor–Eonia), the euro-area yield-curve slope (EuroSwap15y–Euribor), daily changes in VIX and EVZ, daily changes in the iTraxx Europe Non-Financials index, and log returns of the GBP/EUR, CHF/EUR, and NOK/EUR exchange rates. All series are merged by date, cleaned using linear interpolation for isolated gaps, and aligned to CDS trading days. Daily variables enter the forecasting model with a one-day lag to ensure predeterminedness with respect to intraday CDS movements.

3.3.2 Model Specifications

We forecast 30-minute CDS spread changes for 11 countries from 10:00 to 17:00. Ters and Urban (2018) examine various lag lengths and find that results are robust up to five lags. For simplicity, we use a single lag. Accordingly, the models incorporate intraday information lagged by one interval and daily information from the preceding day. The intraday forecasting framework is therefore given by

$$y_{i,d,t} = f\Big(y_{i,d,t-1},\, \mathbf{W}^{\text{intra}}_{d,t-1},\, \mathbf{W}^{\text{daily}}_{d-1},\, \{y_{j,d,t-1}\}_{j=1,j\neq i}^{N}\Big), \qquad i,j = 1,\dots,11, \quad t = 1,\dots,15,$$

where $y_{i,d,t}$ denotes the 30-minute CDS spread change for country i and t indexes 30-minute intervals within trading day d. The vector $\mathbf{W}^{\text{intra}}_{d,t-1}$ contains lagged intraday EuroStoxx regressors, and $\mathbf{W}^{\text{daily}}_{d-1}$ collects lagged daily macro-financial variables.

We organise the forecasting design into four regressor sets:

R1 $= \{y_{i,d,t-1}\}$;

R2 $= \{y_{i,d,t-1},\, \mathbf{W}^{\text{intra}}_{d,t-1}\}$;

R3 $= \{y_{i,d,t-1},\, \mathbf{W}^{\text{intra}}_{d,t-1},\, \mathbf{W}^{\text{daily}}_{d-1}\}$;

R4 $= \{\{y_{j,d,t-1}\}_{j=1}^{N},\, \mathbf{W}^{\text{intra}}_{d,t-1},\, \mathbf{W}^{\text{daily}}_{d-1}\}$.

Estimation employs recursive 24-month rolling windows, with predictions generated for the subsequent month. For example, data from 01/2009–12/2010 (12/2012–11/2014) deliver out-of-sample forecasts for 01/2011 (12/2014). The procedure yields intraday forecasts from January 2011 to December 2014, spanning 947 trading days and 14,205 forecasted observations.

We assess six econometric and machine-learning models for forecasting sovereign CDS spread changes. Table 3.2 summarises the specifications, which follow standard practice in the intraday CDS and asset-pricing literature (Ters and Urban, 2018; Huddleston et al., 2023). The multivariate PVAR-X models include the complete set of lagged dependent variables together with the common controls. Although PVAR-X accommodates cross-equation dependence and cross-country dynamics, adding the cross-country lag block (R4) to the univariate linear specification yields almost identical results; we therefore exclude R4 in the linear model for parsimony. The last-value benchmark provides a naive benchmark, and the GARCH model serves as a standard time-series reference. Our main analysis centres on the linear, PVAR-X, Random Forest, and XGBoost models. Since all models yield conditional mean forecasts, we assess their accuracy using the mean squared prediction error (MSPE).

3.3.3 Results

Table 3.3 reports the out-of-sample MSPEs for the 16 competing specifications over January 2011–December 2014. The benchmark specifications deliver the weakest performance: the last-value predictor performs uniformly poorly, and the GARCH model improves upon it but remains far behind all richer specifications. Augmenting the linear regression with high-frequency EuroStoxx regressors (R2) reduces the MSPE relative to the autoregressive baseline (R1), whereas including daily macro-financial variables (R3) provides no systematic gains at the 30-minute horizon. The PVAR-X model attains the

Table 3.2: Model specifications

Model	Specification
Last-Value	$$\hat{y}_{i,d,t} = y_{i,d,t-1}.$$ Uses R1.
GARCH	Univariate GARCH(1,1) model: $$y_{i,d,t} = \mu_i + \gamma_i y_{i,d,t-1} + \varepsilon_{i,d,t}, \qquad \varepsilon_{i,d,t} \sim N(0, h_{i,d,t}),$$ $$h_{i,d,t} = \omega_i + \alpha_i \varepsilon^2_{i,d,t-1} + \beta_i h_{i,d,t-1}.$$ Uses R1.
Linear	Univariate OLS regressions: $$y_{i,d,t} = \alpha_i + \beta_i y_{i,d,t-1} + \boldsymbol{\delta}_i' \mathbf{W}^{\text{intra}}_{d,t-1} + \boldsymbol{\gamma}_i' \mathbf{W}^{\text{daily}}_{d-1} + \varepsilon_{i,d,t}.$$ Estimated with R1–R3.
PVAR-X	Seemingly unrelated regression for $i = 1, \dots, N$: $$y_{i,d,t} = \alpha_i + \sum_{j=1}^{N} \phi_{ij} y_{j,d,t-1} + \boldsymbol{\delta}_i' \mathbf{W}^{\text{intra}}_{d,t-1} + \boldsymbol{\gamma}_i' \mathbf{W}^{\text{daily}}_{d-1} + \varepsilon_{i,d,t}.$$ Estimated with R1–R3.
Random Forest	$$y_{i,d,t} = f_{\text{RF},i}(\mathbf{X}_{d,t-1}) + \varepsilon_{i,d,t}.$$ Estimated with R1–R4.
XGBoost	$$y_{i,d,t} = f_{\text{XGB},i}(\mathbf{X}_{d,t-1}) + \varepsilon_{i,d,t}.$$ Estimated with R1–R4.

Note: All models yield conditional mean forecasts and are estimated recursively using 24-month windows. Based on the estimated model they generate 1-step ahead forecasts for the subsequent month. The regressor sets R1–R4 are defined in Section 3.3.2 based on the variables presented in Table 3.1. Since PVAR-X is multivariate, all specifications include the full set of country-lags.

lowest MSPE among all specifications, with only minor variation across regressor sets. Among the machine-learning models, Random Forest and XGBoost improve substantially when intraday information and cross-country lags are added; under R4 their performance approaches that of PVAR-X and the linear model. Across the full set of models, intraday information delivers most of the predictive power, while daily variables tend to add noise.

In the following, we focus on the six specifications with the lowest MSPE (marked in bold in Table 3.3); the complete set of $16 \times 15/2 = 120$ pairwise comparisons is provided in Table B.2 in the appendix. The Diebold-Mariano results in Table 3.4 confirm the ordering implied by the MSPEs. All models deliver clear improvements over the last-value and GARCH benchmarks, with rejections that remain robust after Bonferroni

Table 3.3: Mean squared prediction error of all model specifications

Model	**Type**	**R1**	**R2**	**R3**	**R4**
Last-value	*UV*	**15.629**	–	–	–
GARCH	*UV*	**6.749**	–	–	–
Linear	*UV*	6.010	**5.989**	5.993	–
PVAR-X	*MV*	5.826	**5.816**	5.821	–
Random Forest	*UV*	7.316	6.282	5.987	**5.855**
XGBoost	*UV*	6.304	6.283	6.168	**6.051**

Note: Entries are mean squared prediction errors (MSPE) averaged across 11 sovereign CDS series (AT, BE, DE, ES, FI, FR, GR, IE, IT, NL, PT, UK) and 15 intraday forecasts per trading day (10:00, 10:30, ..., 17:00). The out-of-sample period runs from January 2011 to December 2014 (947 trading days; 14,205 forecasts per series). *UV* denotes univariate estimation; *MV* denotes multivariate estimation. Regressor sets R1–R4 are defined in Section 3.3.2. Bold values mark the lowest MSPE within each model. R4 is redundant for PVAR-X owing to its cross-sectional structure.

adjustment for the full set of comparisons. Among the leading specifications, only two contrasts fail to reject equal predictive accuracy at conventional levels: the linear model versus XGBoost and the PVAR-X model versus Random Forest. These correspond to the smallest absolute loss differentials, although the link between $|\bar{d}|$ and the associated p-values is not strictly monotone. Overall, the evidence places PVAR-X and Random Forest in the top tier, with XGBoost and the linear model close behind.

Focusing on the cases in which the data-driven procedure selects $M > 0$, we find that thresholding becomes most relevant when the average loss differential $\bar{d}$ is small, that is, in comparisons where the test is most informative. This relevance arises in the covariance structure of the loss differentials, even though thresholding plays little role for the underlying CDS changes $y_{i,d,t}$ and model forecasts $\hat{y}_{i,d,t}$, as illustrated in Figure 3.1. The observed CDS changes display the familiar segmentation between the highly volatile countries (Portugal, Spain, Italy, and Ireland) and the more stable core sovereigns, a pattern largely mirrored in the forecasts. By contrast, it is only in the squared loss differentials that thresholding becomes necessary: the resulting clusters are neither mutually exclusive nor stable across model pairs. In particular, Finland and the United Kingdom shift between high- and low-volatility groups depending on the models compared. This variation indicates that their linkages are specification-driven rather than structural. The sparsity patterns selected by the thresholded HAC estimator remove numerous noise-driven cross-sectional correlations, most notably those magnified by Portugal's volatility, demonstrating that thresholding is effective even in a cross-section of only eleven sovereigns.

Table 3.4: Diebold-Mariano test results for equal predictive accuracy of selected model pairs

Model m_1	**Model m_2**	$\bar{d}$	$\hat{\sigma}_{\bar{d},NT}$	J_{NT}	**p-value**	p^{BF}	M
Last-value	GARCH	8.880	0.601	14.769	0.000	0.000	0.00
Last-value	Linear$^{\mathrm{R2}}$	9.641	0.600	16.079	0.000	0.000	0.00
Last-value	PVAR-X$^{\mathrm{R2}}$	9.813	0.638	15.369	0.000	0.000	0.00
Last-value	Random Forest$^{\mathrm{R4}}$	9.774	0.624	15.673	0.000	0.000	0.00
Last-value	XGBoost$^{\mathrm{R4}}$	9.579	0.632	15.157	0.000	0.000	0.00
GARCH	Linear$^{\mathrm{R2}}$	0.761	0.052	14.516	0.000	0.000	0.00
GARCH	PVAR-X$^{\mathrm{R2}}$	0.933	0.058	16.098	0.000	0.000	0.00
GARCH	Random Forest$^{\mathrm{R4}}$	0.894	0.050	17.904	0.000	0.000	0.00
GARCH	XGBoost$^{\mathrm{R4}}$	0.699	0.058	11.972	0.000	0.000	0.00
Linear$^{\mathrm{R2}}$	PVAR-X$^{\mathrm{R2}}$	0.172	0.049	3.481	0.000	0.007	**0.33**
Linear$^{\mathrm{R2}}$	Random Forest$^{\mathrm{R4}}$	0.134	0.034	3.938	0.000	0.000	**0.10**
Linear$^{\mathrm{R2}}$	XGBoost$^{\mathrm{R4}}$	-0.062	0.051	-1.224	**0.221**	**1.000**	**0.06**
PVAR-X$^{\mathrm{R2}}$	Random Forest$^{\mathrm{R4}}$	-0.039	0.023	-1.653	**0.098**	**1.000**	**0.16**
PVAR-X$^{\mathrm{R2}}$	XGBoost$^{\mathrm{R4}}$	-0.234	0.029	-8.088	0.000	0.000	0.00
Random Forest$^{\mathrm{R4}}$	XGBoost$^{\mathrm{R4}}$	-0.196	0.027	-7.130	0.000	0.000	**0.11**

Note: Results of best model specifications according to MSPE, see Table 3.3. $\bar{d}$ is the average loss differential between models m_1 and m_2; $\hat{\sigma}_{\bar{d},NT}$ is the hard-thresholded standard error divided by $\sqrt{NT}$; J_{NT} is the corresponding Diebold-Mariano statistic, standard normal under the null of equal predictive accuracy. Bonferroni-adjusted p-values (p^{BF}) correct for 120 pairwise comparisons. Positive $\bar{d}$ indicates higher loss for m_1. Bold entries mark $M > 0$ or non-rejection at the 1% level.

Taken together, the evidence yields two broad conclusions. First, generating sizeable gains at the 30-minute horizon is challenging, and only the strongest specifications improve consistently upon well-performing alternatives. Second, careful treatment of heterogeneous cross-sectional dependence is essential for reliable inference. In particular, thresholding is most informative precisely in the settings where the models are hardest to distinguish. Across all specifications, the overall ranking aligns with the MSPE evidence: PVAR-X and Random Forest form the leading group, XGBoost and the linear models remain competitive, and the GARCH and last-value benchmarks are consistently outperformed across countries and regressor sets.

Robustness

The variance-estimator comparison in Table B.2 confirms that differences across our threshold estimator and the zero-lag, Newey-West, and Driscoll-Kraay alternatives become particularly relevant when the loss differential is small. For contrasts with large $|\bar{d}|$, including all combinations involving the last-value benchmark, GARCH, or other clearly

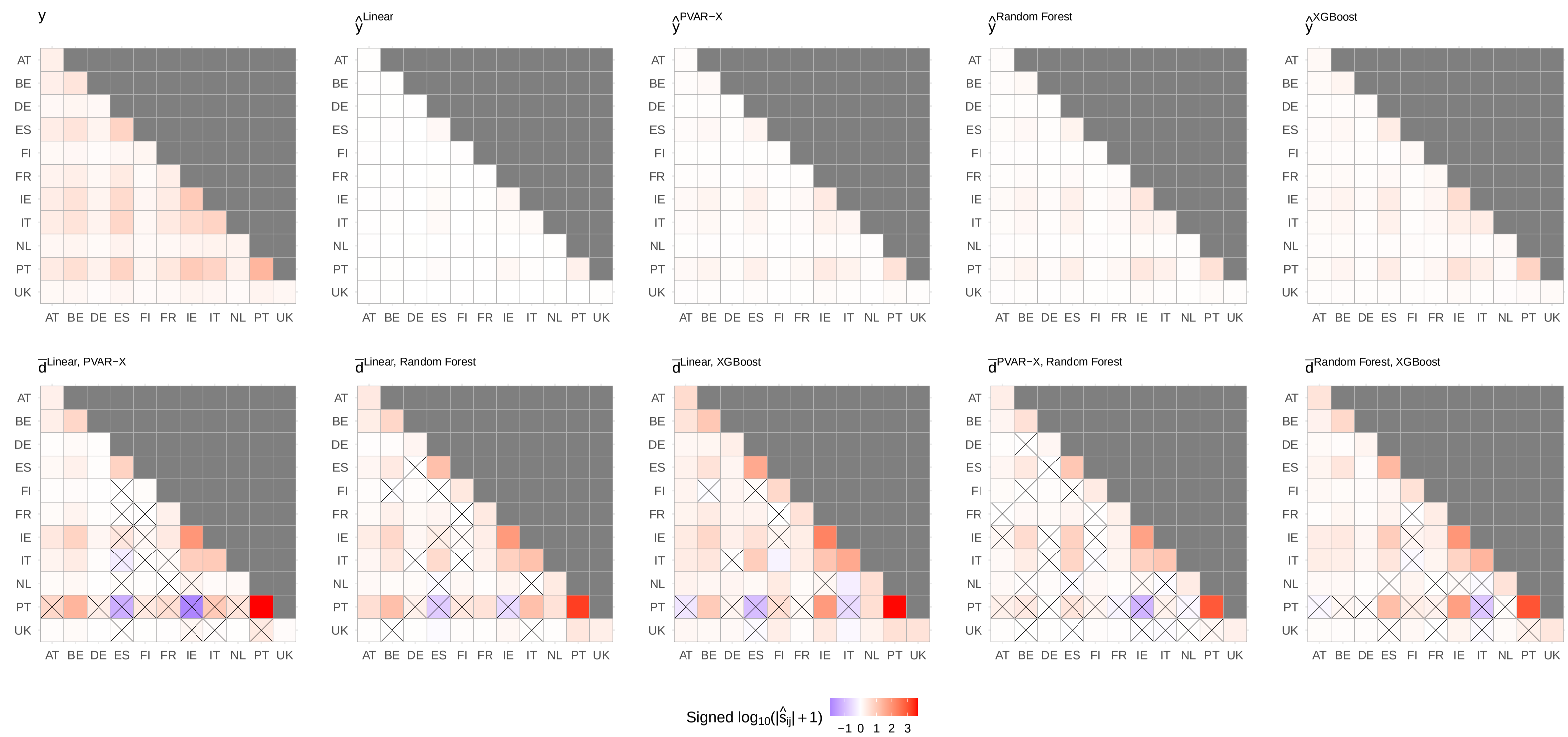

Figure 3.1: Threshold-estimated cross-sectional covariance matrices for average forecast loss differentials of model comparisons with $M > 0$, see Table 3.4, respective forecasts and observed CDS spreads

Note: y denotes the true CDS spreads; $\hat{y}^{\text{Linear}}$, $\hat{y}^{\text{PVAR-X}}$, $\hat{y}^{\text{RF}}$, and $\hat{y}^{\text{XGB}}$ are the corresponding model-based predictions and $\bar{d}^{\text{Linear,PVAR-X}}$, $\bar{d}^{\text{Linear,RF}}$, $\bar{d}^{\text{Linear,XGB}}$, $\bar{d}^{\text{PVAR-X,RF}}$, $\bar{d}^{\text{RF,XGB}}$ denote the average forecast loss differentials. Each panel reports the signed logarithm $\text{sign}(s_{ij})\log_{10}(|\hat{s}_{ij}| + 1)$ of the HAC-estimated contemporaneous covariance $\hat{s}_{ij}$ across countries. The scale mitigates the disproportionate influence of Portugal. Crosses indicate entries set to zero under the thresholded HAC sparsity pattern.

inferior specifications, the four estimators yield virtually identical conclusions, with test statistics far above conventional thresholds and p-values effectively zero. The informative cases arise exclusively among the strongest models and their close variants, where the loss differentials are small and the corresponding covariance estimates become inherently noisy. In these comparisons the thresholding procedure frequently selects a positive bandwidth M, and the resulting standard errors are typically larger than their zero-lag counterparts but broadly similar to the Newey-West and Driscoll-Kraay estimates.

These patterns reflect the pronounced imbalance between the long time-series dimension ($T = 14{,}205$) and the small cross-section ($N = 11$). With temporal dependence dominating cross-sectional dependence, the Newey-West and Driscoll-Kraay standard errors are nearly identical across all comparisons. The thresholded estimates diverge from them only in the few cases with very small $|\bar{d}|$, where cross-country covariance terms are both weak and imprecisely estimated. Trimming these terms yields modest changes in the standard error and slightly different p-values. By contrast, the zero-lag estimator ignores time dependence entirely and is therefore consistently much smaller, often by 30–50%, which inflates test statistics across the board.

Overall, the zero-lag estimator systematically understates uncertainty, whereas the Newey-West, Driscoll-Kraay, and thresholded estimators differ only moderately. Even so, cross-sectional thresholding proves important in the tightest comparisons. Trimming weak cross-country correlations therefore stabilises inference without affecting results elsewhere.

3.4 Conclusion

This chapter develops a Diebold-Mariano type test for equal predictive accuracy that is tailored to large forecast panels characterised by heterogeneous and unknown dependence structures. The proposed procedure combines a HAC-type long-run variance estimator with data-driven cross-sectional covariance thresholding. This construction makes it possible to accommodate rich and potentially overlapping patterns of dependence without requiring the researcher to specify clusters, factors, or network structures in advance. The resulting test statistic is valid under weak conditions, allows both dimensions of the panel to grow, and nests Newey-West and Driscoll-Kraay estimators as limiting cases. In an empirical application to intraday sovereign CDS forecasts, the test distinguishes clearly between models with genuinely different predictive accuracy while remaining stable in comparisons where performance differences are small. Thresholding is most relevant precisely in these latter cases, highlighting its role in mitigating noise in high-dimensional covariance estimation. Our results further show that thresholding is inherently model-comparison specific: cross-sectional dependence in the loss differentials need not mirror that of the underlying data, and observed groupings are neither mutually exclusive nor stable across model pairs. This flexibility is a key advantage of the approach and suggests that thresholding would remain indispensable even if true clusters in the underlying series were known.

Several avenues for further research follow naturally. First, many forecasting settings involve unbalanced or incomplete panels with irregularly spaced observations. Such patterns may arise from the underlying series to be predicted, from differences in forecast availability across units, or from situations in which forecasts are issued by different providers. Extending the estimator to such environments would increase its practical relevance, for instance in applications with asynchronous trading where forecast targets may not align cleanly across units.

Second, while the current framework accommodates flexible forms of temporal dependence, it does not yet allow for structural breaks, regime shifts, or changepoints in the underlying forecasting environment. The same limitation applies to the cross-section, as the present framework assumes that cross-sectional dependence remains stable over time. In many applications, however, the strength and composition of such linkages may themselves evolve. Incorporating these features, for example by drawing on recent advances in two-way clustering and time-varying dependence, would enable the test to adapt to changing dynamics in both dimensions.

Finally, applying the test to substantially larger cross-sections, such as daily stock return forecasts across hundreds of assets, would provide an informative assessment of its performance in genuinely high-dimensional settings. In such panels the number of weak cross-sectional correlations grows rapidly, and covariance thresholding is expected to play an even more prominent role. Exploring these extensions is left for future work.

4 Direction Augmentation in the Evaluation of Armed Conflict Predictions

This chapter is joint work with Johannes Bracher, Fabian Krüger, Sebastian Lerch, and Melanie Schienle, published in *International Interactions* (Bracher et al., 2023)[1], and presented at the *HKMetrics Workshop*, KIT (07/2023), the *MathSEE Symposium*, KIT (09/2023), the *CCEW Symposium*, University of the Bundeswehr, Munich (10/2023), and a workshop for professional conflict forecasters, Berlin (10/2024). It examines how a novel scoring rule (TADDA) introduced in the 2020 VIEWS Prediction Competition shapes forecasters' incentives by favouring conservative, near-zero predictions, and shows why even simple no-change forecasts can outperform more complex models under this metric.

Code and data for the presented results are available at https://github.com/KITmetricslab/tadda as well as in the replication archive at https://dataverse.harvard.edu/dataverse/internationalinteractions.

4.1 The 2020 VIEWS Prediction Competition and Forecast Evaluation

The 2020 VIEWS Prediction Competition (Hegre et al., 2022; Vesco et al., 2022) represents a major effort to improve forecasting capacities in the field of armed conflict studies. It provides a valuable opportunity to compare various statistical and machine learning methods, as well as combined ensemble forecasts.

[1] Bracher, J., L. Rüter, F. Krüger, S. Lerch, and M. Schienle (2023): "Direction Augmentation in the Evaluation of Armed Conflict Predictions," *International Interactions*, 49(6), 989–1004. © 2023 Informa UK Limited, trading as Taylor & Francis Group. Reprinted by permission of Taylor & Francis Group, http://www.tandfonline.com.

The aim of the competition in 2020 was to "improve collective scientific knowledge on forecasting (de-)escalation in Africa" (Vesco et al., 2022, p. 860). Participants were asked to predict the log-change in monthly fatalities due to state-based violence in a given region. In slightly modified notation, the target for a forecast referring to month t and issued at the initial time $t-s$ with a lead time of s months was

$$Y_{t,s} = \log(X_t + 1) - \log(X_{t-s} + 1).$$

Here, X_t is the number of fatalities from state-based conflict in month t as defined in the UCDP data set[2]. A value of $Y_{t,s} = 0$ thus corresponds to no change between months $t-s$ and t, while negative values imply a decrease and positive values an increase in fatalities. Point forecasts were requested for different initial and lead times, and for two forecasting periods. First, forecasts had to be issued for the true future at the designated deadline (30 September 2020), more precisely for the period of October 2020 through March 2021. Second, to facilitate a more comprehensive methods comparison, retrospective predictions for a second, longer prediction period were collected (January 2017 through December 2019; see Table 1 in Hegre et al., 2022). Forecasts for all of Africa could be issued at the country level and for sub-national units defined via a grid[3].

In Vesco et al. (2022, Figure 6) the collected forecasts were assessed via a total of 13 different statistical evaluation scores. These aimed to cover different desirable properties of point forecasts, including their accuracy and positive impact on ensemble forecasts. The main accuracy metrics were the following: (i) the mean squared error (MSE), which had been pre-specified in the announcement of the challenge, and (ii) the newly introduced *Targeted Absolute Distance with Direction Augmentation* (TADDA). TADDA is designed as a "metric specifically tailored to evaluate predictions of changes in fatalities, as it accounts for both the sign and the magnitude of the predictions versus the actual change" (Hegre et al., 2022, p. 542). To this end, TADDA combines a measure of distance between the forecast and the observation and a term penalising forecasts with a different sign from the observed value. See the next section for the exact definition.

[2]UCDP stands for *Uppsala Data Conflict Programme*, see Pettersson and Öberg (2020) and `https://ucdp.uu.se/downloads/` (accessed on 19 July 2023).

[3]PRIO-GRID was chosen as a finer-resolution alternative to the country-level approach. It "is a vector grid network with a resolution of 0.5×0.5 decimal degrees, covering all terrestrial areas of the world", see Tollefsen et al. (2012).

The independent scoring committee (SC) convened to evaluate forecasts ultimately decided to score forecasts mainly based on the MSE, noting that the application of TADDA

> "[...] is somewhat problematic. In particular, with the parameterisation used for the evaluation, in the TADDA score, no model can outperform the no-change model predicting 'no change in violence'. This means that a model with predictions very close to a constant no-change model would be preferred if evaluated according to this score, which the SC did not consider to be a good choice." (Vesco et al., 2022, p. 889)

Indeed, as displayed, e.g., in Figure 2 of Vesco et al. (2022), a *no-change* model systematically predicting $\hat{y}_{t,s} = 0$ achieved the best average TADDA scores of all considered models. This led to the empirical conclusion that "the TADDA score appears to overly favour the simple no-change model" (Vesco et al., 2022, p. 892). We provide a statistical explanation for why a no-change forecast will often achieve better TADDA scores than more sophisticated forecasts, especially if these are designed to minimise the mean squared error. Moreover, we demonstrate empirically that no-change models can be beaten under TADDA by tailoring forecasts to this metric.

Our findings highlight a general principle in forecast evaluation: different scores may favour forecasts with different properties, thus creating incompatible incentives for forecasters. Optimising for one score will not necessarily lead to good performance under another. Different scoring rules should thus only be used simultaneously if their incentives are aligned. Moreover, when choosing a primary scoring rule, e.g., for a forecasting competition, the resulting incentives need to be studied in detail. With respect to the TADDA score we find that it favours rather conservative forecasts close to zero, which in terms of their statistical and practical meaning are difficult to interpret. As we will detail in the discussion, a way to resolve these challenges is to collect and evaluate probabilistic rather than point forecasts.

Similar questions on how sets of rules will shape individuals' responses to a request also appear in various areas of political science and public choice. Voting systems should be constructed such that truthful reporting of preferences is encouraged (Zeckhauser, 1973; Taylor, 2002). Similarly, different auction designs will encourage different bidding schemes, which may be more or less desirable for the principal or seller (Klemperer, 1999, 2002).

The remainder of this chapter is structured as follows. In Section 4.2 we provide the definition of the TADDA score. In Section 4.3 we describe the statistical concept of the *optimal point forecast*, which enables us to study the incentives created by different scoring functions. The incentives implied by the TADDA score are discussed in Section 4.4. In Section 4.5 we provide an empirical illustration of our argument before concluding in Section 4.6 with a discussion of alternative scoring concepts.

4.2 Definition of the TADDA Score

We start by providing the definition of the TADDA score, slightly adapting notation to our purposes. Denoting the outcome of interest by y and the respective point forecast by $\hat{y}$, it is given by

$$\text{TADDA}_\epsilon(\hat{y}, y) = |\hat{y} - y| \quad + \quad a_\epsilon(\hat{y}, y). \tag{4.1}$$

Here, $a_\epsilon(\hat{y}, y)$ with $\epsilon > 0$ is a term reflecting *direction augmentation* and defined as

$$a_\epsilon(\hat{y}, y) = \begin{cases} \hat{y} - \epsilon & \text{if} \quad \hat{y} > \epsilon \text{ and } y < -\epsilon \\ -\epsilon - \hat{y} & \text{if} \quad \hat{y} < -\epsilon \text{ and } y > \epsilon \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

The intuition is that a penalty is applied whenever $\hat{y}$ and y are on opposite sides of a tolerance region $[-\epsilon, \epsilon]$, with ϵ chosen by the analyst. This reflects the idea of asymmetric costs of forecast errors, which are larger if the qualitative trend is predicted incorrectly. Vesco et al. (2022) use $\epsilon = 0.048$, which represents a relative change of 5% in observed fatalities. An illustration of the score with $\epsilon = 0.048$ is provided in Figure 4.1. We added the absolute error (AE) for comparison, which corresponds to the absence of a penalty term $a_\epsilon(\hat{y}, y)$ or a very large value of ϵ.

We note that the above specification of the score is denoted by TADDA1 in Vesco et al. (2022), and is only one of several possible versions. Specifically, the score can also be based on a quadratic rather than absolute distance, and an alternative handling of the tolerance region is implemented in a score called TADDA2. The direction augmentation

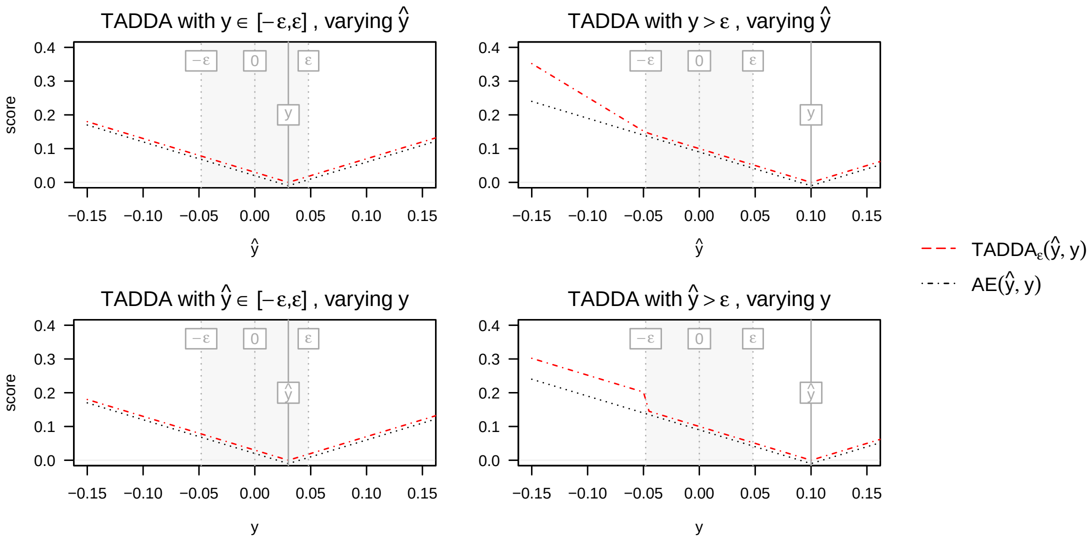

Figure 4.1: Illustration of the absolute error (AE) and TADDA$_\epsilon$ with $\epsilon = 0.048$

Note: Top row: as a function of $\hat{y}$ with fixed observation y. Bottom row: as a function of y with fixed prediction $\hat{y}$. In each row we show one example where $\hat{y}$ or y, respectively, is inside $[-\epsilon, \epsilon]$ and one where it is outside. The lines for the two scores are slightly shifted to avoid over-plotting.

term is then defined as

$$a_\epsilon(\hat{y}, y) = \begin{cases} |\hat{y} - \epsilon| & \text{if } \{\hat{y} \leq \epsilon \text{ and } y > \epsilon\} \text{ or } \{\hat{y} > \epsilon \text{ and } y \in [-\epsilon, \epsilon]\} \\ |-\epsilon - \hat{y}| & \text{if } \{\hat{y} \geq -\epsilon \text{ and } y < -\epsilon\} \text{ or } \{\hat{y} < -\epsilon \text{ and } y \in [-\epsilon, \epsilon]\} \\ 0 & \text{otherwise,} \end{cases}$$

meaning that penalties are applied whenever y and $\hat{y}$ fall on different sides of either $-\epsilon$ or ϵ. Penalties can thus also occur if $-\epsilon \leq \hat{y} \leq \epsilon$, see Figure C.3 for a graphical illustration. To keep our display in the main text concise, we will focus on the score from equations 4.1 and 4.2, which takes the more prominent role in Vesco et al. (2022). Central results on TADDA2 will be stated briefly, but we defer the details to Section E in the appendix, where we also cover the case of quadratic distances.

4.3 Scoring Functions and Incentives

Before we analyse the properties of the TADDA score, we introduce some helpful theoretical notions. Following Gneiting (2011), we conceive the prediction task as a decision problem under uncertainty. The forecaster issues a prediction $\hat{y}$ for a real-valued random variable Y, with the realised value y yet unknown. The assessment of forecast quality is based on a *scoring function* s, which returns a real number based on $\hat{y}$ and y. We orient it such that lower values correspond to better forecasts, meaning that the forecaster chooses her prediction $\hat{y}$ with the aim of minimising her score. While the realised value y of Y is unknown at the time of prediction, we assume that the forecaster can describe her uncertainty about the future via a probability distribution F for Y. When asked for a point prediction, the forecaster then has an incentive to issue the value $\hat{y}_{\mathrm{OPF}}$ which under her predictive distribution F yields the lowest expected score, i.e.,

$$\hat{y}_{\mathrm{OPF}} = \mathrm{argmin}_{\hat{y}\in\mathbb{R}} \mathbb{E}_F[s(\hat{y}, Y)].$$

In the statistical literature this choice of point forecast is often called the *Bayes act* (Gneiting, 2011), but we will simply refer to it as the *optimal point forecast* (OPF). Whenever a certain characteristic, or *functional*, of the predictive distribution F is the OPF under a specific scoring function, we say that the functional is *elicited* by the scoring function. Vice versa, whenever a functional is the OPF under a given scoring function, the scoring function is said to be *strictly consistent* for this functional.

In two well-known cases, the OPF corresponds to measures of central tendency of the predictive distribution F. For the squared error (SE) $s(\hat{y}, y) = (y - \hat{y})^2$, the OPF is given by the mean of the distribution F. For the absolute error (AE) $s(\hat{y}, y) = |y - \hat{y}|$, the OPF corresponds to the median of F. Hence, in the case of a skewed predictive distribution F, the squared and absolute errors usually imply different optimal point forecasts.

There is thus a duality between the scoring function and the functional of the forecaster's predictive distribution which shall be elicited. For instance, a forecaster has an incentive to report a predictive mean when evaluated with the squared error. Conversely, for an evaluator it makes sense to apply the squared error when she knows that predictive means have been reported. By contrast, applying the absolute error for evaluation is incoherent in this situation: Had the forecaster known that the absolute error was applied, she would have reported the median rather than the mean of her predictive distribution. The scoring function a forecaster uses as an optimisation criterion will thus directly impact the characteristics of her forecast, even if the forecaster is not explicitly aware of which functional corresponds to the OPF. As pointed out by (Gneiting, 2011) and (Kolassa, 2020), the definition of a forecasting task should therefore state either (i) one well-chosen scoring function which forecasters should aim to optimise or (ii) the requested functional of forecasters' predictive distributions, which can then be evaluated using one or several strictly consistent scoring functions.

Simulation Example

As an illustrative example we set F to a skew-normal distribution (Azzalini, 2013, Chapter 2) with parameters $\xi = -0.15$, $\omega = 0.4$ and $\alpha = 8$, see Figure 4.2 for the probability density function. As the distribution is skewed, its mean μ and median m differ. They are given by $\mu = 0.167$ and $m = 0.120$, respectively. To compute the expected scores of different point forecasts $\hat{y}$ under F and various scoring functions, we draw $N = 10^6$ independent realisations $y_1, \ldots, y_N$ from F. Using

$$\mathbb{E}_F[s(\hat{y}, Y)] \approx \frac{1}{N} \sum_{i=1}^{N} s(\hat{y}, y_i)$$

for large N, we can then evaluate the respective expected scores in very close approximation. We find that under F, the expected absolute error when reporting the mean is 0.195. When reporting the median instead, it can be lowered to 0.192. Conversely, the expected squared error under F is 0.060 when reporting the mean, but 0.062 for the median. As implied by theory, the predictive median is thus the better choice under the absolute error, and the predictive mean under the squared error (even though the magnitude of the differences may not be impressive).

4.4 Incentives Created by TADDA

A natural question is now what the optimal point forecast under the TADDA score is, i.e., which functional of a forecaster's predictive distribution F it elicits. It turns out that neither the predictive median nor the predictive mean are optimal under TADDA, and both may even be worse choices than a simple zero forecast.

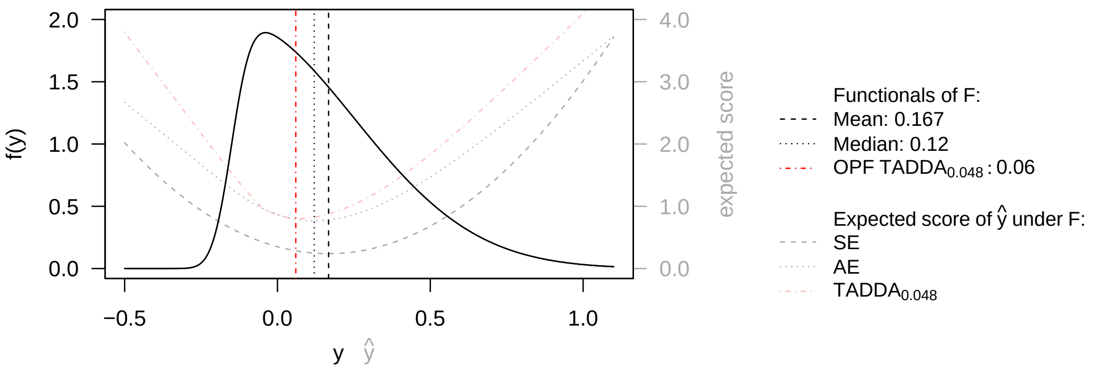

Figure 4.2: Probability density function of a skew-normal distribution with location, scale and shape parameters $\xi = -0.15$, $\omega = 0.4$ and $\alpha = 8$

Note: The mean and median are marked by vertical dashed and dotted lines, respectively (both in black). The red vertical line shows the optimal point forecast under $\text{TADDA}_{0.048}$, see Section 4.4. The light grey and light red curves show the expected values of the SE, AE, and $\text{TADDA}_{0.048}$ as a function of the reported point prediction $\hat{y}$ (see right axis). As implied by theory, the minimum expected scores are achieved by the respective OPFs (mean for SE, median for AE, and OPF $\text{TADDA}_{0.048}$ for $\text{TADDA}_{0.048}$).

The OPF for $\epsilon > 0$ is instead given by

$$\hat{y}_{\text{OPF}} = \begin{cases} F^{-1}\{0.5 \times (1 + \pi_+)\} & \text{if } \pi_- \geq 0.5 \times (1 + \pi_+) & \text{high confidence that } Y < -\epsilon \\ -\epsilon & \text{if } 0.5 < \pi_- < 0.5 \times (1 + \pi_+) & \text{low confidence that } Y < -\epsilon \\ F^{-1}(0.5) = m & \text{if } \pi_- \leq 0.5 \text{ and } \pi_+ \leq 0.5 & \text{unsure about sign of } Y \\ \epsilon & \text{if } 0.5 < \pi_+ \leq 0.5 \times (1 + \pi_-) & \text{low confidence that } Y > \epsilon \\ F^{-1}\{0.5 \times (1 - \pi_-)\} & \text{if } \pi_+ > 0.5 \times (1 + \pi_-) & \text{high confidence that } Y > \epsilon \end{cases} \tag{4.3}$$

Here, we denote by π_- and π_+ the predictive probability that Y is below $-\epsilon$ and above ϵ, respectively, and F^{-1} is the predictive quantile function. The optimal point forecast thus depends on how confident the forecaster is about the sign of Y. If she is unsure with both π_- and π_+ at most 0.5, the optimal point forecast is simply the predictive median m, which in this case is from $[-\epsilon, \epsilon]$. If she is more than 50% confident that $Y > \epsilon$, i.e., $\pi_+ > 0.5$, but no more than a threshold of $0.5 \times (1 + \pi_-)$, the optimal point forecast equals the tolerance value ϵ. And only if this threshold, which is somewhere between 0.5 and 2/3 depending on π_-, is exceeded, is the optimal point forecast larger than ϵ. The forecaster then accepts the risk of a penalty from direction augmentation. However, the optimal point forecast is usually shifted towards ϵ relative to m as it consists of the $0.5 \times (1 - \pi_{-})$ rather than the 0.5 quantile of F. The predictive median is thus generally not an optimal choice in this setting (and not elicited by TADDA). All of this reflects that the forecaster tries to avoid strong penalties from the direction augmentation term. For the case $\pi_{-} > 0.5$, the previous description translates via symmetry. We note that the OPF for TADDA$_\epsilon$ can be computed directly as long as we can evaluate the cumulative distribution function and quantile function of F; generally, no simulation or approximation is needed. Proofs can be found in Section E in the appendix.

For the variation TADDA2$_\epsilon$ of the score, the OPF is provided in Section C.3. While the resulting distinction of cases is even more involved than in equation 4.3, the qualitative behaviour remains similar. If the predictive median m is not contained in $[-\epsilon, \epsilon]$, relative to m the OPF is again shifted towards this interval, and may coincide with one of its ends. A difference is that even if $-\epsilon < m < \epsilon$, the OPF is generally not equal to m and may equal $\pm\epsilon$. We moreover provide the OPF for the TADDA score with quadratic rather than absolute distance in Section C.2 in the appendix. In this case, the OPF is shifted towards the interval $[-\epsilon, \epsilon]$ relative to the predictive mean, but generally does not coincide with one of its ends.

Simulation Example *(Continued)*

We return to the example from the previous section to illustrate our results. For TADDA$_{0.048}$, the optimal point forecast is approximately 0.06. The forecaster is thus sufficiently confident that $Y > 0.048$ to issue a forecast outside of the tolerance region (specifically, she is 62% sure that $Y > \epsilon$). Compared to the predictive mean (0.167) and median (0.120), however, she stays quite close to ϵ. To contrast TADDA and the commonly used absolute and squared errors, Table 4.2 summarises the respective optimal point forecasts (Panel a) and their expected scores under the different metrics (Panel b). We moreover added a forecast of zero, corresponding to the no-change model from Vesco et al. (2022). As implied by theory, the lowest expectation for each score is achieved by the respective optimal point forecast. A central observation is that the zero forecast is a better choice than the predictive mean under TADDA. As most submitted forecasting models were optimised with respect to MSE (Hegre et al. 2022, p. 529) this may explain why none of them was able to outperform the no-change forecast on TADDA. We note that, while not an optimal choice, the predictive median outperforms the zero forecast under TADDA$_{0.048}$. This is plausible as TADDA$_{0.048}$, like the absolute error, is based on the absolute rather than squared distance. In this table we also added results for $\text{TADDA2}_{0.048}$, the variation of $\text{TADDA}_{0.048}$ discussed in more detail in Appendix C.3. The finding that a zero forecast outperforms the predictive mean also holds for this score.

Table 4.1: Expected scores for different combinations of reported functional and scoring function under a skew-normal distribution F with $\xi = -0.15$, $\omega = 0.4$ and $\alpha = 8$ (Figure 4.2)

(a) Optimal point forecasts under different scoring rules

	AE	SE	$\text{TADDA}_{0.048}$	$\text{TADDA2}_{0.048}$
Functional	Median	Mean	See Equation 4.3	See Appendix C.3
Value	0.120	0.167	0.060	0.048

(b) Expected scores for different functionals of the distribution F

Functional	AE	SE	$\text{TADDA}_{0.048}$	$\text{TADDA2}_{0.048}$
Median	**0.192**	0.062	0.207	0.239
Mean	0.195	**0.060**	0.219	0.260
OPF $\text{TADDA}_{0.048}$	0.198	0.071	**0.200**	0.222
OPF $\text{TADDA2}_{0.048}$	0.201	0.074	0.201	**0.220**
Zero	0.216	0.088	0.216	0.256

Note: The best (lowest) expected score for each scoring rule is highlighted in bold. Note that in this table we also added results for $\text{TADDA2}_{0.048}$, a variation of $\text{TADDA}_{0.048}$ discussed in detail in Section C.3 in the appendix.

4.5 Empirical Example

As mentioned in the introduction, the 2020 VIEWS Prediction Competition comprised one true and one retrospective prediction task. Participants were encouraged to use novel data sources, and the organisers provided access to the VIEWS database containing more than 400 variables (Hegre et al., 2022). Several participants exploited additional data sources such as Wikipedia (Oswald and Ohrenhofer, 2022) and online newspapers (Mueller and Rauh, 2022b). They employed a broad range of modelling techniques, such as recurrent neural networks, Markov models, hierarchical hurdle models and gradient boosting, see Table 1 in Vesco et al. (2022). In terms of $\text{TADDA}_{0.048}$, however, none of the models achieved an improvement over the no-change forecast for either of the two prediction periods (compare Footnote 3 in Vesco et al., 2022; we note that in terms of the alternative version $\text{TADDA2}_{0.048}$, some models did).

To illustrate our point we generate a simple probabilistic forecast based on recent observations, from which we derive two sets of point forecasts optimised for SE and

TADDA, respectively. Recall that, for a lead time of s months, the no-change forecast predicts that the number of fatalities X_t in month t is identical to the last known value x_{t-s}. This implies a prediction of $\hat{y}_{t,s} = 0$ for the log-change. For our probabilistic forecast, we extend this principle to the last w observations $\{x_{t-s-i},\ i = 0, \ldots, w-1\}$ and compute the log-changes they would imply relative to the last value x_{t-s}. We denote these by

$$y^*_{t,s,i} = \log(x_{t-s-i} + 1) - \log(x_{t-s} + 1), \quad i = 0, \ldots, w-1. \tag{4.4}$$

We then predict that $Y_{t,s}$ takes on each of the values $y^*_{t,s,i}, i = 1, \ldots, w$ with equal probability $1/w$, i.e., our naïve predictive distribution F is a discrete uniform distribution over the log-changes the w most recent observations would imply. From this distribution, we obtain the optimal point forecasts with respect to the SE and $\text{TADDA}_{0.048}$, i.e., the predictive means and the functionals from equation 4.3. To evaluate quantiles we use the `quantile` function in R (R Core Team, 2021) with default settings (noting that results are robust to the choice of the `type` argument).

An optimal choice for w can be determined by generating forecasts under different values of w for the calibration period (January 2014 through December 2016) and comparing the resulting average scores across all African countries and lead times $s = 2, \ldots, 7$. For $\text{TADDA}_{0.048}$, this leads to a value of $w = 5$ months. Indeed, this choice also yields the best results for the evaluation period. However, as it corresponds to very coarse predictive distributions, we here use $w = 9$ for illustration. The $y^*_{t,s,i}$ can then be seen as the 0%, 12.5%, 25%, ..., 87.5%, 100% quantiles of the predictive distribution. Figure 4.3 illustrates how the predictive distribution for $Y_{t,s}$ arises from the $w = 9$ observations leading up to the time of prediction $t - s$.

Table 4.2 summarises the performance of both types of point forecasts derived from our model, along with the respective scores of the VIEWS ensemble and the no-change model from Vesco et al. (2022). The VIEWS ensemble leads the field in terms of MSE, in particular for longer lead times. The predictive means from our model outperform the corresponding $\text{TADDA}_{0.048}$ OPFs, while the no-change baseline shows the weakest performance. Regarding $\text{TADDA}_{0.048}$, the picture is quite different: while the respective optimal point forecasts from our model outperform the no-change model consistently across all horizons, the predictive means and the VIEWS ensemble yield worse average scores. These results demonstrate that each forecasting approach performs well under the score it was optimised for. The VIEWS ensemble arose from models which target

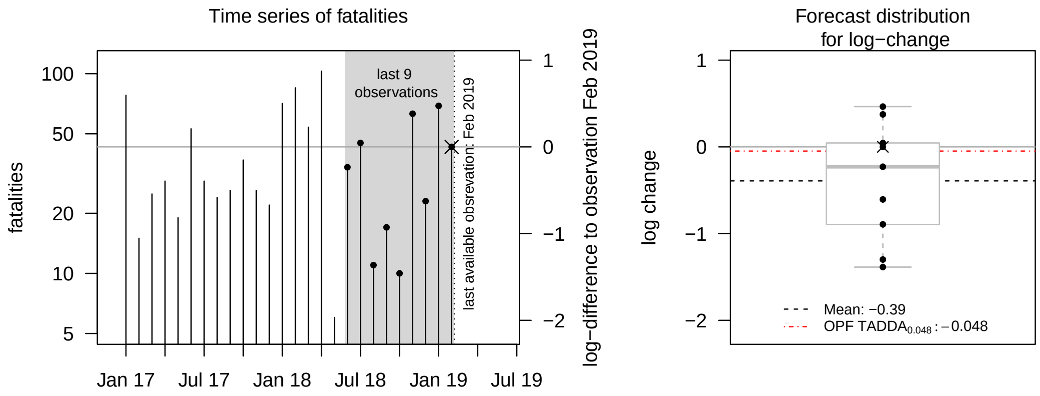

Figure 4.3: Construction of the forecast distribution for $Y_{t,s}$ at time $t-s$ (here: February 2019)
Note: The last $w = 9$ available monthly numbers of fatalities x_{t-s-i}, $i = 0, \ldots, 8$, are extracted (left panel) and for each of them the log-difference $y^*_{t,s,i}$ to the latest observation x_{t-s} (highlighted by a cross) is computed; see equation 4.4. Note that one of them is 0 by construction. The resulting values are used to define a nine-point predictive distribution for $Y_{t,s}$ (right panel). The example data used for the illustration are from Mali. From the left panel both absolute counts (left axis) and log-differences to the most recent observed value from February 2019 (right axis) can be read. The right axis is identical to the one used in the right panel, and dots are at corresponding heights.

MSE, and its weights were likewise determined based on MSE (Hegre et al., 2022). Our simple model is almost competitive with the ensemble when optimised for MSE by using predictive means. The TADDA$_{0.048}$ OPFs, on the other hand, yield only mediocre performance in terms of MSE. Yet, they are the only ones to outperform the no-change model under TADDA$_{0.048}$, i.e., the score they were optimised for.

The fact that the TADDA$_{0.048}$ OPFs from our model outperform the predictive means under TADDA$_{0.048}$, while the opposite is the case under the SE, is robust to the choice of window length $w > 1$. The TADDA$_{0.048}$ OPFs also outperform the no-change model for any window length $w = 2, \ldots, 8$. For $w = 3, \ldots, 7$, the predictive means from our model also achieve slight improvements over the no-change model, indicating that it is not impossible to outperform the no-change model under TADDA while optimising for MSE.

To improve our intuition for the behaviour of the compared forecasting approaches, we conclude by contrasting some of their characteristics. Table 4.3 describes the empirical distributions of the point forecasts and observed log-changes. The most notable pattern is that a majority of observed and predicted log-changes are zero. More precisely, this is the case for 72.0% of the observations, which is one reason why the no-change forecast is not straightforward to beat. Concerning the forecasts, 76.7% of the TADDA$_{0.048}$ OPFs

Table 4.2: Average evaluation scores for the different lead times s and window length $w = 9$ in the prediction period 01/2017–12/2019

	MSE				$\text{TADDA}_{0.048}$			
s	Pred. mean	OPF $\text{TADDA}_{0.048}$	VIEWS	No-change	Pred. mean	OPF $\text{TADDA}_{0.048}$	VIEWS	No-change
2	0.487	0.556	0.504	0.674	0.357	0.335	0.371	0.340
3	0.525	0.614	0.551	0.773	0.368	0.348	0.379	0.365
4	0.556	0.647	0.579	0.841	0.372	0.358	0.394	0.391
5	0.587	0.668	0.548	0.807	0.388	0.366	0.381	0.375
6	0.629	0.717	0.573	0.841	0.399	0.381	0.386	0.384
7	0.664	0.741	0.599	0.864	0.410	0.385	0.400	0.389
	0.575	0.657	0.559	0.800	0.382	0.362	0.385	0.374

Note: The table includes scores for the predictive means and optimal point forecasts for $\text{TADDA}_{0.048}$ from our model. For comparison we also included the no-change model and the VIEWS ensemble (Table 2 in Vesco et al. 2022). The bottom line contains the column means.

and 63.9% of the predictive means from our model are zero. The $\text{TADDA}_{0.048}$ OPFs are much less frequently outside the interval $[-0.048, 0.048]$ than the predictive means. This is beneficial for their TADDA scores, as they avoid the risk of penalties from direction augmentation. The no-change forecasts, too, benefit from this characteristic of the score.

Table 4.3: Empirical quantiles of the predictive means, $\text{TADDA}_{0.048}$ OPF and true log-changes across all lead times $s = 2, \ldots, 7$ in the prediction period 01/2017–12/2019

	5%	10%	15%	20%	$\cdots$	75%	80%	85%	90%	95%
Mean	–1.038	–0.510	–0.068	0	$\cdots$	0	0.033	0.256	0.462	0.877
OPF $\text{TADDA}_{0.048}$	–1.012	–0.211	0	0	$\cdots$	0	0	0	0	0.663
True log-changes	–1.386	–0.588	0	0	$\cdots$	0	0	0	0.693	1.447

4.6 Discussion

The theoretical and empirical results from the previous sections underscore the duality between scoring functions and optimal point forecasts. This has two main implications. Firstly, whenever a scoring rule is chosen, it should be assessed whether the functional of a forecaster's predictive distribution it elicits is of interest. As an example, expected (mean) costs may be relevant in financial contexts, and can be elicited via the squared error. Concerning the TADDA score, it is unclear whether the hard-to-interpret functional provided in equation 4.3 is of practical relevance. Given that the resulting point forecasts in absolute value rarely exceed the tolerance value ϵ, they may be considered overly conservative in the sense that they avoid strong statements about upcoming changes. This is echoed in the judgement of the independent scoring committee cited in Section 4.1. Secondly, while applying different evaluation scores in parallel can yield a more comprehensive picture of performance, it should be ensured that they all elicit the same functional (see also Gneiting, 2011 and Kolassa, 2020). Otherwise, forecasters lack a clear objective and need to make a choice on which score to prioritise. In the context of the VIEWS challenge, it seems that an important reason for the modest performance in terms of TADDA was that forecasts were optimised for MSE.

A natural question is whether other scores can be conceived which reward point forecasts for having the correct sign, but elicit standard and interpretable functionals of the forecasters' predictive distributions. General construction principles for scoring functions to elicit predictive means have been established in Ehm et al. (2016). It seems feasible to construct variations of TADDA which elicit predictive means or medians, but a detailed discussion is outside the scope of this note and will be provided elsewhere. Another possibility is to evaluate a model's point forecasts via the squared or absolute error, and complement this with an assessment of the model's predicted probability of a positive outcome. The latter could be assessed using the Brier score, a widely employed scoring rule for binary targets (Gneiting and Raftery, 2007).

Finally, a more general alternative to the evaluation of point forecasts as performed in the VIEWS challenge is to collect and score probabilistic forecasts. This way, no choice concerning an appropriate functional to summarise predictive distributions would be necessary. The potential of a probabilistic approach has already been evoked in the outlook of Vesco et al. (2022), and Brandt et al. (2022) provide some results based on the continuous ranked probability score (CRPS). There is a rich body of literature on methods for probabilistic forecast evaluation (Gneiting and Katzfuss, 2014), and these are widely used in challenge-based formats, e.g., in meteorological (Vitart et al., 2022), epidemiological (Bracher et al., 2021) and energy forecasting (Hong et al., 2016).

5 Contribution to the 2023/24 VIEWS Prediction Challenge

This chapter is a close adaptation of the documentation of our contribution to the VIEWS Prediction Challenge 2023/2024[1], co-authored with Tobias Bodentien (Bodentien and Rüter, 2024). We compare three approaches to probabilistic modelling of monthly conflict-related fatalities at the country level, namely a negative binomial specification, a hurdle model, and neural networks. Our submission was included in the competition's summary article published in the *Journal of Peace Research* (Hegre et al., 2025) and presented at the *2023/24 VIEWS Prediction Competition Workshop* at the German Federal Foreign Office (10/2023) and at the *CCEW Symposium* at the University of the Bundeswehr, Munich (09/2024). The challenge received financial support from the German Ministry for Foreign Affairs.

The code and data for this project are available at `https://github.com/toboden/ConflictPredProbabilistic`.

5.1 Introduction

The 2023/24 VIEWS Prediction Challenge invited teams to submit probabilistic forecasts of global monthly fatalities from organised political violence, using data from the Uppsala Conflict Data Program (UCDP; Davies et al., 2024), across a prospective period from July 2024 to June 2025 (Task 1) and for each calendar year 2018 to 2023 (Task 2). Participants provided forecasts represented by predictive samples on a country level or on a subnational spatial grid-based level, based on information available up to and including October of the previous year and thus with lead times ranging from three to fifteen months. Accordingly, even forecasts for past periods rely exclusively on information that

[1] `https://VIEWSforecasting.org/research/prediction-challenge-2023`, accessed on 29 November 2025.

Table 5.1: Characteristics of our modelling approaches and their flexibility

	Neg. Bin. Model	Hurdle Model	Neural Networks
Overdispersion	✓	✓	✓
Zero-Inflation		✓	✓
Spatio-Temporal Dependencies			✓
Flexibility	Low	Middle	High

would have been available in real time. The multi-stage challenge concluded in June 2025, with running evaluations released throughout and forecasts assessed primarily using the continuous ranked probability score (CRPS). More details on the challenge format are provided in Section 6.2.

Compared to other approaches that obtain a probabilistic forecast by adding uncertainty estimates to a given point forecast, we model the predictive distribution of country-month fatalities directly. For model selection purposes, we compare three approaches that differ in their levels of flexibility to incorporate specific characteristics inherent in the data, see Table 5.1 for an overview and Hegre et al. (2025) for more details on these data properties.

First, we utilise a negative binomial distribution (NB) to account for the overdispersion inherent in fatality data. Its parameters are estimated via empirical moments of each country's past w fatalities. By construction this approach is unable to predict conflict outbreak in previously peaceful regions. Second, we employ a hurdle model that additionally accounts for zero-inflation by modelling the distribution of zeros separately using a Bernoulli variable while positive numbers of fatalities are modelled via a truncated negative binomial distribution. Again, the respective model parameters are estimated based on past fatalities. Third, we flexibly incorporate additional feature variables provided by the VIEWS team using feed-forward neural networks to further include spatio-temporal dependencies. In all three cases, we tune the hyperparameters in such a way that the average CRPS is minimised. The same criterion is used for the ensuing model selection.

We find that the simple NB approach outperforms the two more involved alternatives in terms of the average CRPS across all countries and years for Task 2, as available at the beginning of the competition, which comprises monthly forecasts for 2018 to 2021 based on data up to and including October of the respective previous year. Since quantiles contain more information on the underlying estimated distribution than random samples as there is no stochasticity involved, we therefore submitted the $0.1, 0.2, \ldots, 99.9\%$-quantiles of the estimated NBs as our predictive samples. The resulting model is simple, straightforward, transparent, and easy to interpret. It naturally captures the conflict-trap dynamic, where countries with recent violence are more likely to experience further violence, yet is by construction unable to predict the outbreak of conflicts or identify trends that have not previously occurred.

At the time of submission of this thesis, the joint publication of all participants' model results by the VIEWS organisers is still pending.

This summary is structured as follows. We first review the characteristics of the data and the resulting challenges for the prediction task in Section 5.2. We then describe our three modelling approaches in Section 5.3, followed by a presentation of our results in Section 5.4. Section 5.5 concludes.

5.2 Data

We base our predictions solely on the data provided by the VIEWS team. At the country-month level, these data comprise monthly fatalities as well as observations of 123 additional feature variables for 191 countries, 93 of which have zero recorded historic fatalities as of October 2022. The prediction target, that is, the number of fatalities due to state-based conflict, exhibits certain characteristics that require appropriate modelling strategies. First, the data are count data and hence integer-valued. Second, they are overdispersed meaning that the variance in the data is often higher than expected by a simple model, for instance, a Poisson distribution. Third, due to many non-conflict countries and peaceful months, the data are (fortunately) zero-inflated. The additional feature variables, comprising lagged, spatial conflict data, aspects of democracy and development indicators, amongst others, are divided into three main categories *conflicthistory*, *vdem* and *wdi*[2].

For the visualisation of the results, we divide all 98 countries with at least one month of non-zero reported fatalities into three conflict levels depending on the average number of fatalities $\overline{Y}_i$ of country i between January 1990 and October 2022, the training period for Task 2 (which partially overlaps with the period to be predicted):

[2]See Section 4.1.1 in https://www.diva-portal.org/smash/get/diva2:1667048/FULLTEXT01.pdf for further details. A complete list of all features is available at https://viewsforecasting.org/wp-content/uploads/cm_features_competition.pdf. Both documents were retrieved on 29 November 2025.

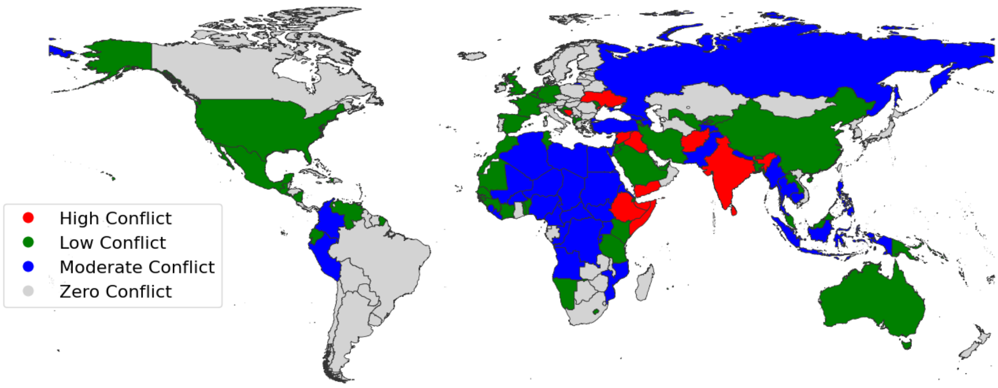

Figure 5.1: Categorisation of countries into zero, low, moderate, and high levels of conflict according to the definition in Section 5.2

1. **Low conflict level** ($0 < \overline{Y}_i \leq 5$), 50 countries

2. **Moderate conflict level** ($5 < \overline{Y}_i \leq 100$), 38 countries

3. **High conflict level** ($\overline{Y}_i > 100$), 10 countries

Figure 5.1 yields an illustration of the different categories.

5.3 Modelling Approaches

We model the predictive distribution of the s-step-ahead ($s = 3, ..., 14$) number of monthly fatalities $Y_{i,t+s}$ of country $i = 1, ..., 191$ issued at time $t \in \{10/2017, 10/2018, ..., 10/2022\}$ using three different modelling approaches. Their underlying concepts are outlined below, while technical details on the modelling and estimation procedures are provided in Appendix D.1.

5.3.1 Negative Binomial Distribution

The negative binomial distribution has been deployed in multiple research fields to model overdispersed count data, for example, in sociology, epidemiology, and ecology (Moghimbeigi et al., 2008; Ver Hoef and Boveng, 2007). We estimate its parameters from the past w observations as described in Appendix D.1.1.

To determine the optimal hyperparameter w^*, we minimise the average CRPS over $w \in \{2, ..., 24\}$ across all countries i and months t contained in the training data of

Table 5.2: Four approaches to selecting the optimal window length w^*

Variant	Minimisation Target	Optimal Window
1	Average CRPS across all countries and lead times	w_1^*
2	Average CRPS for each country i	w_{2i}^* $(i = 1, \dots, 191)$
3	Average CRPS for each lead time s	w_{3s}^* $(s = 3, \dots, 14)$
4	Average CRPS for each country i and lead time s	w_{4is}^* $(i = 1, \dots, 191;\ s = 3, \dots, 14)$

Task 2. For a given w, the CRPS for observation $y_{i,t+s}$ is computed based on the 0.1%, 0.2%,...,99.9%-quantiles of the respective negative binomial distribution. As shown in Table 5.2, we evaluate four alternative approaches to determine the optimal w, selecting the best-performing option for our model configuration.

5.3.2 Hurdle Model

As an extension of the NB approach, we estimate a hurdle model, first introduced by Cragg (1971), to explicitly account for excess zeros in the data. In contrast to the NB approach, it includes an additional Bernoulli component that separately models the probability of observing any fatalities in a given country, while positive counts are handled by a truncated negative binomial distribution (TNB). Technical details are provided in Appendix D.1.2, and the optimal window length w^* is determined in the same way as for the NB approach, see Section 5.3.1 and Table 5.2 for reference.

5.3.3 Neural Networks

To assess the value of incorporating additional feature data through a flexible, data-driven framework, we compare the two aforementioned approaches with feed-forward neural networks.

We model predictive distributions of fatalities by training separate neural networks for each country $i = 1, ..., 191$ and each lead time $s = 3, ..., 14$. As in the other models, we use the number of conflict deaths in the previous w months as inputs and additionally include a feature set f_{set}. To avoid extensive computations and overfitting, the output layer contains $n_{\text{output}} = 200$ rather than $1,000$ neurons, each corresponding to one draw from the predictive distribution. To ensure non-negative, integer-valued predictions, outputs are ReLU-transformed and rounded to integers. For CRPS optimisation, we use the energy form of the metric as the loss function (Gneiting and Raftery, 2007, p. 367).

The model's hyperparameters are the number of previous months used for the prediction w, the number of hidden layers h, the total number of hidden-layer neurons N_h, the learning rate l, the batch size b, the dropout rate d, the number of epochs e, and the selected feature subset f_{set}. The set of hyperparameters $(w, h, N_h, l, b, d, e, f_{\text{set}})$ is jointly determined for each country via a random search of 20 trials (Bergstra and Bengio, 2012). Hyperparameter tuning is performed only for countries with non-zero fatalities; for countries without any reported fatalities, we issue predictive samples consisting solely of zeros. In all remaining cases, one neural network per country and lead time is trained on the Task 2 training data using the selected hyperparameters.

For each subtask in Task 2, we split the data into 70% training and 30% validation sets, following standard practice (Joseph, 2022, p. 531), with the most recent observations reserved for validation. Further details on the network architecture and tuning procedure are provided in Appendix D.1.3.

5.4 Results

5.4.1 Optimal Window Length for the NB and the Hurdle Model

The distinction between the NB distribution and the hurdle model as well as the four variants of determining w^*, see Table 5.2, results in eight different models. Averaging over all countries and all six test windows 2018–2023 yields the CRPS values presented in Table 5.3. With respect to w, we find that for countries with non-zero conflicts, larger values tend to be optimal in all model variants. For countries with no reported conflicts, the choice of w does not affect the results; to maintain consistency in the implementation, we choose the smallest w, namely $w = 2$.

We find that the more elaborate hurdle model is not able to outperform the NB model in any of the four variants when evaluated by the average CRPS. For the NB approach, optimising w for each lead time $s = 3, ..., 14$ yields the best results, whereas additional discrimination by country offers no further gains. As NB Variant 3, NB Variant 1, and Hurdle Variant 1 yield very similar CRPS values, this suggests that the gains from optimising the window length by lead time are modest and that all three specifications provide a comparable level of predictive accuracy. In the following, we report results for the two best-performing specifications, namely NB Variant 3 and Hurdle Variant 1.

Table 5.3: Average CRPS across the eight NB or hurdle model variants over the six test windows (2018–2023)

	NB		Hurdle Model	
Variant	Average CRPS	$(\bar{w}^*,\, sd_{w^*})$	Average CRPS	$(\bar{w}^*,\, sd_{w^*})$
1 with w_1^*	56.283	(16.50, 8.40)	**56.662**	(16.50, 8.40)
2 with w_{2i}^*	77.112	(5.93, 7.83)	77.470	(5.90, 7.67)
3 with w_{3s}^*	**56.110**	(18.62, 7.92)	58.650	(17.85, 8.55)
4 with w_{4is}^*	69.246	(5.52, 7.36)	74.685	(5.32, 6.98)

Note: For each variant, we report the mean CRPS as well as the empirical mean $\bar{w}^*$ and standard deviation sd_{w^*} of the optimal window lengths obtained across 191 countries and the six test periods. The lowest CRPS values in each model class are highlighted in bold. Optimising over the data available at the start of the competition, that is, 2018–2021, yields the same optimal model specifications.

5.4.2 Model Performance

While the CRPS was communicated as the main evaluation metric, the Ignorance Score (IGN), also known as the Log Score, and the Mean Interval Score (MIS) were specified as secondary metrics; see the invitation to the prediction competition (Hegre et al., 2023). We therefore report their values here as well, although they did not influence the optimisation procedure. The results of our models are presented in Table 5.4, alongside those of two VIEWS benchmarks: the Conflictology benchmark (`bm_conflictology_country12`), which is the empirical distribution of the past twelve observed values, and the Last Poisson benchmark (`bm_last_historical`), which generates forecasts from a Poisson distribution using the most recent observation as its parameter (Hegre et al., 2025).

We find that all three of our models outperform the VIEWS Last Poisson benchmark on average across all years and metrics. However, the Conflictology benchmark remains the best-performing model overall in terms of CRPS (49.36) and MIS (873.53). NB Variant 3 performs only slightly worse in terms of overall average CRPS (56.11) and achieves the best IGN (0.61), both for individual years and in the aggregate. It also attains the lowest MIS values in two of the six years. By contrast, the more advanced Hurdle Variant 1 is not optimal with respect to either CRPS or IGN in any evaluation year.

The most complex approach, the neural networks (NNs), although yielding the lowest overall average CRPS among our three models (52.72), fail to outperform the Conflictology benchmark (49.36). Aside from the CRPS in 2022 and 2023 and the MIS in 2023, all

Table 5.4: Average evaluation metrics for our models and the two VIEWS benchmarks across all countries and all forecast periods of Task 2

Model	**Metric**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2018–23**
VIEWS Conflictology	CRPS	14.483	**9.146**	21.339	76.850	123.995	**50.357**	**49.362**
	IGN	0.640	0.610	0.567	0.686	0.695	0.682	0.647
	MIS	186.554	**89.058**	344.964	1,435.555	2,142.128	1,042.916	**873.529**
VIEWS Last Poisson	CRPS	20.174	9.480	23.698	85.606	131.017	678.960	158.157
	IGN	1.198	1.046	1.110	1.228	1.124	1.125	1.139
	MIS	380.623	172.686	455.806	1,690.711	2,599.278	13,523.463	3,137.095
NB Variant 3	CRPS	**14.084**	11.129	**20.640**	**76.577**	125.616	88.614	56.110
	IGN	**0.598**	**0.558**	**0.545**	**0.634**	**0.679**	**0.670**	**0.614**
	MIS	**145.102**	113.754	**317.478**	1,405.207	2,177.885	1,234.818	899.040
Hurdle Variant 1	CRPS	14.360	11.126	20.682	77.012	128.842	87.951	56.662
	IGN	0.752	0.654	0.651	0.728	0.805	0.783	0.723
	MIS	146.887	112.319	318.610	**1,398.950**	2,309.887	1,245.483	922.023
Neural Network	CRPS	16.529	16.236	25.293	81.838	**121.242**	55.157	52.716
	IGN	1.072	1.061	1.072	1.147	1.157	1.178	1.115
	MIS	227.523	229.862	406.504	1,548.841	2,370.910	**1,038.628**	970.378

Note: The rightmost column reports the average across the six forecast years.

other metrics are consistently higher for the NNs than for NB Variant 3 and Hurdle Variant 1, both by year and across forecasting horizons. This pattern is mirrored in the monthly CRPS values for 2018 to 2021 across the conflict levels introduced in Section 5.2, see Figure 5.2. Note the difference in scale between conflict levels, with high-intensity regions contributing disproportionately to the overall CRPS. This point is discussed in more detail in Chapter 6. These results therefore provide no evidence that the NNs systematically outperform the other approaches in the majority of cases.

At the same time, the country-level results shown in Figure 5.3 clearly indicate that there is no single best-performing model. The best performing model specification varies considerably across years and countries, with no approach consistently dominating either over time or geographically. This reflects both the limited availability of data and the inherent complexity of conflict dynamics, which are often difficult to capture through observable variables (e.g., Bazzi et al., 2022). The CRPS patterns are driven by temporal fluctuations in conflict intensity, which affect all models similarly (Figure 5.2).

Since the data for 2022 and 2023 was made available only shortly before the submission deadline of Task 1, our model selection was based on the information available at the start of the prediction competition, that is, the years 2018 to 2021. In this period, the NB Variant 3 outperformed our other approaches. Accordingly, we submitted probabilistic forecasts for 2024/25 generated by NB Variant 3. Other advantages of the NB approach

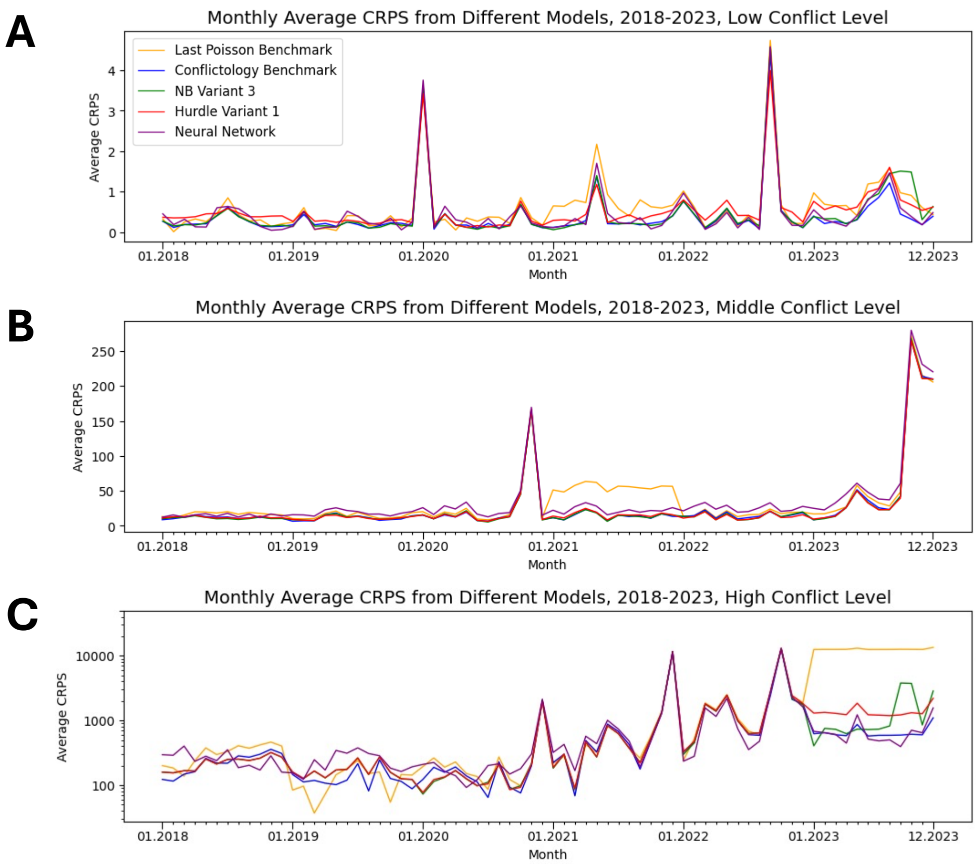

Figure 5.2: Average monthly CRPS across all models for Task 2, disaggregated by conflict level as defined in Section 5.2

Note: A: low conflict countries; B: moderate conflict countries; C: high conflict countries.

are its straightforward interpretability and superior performance relative to the IGN. In Hegre et al. (2025), it is referred to as `bodentien_rueter_negbin`.

5.5 Conclusion

We model the predictive distribution of monthly fatalities at the country level using a negative binomial distribution, a hurdle model that combines a truncated negative binomial distribution with a Bernoulli variable to account for excess zeros, and feed-forward neural networks. The first two models are estimated solely from past observations of fatalities in the country for which predictions are generated. The neural networks incorporate additional feature data provided by the VIEWS organisers, including spatial

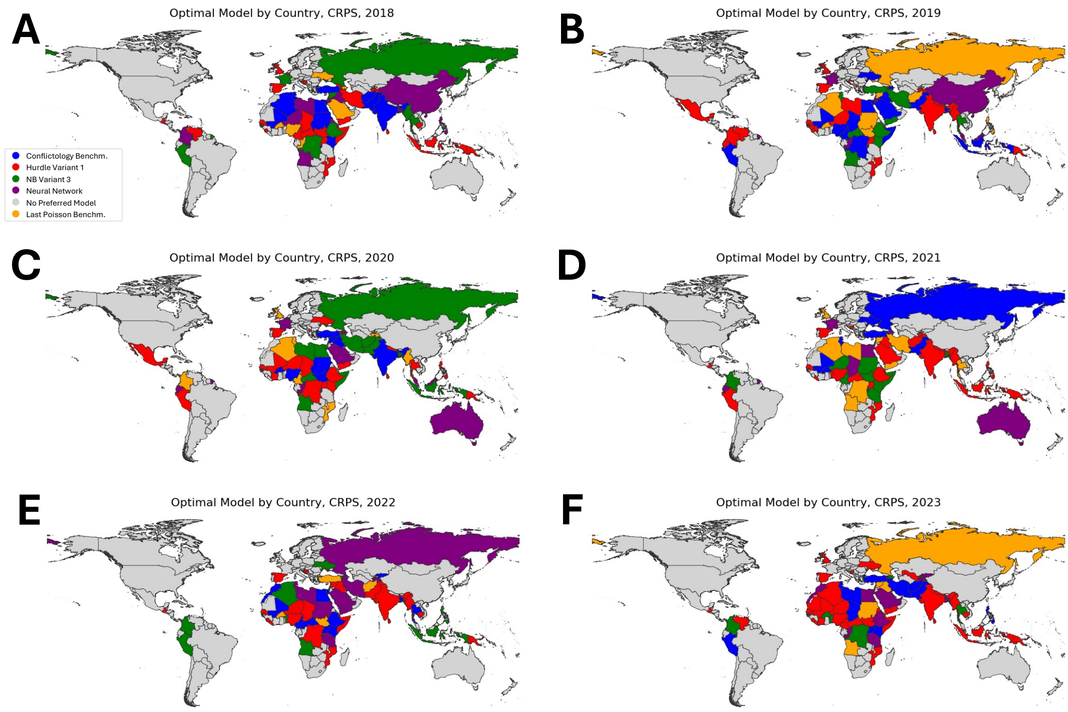

Figure 5.3: Optimal model per country based on the average CRPS for each year within Task 2

Note: Panels A–F correspond to the years 2018–2023, respectively. *No preferred model* denotes countries with zero predicted and observed fatalities, for which all model predictions coincide.

and temporal lags of fatalities as well as economic and development indicators, among others. In total, we thereby account for overdispersion, zero-inflation, and spatio-temporal dependencies in the conflict data. We optimise for the CRPS and find that none of our models is able to outperform the VIEWS Conflictology benchmark in terms of average CRPS across all test years. We select the optimal model based on the data available at the start of the competition, that is, the years 2018 to 2021, and thus issue predictions for the true future with the NB model. This model offers additional advantages: it is straightforward, simple, and transparent, and it outperforms the NNs in terms of the secondary metrics IGN and MIS. Nonetheless, the NNs perform best in terms of the average CRPS for the years 2018 to 2023 and therefore represent a promising approach that could be further improved through various enhancements, such as integrating additional data from alternative sources.

6 Challenges in Evaluating Conflict Fatality Forecasts from an Onset Perspective

This chapter documents joint work with Tobias Bodentien, Johannes Bracher, and Melanie Schienle, presented at the *Symposium on Crisis Early Warning*, German Federal Foreign Office (09/2025) and as an invited talk at the *2023/24 VIEWS Prediction Competition Closing Event* (online, 12/2025). It introduces a streamlined framework to evaluate how well probabilistic fatality models capture conflict onset and applies it to submissions from the 2023/24 VIEWS Prediction Challenge. The results show that although many models distinguish higher from lower onset risk, calibration remains mixed, pointing to the need for approaches that more directly target onset prediction.

The data used in this project were retrieved from a Dropbox folder https://tinyurl.com/VIEWSdata as referenced in the official GitHub repository of the VIEWS Prediction Competition: https://github.com/prio-data/prediction_competition_2023, both accessed on 12 December 2025. The code of our analysis is available at https://github.com/KITmetricslab/conflict_onset.

6.1 Introduction

Conflict forecasting has developed around several interrelated but distinct policy purposes, and useful prediction systems should be informative on multiple aspects simultaneously:

(i) anticipating the outbreak of conflict in previously peaceful areas;

(ii) detecting shifting trends in ongoing conflicts, i.e., escalation and de-escalation;

(iii) predicting the magnitude of conflict in terms of fatalities from state-based violence.

Each of these perspectives is linked to specific types of decisions. Forecasts of conflict outbreak are crucial for early-warning systems, peacekeeping, and prevention strategies aimed at avoiding the 'conflict trap' (Collier and Sambanis, 2002; Mueller and Rauh, 2022a), in which violence tends to persist once begun. Forecasts of changes in conflict intensity matter for the allocation of de-escalation efforts and humanitarian relief, while accurate estimates of conflict magnitude help guide the distribution of resources such as medical support and emergency aid (Cederman and Weidmann, 2017; Hegre et al., 2017; Bazzi et al., 2022). Similar distinctions between occurrence, relative change and magnitude also arise in other fields such as infectious disease (Bosse et al., 2023) and earthquake forecasting (Brehmer et al., 2024).

Conceptually, probabilistic forecasts of conflict magnitude are the most general, as they imply statements on onset events and (de)escalation of ongoing conflicts. In this chapter, however, we argue that despite this nesting, the evaluation of conflict forecasts is enhanced by addressing the different tasks explicitly. Each of them highlights different aspects of predictive performance, and different types of models may be best equipped to perform well. For instance, the aggregated predictive performance a model achieves in terms of conflict magnitude is often dominated by how well it anticipates the further course of high-intensity conflicts. This may obscure how well conflict onsets are anticipated, which initially are often limited to lower numbers of fatalities. Evaluations of binary conflict onset predictions, on the other hand, are obviously blind to the magnitude of conflicts. Predictions of (de)escalation in terms of relative growth or decline in fatalities take an intermediate role, where conflicts of low and high intensity receive more equal weight in the evaluation.

In this chapter we lay out a coherent evaluation framework to compare prediction models in terms of the three aforementioned tasks. It is based on the Brier score (BS) and the continuous ranked probability score (CRPS), which are widely used performance metrics for binary and count-valued settings, respectively (Gneiting and Raftery, 2007). The CRPS, arising by aggregation of BS values achieved for binarised predictions at varying cut-offs (see Section 6.3.1), with different threshold weighting schemes (Gneiting, 2011), can be defined to tailor the evaluation to tasks (i)–(iii). For onset predictions, we moreover discuss state-of-the-art techniques (Dimitriadis et al., 2024) to assess how overall performance arises from forecast discrimination (the ability to distinguish high-risk from low-risk situations) and calibration (whether predicted probabilities match observed frequencies). The latter is rarely considered in the conflict forecasting literature, but turns out to be a particularly challenging task.

We illustrate our approach using forecasts from the 2023/24 VIEWS Prediction Challenge (Hegre et al., 2025), which is the latest in a series of competitive formats led by teams at the Universities of Oslo and Uppsala. The VIEWS challenges have had a lasting impact on the field of conflict forecasting by leveraging contributions from research teams across numerous institutions and fields. Reflecting the ambition to provide policymakers with increasingly rich information, the history of VIEWS shows a progressive broadening of prediction targets, paralleling our outline of tasks (i)–(iii) above. Early versions focused on binary forecasts of whether a given geography would experience conflict in a given time period (Hegre et al., 2019, 2021), thus emphasising onset prediction. This was followed by models predicting (de)escalation, i.e., relative changes in conflict intensity, expressed as point forecasts of log-changes in fatalities (Hegre et al., 2022; Vesco et al., 2022). Ultimately, the 2023/24 edition (Hegre et al., 2025) moved to full probabilistic forecasts of fatality counts, the CRPS-based evaluation thus focusing on conflict magnitude. In this chapter we demonstrate how, thanks to the general format in which forecasts were collected in this latest effort, this approach can be complemented by tailored evaluations covering conflict onset and (de)escalation, as considered in earlier editions of the challenge. In the present version of this work, we focus exclusively on the empirical evaluation of onset predictions, outlining the extension of the framework to (de)escalation from a theoretical perspective while leaving its empirical implementation to future versions.

Our empirical results confirm the theoretical expectation that onset prediction plays a minor role in CRPS-based evaluation of conflict magnitude. Indeed, there are non-negligible differences in the model rankings for the two targets. We attempt to link this to whether the different approaches handle conflict onset explicitly via a hurdle-like modelling strategy, but the results remain inconclusive. Delving deeper into the performance of binary onset forecasts, we find that while even rather simple models have good discrimination ability, some approaches struggle to achieve good calibration.

The remainder of the chapter proceeds as follows. Section 6.2 introduces the 2023/24 VIEWS prediction challenge and the different prediction targets we will consider. In Section 6.3 we develop our evaluation framework, before applying it to the forecasts from the VIEWS challenge in Section 6.4. Section 6.5 concludes with a discussion.

6.2 The 2023/24 VIEWS Prediction Challenge

6.2.1 Challenge Format and Timeline

The 2023/24 VIEWS Prediction Challenge (Hegre et al., 2025) invited research teams to provide probabilistic forecasts of the monthly number of fatalities from organised political violence, as recorded by the Uppsala Conflict Data Program (UCDP; Davies et al., 2024). Participants were asked to generate retrospective predictions for six historical test windows (2018–2023) as well as prospective ones for the period July 2024 to June 2025. Our analysis uses only the retrospective forecasts, all of which were based on data available up to October of the preceding year. Accordingly, predictions for January had a lead time of three months, for February four months, and so on.

Predictions could be made at the country level and for finer PRIO-GRID cells, of which we only consider the former. Most model forecasts (and all considered here) were submitted in the form of 1000 predictive samples. In total, predictions from 16 different models were collected.

We note that despite the integer-valued prediction target, some teams submitted real-valued samples. As our methodological framework assumes integer-valued targets and forecasts, we rounded all samples to the nearest whole number for our analyses.

The challenge unfolded in multiple stages, with final submissions for the retrospective and true-future predictions due on 23 June 2024. From September 2024 to June 2025, running evaluations of submitted forecasts were released, before the competition concluded with the end of the forecasting window on 30 June 2025. As detailed in Section 6.3, the primary evaluation metric was pre-defined as the continuous ranked probability score (CRPS).

6.2.2 Targets and Notation

The 2023/24 VIEWS Prediction Challenge was focused on forecasts of conflict magnitude. However, due to the detailed submission format, predictions of conflict onset and (de)-escalation can be derived. In this section we provide the respective definitions, which draw on specifications from previous VIEWS challenges.

The Original Target: Conflict Magnitude. For a given country and a month t, we denote the monthly number of fatalities based on the UCDP data by X_t. As outlined in Section 6.2.1, predictions at lead times $h = 3, \ldots, 14$ months ahead (i.e., based on data up to X_{t-h}) are considered. Each forecast is given by a cumulative distribution function (CDF) $F_{t,h}$, or equivalently a probability mass function (PMF) $f_{t,h}$. The latter assigns probabilities to the possible values $0, 1, 2, \ldots$ of X_t. In practice, the distribution is represented by 1000 samples, but we here treat it as a general integer-valued probability distribution.

Conflict Onset. Conflict onset represents a transition from a peaceful state to conflict during a given time period. Loosely following Randahl and Vegelius (2024), we treat onset as one of four possible developments between a reference period $t - h$, which by default is the latest observed one, and a period of interest t. For a specific threshold $k > 0$, we therefore obtain

$$W_{t,h}^{(k)} = \begin{cases} \text{continued peace} & \text{if } X_t \leq k \text{ and } X_{t-h} \leq k \\ \text{onset} & \text{if } X_t > k \text{ and } X_{t-h} \leq k \\ \text{conflict end} & \text{if } X_t \leq k \text{ and } X_{t-h} > k \\ \text{continued conflict} & \text{if } X_t > k \text{ and } X_{t-h} > k. \end{cases} \tag{6.1}$$

We introduce separate notation for this categorical variable as it will serve for descriptive purposes and stratification of evaluation results in Section 6.4.

An onset prediction is issued when $X_{t-h} \leq k$, yielding the binary target

$$Y_{t,h}^{(k)} = \begin{cases} 0 & \text{continued peace,} \\ 1 & \text{onset.} \end{cases} \tag{6.2}$$

The onset forecast is given by a single probability which is derived from the full predictive distribution $F_{t,h}$ as

$$p_{t,h}^{(k)} = \text{Prob}_{F_{t,h}}\left(Y_{t,h}^{(k)} = 1\right) = \text{Prob}_{F_{t,h}}(X_t > k). \tag{6.3}$$

As in the VIEWS challenge, all retrospective forecasts were based on data available up to October of the previous year; in practice the reference month $t - h$ is always October.

It is not obvious how to choose the threshold k. Hegre et al. (2019) used $k = 0$, i.e., a single fatality sufficed to declare a country or PRIO-GRID cell as in conflict. In Hegre et al. (2021) this was replaced by $k = 24$ for the country level. We here use $k = 0$ as our default, but will explore results for $k = 24$ in a future version of this work. We also note that our one-month reference period may not always be appropriate. Stricter onset definitions, e.g., based on the 12 previous months, will also be explored in future updates.

(De)Escalation. Following Hegre et al. (2022), relative changes in conflict intensity are defined via log-changes:

$$Z_{t,h} = \log\left(\frac{X_t + 1}{X_{t-h} + 1}\right) = \log(X_t + 1) - \log(X_{t-h} + 1). \tag{6.4}$$

Here, h is again the prediction horizon, meaning that $Z_{t,h}$ is the log-change with respect to the last available count X_{t-h} at the time of prediction. The forecast distribution for $Z_{t,h}$, denoted by $\tilde{F}_{t,h}$, is implied by the magnitude prediction $F_{t,h}$ and the value of X_{t-h}. We note that the addition of 1 to X_t and X_{t-h} serves to avoid undefined values of the logarithm in case of zeros, though this approach is somewhat arbitrary (Bosse et al., 2023).

6.2.3 Selected Models

The purpose of the present work is the development of an evaluation approach rather than a comprehensive evaluation of the VIEWS challenge (which will be provided elsewhere). We therefore only consider 6 of the 16 submitted models, along with two benchmarks provided by the VIEWS team. Table 6.1 provides brief descriptions and references to model documentations collected by VIEWS. We also state whether the occurrence of conflict (i.e., whether zero or a positive number of fatalities occurs) is handled explicitly, e.g., using some form of hurdle model.

6.3 A Coherent Framework to Evaluate Forecasts of Conflict Onset, (De)Escalation and Magnitude

6.3.1 Proper Scoring Rules

Proper scoring rules are widely used measures of predictive performance encouraging *honest* forecasting (Gneiting and Raftery, 2007). We here detail how the evaluation of

Table 6.1: Selected models for our analysis from the 2023/24 VIEWS Prediction Challenge

Identifier	Reference	Description	Hurdle
`BDT P GLMM`	Brandt (2024)	Poisson generalised linear mixed model.	No
`CFLT RF`	Málaga et al. (2024)	Two-step random forest method using newspaper text.	Yes
`MT ZeroInfl GAM`	Muchlinski and Thornhill (2024)	Zero-Inflated Poisson Generalised Additive Model.	Yes
`PACE ShapeFinder`	Schincariol et al. (2024)	Matching of similar historical sequences.	No
`RV O MM`	Randahl and Vegelius (2024)	Observed Markov model.	Yes
`UNITO NB Transformer`	Macis et al. (2024)	Temporal transformer model.	No
`VIEWS Conflictology`	Hegre et al. (2025)	Uniform over last 12 observed values.	No
`VIEWS Zero`	Hegre et al. (2025)	Always predicts zero.	No

Note: The final column denotes whether an explicit mechanism to handle zero values is included.

conflict onset, (de)escalation and magnitude forecasts is intertwined for common scoring rules, enabling the set-up of an overarching evaluation framework.

Brier Score (BS) for Onset Predictions. The Brier score is widely used for binary prediction targets and is thus a natural choice for forecasts of conflict onset. It has been used as a secondary metric in Hegre et al. (2019) and Hegre et al. (2021), and is given by

$$\mathrm{BS}(p, y) = (p - y)^2.$$

Here, $y \in \{0, 1\}$ is the observed value, and p is the predicted probability. For onset predictions, these are defined by Equations 6.2 and 6.3, respectively.

CRPS for Magnitude Predictions. The CRPS is a common scoring rule for quantitative outcomes. It serves as the main metric to assess forecasts of conflict magnitude in the 2023/24 VIEWS Prediction Challenge (Hegre et al., 2025) and has previously been discussed for conflict forecasts in Brandt et al. (2014) and Brandt et al. (2022). For a predictive CDF F and outcome x, it is given by

$$\mathrm{CRPS}(F, x) = \int_{-\infty}^{\infty} [F(a) - \mathbf{1}(a \geq x)]^2 \mathrm{d}a. \tag{6.5}$$

The CRPS is a probabilistic generalisation of the absolute error and is thus negatively oriented (smaller values are better). It has the same unit as the predicted variable, and quantifies “how far” the observation is from the forecast, accounting for the uncertainty of the latter. The (expected) CRPS typically “scales” with the order of magnitude of the quantity to predict (Bosse et al., 2023), meaning that average CRPS values emphasise performance for high-intensity conflict zones.

CRPS for (De)Escalation. To assess the predictive performance for (de)escalation, we apply the CRPS to predictions of log-changes as defined in Equation 6.4. A similar approach has been discussed in an application to disease forecasting by Bosse et al. (2023). The resulting score measures how well the monthly rate of growth or decline in fatalities is predicted. Bosse et al. (2023) show that using log-changes under certain conditions “stabilises” expected CRPS scores, which become independent of the expected conflict magnitude. Regions and months with different magnitudes of conflict thus have a more uniform impact on average scores.

We note that in Hegre et al. (2022), (de)escalation forecasts were deterministic rather than probabilistic. Evaluation was based on squared errors and a metric called *TADDA* emphasising the signs of log changes. In the point forecast setting, however, this parallel use of different scoring metrics led to some difficulties, as outlined in Chapter 4 and published in Bracher et al. (2023).

Unified Perspective: The Threshold-Weighted CRPS. The BS and CRPS are closely linked, enabling us to set up a common scoring framework for all targets. Using $y^{(a)} = \mathbf{1}(x > a)$ and $p_a = \mathrm{Prob}_F(x > a)$ as in Equations 6.2 and 6.3, we can write

$$\mathrm{CRPS}(F, x) = \int_{-\infty}^{\infty} \mathrm{BS}(p_a, y^{(a)}) \mathrm{d}a. \tag{6.6}$$

The CRPS is hence nothing but an aggregation of Brier scores for binary prediction targets at all possible cut-offs (Gneiting and Raftery, 2007; Brandt et al., 2014). For the magnitude target X_t with support $0, 1, 2, \ldots$, this simplifies to the ranked probability score,

$$\mathrm{RPS}(F, x) = \sum_{a=0}^{\infty} \mathrm{BS}(p_a, y^{(a)}), \tag{6.7}$$

meaning that the Brier score $\mathrm{BS}(p_k, y^{(k)})$ for conflict onset (with default $k = 0$) is one of many additive components entering into the (C)RPS of the magnitude forecast. The above relationships 6.6 and 6.7 are detailed in Appendix E.2.

In the standard CRPS used for magnitude forecasts, all cut-off values receive equal weight, but the extension to a *threshold-weighted* CRPS (Gneiting, 2011),

$$\mathrm{twCRPS}(F, x) = \int_{-\infty}^{\infty} w(a) \times \mathrm{BS}(p_a, y^{(a)})\mathrm{d}a,$$

is straightforward. The weighting shifts the focus of the CRPS to certain cut-off values, and it turns out that the evaluation strategies outlined above for onset and (de)escalation are just the results of different weighting schemes. The Brier score for the binarised onset target, Equation 6.2, obviously results for $w(a) = 1$ if $a = k$ and 0 otherwise. For the CRPS of the log-change target in Equation 6.4, it can be shown that

$$\mathrm{CRPS}(\tilde{F}, z) = \mathrm{twCRPS}(F, x)$$

with $\tilde{F}$ the predictive distribution of the log-change z, and weighting function

$$w(a) = 1/(a + 1).$$

The derivation, based on Allen et al. (2023), will be provided in a future manuscript version.

We thus have a unified framework to evaluate forecasts of conflict magnitude, (de)escalation and onset. This is appealing as we can explore evaluation results jointly for all targets. In Section 6.4, we will use the nesting of the Brier score into the CRPS from Equation 6.7 to assess the importance of onsets in evaluations for conflict magnitude. Moreover, in future extensions we will visually compare average Brier scores of different models across cut-off values a (Gneiting, 2011). If these average BS curves of two models do not cross, one model is superior irrespective of the weighting scheme, and thus for onset, (de)escalation and magnitude targets. If the curves do cross, this explains flipped rankings for different tasks, and thresholds with poor performance can be targeted in model improvement.

6.3.2 Calibration and Discrimination of Onset Forecasts

Score-based evaluation is useful for ranking different models in terms of their overall performance. However, aggregate results may be opaque, and rankings imply little about whether any of the considered models actually works. Building on Dimitriadis et al. (2024), we therefore employ state-of-the-art tools to assess the *discrimination* and *calibration* of conflict onset forecasts in more detail. This is motivated by the fact that in the existing conflict forecasting literature, there is a strong focus on discriminative ability, while calibration is rarely considered.

Discrimination. *Discrimination* refers to the ability to distinguish between instances in which events (conflicts) do or do not occur. Obviously, predictive probabilities should be higher in the former case. Common discrimination measures are the areas under the Receiver Operating Characteristic (ROC) (AUC; e.g., Mueller and Rauh, 2022a) and the Precision–Recall curve (Hegre et al., 2021). Following Dimitriadis et al. (2024), we consider full ROC curves, i.e., plot true positive rates, also called hit rate (HR; sensitivity) against false positive rates, also known as false alarm rate (FAR; $1-$ sensitivity) for different cut-offs on the predictive probability p. An ideal ROC curve runs close to the top-left corner of the graph, meaning that high HR and low FAR can be achieved jointly.

Calibration. *Calibration* refers to whether the predictive probabilities issued by a model are aligned with the observed empirical relative frequencies. For instance, among all country-months assigned a conflict probability of $p_F = 0.25$, a quarter should actually experience conflict. This property can be assessed using reliability curves, which display the predicted probability p against the observed frequencies based on a suitably chosen binning system (see Dimitriadis et al., 2024 for details). An ideal calibration curve follows the diagonal, as predicted probabilities and observed frequencies should agree.

6.4 Results

As noted in the introduction, this version focuses on comparing magnitude and onset prediction, with escalation and de-escalation results to be added in future iterations of the project.

6.4.1 Global Distribution of Conflict Magnitude and Onset

We first summarise overall trends in monthly numbers of fatalities from organised political violence according to the UCDP data set. Figure 6.1 (top panel) shows the global monthly counts decomposed into the five most affected countries and a residual category for the remaining countries. It can be seen that Ethiopia, Ukraine, Afghanistan, Yemen and

Syria (in this order) account for a very large share of fatalities. The middle panel displays the monthly number of conflict onsets according to the definition from Equation 6.2 with $k = 0$. The bottom panel shows the coding of months based on the more detailed definition given in Equation 6.1 for a few selected countries. As onset events are always defined relative to the preceding month of October (highlighted in yellow), conflicts which start and then persist can count as onsets in multiple consecutive months. Indeed, the total of 505 onset events counted in the figure refer to only 114 country / year combinations from 47 countries (see Supplementary Figure E.1). For some of the onset events this definition may be questionable. For instance, in Myanmar every month of 2020 is considered an “onset” as October 2019 was peaceful. However, as most other months of 2019 were not, the fatalities in 2020 may be better seen as a continuation of an ongoing conflict.

A further caveat concerns the large fatality spikes for Ethiopia in Figure 6.1. These stem from the earlier VIEWS procedure in which fatalities from multi-month UCDP events were assigned to the final month of each episode. This aggregation created artificial spikes when long events ended, such as the October 2022 spike from an August–October event and the elevated 2021 figures from a year-long episode, and can also yield misleading onsets. Following feedback to our talk at the *Symposium on Crisis Early Warning*, VIEWS has recently revised its pipeline to distribute such events across their true duration[1]. Because the 2023/24 Prediction Challenge relied on the earlier end-of-period-assigned data, we employ the same version here, and the results should be interpreted accordingly.

6.4.2 Mean CRPS of Magnitude Forecasts Is Dominated by High-Intensity Conflicts

We now move to the evaluation of the eight selected models from Section 6.2.3. Figure 6.2 (top panel) displays the sum of monthly CRPS values per country, averaged across models. The bars are split into the most influential contributions to the overall CRPS which results in the same set of countries as in Figure 6.1. Some parallels between the two figures are striking, especially concerning the early phases of the conflicts in Ethiopia and Ukraine. They reflect that in October 2021 (and in the case of Ethiopia, also October 2020) the models did not anticipate the massive (artificial) increases in fatalities, leading

[1] https://viewsforecasting.org/news/improved-handling-of-ucdp-summary-events-in-views-forecasts-and-api-data/, accessed 25 November 2025.

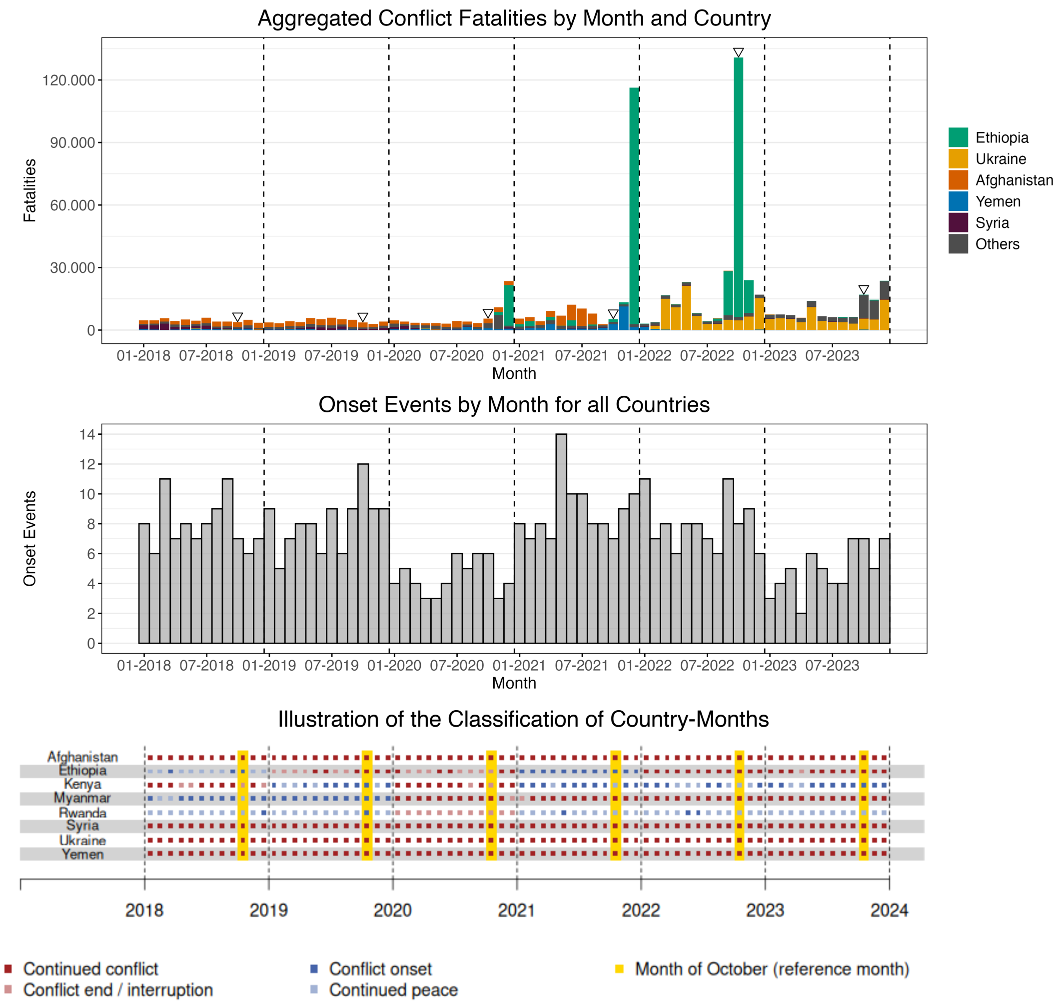

Figure 6.1: Monthly totals of fatalities and onset events

Note: Top: Global totals of monthly fatalities, highlighting the five most-affected countries and aggregating others. Triangles mark the October reference month used in onset definitions. Middle: Monthly counts of onset events using $k = 0$. Bottom: Classification of country-months according to Equation 6.1 for selected months. The spikes observed for Ethiopia stem from an earlier aggregation practice that assigned fatalities from multi-month events to the final month.

to CRPS values close to the observed values (remember that CRPS reflects roughly the absolute distance between prediction and observation). The large CRPS values for Ethiopia in 2023, on the other hand, are the result of overprediction: given the high numbers of fatalities in October 2022, the models anticipated high numbers of fatalities which did not materialise.

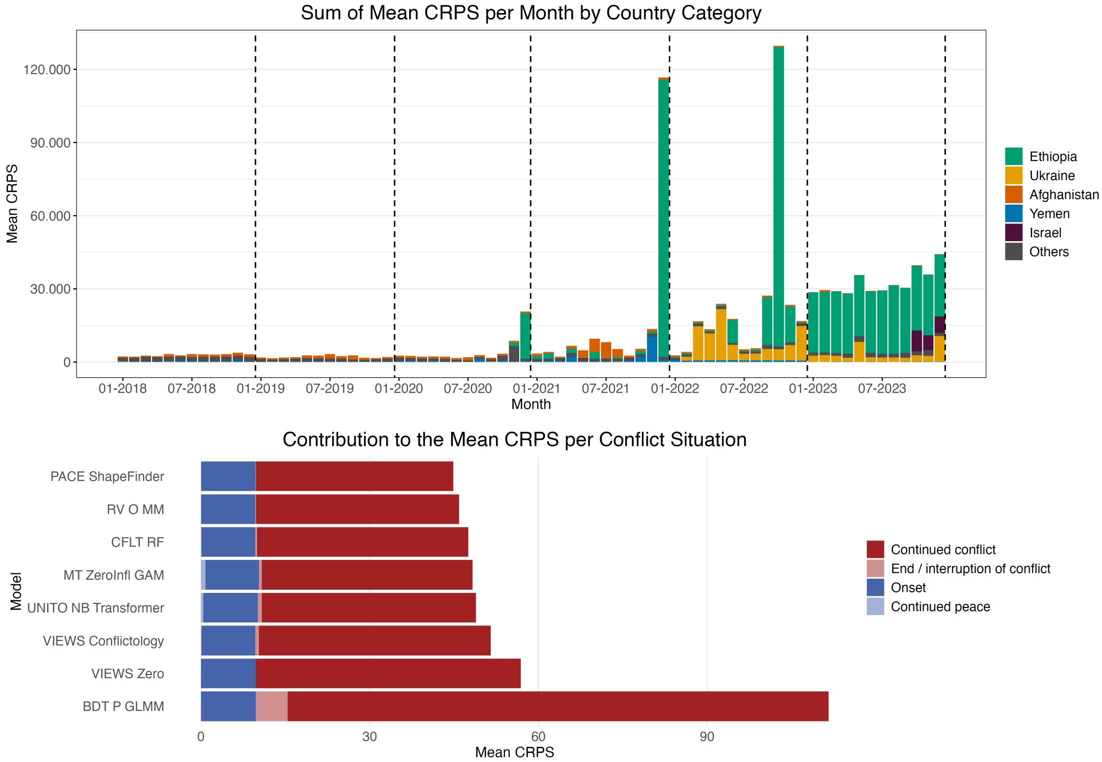

Figure 6.2: CRPS results for conflict magnitude

Note: Top: Country-level CRPS contributions averaged across models, with five largest contributors highlighted. Bottom: Mean CRPS values by contributions of different conflict situations as defined in Equation 6.1.

The bottom panel of Figure 6.2 summarises the contributions of different conflict situations, categorised using Definition 6.1, to the overall mean CRPS. Blue segments represent performance in settings with peace in the preceding month of October, while red ones indicate previous conflict. For descriptive purposes, we also show whether conflict occurred (dark) or not (light) in the respective months. However, light and dark segments should always be considered in conjunction, as stratifying evaluation results by the outcome can lead to skewed conclusions (Lerch et al., 2017). The overall mean CRPS is clearly dominated by settings with previous conflict and especially cases where conflict continues (dark red segments). These of course mainly represent the countries highlighted in the top panel of Figure 6.2. Somewhat surprisingly, the average CRPS performance in settings with previous peace is very similar across models, an aspect we will discuss in the following section.

6.4.3 Binary Onset Targets Represent a Distinct Dimension of Performance

Figure 6.3 magnifies the blue onset and peace components from Figure 6.2, and highlights the contribution of the Brier score at threshold $k = 0$ in grey/black (based on Equation 6.7). This contribution can be interpreted as the ability to predict *occurrence of onset* in settings of previous peace (according to our binarised definition), while the remaining blue segments represent the ability to predict *onset magnitude given onset occurs*. The blue segments dominate average scores, meaning that even when conditioning on previous peace, average CRPS mainly reflects magnitudes of conflict.

We moreover note that the average CRPS values achieved by different models are strikingly similar. This occurs because Ethiopia reported no fatalities in October 2020, meaning that the large number recorded in 2021 is captured within the onset-magnitude component. As no model anticipated this massive rise in (aggregated) fatalities, they

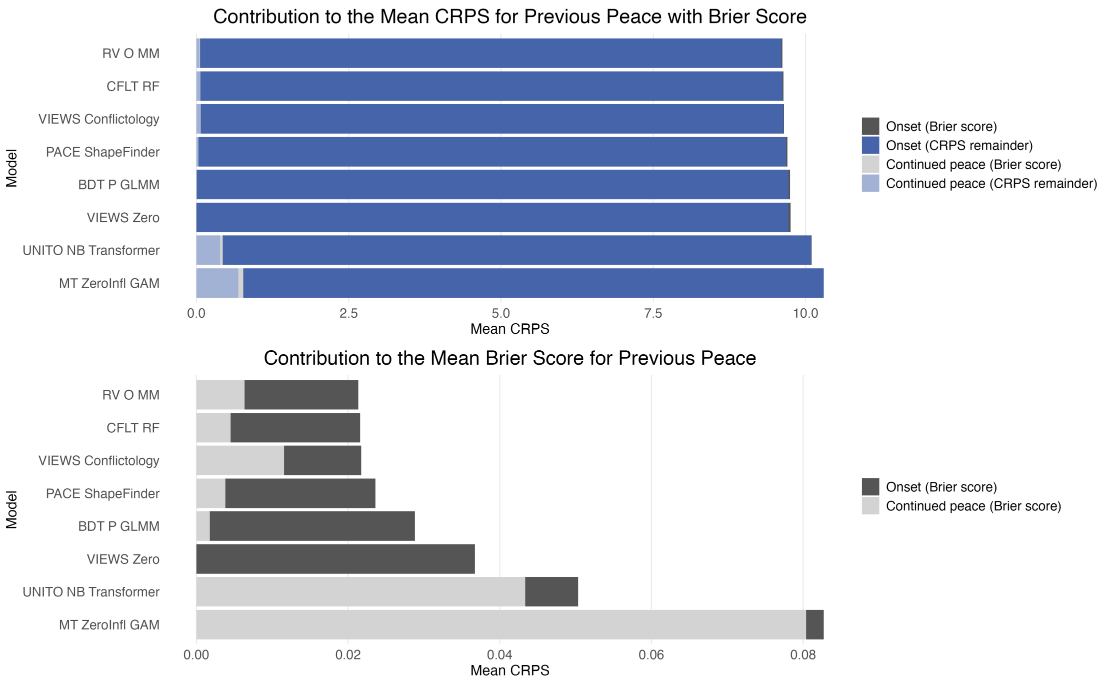

Figure 6.3: CRPS and Brier Score evaluation under previous peace

Note: Top: Mean CRPS restricted to settings with previous peace thereby magnifying the respective bars from Figure 6.2. The contribution of the Brier score for the binary onset target according to Equation 6.7 is highlighted in grey (continued peace) and black (onset; both hard to discern visually). Bottom: Zoom-in on the average Brier scores (grey and black segments from the top plot).

all receive almost identically poor CRPS values, which dominate and blur the overall evaluation.

To avoid this conflation of onset occurrence and magnitude, the bottom panel of Figure 6.3 magnifies the Brier scores for our binary onset target from Equation 6.2. Interestingly, relative to the top panel, the model ranking remains identical. Compared to Figure 6.2, however, we can see distinct changes. Most notably, the `UNITO NB Transformer` and `MT ZeroInfl GAM` models, which achieved average ranks for conflict magnitude, struggle with the onset target. While `UNITO NB Transformer` may be affected by its lack of explicit zero handling (see Table 6.1), the cause is unclear for the `MT ZeroInfl GAM`, which does include zero-inflation covariates. Interestingly, although the `VIEWS Conflictology` baseline was surpassed by most models in terms of magnitude CRPS, it ranks third in this setting and performs close to the top models. It represents an empirical distribution of the last twelve months and thus only predicts "onsets" in cases of recently cooled-down conflicts. While this indicates that there is limited predictability beyond such relapses, it also raises questions regarding our definition of onset events, see the discussion in Section 6.5.

6.4.4 Calibration of Onset Forecasts Is Challenging

To shed light on how overall predictive performance for conflict onset arises, we conclude with an analysis of the models' discrimination and calibration. The top panel of Figure 6.4 shows the ROC curves for the binary onset target defined in Equation 6.2, restricted to settings with previous peace as in Figure 6.3. All models show rather high AUC values between 0.92 and 0.96 (shown in the legend), apart from the `VIEWS Zero` model, which, as it always predicts zero, has no discriminatory power. However, the high AUC of the `VIEWS Conflictology` means that much of this discriminatory ability (AUC 0.94) is attainable purely based on recent conflict history. Additional discriminatory power seems to be hard to achieve.

The bottom panels of Figure 6.4 show calibration plots for the eight considered models. The models `CLFT RF`, `PACE ShapeFinder` and `RV O MM` are rather well-calibrated, explaining their favourable ranking in Figure 6.3. The same holds for the `VIEWS Conflictology`. The remaining models suffer from calibration issues. While `BDT P GLMM` assigns predictive probabilities that are too low (conflicts occur more often than expected) the opposite holds for `MT ZeroInfl GAM` and `Unito NB Transformer`. For the latter two, even in

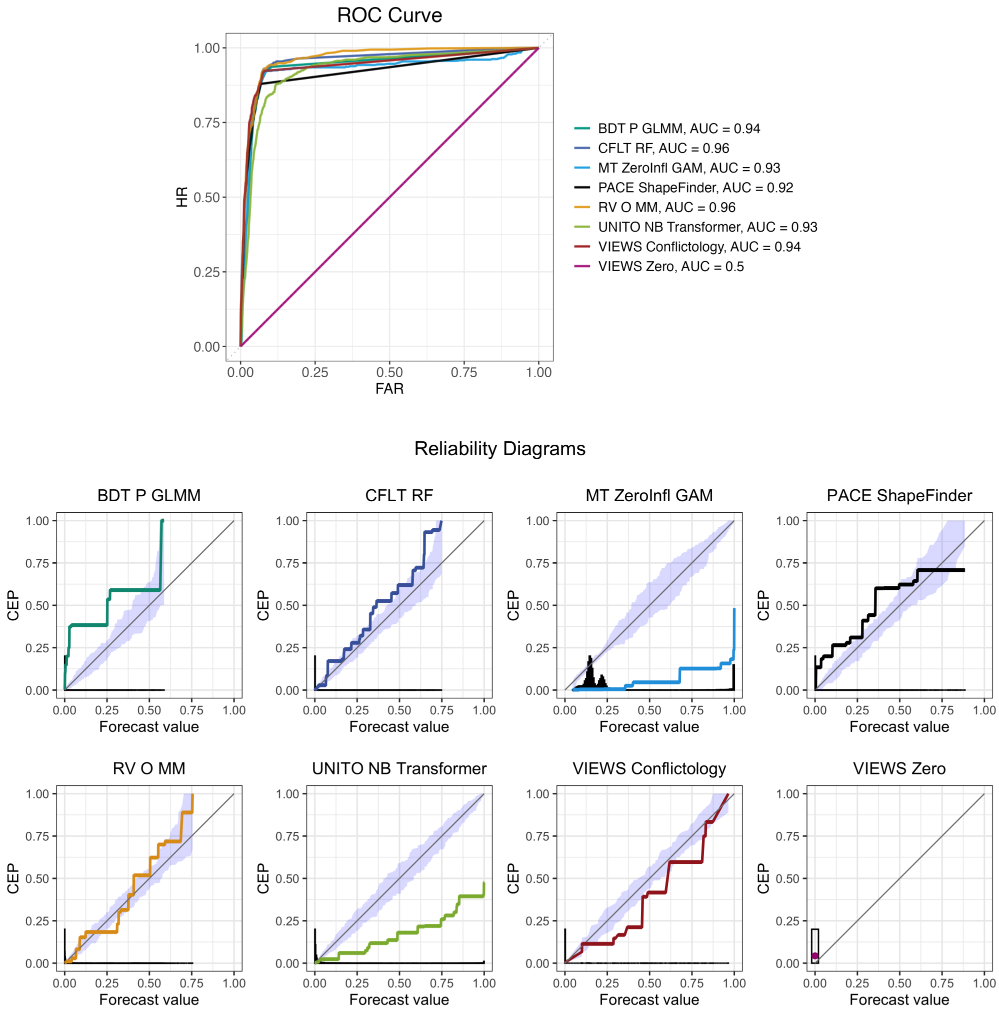

Figure 6.4: Discrimination and calibration of onset forecasts

Note: Top: ROC curves for all models, with AUC values mentioned in the legend. Bottom: Calibration curves with consistency bands (light blue). Calibration curves should be close to the diagonal, with curves outside the consistency bands indicating lack of calibration.

settings where conflict onset is predicted to occur with a probability close to one, this materialises in less than half of the cases. This explains their drop in performance in Figure 6.3.

6.5 Discussion

We propose a coherent evaluation framework to assess conflict forecasts with respect to their ability to predict conflict magnitude, (de)escalation, and onset. Our empirical results show that, although onset evaluation is conceptually nested within magnitude evaluation, in practice the two provide complementary insights into model performance. We also find that some onset forecasts suffer from miscalibration while still displaying good discrimination, the latter even being achieved by relatively simple baselines, with further gains proving difficult.

We emphasise that we by no means advocate abandoning or replacing the magnitude target used in the 2023/24 VIEWS Prediction Challenge. Challenge formats require a clear choice of a primary target and evaluation metric, and strong arguments support the magnitude target, particularly given the importance of high-intensity events. Our aim is instead to show that complementary evaluations for different targets can offer useful perspectives on predictive performance in different contexts.

These complementary analyses are feasible in a conceptually coherent way due to the detailed probabilistic submission framework of the 2023/24 VIEWS Prediction Challenge. In point-forecast settings, evaluating forecasts with multiple metrics is difficult because it can create conflicting incentives for forecasters (Bracher et al., 2023). In the probabilistic setting, this issue does not arise as long as all scores are *proper* (Gneiting and Raftery, 2007). In our case, a forecaster minimising CRPS implicitly optimises the Brier score at threshold $k = 0$, making it legitimate to assess performance under this metric.

A limitation of our approach is the challenge of defining a binary onset target. We use a sensitive definition based on a threshold of one fatality per country-month and a one-month reference period for determining the previous conflict state. In some cases (e.g., Ethiopia and Myanmar), this leads to classifications that may be questioned. A related limitation is that forecasts were released annually, which made the definition of reference months across lead times somewhat peculiar. Moreover, we note that artificial end-of-period spikes in the fatality data, stemming from an earlier aggregation of multi-month events, also affect onset classifications and evaluation results.

In future iterations, we will explore alternative thresholds and reference periods and integrate the (de)escalation target into the empirical analysis. We further plan to extend the analysis to probabilistic forecasting in epidemiology, where analogous concepts such as disease outbreak and relative changes in the course of disease dynamics are also relevant.

Appendices

A Appendix to Chapter 2: Model Determination for High-Dimensional Longitudinal Data with Missing Observations: An Application to Microfinance Data

A.1 Details on the Construction of the Panel Data Set

We construct the balanced panel used in this work as follows. First, the *Financial Performance Data Set in USD* and the *Social Performance Data Set* from the MIX are merged with an inner join regarding the variables *MFI ID*, *MFI Name*, *Fiscal Year*, *As of Date*, and *Period Type*. The resulting data set comprises annual data of 1026 MFIs and observations of 472 variables from 2007 to 2018. From these data, we select the largest panel that includes as many years as possible. We obtain $w^* = [2009, 2014]$ as described in Section 2.2. The pre-processing is continued with all variables in the merged data set and only those $N_6 = 213$ MFIs and years contained in w^*.

Subsequently, all variables with more than 50% missing observations in the chosen window are dropped. This results in the removal of 148 variables including all variables from the financial performance categories *Deposit Products*, *Digital Delivery Channels*, *Enterprises Financed*, *Job Creation*, *Non-financial Services*, *Poverty Outreach* and *Products*.

The remaining 324 variables are examined in more depth. The main objective of this work requires the data set to be comprehensive in such a way that it contains all potentially meaningful and important variables and no redundant or uninformative data. Moreover, the data must be comparable between MFIs of different sizes. All these aspects are addressed in an appropriate manner. While some variables are removed, others are transformed or simply kept for the subsequent analysis. All steps undertaken are documented in Table A.2, which also provides definitions of all included variables. Any values deemed implausible (e.g., percentages greater than 1) are excluded. The dummy variables in the social performance data are transformed according to the procedure described below.

Transformation of Social Performance Dummies

Much of the social performance data is recorded using multiple dummy variables. As a consequence, many columns of the original data set contain redundant data. The problem can be illustrated with the following example from the data. Within the category *Client Protection*, several variables regarding debt collection practices are defined as stated in Table A.1 where each row represents one variable.

Table A.1: All variables concerning debt collection practices in the raw data set

Clear Debt Collection Practices	No		
— " —	Partially		
— " —	— " —	Clear Sanctions for Violations of Debt Collection Practices	No
— " —	— " —	— " —	Partially
— " —	— " —	— " —	Unknown
— " —	— " —	— " —	Yes
— " —	Unknown		
— " —	Yes		
— " —	— " —	Clear Sanctions for Violations of Debt Collection Practices	No
— " —	— " —	— " —	Partially
— " —	— " —	— " —	Unknown
— " —	— " —	— " —	Yes

As a result, the original data set contains 12 different variables that essentially carry the same or at least very similar information. Such variables are reduced to one single dummy variable. In this example, the new variable is defined as

$$\textit{Clear Debt Collection Practices} = \begin{cases} 1, & \text{if } \textit{Clear Debt Collection Practices} \triangleright \textit{Yes} = 1, \\ 0, & \text{if } \textit{Clear Debt Collection Practices} \triangleright \textit{No} = 1, \\ NA, & \text{otherwise.} \end{cases}$$

Additional sub-variables, such as those providing details on *Clear Sanctions for Violations of Debt Collection Practices*, are deleted. The original data set contains very few observations of variables that are similar (in the way that they contain information on partially fulfilled conditions) to *Clear Debt Collection Practices* ▷ *Partially* with value 1. Hence, they are discarded and included as *NA* in the new dummy variables.

Table A.2: Details on the construction of the financial and social performance data of the balanced and unbalanced panel data set

Original Variable	Action	New Variable	Definition of Reason for Deletion
		Financial Performance Variables	
Balance Sheet			
Assets	K	Assets	Total value of resources controlled by the MFI as a result of past events and from which future economic benefits are expected to flow to the MFI; for calculation purposes, assets are the sum of each individual asset account
Average Assets	D		Redundant, similar to *Assets*
Borrowings	T	Borrowings/Liabilities	*Borrowings/Liabilities*; Borrowings denote the principal balance for all funds received through a loan agreement, may include bonds or similar debt securities issued and credit lines
Average Assets	D		Redundant, similar to *Assets*
Borrowings	T	Borrowings/Liabilities	*Borrowings/Liabilities*; Borrowings denote the principal balance for all funds received through a loan agreement, may include bonds or similar debt securities issued and credit lines
Cash and Cash Equivalents	D		Redundant: contained in *% Assets: Non-Earn. Liq. Ass.*
Donated Equity	T	Donated Equity/Equity	*Donated Equity/Equity*; Donated equity denotes accumulated donations to MFI
Equity	K	Equity	Residual interest in the assets of the financial institution after deducting all its liabilities
Impairment Loss Allowance	T	Imp. Loss Allow./GLP	*Impairment Loss Allowance/GLP*
Liabilities	K	Liabilities	Total value of present obligations of the MFI arising from past events, the settlement of which is expected to result in an outflow from the financial institution of resources embodying economic benefits
	T	Liabilities/Assets	*Liabilities/Assets*
Liabilities and Equity	D		Redundant: equals *Assets*
Net Fixed Assets	D		Uninformative
Net Loan Portfolio	T	Net GLP/GLP	*Net Loan Portfolio/GLP*; Net loan portfolio denotes the value of loan portfolio net of impairment loss allowance and unearned income and discount (when applicable)
Other Assets	D		Uninformative
Other Equity	D		Uninformative
Other Liabilities	D		Uninformative
Other Short Term Financial Liabilities	D		Uninformative
Paid-In Capital	T	Paid in Capital/Equity	*Paid-In Capital/Equity*; Paid-in capital denotes capital paid by shareholders or members at the par value the shares
Retained Earnings	T	Retained Earnings/Equity	*Retained Earnings/Equity*
Subordinated Debt	T	Subord. Debt/Equity	*Subordinated Debt /Liabilities*
Clients			
Gross Loan Portfolio (GLP)	K	Gross Loan Portfolio (GLP)	All outstanding principals due for all outstanding client loans; includes current, delinquent, and renegotiated loans, but not loans that have been written off
GLP ▷ Gender ▷ Female	T	Av. Loan Size: Female	*GLP ▷ Gender ▷ Female/NLO ▷ Gender ▷ Female*

(Table continued on next page)

Table A.2: Details on the construction of the financial and social performance data *(continued)*

Original Variable	Action	New Variable	Definition of Reason for Deletion
GLP ▷ Gender ▷ Legal Entity	T	Av. Loan Size: Legal Entity	*GLP ▷ Gender ▷ Legal Entity/NLO ▷ Gender ▷ Legal Entity*
GLP ▷ Gender ▷ Male	T	Av. Loan Size: Male	*GLP ▷ Gender ▷ Male/NLO ▷ Gender ▷ Male*
GLP ▷ Location ▷ Rural	T	Av. Loan Size: Rural	*GLP ▷ Location ▷ Rural/NLO ▷ Location ▷ Rural*
GLP ▷ Location ▷ Urban	T	Av. Loan Size: Urban	*GLP ▷ Location ▷ Urban/NLO ▷ Location ▷ Urban*
GLP ▷ Relationship ▷ External Customers	T	Av. Loan Size: Ext. Customers	*GLP ▷ Relationship ▷ External Customers/NLO ▷ Relationship ▷ External Customers*
GLP ▷ Relationship ▷ Management and Staff	T	Av. Loan Size: Man. & Staff	*GLP ▷ Relationship ▷ Management and Staff/NLO ▷ Relationship ▷ Management and Staff*
Number of Active Borrowers (NAB)	K	# Active Borrowers	Number of individuals who currently have an outstanding loan balance with the financial institution or are primarily responsible for repaying any portion of the GLP
	T	log(NAB)	log(Number of Active Borrowers)
NAB ▷ Gender ▷ Female	T	% Borrowers: Female	*NAB ▷ Gender ▷ Female/Number of Active Borrowers*
NAB ▷ Gender ▷ Legal Entity	T	% Borrowers: Legal Entity	*NAB ▷ Gender ▷ Legal Entity/Number of Active Borrowers*
NAB ▷ Gender ▷ Male	T	% Borrowers: Male	*NAB ▷ Gender ▷ Male/Number of Active Borrowers*
NAB ▷ Location ▷ Rural	T	% Borrowers: Rural	*NAB ▷ Location ▷ Rural/Number of Active Borrowers*
NAB ▷ Location ▷ Urban	T	% Borrowers: Urban	*NAB ▷ Location ▷ Urban/Number of Active Borrowers*
NAB ▷ Relationship ▷ External Customers	T	% Borrowers: Ext. Customers	*NAB ▷ Relationship ▷ External Customers/Number of Active Borrowers*
NAB ▷ Relationship ▷ Management and Staff	T	% Borrowers: Man. & Staff	*NAB ▷ Relationship ▷ Management and Staff/Number of Active Borrowers*
Number of Loans Outstanding (NLO)	K	# Loans Outstanding	The number of loans in the GLP; for financial institutions using a group lending methodology, the number of loans should refer to the number of individuals receiving loans as part of a group or as part of a group loan
NLO ▷ Gender ▷ Female	D		Redundant, since *% Borrowers: Female* is included
NLO ▷ Gender ▷ Legal Entity	D		Redundant, since *% Borrowers: Legal Entity* is included
NLO ▷ Gender ▷ Male	D		Redundant, since *% Borrowers: Male* is included
NLO ▷ Location ▷ Rural	D		Redundant, since *% Borrowers: Rural* is included
NLO ▷ Location ▷ Urban	D		Redundant, since *% Borrowers: Urban* is included
NLO ▷ Relationship ▷ External Customers	D		Redundant, since *% Borrowers: Ext. Customers* is included
NLO ▷ Relationship ▷ Management And Staff	D		Redundant, since *% Borrowers: Man. & Staff* is included
Number of New Borrowers	T	# New Borrowers/NAB	The number of borrowers during the current reporting period who have either never received a loan from the institution or who have not received a loan in three years or more/NAB
Credit Products			
GLP ▷ Cred. Prod. ▷ Enterprise Finance	T	Av. Loan Size: Enterpr. Fin.	*GLP ▷ Cred. Prod. ▷ Enterprise Finance/NLO ▷ Cred. Prod. ▷ Enterprise Finance*
	T	% GLP: Enterprise Finance	*GLP ▷ Cred. Prod. ▷ Enterprise Finance/GLP*
GLP ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Large Corporations	T	Av. Loan Size: Large Corp.	*GLP ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Large Corporations/NLO ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Large Corporations*
	T	% GLP: Large Corporations	*GLP ▷ Cred. Prod. ▷ Enterprise Finance ▷ Large Corporations / GLP*
GLP ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Small, Med. Ent.	T	Av. Loan Size: Sm. & Med. Enterpr.	*GLP ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Small, Med. Ent./NLO ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Small, Med. Ent.*

(Table continued on next page)

Table A.2: Details on the construction of the financial and social performance data *(continued)*

Original Variable	Action	New Variable	Definition of Reason for Deletion
	T	% GLP: Small & Med. Enterpr.	*GLP ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Small, Med. Ent./GLP*
GLP ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Microenterprise	T	Av. Loan Size: Microenterpr.	*GLP ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Microenterprise/NLO ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Microenterprise*
	T	% GLP: Microenterprise	*GLP ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Microenterprise/GLP*
GLP ▷ Cred. Prod. ▷ Household Financing	T	Av. Loan Size: Household Fin.	*GLP ▷ Cred. Prod. ▷ Household Financing/NLO ▷ Cred. Prod. ▷ Household Financing*
	T	% GLP: Household Financing	*GLP ▷ Cred. Prod. ▷ Household Financing/GLP*
GLP ▷ Cred. Prod. ▷ Househ. Fin. ▷ Consumption	T	Av. Loan Size: Consumption	*GLP ▷ Cred. Prod. ▷ Househ. Fin. ▷ Consumption/NLO ▷ Cred. Prod. ▷ Househ. Fin. ▷ Consumption*
	T	% GLP: Consumption	*GLP ▷ Cred. Prod. ▷ Househ. Fin. ▷ Consumption/GLP*
GLP ▷ Cred. Prod. ▷ Househ. Fin. ▷ Mortgage/Housing	T	Av. Loan Size: Mortgage, Housing	*GLP ▷ Cred. Prod. ▷ Househ. Fin. ▷ Mortgage, Housing/NLO ▷ Cred. Prod. ▷ Househ. Fin. ▷ Mortgage, Housing*
	T	% GLP: Mortgage, Housing	*GLP ▷ Cred. Prod. ▷ Househ. Fin. ▷ Mortgage, Housing/GLP*
GLP ▷ Cred. Prod. ▷ Househ. Fin. ▷ Other Househ. Fin.	T	Av. Loan Size: Other Househ. Fin.	*GLP ▷ Cred. Prod. ▷ Househ. Fin. ▷ Other Househ. Fin./NLO ▷ Cred. Prod. ▷ Househ. Fin. ▷ Other Househ Fin.*
	T	% GLP: Other Household Fin.	*GLP ▷ Cred. Prod. ▷ Household Financing ▷ Other Household Financing/GLP*
GLP ▷ Methodology ▷ Individual	T	Av. Loan Size: Individual	*GLP ▷ Methodology ▷ Individual/NLO ▷ Methodology ▷ Individual*
	T	% GLP: Individual	*GLP ▷ Methodology ▷ Individual/GLP*
GLP ▷ Methodology ▷ Solidarity Group	T	Av. Loan Size: Solidarity Group	*GLP ▷ Methodology ▷ Solidarity Group/NLO ▷ Methodology ▷ Solidarity Group*
	T	% GLP: Solidarity Group	*GLP ▷ Methodology ▷ Solidarity Group/GLP*
GLP ▷ Methodology ▷ Village Banking SHG	T	Av. Loan Size: Village (SHG)	*GLP ▷ Methodology ▷ Village Banking SHG/NLO ▷ Methodology ▷ Village Banking SHG*
	T	% GLP: Village Banking (SHG)	*GLP ▷ Methodology ▷ Village Banking SHG/GLP*
NLO ▷ Cred. Prod. ▷ Enterprise Finance	D		Redundant, since *% GLP: Enterprise Finance* is included
NLO ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Large Corporations	D		Redundant, since *% GLP: Large Corporations* is included
NLO ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Small, Med. Ent.	D		Redundant, since *% GLP: Sm. & Med. Enterpr.* is included
NLO ▷ Cred. Prod. ▷ Enterpr. Fin. ▷ Microenterprise	D		Redundant, since *% GLP: Microenterpr.* is included
NLO ▷ Cred. Prod. ▷ Household Financing	D		Redundant, since *% GLP: Household Fin.* is included
NLO ▷ Cred. Prod. ▷ Household Fin. ▷ Consumption	D		Redundant, since *% GLP: Consumption* is included
NLO ▷ Cred. Prod. ▷ Househ. Fin. ▷ Mortgage/Housing	D		Redundant, since *% GLP: Mortgage, Housing* is included
NLO ▷ Cred. Prod. ▷ Househ Fin. ▷ Other Househ. Fin.	D		Redundant, since *% GLP: Other Househ. Fin.* is included
NLO ▷ Methodology ▷ Individual	D		Redundant, since *% GLP: Individual* is included

(Table continued on next page)

Table A.2: Details on the construction of the financial and social performance data *(continued)*

Original Variable	Action	New Variable	Definition of Reason for Deletion
NLO ▷ Methodology ▷ Solidarity Group	D		Redundant, since *% GLP: Solidarity Group* is included
NLO ▷ Methodology ▷ Village Banking SHG	D		Redundant, since *% GLP: Village Banking (SHG)* is included
Delinquency			
GLP ▷ Del. ▷ >1 Month	D		Redundant, since *% GLP: 1-3 Months Delinquent* and *% GLP: >3 Months Delinquent* are included
GLP ▷ Del. ▷ >1 Month ▷ 1-3 Months	T	% GLP: 1-3 Months Delinquent	GLP ▷ Delinquency ▷ >1 Month ▷ 1-3 Months/GLP
GLP ▷ Del. ▷ >1 Month ▷ >3 Months	T	% GLP: >3 Months Delinquent	GLP ▷ Delinquency ▷ >1 Month ▷ >3 Months/GLP
GLP ▷ Del. ▷ >1 Month ▷ >3 Months ▷ 3-6 Months	D		Redundant, since *% GLP: 1-3 Months Delinquent* and *% GLP: >3 Months Delinquent* are included
GLP ▷ Del. ▷ >1 Month ▷ >3 Months ▷ >6 Months	D		Redundant, since *% GLP: 1-3 Months Delinquent* and *% GLP: >3 Months Delinquent* are included
GLP ▷ Del. ▷ Renegotiated Loans	T	% GLP: Renegotiated Loans	GLP ▷ Delinquency ▷ Renegotiated Loans/GLP
Write-Offs	D		Redundant, since *Write-Off Ratio* is included
Financial Performance			
Operational Self-Sufficiency (OSS)	K	Operational Self-Sufficiency	*Financial Revenue/(Financial Expense on Funding Liabilities + Net Impairment Loss + Operating Expense)*
Return on Assets	K	Return on Assets	*(Net Operating Income − Taxes)/Average Total Assets*
Return on Equity	K	Return on Equity	*(Net Operating Income − Taxes)/Average Total Equity*
Financing Structure			
Capital/Asset Ratio	K	Capital/Assets	*Total Equity/Total Assets*
Debt to Equity Ratio	K	Debt/Equity	*Total Liabilities/Total Equity*
Deposits to Loans	K	Deposits/Loans	*Deposits/GLP*
Deposits to Total Assets	K	Deposits/Total Assets	*Deposits/Total Assets*
GLP to Total Assets	K	GLP/Total Assets	*GLP/Total Assets*
Income Statement			
Administrative Expense	D		Redundant, since *Administrative Exp./Assets* is included
Depreciation and Amortisation Expense	D		Uninformative
Donations	D		Very few observations $\neq 0$
Fee and Commission Income on Loan Portfolio	D		Not clearly defined
Financial Expense on Funding Liabilities	T	Fin. Exp. Fund. Liab./Assets	*Financial Expense on Funding Liabilities/Assets*
Financial Revenue	D		Redundant, since *Financial Revenue/Assets* is included
Financial Revenue from Loans	D		Redundant, since contained in *Yield on GLP (Real)*
Impairment Loss (Reversal of Impairment Loss), GLP	D		Redundant, since contained in *Loan Impairm. Prov./Assets*
Income from Penalty Fees on Loan Portfolio	T	Income on Pen. Fees/GLP	*Income from Penalty Fees on GLP/GLP*

(Table continued on next page)

Table A.2: Details on the construction of the financial and social performance data *(continued)*

Original Variable	Action	New Variable	Definition of Reason for Deletion
Interest Expense on Borrowings	T	Interest Exp. on Borr./Borr.	*Interest Expense on Borrowings/(Borrowings per Liabilities · Liabilities)*
Interest Expense on Deposits	D		Uninformative
Interest Income on Loan Portfolio	T	Interest Income on GLP/GLP	*Interest Income on GLP/GLP*
Net Impairment Loss, GLP	D		Redundant, since contained in *Loan Loss Rate*
Net Non-Operating Income	D		Uninformative
Non-Operating Expense	D		Uninformative
Non-Operating Income	D		Uninformative
Operating Expense	D		Redundant, since contained in *Operating Expense/GLP*
Other Financial Expense	D		Uninformative
Other Financial Revenue	D		Uninformative
Other Income	D		Uninformative
Other Income from Operations	D		Uninformative
Personnel Expense	D		Redundant, since contained in *Personnel Expense/GLP*
Profit (Loss)	D		Redundant, since contained in *Profit Margin*
Recoveries on Loans Written Off	D		Redundant, since contained in *Loan Loss Rate*
Tax Expense	T	Tax Expense/Assets	*Tax Expense/Assets*
Income Statement Subtotals			
Net Income after Taxes and before Donations	D		Uninformative
Net Income before Taxes and Donations	D		Uninformative
Net Operating Income	D		Redundant, since contained in *Profit Margin*
Infrastructure			
Loan Officers	T	% Staff: Loan Officers	*Loan Officers/Personnel*
Loan Officers ▷ Gender ▷ Female	D		Redundant, since contained in *% Loan Officers: Female*
Number of Board Members	T	% Staff: Board Members	*Loan Officers/Personnel*
Number of Board Members ▷ Gender ▷ Female	D		Redundant, since contained in *% Board Members: Female*
Number of Managers	T	% Staff: Managers	*Loan Officers/Personnel*
Number of Managers ▷ Gender ▷ Female	D		Redundant, since contained in *% Managers: Female*
Number of Staff Employed for One Year or More	T	% Staff: Employed > 1 Year	*Loan Officers/Personnel*
Number of Staff Exiting the Organization during Period	T	% Staff: Leaving During Period	*Number of Staff Exiting the Organization during Period/Personnel*
Offices	T	Borrowers/Office	*# Active Borrowers/Office*
	T	Personnel/Office	*Personnel/Office*
Other Points of Service	D		Uninformative
Personnel	K	# Personnel	Number of individuals who are actively employed by an entity; includes contract employees or advisors who dedicate a substantial portion of their time to the entity, even if they are not on the entity's employees roster
Personnel ▷ Gender ▷ Female	D		Redundant, since contained in *% Staff: Female*

(Table continued on next page)

Table A.2: Details on the construction of the financial and social performance data *(continued)*

Original Variable	Action	New Variable	Definition of Reason for Deletion
Outreach			
Average Equity	D		Redundant, since similar to *Equity*
Average GLP	D		Redundant, since similar to *GLP*
Average Loan Balance per Borrower	D		Redundant, since included in *Av. Loan Bal./GNI p. c.*
Av. Loan Bal. per Borrower/GNI per Capita (ALBG)	K	Av. Loan Bal./GNI p. c.	*Average Loan Balance per Borrower/GNI per Capita*
	T	− log(ALBG)	*−log(Av. Loan Bal./GNI p. c.)*
Average Number of Active Borrowers	D		Redundant, since similar to *# Active Borrowers*
Average Number of Loans Outstanding	D		Redundant, since similar to *# Loans Outstanding*
Average Outstanding Balance	D		Redundant, since contained in *Av. Outst. Bal./GNI p. c.*
Average Outstanding Balance/GNI per Capita	K	Av. Outst. Bal./GNI p. c.	*Av. Outstanding Balance/GNI per Capita*
Percent of Female Borrowers	D		Redundant, since contained in *% Borrowers: Female*
Productivity & Efficiency			
Average Salary/GNI per Capita	K	Av. Salary/GNI p. c.	*Average Personnel Expense/GNI per Capita*
Borrowers per Loan Officer	K	Borrowers/Loan Officer	*NAB/Number of Loan Officers*
Borrowers per Staff Member	K	Borrowers/Staff Member	*NAB/Number of Personnel*
Cost per Borrower	K	Cost per Borrower	*Operating Expense/Average Number of Active Borrowers*
Cost per Loan	K	Cost per Loan	*Operating Expense / Average NLO*
Deposit Accounts per Staff Member	K	Deposit Accounts/Staff	*Number of Deposit Accounts/Number of Personnel*
Depositors per Staff Member	K	Depositors/Staff	*Number of Depositors/Number of Personnel*
Loans per Loan Officer	D		Redundant, since highly correlated with *Borrowers/Loan Officer*
Loans per Staff Member	D		Redundant, since highly correlated with *Borrowers/Staff Member*
Operating Expense/Loan Portfolio	K	Operating Expense/GLP	*Operating Expense/Average GLP*
Personnel Allocation Ratio	D		Redundant, equals *% Staff: Loan Officers*
Personnel Expense/Loan Portfolio	K	Personnel Expense/GLP	*Personnel Expense/Average GLP*
Revenue & Expenses			
Administrative Expense/Assets	K	Administrative Exp./Assets	*Administrative Expense/Average Total Assets*
Financial Expense/Assets	K	Financial Expense/Assets	*Financial Expense/Average Total Assets*
Financial Revenue/Assets	K	Financial Revenue/Assets	*Financial Revenue/Average Total Assets*
Operating Expense/Assets	K	Operating Expense/Assets	*Operating Expense/Average Total Assets*
Personnel Expense/Assets	K	Personnel Expense/Assets	*Personnel Expense/Average Total Assets*
Profit Margin	K	Profit Margin	*Net Operating Income/Financial Revenue*
Provision for Loan Impairment/Assets	K	Loan Impairm. Prov./Assets	*Impairment Losses on Loans/Average Total Assets*
Yield on Gross Portfolio (Nominal)	D		Redundant, since included in *Yield on GLP (Real)*

(Table continued on next page)

Table A.2: Details on the construction of the financial and social performance data *(continued)*

Original Variable	Action	New Variable	Definition of Reason for Deletion
Yield on Gross Portfolio (Real)	K	Yield on GLP (Real)	*(Financial Revenue from Loan Portfolio/Average GLP − Inflation Rate) / (1 + Inflation Rate)*
Risk & Liquidity			
Loan Loss Rate	K	Loan Loss Rate	*(Write-Offs − Value of Loans Recovered)/Average GLP*
Non-Earning Liquid Assets as a % of Total Assets	K	% Assets: Non-Earn. Liq. Ass.	*Cash and Cash Equivalents/Total Assets*
Portfolio at Risk ▷ 30 Days	K	Portfolio at Risk: > 30 Days	*(Portfolio Overdue > 30 Days + Renegotiated Portfolio)/GLP*
Portfolio at Risk ▷ 90 Days	K	Portfolio at Risk: > 90 Days	*(Portfolio Overdue > 90 Days + Renegotiated Portfolio)/GLP*
Risk Coverage	K	Risk Coverage	*Impairment Loss Allowance/Portfolio at Risk > 30 Days*
Write-Off Ratio	K	Write-Off Ratio	*Write-Offs/Average GLP*
Social Performance			
Borrower Retention Rate	K	Borrower Retention Rate	NAB at end of the reporting period (RP) divided by NAB at beginning of the RP and new borrowers during the RP
Percent of Female Board Members	K	% Board Members: Female	Number of Female Board Members per Board Member
Percent of Female Loan Officers	K	% Loan Officers: Female	Number of Female Loan Officers per Loan Officer
Percent of Female Managers	K	% Managers: Female	Number of Female Managers per Manager
Percent of Female Staff	K	% Staff: Female	Number of Female Staff per Staff Member
Staff Turnover Rate	D		Redundant, equals *% Staff: Leaving During Period*
		Social Performance Variables	
Audit & Rating			
All variables (18)	D		Very few observations = 1
Client Protection			
Clear Debt-Collection Practices ▷ [...] (4)	T	Client Prot.: Clear Debt Coll.	= 1, if clear debt collection practices are in place
Client Protection Assessment ▷ [...] (2)	D		Very few observations = 1
Full Disclosure of Prices, Terms, Conditions ▷ [...] (4)	T	Client Prot.: Full Disclosure	= 1, if prices, terms and conditions are fully disclosed
Functioning Client Complaint Mechanism ▷ [...] (4)	T	Client Prot.: Complaint Mech.	= 1, if the MFI has a functioning complaint mechanism
Interest Rate Calculation Method(s) ▷ [...] (2)	T	Client Prot.: Decl. Bal. Method	= 1, if declining balance interest method is used; = 0, if flat interest method is used
Internal Audits Verify Over-Intebt. Prev. ▷ [...] (4)	T	Client Prot.: Over-Indebt. Prev.	= 1, if internal audits verify over-indebtedness prevention
Privacy Data Clause in Loan Contracts ▷ [...] (4)	T	Client Prot.: Privacy Data Clause	= 1, if a privacy data clause is included in load contracts
Robust Repayment Evaluation ▷ [...] (4)	T	Client Prot.: Rob. Payment Eval.	= 1, if the MFI supports good repayment capacity analysis of its clients

(Table continued on next page)

Table A.2: Details on the construction of the financial and social performance data *(continued)*

Original Variable	Action	New Variable	Definition of Reason for Deletion
Written Policy on Client Collection Practices ▷ [. . .] (2)	D		All observations = 0
Environment			
All variables (5)	D		Uninformative
Governance & HR			
Bases for Staff Incentives ▷ None of the Above	D		Uninformative
Bases for Staff Incentives ▷ Num. of Clients ▷ [. . .] (5)	T	Incentives: # Clients	= 1, if the MFI has staff incentives related to the number of clients
Bases for Staff Incentives ▷ Portfolio Quality	K	Incentives: Portf. Qual.	= 1, if the MFI has staff incentives related to the portfolio quality
Bases for Staff Incentives ▷ Quality of Interaction with Clients Based on Client Feedback Mechanism	K	Incentives: Interaction Qual.	= 1, if the MFI has staff incentives related to the quality of the interaction with clients based on client feedback mechanism
Bases for Staff Incentives ▷ Quality of Soc. Data Coll.	K	Incentives: Data-Collect. Qual.	= 1, if the MFI has staff incentives related to the quality of the social data collected
Board Member with SP Educ., Work Exper. ▷ [. . .] (3)	D		Very few observations = 1
Board Orient. on Social Mission and Goals ▷ [. . .] (3)	T	Board: Social Training	= 1, if during orientation, board members are provided with training on the MFI's social mission and goals
Human Resource Policies in Place ▷ [. . .] (6)	D		After transformation as described in (A.1) more than 50% missing
SPM Champion, SPM Committee on Board ▷ [. . .] (3)	D		After transformation as described in (A.1) more than 50% missing
Products & Services			
Compulsory Insurance ▷ [. . .] (3)	T	Offers Compulsory Insurance	= 1, if there is any insurance offered by the institution or its partners that some or all clients are required to have, usually in order to access other financial products/services
Credit Product Offering ▷ [. . .] (13)	T	Offers Non-Inc. Gen. Loans Also	= 1, if the MFI offers income and non-income generating loans
Enterprise Services (Non-financial) ▷ [. . .] (3)	T	Offers Enterprise Services	= 1, if the MFI offers non-financial enterprise services
Health Services (Non-financial) ▷ [. . .] (3)	T	Offers Health Services	= 1, if the MFI offers non-financial health services
Offers Other Financial Services ▷ [. . .] (11)	T	Offers Other Fin. Services	= 1, if the MFI offers other financial services
Other Education Services (Non-financial) ▷ [. . .] (8)	T	Offers Education Services	= 1, if the MFI offers other non-financial education services
Savings Product Offering ▷ [. . .] (3)	T	Offers Savings Products	= 1, if the MFI offers savings products
Voluntary Insurance ▷ [. . .] (3)	T	Offers Voluntary Insurance	= 1, if the MFI offers voluntary insurance products
Women's Empow. Services (Non-financial) ▷ [. . .] (3)	T	Offers Women's Empow. Serv.	= 1, if the MFI offers women's empow. services such as leadership training, education, counselling and legal support

(Table continued on next page)

Table A.2: Details on the construction of the financial and social performance data *(continued)*

Original Variable	Action	New Variable	Definition of Reason for Deletion
Social Goals			
Development Goals ▷ Access to Water and Sanitation	T	Goals: Health Infrastructure	= 1, if the MFI follows at least one of these goals: access to water and sanitation, health improvement and housing
Development Goals ▷ Health Improvement	T		see *Goals: Health Infrastructure*
Development Goals ▷ Housing	T		see *Goals: Health Infrastructure*
Development Goals ▷ Children's Schooling	T	Goals: Education Opport.	= 1, if the MFI follows at least one of these goals: children's schooling, improv. of adult education and youth opp.
Development Goals ▷ Improvement of Adult Education	T		see *Goals: Education Opport.*
Development Goals ▷ Youth Opportunities	T		see *Goals: Education Opport.*
Development Goals ▷ Dev. of Start-Up Enterprises	T	Goals: Econ. Improvement	= 1, if the MFI follows at least one of these goals: development of startup enterprises, employment generation, growth of existing businesses, increased access to financial services and poverty reduction
Development Goals ▷ Employment Generation	T		*Goals: Econ. Improvement*
Development Goals ▷ Growth of Existing Businesses	T		*Goals: Econ. Improvement*
Development Goals ▷ Incr. Access to Financial Services	T		*Goals: Econ. Improvement*
Development Goals ▷ Poverty Reduction	T		*Goals: Econ. Improvement*
Development Goals ▷ Gender Equal., Women's Empow.	K	Goals: Women's Empow.	= 1, if the MFI follows the goal of gender equality and women's empowerment
Development Goals ▷ None of the Above	D		Uninformative
Measures Client Poverty ▷ [...] (13)	T	Measures Client Poverty	= 1, if the MFI measures client poverty
Poverty Target ▷ [...] (4)	T	Has Poverty Target	= 1, if the MFI targets the poor (low income, poor and/or very poor)
Target Market ▷ Adolescents and Youth (Below 18)	K	Target Market: Youth	= 1, if adolescents and youth (below 18) represent the MFI's target market
Target Market ▷ Clients Living in Urban Areas	K	Target Market: Rural	= 1, if clients living in rural areas represent the MFI's target market
Target Market ▷ Clients Living in Urban Areas	K	Target Market: Urban	= 1, if clients living in urban areas represent the MFI's target market
Target Market ▷ None of the Above	D		Uninformative
Target Market ▷ Women	K	Target Market: Women	= 1, if women represent the MFI's target market

Note: Variables are either kept (K), transformed (T) or deleted (D). The operator "▷" functions as a subcategory indicator in the variable names similar to the original data set. The colours of the new variables indicate the factors they belong to, see Figure 2.1. The descriptions of the variables are taken from the file *Financial Performance Field Definitions.xlsx* and *Social Performance Field Definitions.xlsx*[2].

A.2 Descriptive Statistics

Table A.3: Descriptive statistics of the balanced panel data

Variable	Mean	SD	Min.	Q1	Median	Q3	Max.	%NA (N_I)
Costs								
Administrative Exp./Assets	0.070	0.054	0.000	0.041	0.056	0.084	0.554	0.042 (0)
Cost per Borrower	215.845	237.845	5.000	93.000	164.000	271.000	4003.000	0.052 (0)
Cost per Loan	198.964	206.855	5.000	89.500	160.000	254.000	4003.000	0.052 (0)
Fin. Exp. Fund. Liab./Assets	0.054	0.033	-0.370	0.034	0.052	0.070	0.287	0.106 (0)
Financial Expense/Assets	0.058	0.033	-0.335	0.038	0.058	0.076	0.277	0.038 (0)
Interest Exp. on Borr./Borr.	0.094	0.095	0.000	0.060	0.080	0.107	1.876	0.063 (0)
Operating Expense/Assets	0.170	0.108	0.009	0.098	0.144	0.212	0.954	0.034 (0)
Operating Expense/GLP	0.225	0.164	0.012	0.122	0.182	0.265	1.543	0.029 (0)
Financial Performance								
Financial Revenue/Assets	0.273	0.110	0.018	0.197	0.254	0.336	0.779	0.033 (0)
Operational Self-Sufficiency	1.171	0.347	0.254	1.033	1.131	1.266	6.197	0.025 (0)
Profit Margin	0.094	0.235	-2.943	0.032	0.115	0.210	0.839	0.025 (0)
Return on Assets	0.022	0.062	-0.467	0.005	0.022	0.050	0.285	0.033 (0)
Return on Equity	0.096	0.864	-15.815	0.022	0.097	0.183	21.727	0.033 (0)
Yield on GLP (Real)	0.244	0.137	-0.109	0.150	0.219	0.311	0.954	0.028 (0)
Financing Structure								
Assets	145875055	471699925	0.000	7748768	25703522	89587892	6379061427	0.012 (0)
Borrowings/Liabilities	0.708	1.732	-15.490	0.394	0.822	0.927	57.734	0.027 (0)
Capital/Assets	0.299	0.208	-0.115	0.154	0.228	0.377	1.046	0.016 (0)
Debt/Equity	4.118	5.798	-52.200	1.585	3.330	5.415	91.320	0.021 (0)
Deposits/Loans	0.253	0.422	0.000	0.000	0.000	0.424	2.459	0.047 (0)
Deposits/Total Assets	0.174	0.267	0.000	0.000	0.000	0.330	0.860	0.050 (0)
Donated Equity/Equity	0.149	0.627	-10.622	0.000	0.000	0.152	6.818	0.225 (0)
Equity	27084545	90018899	-1.39M	2186532	6687075	19452421	1355761891	0.013 (0)
Liabilities	119117186	389128592	-873595	4138758	17663752	72606668	5292796853	0.017 (0)
Liabilities/Assets	0.699	0.209	-0.046	0.616	0.771	0.845	1.115	0.017 (0)
Paid in Capital/Equity	0.498	0.599	0.000	0.017	0.401	0.749	6.035	0.053 (0)
Subord. Debt/Equity	0.123	0.323	0.000	0.000	0.000	0.116	3.027	0.437 (0)
% Assets: Non-Earn. Liq. Ass.	0.138	0.115	0.000	0.060	0.108	0.187	0.906	0.026 (0)
Operational Aspects								
Av. Loan Bal./GNI p. c.	0.486	0.680	0.019	0.125	0.273	0.540	6.402	0.027 (0)
–log(Av. Loan Bal./GNI p. c.)	0.486	0.680	0.019	0.125	0.273	0.540	6.402	0.027 (0)
Av. Loan Size: Consumption	1079.433	2242.712	0.000	0.000	586.666	1335.592	43266.929	0.253 (0)
Av. Loan Size: Enterpr. Fin.	1775.797	8286.885	0.000	327.499	834.368	1778.467	267879	0.113 (0)
Av. Loan Size: Ext. Customers	1905.953	13276.366	0.000	347.977	848.579	1641.770	312840	0.275 (0)
Av. Loan Size: Female	1167.258	1894.212	1.282	301.417	726.389	1462.084	36629.286	0.188 (0)
Av. Loan Size: Household Fin.	1468.731	2611.795	0.000	112.985	822.570	1739.107	43266.929	0.172 (0)
Av. Loan Size: Individual	1903.496	4828.446	0.000	598.187	1151.442	1940.419	84483.000	0.153 (0)
Av. Loan Size: Large Corp.	71320.663	718619	0.000	0.000	0.000	0.000	13971589	0.397 (0)
Av. Loan Size: Legal Entity	16668.559	59861.489	0.000	0.000	0.000	5577.000	507244	0.378 (0)
Av. Loan Size: Male	1580.074	2209.341	0.000	484.581	1075.460	1905.439	33884.810	0.225 (0)
Av. Loan Size: Man. & Staff	2661.027	7427.704	0.000	0.000	922.132	2914.623	161351	0.379 (0)
Av. Loan Size: Microenterpr.	1259.924	1786.901	0.000	322.698	763.962	1542.150	19564.523	0.121 (0)
Av. Loan Size: Mortgage, Hous.	3949.476	11382.678	0.000	0.000	591.801	3014.007	187131	0.290 (0)
Av. Loan Size: Other Hous. Fin.	499.753	3052.449	0.000	0.000	0.000	119.300	78264.918	0.358 (0)
Av. Loan Size: Rural	1218.178	2058.439	0.000	285.413	736.877	1570.433	50000.000	0.228 (0)
Av. Loan Size: Sm. & Med. Ent.	8042.001	26370.362	0.000	0.000	0.000	4042.986	398981	0.445 (0)

(Table continued on next page)

Table A.3: Descriptive statistics of the balanced panel data *(continued)*

Variable	Mean	SD	Min.	Q1	Median	Q3	Max.	%NA (N_I)
Av. Loan Size: Solidarity Group	337.440	719.062	0.000	0.000	145.998	392.484	11129.320	0.214 (0)
Av. Loan Size: Urban	1567.969	2416.257	0.000	350.119	900.433	1844.695	26925.000	0.219 (0)
Av. Loan Size: Village (SHG)	215.923	861.000	0.000	0.000	0.000	156.575	8793.549	0.290 (0)
Av. Outst. Bal./GNI p. c.	0.456	0.636	0.019	0.117	0.260	0.511	6.105	0.031 (0)
Borrower Retention Rate	0.768	0.125	0.315	0.706	0.774	0.844	1.153	0.479 (0)
Client Prot.: Clear Debt Coll.	0.884	0.320	0.000	1.000	1.000	1.000	1.000	0.224 (0)
Client Prot.: Complaint Mech.	0.823	0.382	0.000	1.000	1.000	1.000	1.000	0.217 (0)
Client Prot.: Decl. Bal. Method	0.722	0.448	0.000	0.000	1.000	1.000	1.000	0.051 (0)
Client Prot.: Full Disclosure	0.989	0.105	0.000	1.000	1.000	1.000	1.000	0.085 (0)
Client Prot.: Over-Indebt. Prev.	0.882	0.323	0.000	1.000	1.000	1.000	1.000	0.477 (0)
Client Prot.: Privacy Data Clause	0.974	0.160	0.000	1.000	1.000	1.000	1.000	0.193 (0)
Client Prot.: Rob. Payment Eval.	0.980	0.141	0.000	1.000	1.000	1.000	1.000	0.068 (0)
Deposit Accounts/Staff	113.140	220.601	0.000	0.000	0.000	148.000	1927.000	0.189 (0)
GLP/Total Assets	0.805	0.294	0.118	0.727	0.820	0.892	9.834	0.015 (0)
Gross Loan Portfolio (GLP)	120082992	415099262	125432	5725471	20511460	72426248	6052089498	0.002 (0)
Imp. Loss Allow./GLP	0.039	0.032	0.000	0.017	0.033	0.054	0.279	0.025 (1)
Income on Pen. Fees/GLP	0.004	0.013	0.000	0.000	0.000	0.003	0.230	0.448 (0)
Interest Income on GLP/GLP	0.275	0.137	0.000	0.179	0.250	0.337	1.024	0.030 (0)
Loan Impairm. Prov./Assets	0.017	0.024	-0.065	0.005	0.012	0.023	0.395	0.036 (0)
Loan Loss Rate	0.017	0.031	-0.132	0.000	0.007	0.022	0.426	0.041 (0)
Measures Client Poverty	0.670	0.471	0.000	0.000	1.000	1.000	1.000	0.223 (0)
Net GLP/GLP	0.943	0.081	0.000	0.939	0.962	0.980	1.037	0.020 (0)
Portfolio at Risk: > 30 Days	0.048	0.063	0.000	0.012	0.032	0.062	0.924	0.061 (2)
Portfolio at Risk: > 90 Days	0.035	0.055	0.000	0.007	0.021	0.043	0.902	0.065 (2)
Retained Earnings/Equity	0.197	0.905	-17.648	0.049	0.190	0.575	11.622	0.030 (0)
Risk Coverage	7.370	77.566	0.000	0.632	1.013	1.641	1960.157	0.082 (1)
Tax Expense/Assets	0.005	0.012	-0.108	0.000	0.002	0.009	0.145	0.087 (0)
Write-Off Ratio	0.021	0.032	-0.011	0.002	0.010	0.027	0.426	0.077 (0)
# Active Borrowers	193978	783059	16.000	7247.500	25813.000	102302	8166287	0.024 (0)
log(NAB)	10.201	1.887	2.773	8.888	10.159	11.536	15.916	0.024 (0)
# Loans Outstanding	192522	792082	16.000	7295.250	25885.000	106668	8653095	0.028 (0)
# New Borrowers/NAB	0.401	0.216	0.000	0.269	0.368	0.501	1.461	0.470 (0)
% Borrowers: Ext.Customers	0.983	0.113	0.000	0.996	0.999	1.000	1.000	0.254 (1)
% Borrowers: Female	0.629	0.241	0.019	0.443	0.594	0.851	1.000	0.100 (2)
% Borrowers: Legal Entity	0.010	0.062	0.000	0.000	0.000	0.000	0.928	0.325 (0)
% Borrowers: Male	0.375	0.231	0.000	0.162	0.411	0.553	0.981	0.160 (0)
% Borrowers: Man. & Staff	0.009	0.072	0.000	0.000	0.001	0.005	1.000	0.351 (0)
% Borrowers: Rural	0.531	0.313	0.000	0.261	0.571	0.803	1.000	0.189 (0)
% Borrowers: Urban	0.480	0.316	0.000	0.201	0.440	0.751	1.000	0.184 (0)
% GLP: 1-3 Months Delinquent	0.013	0.030	0.000	0.003	0.008	0.016	0.873	0.070 (0)
% GLP: >3 Months Delinquent	0.025	0.034	0.000	0.005	0.015	0.033	0.372	0.254 (2)
% GLP: Consumption	0.089	0.139	0.000	0.000	0.021	0.131	1.000	0.209 (1)
% GLP: Enterprise Finance	0.836	0.222	0.000	0.749	0.928	1.000	1.000	0.104 (18)
% GLP: Household Financing	0.153	0.192	0.000	0.000	0.076	0.246	1.000	0.132 (1)
% GLP: Individual	0.671	0.385	0.000	0.282	0.895	1.000	1.000	0.135 (2)
% GLP: Large Corporations	0.020	0.064	0.000	0.000	0.000	0.000	0.556	0.314 (0)
% GLP: Microenterprise	0.783	0.260	0.000	0.631	0.890	0.999	1.000	0.102 (4)
% GLP: Mortgage, Housing	0.051	0.082	0.000	0.000	0.007	0.071	0.589	0.241 (1)
% GLP: Other Household Fin.	0.029	0.105	0.000	0.000	0.000	0.000	1.000	0.290 (1)
% GLP: Renegotiated Loans	0.011	0.035	0.000	0.000	0.000	0.007	0.719	0.158 (0)
% GLP: Small & Med. Enterpr.	0.079	0.169	0.000	0.000	0.000	0.052	0.999	0.387 (0)
% GLP: Solidarity Group	0.238	0.355	0.000	0.000	0.024	0.412	1.000	0.158 (1)

(Table continued on next page)

Table A.3: Descriptive statistics of the balanced panel data *(continued)*

Variable	Mean	SD	Min.	Q1	Median	Q3	Max.	%NA (N_I)
% GLP: Village Banking (SHG)	0.105	0.246	0.000	0.000	0.000	0.013	1.000	0.210 (1)
Organisational Strategy								
Board: Social Training	0.830	0.376	0.000	1.000	1.000	1.000	1.000	0.305 (0)
Goals: Econ. Improvement	0.897	0.304	0.000	1.000	1.000	1.000	1.000	0.048 (0)
Goals: Education Opport.	0.462	0.499	0.000	0.000	0.000	1.000	1.000	0.048 (0)
Goals: Health Infrastructure	0.565	0.496	0.000	0.000	1.000	1.000	1.000	0.048 (0)
Goals: Women's Empow.	0.602	0.490	0.000	0.000	1.000	1.000	1.000	0.048 (0)
Has Poverty Target	0.921	0.270	0.000	1.000	1.000	1.000	1.000	0.142 (0)
Incentives: Data-Collect. Qual.	0.156	0.363	0.000	0.000	0.000	0.000	1.000	0.000 (0)
Incentives: Interaction Qual.	0.258	0.438	0.000	0.000	0.000	1.000	1.000	0.000 (0)
Incentives: Portf. Qual.	0.761	0.426	0.000	1.000	1.000	1.000	1.000	0.000 (0)
Incentives: # Clients	0.721	0.449	0.000	0.000	1.000	1.000	1.000	0.000 (0)
Target Market: Rural	0.812	0.391	0.000	1.000	1.000	1.000	1.000	0.048 (0)
Target Market: Urban	0.772	0.419	0.000	1.000	1.000	1.000	1.000	0.048 (0)
Target Market: Women	0.825	0.380	0.000	1.000	1.000	1.000	1.000	0.048 (0)
Target Market: Youth	0.205	0.404	0.000	0.000	0.000	0.000	1.000	0.048 (0)
Personnel Structure								
Av. Salary/GNI p. c.	3.857	2.584	0.000	2.000	3.000	5.000	18.000	0.075 (0)
Borrowers/Loan Officer	381.506	1966.840	16.000	199.000	278.000	381.000	67418.000	0.067 (0)
Borrowers/Office	1836.485	3619.137	16.000	720.016	1214.400	2203.250	112698	0.042 (0)
Borrowers/Staff Member	139.785	102.755	4.000	78.000	114.000	169.000	928.000	0.037 (0)
Depositors/Staff	89.776	174.291	0.000	0.000	0.000	125.000	1692.000	0.186 (0)
Personnel/Office	13.404	11.454	0.160	7.000	10.455	15.797	142.045	0.035 (0)
Personnel Expense/Assets	0.100	0.064	0.000	0.055	0.086	0.129	0.606	0.053 (0)
Personnel Expense/GLP	0.131	0.095	0.000	0.069	0.108	0.162	0.980	0.050 (0)
# Personnel	855.004	2123.995	4.000	75.000	232.000	826.000	22733.000	0.023 (0)
% Board Members: Female	0.324	0.235	0.000	0.167	0.286	0.429	1.000	0.354 (0)
% Loan Officers: Female	0.364	0.274	0.000	0.125	0.346	0.526	1.000	0.322 (0)
% Managers: Female	0.359	0.270	0.000	0.143	0.333	0.500	1.000	0.326 (0)
% Staff: Board Members	0.083	0.139	0.000	0.007	0.030	0.091	1.000	0.310 (10)
% Staff: Employed > 1 Year	0.692	0.226	0.000	0.600	0.732	0.848	1.000	0.414 (4)
% Staff: Female	0.449	0.191	0.000	0.336	0.463	0.561	0.990	0.286 (2)
% Staff: Leaving During Period	0.211	0.178	0.000	0.091	0.172	0.275	1.255	0.344 (0)
% Staff: Loan Officers	0.446	0.145	0.001	0.344	0.437	0.537	0.975	0.056 (0)
% Staff: Managers	0.085	0.087	0.002	0.025	0.060	0.117	1.000	0.286 (0)
Services								
Offers Compulsory Insurance	0.566	0.496	0.000	0.000	1.000	1.000	1.000	0.214 (0)
Offers Education Services	0.625	0.484	0.000	0.000	1.000	1.000	1.000	0.105 (0)
Offers Enterprise Services	0.481	0.500	0.000	0.000	0.000	1.000	1.000	0.093 (0)
Offers Health Services	0.308	0.462	0.000	0.000	0.000	1.000	1.000	0.120 (0)
Offers Non-Inc. Gen. Loans Also	0.790	0.408	0.000	1.000	1.000	1.000	1.000	0.000 (0)
Offers Other Fin. Services	0.534	0.499	0.000	0.000	1.000	1.000	1.000	0.002 (0)
Offers Savings Products	0.412	0.492	0.000	0.000	0.000	1.000	1.000	0.094 (0)
Offers Voluntary Insurance	0.356	0.479	0.000	0.000	0.000	1.000	1.000	0.267 (0)
Offers Women's Empow. Serv.	0.525	0.500	0.000	0.000	1.000	1.000	1.000	0.102 (0)

Note: The balanced panel contains 1278 observations of 213 MFIs from 2009 to 2014. $Q1$ and $Q3$ denote the first (third) quartile. $\%NA$ gives the resulting proportion of missing values and N_I is the number of implausible values that were manually removed from the analysis.

Table A.4: Descriptive statistics of the unbalanced panel data

Variable	Mean	SD	Min.	Q1	Median	Q3	Max.	%NA (N_I)
Costs								
Administrative Exp./Assets	0.077	0.075	-0.050	0.037	0.056	0.090	1.179	0.105 (0)
Cost per Borrower	239.261	733.423	-339.0	45.000	146.000	268.000	34387.000	0.114 (0)
Cost per Loan	218.831	673.751	-312.0	45.000	143.000	254.000	32127.000	0.133 (0)
Fin. Exp. Fund. Liab./Assets	0.054	0.039	-0.370	0.031	0.051	0.071	0.605	0.108 (0)
Financial Expense/Assets	0.058	0.039	-0.335	0.033	0.056	0.079	0.878	0.096 (0)
Interest Exp. on Borr./Borr.	0.131	1.562	-0.121	0.056	0.080	0.113	90.911	0.106 (0)
Operating Expense/Assets	0.183	0.161	-0.120	0.092	0.137	0.220	3.091	0.093 (0)
Operating Expense/GLP	0.261	0.537	-0.143	0.115	0.178	0.291	26.522	0.078 (0)
Financial Performance								
Financial Revenue/Assets	0.275	0.139	0.000	0.188	0.242	0.330	2.152	0.092 (0)
Operational Self-Sufficiency	1.152	0.947	-47.845	1.016	1.125	1.271	15.619	0.050 (0)
Profit Margin	0.024	0.780	-23.081	0.016	0.112	0.214	3.002	0.051 (1)
Return on Assets	0.012	0.124	-3.453	0.003	0.022	0.048	0.587	0.093 (0)
Return on Equity	0.131	4.817	-31.692	0.015	0.094	0.194	276.974	0.093 (0)
Yield on GLP (Real)	0.246	0.177	-0.216	0.136	0.207	0.316	2.772	0.094 (0)
Financing Structure								
Assets	99337082	358869905	0.000	3976317	14000117	54218327	6379061427	0.033 (0)
Borrowings/Liabilities	0.632	1.056	-15.490	0.303	0.755	0.925	57.734	0.061 (0)
Capital/Assets	0.302	0.273	-1.871	0.150	0.235	0.406	7.116	0.039 (0)
Debt/Equity	4.623	25.721	-219.24	1.372	3.125	5.540	1314.180	0.044 (0)
Deposits/Loans	0.692	22.380	0.000	0.000	0.000	0.506	1326.968	0.084 (0)
Deposits/Total Assets	0.199	0.271	0.000	0.000	0.000	0.377	1.740	0.094 (0)
Donated Equity/Equity	0.192	0.902	-10.622	0.000	0.000	0.181	31.525	0.261 (0)
Equity	18946198	68282603	-112M	1049344	3756582	12372410	1355761891	0.034 (0)
Liabilities	80714950	300675327	-873595	2182496	9516400	40580306	5292796853	0.038 (0)
Liabilities/Assets	0.700	0.251	-0.046	0.594	0.765	0.850	2.911	0.038 (0)
Paid in Capital/Equity	0.468	3.278	-91.559	0.004	0.390	0.749	98.477	0.118 (0)
Subord. Debt/Equity	0.190	2.446	-30.950	0.000	0.000	0.097	92.289	0.511 (0)
% Assets: Non-Earn. Liq. Ass.	0.141	0.118	0.000	0.058	0.112	0.191	0.906	0.054 (1)
Operational Aspects								
Av. Loan Bal./GNI p. c.	0.734	6.762	0.001	0.114	0.254	0.608	314.829	0.062 (0)
–log(Av. Loan Bal./GNI p. c.)	0.734	6.762	0.001	0.114	0.254	0.608	314.829	0.062 (0)
Av. Loan Size: Consumption	1638.918	33973.941	0.000	0.000	326.047	1104.194	1664883	0.373 (0)
Av. Loan Size: Enterpr. Fin.	32912.708	1477197	0.000	202.747	545.855	1571.824	81475549	0.177 (0)
Av. Loan Size: Ext. Customers	15207.429	656034	0.000	218.777	630.122	1552.782	33139368	0.336 (0)
Av. Loan Size: Female	10802.307	480213	0.000	204.782	542.568	1363.410	25601792	0.259 (0)
Av. Loan Size: Household Fin.	6071.633	250190	0.000	0.000	556.949	1581.816	13304450	0.265 (0)
Av. Loan Size: Individual	8470.136	271966	0.000	350.751	958.311	1822.463	13304450	0.253 (0)
Av. Loan Size: Large Corp.	196356	6711039	0.000	0.000	0.000	0.000	298311856	0.484 (0)
Av. Loan Size: Legal Entity	259324	10863363	0.000	0.000	0.000	5438.531	488853296	0.473 (0)
Av. Loan Size: Male	12908.885	556334	0.000	327.653	886.468	1851.462	28555736	0.314 (0)
Av. Loan Size: Man. & Staff	139782	6163441	0.000	0.000	609.641	2592.123	278109344	0.471 (0)
Av. Loan Size: Microenterpr.	2890.337	90831.679	0.000	199.227	511.708	1352.959	5085206	0.184 (0)
Av. Loan Size: Mortgage, Hous.	3237.727	9448.588	0.000	0.000	145.989	2175.760	187131	0.399 (0)
Av. Loan Size: Other Hous. Fin.	11008.780	484076	0.000	0.000	0.000	174.967	22558854	0.435 (0)
Av. Loan Size: Rural	1876.984	16346.231	0.000	178.279	513.124	1404.163	726426	0.309 (0)
Av. Loan Size: Sm. & Med. Ent.	34665.040	1234407	0.000	0.000	0.000	4515.635	54620289	0.491 (0)
Av. Loan Size: Solidarity Group	1349.567	28279.829	0.000	0.000	130.665	311.126	1272396	0.321 (0)
Av. Loan Size: Urban	17022.950	771073	0.000	210.577	669.876	1740.305	39820368	0.306 (0)

(Table continued on next page)

Table A.4: Descriptive statistics of the unbalanced panel data *(continued)*

Variable	Mean	SD	Min.	Q1	Median	Q3	Max.	%NA (N_I)
Av. Loan Size: Village (SHG)	526.711	7620.074	0.000	0.000	0.000	0.000	222135	0.405 (0)
Av. Outst. Bal./GNI p. c.	0.620	3.077	0.001	0.107	0.238	0.575	112.774	0.072 (0)
Borrower Retention Rate	0.759	0.156	0.000	0.685	0.772	0.848	1.829	0.530 (0)
Client Prot.: Clear Debt Coll.	0.892	0.310	0.000	1.000	1.000	1.000	1.000	0.206 (0)
Client Prot.: Complaint Mech.	0.822	0.383	0.000	1.000	1.000	1.000	1.000	0.209 (0)
Client Prot.: Decl. Bal. Method	0.693	0.461	0.000	0.000	1.000	1.000	1.000	0.084 (0)
Client Prot.: Full Disclosure	0.982	0.131	0.000	1.000	1.000	1.000	1.000	0.124 (0)
Client Prot.: Over-Indebt. Prev.	0.892	0.311	0.000	1.000	1.000	1.000	1.000	0.331 (0)
Client Prot.: Privacy Data Clause	0.944	0.230	0.000	1.000	1.000	1.000	1.000	0.198 (0)
Client Prot.: Rob. Payment Eval.	0.973	0.162	0.000	1.000	1.000	1.000	1.000	0.117 (0)
Deposit Accounts/Staff	166.903	540.978	0.000	0.000	0.000	201.000	17149.000	0.216 (0)
GLP/Total Assets	0.821	0.621	0.001	0.707	0.810	0.888	20.503	0.039 (0)
Gross Loan Portfolio (GLP)	80145135	299841064	27.000	2990192	10419495	41555434	6052089498	0.006 (0)
Imp. Loss Allow./GLP	0.041	0.058	0.000	0.013	0.029	0.051	1.000	0.060 (8)
Income on Pen. Fees/GLP	0.005	0.016	-0.001	0.000	0.000	0.003	0.300	0.497 (0)
Interest Income on GLP/GLP	0.277	0.168	0.000	0.174	0.239	0.329	2.767	0.057 (0)
Loan Impairm. Prov./Assets	0.018	0.038	-0.336	0.003	0.010	0.023	0.623	0.103 (0)
Loan Loss Rate	0.025	0.471	-0.532	0.000	0.005	0.019	25.708	0.106 (0)
Measures Client Poverty	0.633	0.482	0.000	0.000	1.000	1.000	1.000	0.168 (0)
Net GLP/GLP	0.941	0.183	0.000	0.940	0.967	0.984	9.890	0.044 (2)
Portfolio at Risk: > 30 Days	0.059	0.102	0.000	0.011	0.032	0.065	1.000	0.108 (3)
Portfolio at Risk: > 90 Days	0.043	0.089	0.000	0.006	0.020	0.045	0.999	0.113 (3)
Retained Earnings/Equity	0.180	4.950	-191.25	0.037	0.195	0.577	113.019	0.061 (0)
Risk Coverage	6.733	65.283	0.000	0.562	0.954	1.558	1960.157	0.165 (8)
Tax Expense/Assets	0.005	0.012	-0.108	0.000	0.001	0.008	0.145	0.153 (0)
Write-Off Ratio	0.021	0.044	-0.011	0.001	0.008	0.025	0.871	0.160 (1)
# Active Borrowers	137047	570871	3.000	4923.250	17666.000	70320.500	8166287	0.040 (0)
log(NAB)	9.793	2.031	1.099	8.502	9.779	11.161	15.916	0.040 (0)
# Loans Outstanding	143467	594428	3.000	5041.500	18108.000	73819.000	8653095	0.050 (0)
# New Borrowers/NAB	324.949	14187.648	0.000	0.246	0.374	0.530	621348	0.501 (0)
% Borrowers: Ext.Customers	0.980	0.122	0.000	0.995	1.000	1.000	1.000	0.311 (5)
% Borrowers: Female	0.656	0.268	0.000	0.442	0.649	0.930	1.000	0.140 (4)
% Borrowers: Legal Entity	0.022	0.104	0.000	0.000	0.000	0.001	1.000	0.421 (1)
% Borrowers: Male	0.355	0.253	0.000	0.106	0.370	0.551	1.000	0.235 (0)
% Borrowers: Man. & Staff	0.012	0.085	0.000	0.000	0.001	0.006	1.000	0.442 (0)
% Borrowers: Rural	0.547	0.321	0.000	0.285	0.595	0.811	1.000	0.254 (0)
% Borrowers: Urban	0.483	0.324	0.000	0.211	0.438	0.766	1.000	0.255 (1)
% GLP: 1-3 Months Delinquent	0.017	0.042	0.000	0.002	0.007	0.017	1.000	0.124 (1)
% GLP: >3 Months Delinquent	0.035	0.079	0.000	0.005	0.016	0.036	0.989	0.265 (3)
% GLP: Consumption	0.092	0.169	0.000	0.000	0.009	0.120	1.000	0.325 (2)
% GLP: Enterprise Finance	0.835	0.254	0.000	0.767	0.958	1.000	1.000	0.158 (35)
% GLP: Household Financing	0.164	0.238	0.000	0.000	0.057	0.241	1.000	0.216 (2)
% GLP: Individual	0.640	0.413	0.000	0.152	0.898	1.000	1.000	0.236 (5)
% GLP: Large Corporations	0.021	0.086	0.000	0.000	0.000	0.000	1.000	0.409 (0)
% GLP: Microenterprise	0.771	0.295	0.000	0.628	0.913	1.000	1.000	0.159 (9)
% GLP: Mortgage, Housing	0.054	0.110	0.000	0.000	0.000	0.061	1.000	0.349 (2)
% GLP: Other Household Fin.	0.044	0.147	0.000	0.000	0.000	0.003	1.000	0.371 (2)
% GLP: Renegotiated Loans	0.010	0.045	0.000	0.000	0.000	0.004	0.994	0.233 (0)
% GLP: Small & Med. Enterpr.	0.098	0.200	0.000	0.000	0.000	0.091	1.000	0.437 (1)
% GLP: Solidarity Group	0.296	0.399	0.000	0.000	0.027	0.649	1.000	0.273 (7)
% GLP: Village Banking (SHG)	0.105	0.265	0.000	0.000	0.000	0.000	1.000	0.339 (2)

(Table continued on next page)

Table A.4: Descriptive statistics of the unbalanced panel data *(continued)*

Variable	Mean	SD	Min.	Q1	Median	Q3	Max.	%NA (N_I)
Organisational Strategy								
Board: Social Training	0.763	0.425	0.000	1.000	1.000	1.000	1.000	0.243 (0)
Goals: Econ. Improvement	0.926	0.262	0.000	1.000	1.000	1.000	1.000	0.027 (0)
Goals: Education Opport.	0.453	0.498	0.000	0.000	0.000	1.000	1.000	0.027 (0)
Goals: Health Infrastructure	0.516	0.500	0.000	0.000	1.000	1.000	1.000	0.027 (0)
Goals: Women's Empow.	0.569	0.495	0.000	0.000	1.000	1.000	1.000	0.027 (0)
Has Poverty Target	0.880	0.324	0.000	1.000	1.000	1.000	1.000	0.087 (0)
Incentives: Data-Collect. Qual.	0.151	0.358	0.000	0.000	0.000	0.000	1.000	0.002 (0)
Incentives: Interaction Qual.	0.254	0.435	0.000	0.000	0.000	1.000	1.000	0.002 (0)
Incentives: Portf. Qual.	0.771	0.420	0.000	1.000	1.000	1.000	1.000	0.002 (0)
Incentives: # Clients	0.722	0.448	0.000	0.000	1.000	1.000	1.000	0.002 (0)
Target Market: Rural	0.808	0.394	0.000	1.000	1.000	1.000	1.000	0.027 (0)
Target Market: Urban	0.752	0.432	0.000	1.000	1.000	1.000	1.000	0.027 (0)
Target Market: Women	0.829	0.377	0.000	1.000	1.000	1.000	1.000	0.027 (0)
Target Market: Youth	0.193	0.395	0.000	0.000	0.000	0.000	1.000	0.027 (0)
Personnel Structure								
Av. Salary/GNI p. c.	4.034	4.768	-2.000	2.000	3.000	5.000	121.000	0.147 (0)
Borrowers/Loan Officer	345.380	1191.839	3.000	175.000	260.000	377.750	67418.000	0.084 (0)
Borrowers/Office	1654.143	2677.092	3.000	639.562	1138.882	1950.440	112698	0.066 (0)
Borrowers/Staff Member	137.819	108.766	1.000	70.000	112.000	174.000	1311.000	0.060 (0)
Depositors/Staff	132.493	499.201	0.000	0.000	0.000	170.750	17149.000	0.208 (0)
Personnel/Office	13.007	11.682	0.160	6.400	9.648	15.122	142.045	0.064 (0)
Personnel Expense/Assets	0.105	0.100	-0.070	0.051	0.080	0.128	1.912	0.113 (0)
Personnel Expense/GLP	0.144	0.193	-0.083	0.064	0.102	0.167	4.461	0.099 (0)
# Personnel	682.640	1822.131	2.000	51.250	167.000	534.000	26749.000	0.042 (0)
% Board Members: Female	0.311	0.240	0.000	0.143	0.286	0.429	1.000	0.379 (0)
% Loan Officers: Female	0.358	0.286	0.000	0.111	0.320	0.523	1.000	0.339 (0)
% Managers: Female	0.336	0.288	0.000	0.098	0.273	0.500	1.000	0.342 (0)
% Staff: Board Members	0.094	0.150	0.000	0.011	0.036	0.107	1.000	0.332 (57)
% Staff: Employed > 1 Year	0.691	0.255	0.000	0.578	0.746	0.876	1.000	0.419 (11)
% Staff: Female	0.422	0.221	0.000	0.269	0.425	0.558	1.000	0.300 (2)
% Staff: Leaving During Period	0.219	0.221	0.000	0.077	0.162	0.285	2.471	0.355 (0)
% Staff: Loan Officers	0.453	0.158	0.000	0.346	0.455	0.558	1.000	0.074 (2)
% Staff: Managers	0.116	0.106	0.000	0.035	0.091	0.167	1.000	0.295 (0)
Services								
Offers Compulsory Insurance	0.566	0.496	0.000	0.000	1.000	1.000	1.000	0.168 (0)
Offers Education Services	0.599	0.490	0.000	0.000	1.000	1.000	1.000	0.085 (0)
Offers Enterprise Services	0.458	0.498	0.000	0.000	0.000	1.000	1.000	0.075 (0)
Offers Health Services	0.244	0.430	0.000	0.000	0.000	0.000	1.000	0.091 (0)
Offers Non-Inc. Gen. Loans Also	0.714	0.452	0.000	0.000	1.000	1.000	1.000	0.001 (0)
Offers Other Fin. Services	0.460	0.498	0.000	0.000	0.000	1.000	1.000	0.004 (0)
Offers Savings Products	0.447	0.497	0.000	0.000	0.000	1.000	1.000	0.064 (0)
Offers Voluntary Insurance	0.331	0.471	0.000	0.000	0.000	1.000	1.000	0.199 (0)
Offers Women's Empow. Serv.	0.480	0.500	0.000	0.000	0.000	1.000	1.000	0.093 (0)

Note: The unbalanced panel contains 3846 observations of 1026 MFIs over the years 2007 and 2018. *Q1* and *Q3* denote the first (third) quartile. *%NA* gives the resulting proportion of missing values and N_I is the number of implausible values that were manually removed from the analysis.

A.3 Detailed aMIRL Results

Table A.5: Detailed aMIRL results for OSS

Variable	BIC			AIC			C_p		
	$\hat{\pi}$	b_{aMIRL}	$\text{sd}_{b_{\text{aMIRL}}}$	$\hat{\pi}$	b_{aMIRL}	$\text{sd}_{b_{\text{aMIRL}}}$	$\hat{\pi}$	b_{aMIRL}	$\text{sd}_{b_{\text{aMIRL}}}$
Costs									
Cost per Borrower	0.784	−0.016	0.031	0.780	−0.044	0.061	0.799	−0.041	0.061
Fin. Exp. Fund. Liab. / Assets	0.188	0.002	0.027	0.701	0.051•••	0.081	0.622	0.043	0.077
Financial Expense / Assets	0.998	−0.095	0.071	0.982	−0.183	0.108	0.980	−0.172	0.107
Operating Expense / Assets	0.696	−0.042	0.154	0.866	−0.229	0.232	0.864	−0.211	0.230
Operating Expense / GLP	0.891	0.053	0.126	0.826	0.111	0.175	0.836	0.102	0.171
Financial Performance									
Financial Revenue / Assets	0.931	0.064	0.161	0.972	0.263	0.201	0.970	0.243	0.209
Profit Margin	1.000	0.481	0.127	1.000	0.484	0.118	1.000	0.482	0.122
Return on Assets	0.936	0.154	0.133	0.913	0.070	0.143	0.919	0.079	0.148
Return on Equity	0.939	−0.028	0.045	0.508	−0.015	0.033	0.602	−0.018	0.037
Financing Structure									
Deposits / Loans	0.538	0.010	0.034	0.713	0.060	0.081	0.662	0.050	0.080
Operational Aspects									
Av. Loan Size: Urban	0.942	−0.095•••	0.098	0.907	−0.106•••	0.113	0.907	−0.103•••	0.110
Loan Impairm. Prov. / Assets	0.940	0.028•••	0.059	0.477	0.001°	0.050	0.556	0.001°	0.053
Tax Expense / Assets	0.998	0.078	0.046	0.872	0.059	0.045	0.916	0.063	0.045
Write-Off Ratio	0.904	0.026	0.047	0.776	0.053•••	0.082	0.779	0.047•••	0.075
Personnel Structure									
% Staff: Female	0.343	−0.003	0.023	0.809	−0.073•••	0.082	0.778	−0.066	0.080
% Staff: Managers	0.965	−0.087•••	0.079	0.863	−0.075•••	0.081	0.901	−0.078•••	0.081
Optimal threshold $\hat{\pi}^*$		0.936			0.701			0.779	
# variables in the final model		8			14			11	
Overall $R^2_{\text{av.}}$ (adjusted $\bar{R}^2_{\text{av.}}$)		0.878 (0.853)			0.880 (0.854)			0.878 (0.853)	
Within $R^2_{\text{av.}}$ (adjusted $\bar{R}^2_{\text{av.}}$)		0.541 (0.445)			0.547 (0.450)			0.541 (0.444)	

Note: The table contains the empirical selection probabilities and aMIRL coefficients of all variables that were selected to at least one of the three final models. $\text{sd}_{b_{\text{aMIRL}}}$ denotes the empirical standard deviation of the initial estimates that the aMIRL estimates are computed from, see Equation 2.5. All effects are significant at the 1% level based on bootstrap confidence intervals, unless stated otherwise: ° implies that that the respective effect is not significant at the 10% level. The coloured bullets ••• indicate the signs (• positive, • negative) of the postimputation and postselection quantile regression estimates for $\tau = 0.25, 0.5, 0.75$, where at least one of them differs from that of the aMIRL estimate, see Section 2.4.3. Grey font symbolises that the respective variable is not included in the final model. $R^2_{\text{av.}}$ ($\bar{R}^2_{\text{av.}}$) denotes the average (adjusted) R^2 across all 10 imputed data sets.

Table A.6: Detailed aMIRL results for log(NAB)

Variable	BIC $\hat{\pi}$	BIC b_{aMIRL}	BIC $\text{sd}_{b_{\text{aMIRL}}}$	AIC $\hat{\pi}$	AIC b_{aMIRL}	AIC $\text{sd}_{b_{\text{aMIRL}}}$	C_p $\hat{\pi}$	C_p b_{aMIRL}	C_p $\text{sd}_{b_{\text{aMIRL}}}$
Costs									
Cost per Borrower	0.850	−0.026	0.028	0.630	−0.024	0.027	0.688	−0.028	0.028
Fin. Exp. Fund. Liab. / Assets	0.242	0.001	0.003	0.627	0.011	0.024	0.568	0.009	0.021
Financial Expense / Assets	0.056	0.000	0.002	0.682	−0.015	0.032	0.620	−0.012	0.028
Financial Performance									
Return on Assets	0.857	0.026	0.018	0.742	0.024	0.022	0.716	0.024	0.021
Yield on GLP (Real)	0.040	0.000°	0.001	0.530	0.005	0.012	0.479	0.005	0.012
Financing Structure									
Capital / Assets	0.747	−0.037	0.038	0.755	−0.053	0.107	0.698	−0.047	0.112
Liabilities	0.473	0.008	0.024	0.579	0.022	0.072	0.499	0.017	0.049
Operational Aspects									
Av. Loan Bal. / GNI p. c.	1.000	−0.098	0.038	0.992	−0.201	0.112	0.995	−0.193	0.111
Av. Outst. Bal. / GNI p. c.	0.042	0.001^{**}	0.024	0.917	0.097	0.118	0.899	0.092	0.117
Deposit Accounts / Staff	0.962	0.043	0.020	0.661	0.032	0.027	0.703	0.034	0.026
Loan Impairm. Prov. / Assets	0.892	0.019	0.014	0.609	0.015	0.015	0.665	0.016	0.015
Portfolio at Risk: > 30 Days	0.031	0.000	0.002	0.670	0.000°	0.038	0.511	−0.004	0.013
% Borrowers: Urban	0.933	−0.033	0.024	0.704	−0.026	0.024	0.766	−0.029	0.023
% GLP: Enterprise Finance	0.007	0.000°	0.000	0.823	0.023	0.033	0.775	0.018	0.030
% GLP: Microenterprise	0.526	−0.006	0.009	0.907	−0.043	0.043	0.859	−0.035	0.040
% GLP: Small & Med. Enterpr.	0.270	0.001	0.004	0.556	−0.010	0.021	0.419	−0.007	0.019
% GLP: Solidarity Group	0.693	0.009	0.018	0.528	0.013	0.021	0.561	0.012	0.021
Organisational Strategy									
Board: Social Training	0.808	−0.017	0.011	0.398	−0.009	0.012	0.457	−0.010	0.012
Goals: Econ. Improvement	0.751	−0.012	0.015	0.508	−0.011	0.014	0.549	−0.012	0.015
Goals: Health Infrastructure	0.821	0.016	0.014	0.524	0.012	0.014	0.595	0.014	0.014
Incentives: Portf. Qual.	0.874	0.023	0.014	0.483	0.016	0.018	0.556	0.018	0.018
Incentives: # Clients	0.726	0.012	0.014	0.385	0.010	0.015	0.409	0.010	0.015
Target Market: Youth	0.906	0.025	0.012	0.485	0.014	0.016	0.572	0.017	0.016
Personnel Structure									
Av. Salary / GNI p. c.	0.416	0.004	0.015	0.757	0.032	0.025	0.820	0.033	0.024
Borrowers / Office	0.852	0.018	0.029	0.420	0.009	0.024	0.474	0.009	0.023
Borrowers / Staff Member	1.000	0.231	0.049	0.997	0.229	0.046	0.998	0.231	0.045
Personnel Expense / GLP	0.227	−0.002	0.011	0.710	−0.028	0.036	0.556	−0.020	0.031
# Personnel	1.000	0.272	0.075	0.998	0.270	0.084	0.999	0.272	0.081
% Staff: Board Members	1.000	−0.168	0.052	0.992	−0.179	0.057	0.999	−0.178	0.054
% Staff: Managers	0.849	−0.017	0.016	0.505	−0.011	0.014	0.576	−0.013	0.015
Services									
Offers Other Fin. Services	0.762	0.017	0.021	0.523	0.016	0.019	0.614	0.019	0.019

(Table continued on next page)

Table A.6: Detailed aMIRL results for log(NAB) *(continued)*

Variable	BIC			AIC			C_p		
	$\hat{\pi}$	b_{aMIRL}	$\text{sd}_{b_{\text{aMIRL}}}$	$\hat{\pi}$	b_{aMIRL}	$\text{sd}_{b_{\text{aMIRL}}}$	$\hat{\pi}$	b_{aMIRL}	$\text{sd}_{b_{\text{aMIRL}}}$
Optimal threshold $\hat{\pi}^*$		0.726			0.483			0.549	
# variables in the final model		19			28			24	
Overall $R^2_{\text{av.}}$ (adjusted $\bar{R}^2_{\text{av.}}$)		0.979 (0.975)			0.979 (0.975)			0.979 (0.975)	
Within $R^2_{\text{av.}}$ (adjusted $\bar{R}^2_{\text{av.}}$)		0.529 (0.425)			0.528 (0.418)			0.526 (0.418)	

Note: The table contains the empirical selection probabilities and aMIRL coefficients of all variables that were selected to at least one of the three final models. $\text{sd}_{b_{\text{aMIRL}}}$ denotes the empirical standard deviation of the initial estimates that the aMIRL estimates are computed from, see Equation 2.5. All effects are significant at the 1% level based on bootstrap confidence intervals, unless stated otherwise: ** denotes that the respective effect is significant at the 5% level, $^{\circ}$ implies that that the respective effect is not significant at the 10% level. Their signs coincide with those of the postimputation and postselection quantile regression estimates for $\tau = 0.25, 0.5, 0.75$, see Section 2.4.3. Grey font symbolises that the respective variable is not included in the final model. $R^2_{\text{av.}}$ ($\bar{R}^2_{\text{av.}}$) denotes the average (adjusted) R^2 across all 10 imputed data sets.

Table A.7: Detailed aMIRL results for –log(ALBG)

Variable	BIC			AIC			C_p		
	$\hat{\pi}$	b_{aMIRL}	$\text{sd}_{b_{\text{aMIRL}}}$	$\hat{\pi}$	b_{aMIRL}	$\text{sd}_{b_{\text{aMIRL}}}$	$\hat{\pi}$	b_{aMIRL}	$\text{sd}_{b_{\text{aMIRL}}}$
Costs									
Cost per Borrower	0.927	−0.057	0.048	0.830	−0.069	0.059	0.859	−0.071	0.056
Cost per Loan	0.944	−0.060	0.030	0.680	−0.045	0.043	0.748	−0.045	0.041
Fin. Exp. Fund. Liab. / Assets	0.998	0.171	0.060	0.982	0.180	0.053	0.991	0.184	0.050
Financial Expense / Assets	0.997	−0.179	0.065	0.987	−0.189	0.058	0.992	−0.193	0.055
Operating Expense / GLP	0.437	0.009	0.025	0.640	0.037	0.083	0.647	0.034	0.074
Financial Performance									
Profit Margin	0.779	−0.021	0.029	0.682	−0.031	0.039	0.724	−0.031	0.038
Return on Assets	0.297	0.004	0.020	0.736	0.034	0.041	0.702	0.033	0.040
Return on Equity	0.695	0.019	0.021	0.375	0.009	0.017	0.392	0.010	0.017
Financing Structure									
Assets	0.440	−0.012	0.025	0.865	−0.071	0.139	0.812	−0.052	0.094
Operational Aspects									
Av. Loan Size: Microenterpr.	0.828	−0.037	0.030	0.575	−0.029	0.032	0.636	−0.031	0.033
Av. Loan Size: Urban	0.783	−0.032	0.026	0.544	−0.027	0.031	0.556	−0.027	0.031
Borrower Retention Rate	0.681	0.014	0.015	0.533	0.013	0.015	0.588	0.015	0.016
GLP / Total Assets	0.837	−0.039	0.042	0.636	−0.033	0.043	0.705	−0.036	0.043
Gross Loan Portfolio (GLP)	0.153	−0.002	0.011	0.597	0.021	0.134	0.378	0.014	0.102
Interest Income on GLP / GLP	0.937	0.050	0.032	0.776	0.048	0.039	0.808	0.050	0.038
Portfolio at Risk: > 90 Days	0.514	0.009	0.019	0.849	0.029	0.038	0.854	0.034	0.038
# Active Borrowers	0.301	−0.010	0.032	0.823	−0.091	0.095	0.821	−0.083	0.094

(Table continued on next page)

Table A.7: Detailed aMIRL results for –log(ALBG) *(continued)*

Variable	BIC $\hat{\pi}$	BIC b_{aMIRL}	BIC $\text{sd}_{b_{\text{aMIRL}}}$	AIC $\hat{\pi}$	AIC b_{aMIRL}	AIC $\text{sd}_{b_{\text{aMIRL}}}$	C_p $\hat{\pi}$	C_p b_{aMIRL}	C_p $\text{sd}_{b_{\text{aMIRL}}}$
# New Borrowers / NAB	0.504	0.007	0.016	0.535	0.010	0.015	0.563	0.012	0.016
% Borrowers: Female	0.777	0.040	0.046	0.427	0.030	0.061	0.476	0.028	0.068
% Borrowers: Male	0.737	−0.029	0.040	0.510	−0.022	0.060	0.580	−0.029	0.066
% GLP: Individual	0.053	0.000°	0.008	0.717	0.028	0.056	0.668	0.027	0.053
% GLP: Renegotiated Loans	0.257	−0.002	0.012	0.590	−0.015	0.025	0.597	−0.016	0.027
% GLP: Solidarity Group	0.416	0.006	0.015	0.707	0.041	0.056	0.708	0.038	0.053
% GLP: Village Banking (SHG)	0.528	0.011	0.018	0.685	0.039	0.045	0.711	0.039	0.043
Personnel Structure									
Av. Salary / GNI p. c.	0.999	−0.187	0.029	0.982	−0.194	0.040	0.991	−0.195	0.035
Borrowers / Staff Member	1.000	0.222	0.047	0.995	0.247	0.061	0.996	0.246	0.060
Personnel Expense / Assets	0.979	0.096	0.055	0.887	0.102	0.080	0.913	0.101	0.072
Personnel Expense / GLP	0.987	0.142	0.065	0.868	0.130	0.100	0.917	0.132	0.093
# Personnel	0.013	0.000°	0.002	0.836	0.058	0.064	0.771	0.050	0.063
% Staff: Managers	0.840	−0.028	0.027	0.582	−0.021	0.025	0.646	−0.023	0.025
Optimal threshold $\hat{\pi}^*$		0.681			0.544			0.556	
# variables in the final model		18			25			27	
Overall $R^2_{\text{av.}}$ (adjusted $\bar{R}^2_{\text{av.}}$)		0.975 (0.969)			0.975 (0.969)			0.975 (0.970)	
Within $R^2_{\text{av.}}$ (adjusted $\bar{R}^2_{\text{av.}}$)		0.432 (0.307)			0.438 (0.310)			0.447 (0.320)	

Note: The table contains the empirical selection probabilities and aMIRL coefficients of all variables that were selected to at least one of the three final models. $\text{sd}_{b_{\text{aMIRL}}}$ denotes the empirical standard deviation of the initial estimates that the aMIRL estimates are computed from, see Equation 2.5. All effects are significant at the 1% level based on bootstrap confidence intervals, unless stated otherwise: ° implies that that the respective effect is not significant at the 10% level. Their signs coincide with those of the postimputation and postselection quantile regression estimates for $\tau = 0.25, 0.5, 0.75$, see Section 2.4.3. Grey font symbolises that the respective variable is not included in the final model. $R^2_{\text{av.}}$ ($\bar{R}^2_{\text{av.}}$) denotes the average (adjusted) R^2 across all 10 imputed data sets.

A.4 Detailed MIRL Results

Table A.8: Detailed MIRL results on the balanced and unbalanced panel data for OSS

	Balanced Panel									Unbalanced Panel								
	BIC			AIC			C_p			BIC			AIC			C_p		
Variable	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$
Costs																		
Cost per Borrower	0.019	0.000°	0.000	0.963	0.225	0.172	0.953	0.232	0.170	0.051	0.000	0.007	0.384	0.016	0.121	0.400	0.016	0.126
Cost per Loan	0.986	−0.037	0.032	0.997	−0.283	0.171	0.993	−0.290	0.169	0.433	−0.003	0.015	0.723	−0.029	0.120	0.705	−0.029	0.124
Fin. Exp. Fund. Liab. / Assets	0.145	0.000	0.006	0.839	0.122	0.113	0.810	0.118	0.112	0.841	−0.022	0.057	0.743	−0.039	0.075	0.732	−0.036	0.069
Financial Expense / Assets	0.998	−0.057	0.029	0.974	−0.248	0.131	0.978	−0.243	0.131	0.255	−0.001°	0.042	0.500	0.013	0.085	0.491	0.009	0.084
Operating Expense / Assets	0.803	−0.023	0.035	0.918	−0.298	0.198	0.898	−0.293	0.201	0.791	−0.231	1.170	0.912	−0.209	0.773	0.891	−0.202	0.744
Financial Performance																		
Financial Revenue / Assets	0.001	0.000°	0.000	0.953	0.303	0.200	0.955	0.300	0.198	0.960	0.112	0.268	0.925	0.125	0.324	0.905	0.132	0.336
Profit Margin	1.000	0.768	0.111	1.000	0.862	0.113	1.000	0.859	0.113	0.981	0.072	0.075	0.846	0.062	0.071	0.841	0.059	0.070
Return on Assets	0.953	−0.039	0.118	0.981	−0.292	0.149	0.977	−0.288	0.154	0.987	0.122	0.267	0.948	0.123	0.325	0.952	0.117	0.334
Financing Structure																		
Capital / Assets	0.949	0.074	0.082	0.818	0.093	0.076	0.832	0.094	0.081	0.442	0.004	0.019	0.201	0.006	0.025	0.188	0.005	0.023
Deposits / Loans	0.555	0.009	0.051	0.959	0.178	0.090	0.957	0.175	0.090	0.547	−0.008	0.108	0.505	−0.003°	0.168	0.524	−0.004°	0.174
Deposits / Total Assets	0.750	−0.014	0.054	0.981	−0.206	0.087	0.975	−0.203	0.089	0.838	0.019	0.041	0.770	0.039	0.047	0.761	0.037	0.048
Paid in Capital / Equity	0.893	0.016	0.031	0.616	0.031	0.032	0.658	0.034	0.032	0.170	0.000	0.005	0.157	−0.002	0.011	0.200	−0.003	0.011
% Assets: Non-Earn. Liq. Ass.	0.982	0.038	0.031	0.600	0.023	0.027	0.627	0.024	0.027	0.898	−0.022	0.029	0.646	−0.024	0.029	0.625	−0.021	0.027
Operational Aspects																		
Deposit Accounts / Staff	0.982	−0.034	0.041	0.818	−0.056	0.052	0.828	−0.058	0.053	0.017	0.000*	0.001	0.109	−0.001	0.006	0.100	−0.001	0.005
Loan Impairm. Prov. / Assets	0.996	0.068	0.062	0.572	0.018	0.049	0.602	0.020	0.051	0.995	−0.083	0.083	0.914	−0.089	0.090	0.925	−0.093	0.094
Tax Expense / Assets	0.968	0.031	0.034	0.325	0.001°	0.026	0.367	0.002**	0.027	0.998	0.099	0.053	0.917	0.086	0.060	0.918	0.086	0.060
% GLP: Village Banking (SHG)	0.977	−0.027	0.029	0.791	−0.046	0.029	0.791	−0.046	0.030	0.578	−0.006	0.020	0.555	−0.011	0.029	0.526	−0.010	0.025
Organisational Strategy																		
Board: Social Training	0.952	−0.026	0.026	0.395	−0.014	0.023	0.426	−0.015	0.024	0.061	0.000°	0.005	0.245	0.001	0.012	0.240	0.001	0.012

(Table continued on next page)

Table A.8: Detailed MIRL results for OSS *(continued)*

	Balanced Panel									Unbalanced Panel								
	BIC			AIC			C_p			BIC			AIC			C_p		
Variable	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$
Target Market: Youth	0.959	0.024	0.029	0.686	0.036	0.029	0.672	0.036	0.030	0.062	0.000°	0.000	0.144	0.001	0.006	0.133	0.001	0.005
Personnel Structure																		
Borrowers / Staff Member	0.484	0.003	0.012	0.747	0.038	0.037	0.736	0.036	0.036	0.739	0.014	0.028	0.797	0.038	0.031	0.822	0.041	0.031
Personnel Expense / Assets	0.729	−0.011	0.028	0.355	−0.012	0.038	0.361	−0.013	0.039	0.919	0.077	0.624	0.900	0.080	0.365	0.893	0.066	0.349
# Personnel	0.967	0.038	0.035	0.690	0.054	0.063	0.677	0.049	0.061	0.587	0.007	0.018	0.656	0.021	0.035	0.672	0.023	0.039
% Board Members: Female	0.992	−0.040	0.030	0.533	−0.023	0.026	0.538	−0.023	0.026	0.601	−0.007	0.017	0.345	−0.006	0.016	0.329	−0.005	0.016
% Staff: Board Members	1.000	0.142	0.065	0.949	0.146	0.066	0.964	0.148	0.063	0.970	0.040	0.064	0.695	0.034	0.064	0.696	0.035	0.063
% Staff: Leaving During Period	0.977	0.031	0.033	0.536	0.021	0.027	0.581	0.022	0.027	0.772	−0.015	0.050	0.408	−0.007	0.037	0.403	−0.007	0.040
% Staff: Managers	0.986	0.050	0.032	0.651	0.039	0.038	0.658	0.039	0.038	0.800	0.015	0.039	0.460	0.011	0.037	0.447	0.010	0.036
Services																		
Offers Other Fin. Services	0.971	−0.025	0.024	0.583	−0.025	0.025	0.567	−0.024	0.025	0.320	−0.002	0.011	0.189	−0.002	0.011	0.178	−0.002	0.009
Optimal threshold $\hat{\pi}^*$	0.949			0.651			0.658			0.981			0.900			0.891		
# variables in the final model	18			18			19			4			6			6		
Overall $R^2_{av.}$ (adjusted $\bar{R}^2_{av.}$)	0.623 (0.617)			0.630 (0.625)			0.632 (0.627)			0.069 (0.068)			0.073 (0.072)			0.073 (0.071)		

Note: The table contains the empirical selection probabilities and MIRL coefficients of all variables that were selected to at least one of the six final models. sd_{MIRL} denotes the empirical standard deviation of the initial estimates that the MIRL estimates are computed from. All effects are significant at the 1% level based on bootstrap confidence intervals, unless otherwise stated: ** (*) denotes that the respective effect is significant at the 5% (10%) level, ° implies that it is not significant at the 10% level. Grey font symbolises that the respective variable is not included in the final model. $R^2_{av.}$ ($\bar{R}^2_{av.}$) denotes the average (adjusted) R^2 across all 10 imputed data sets.

Table A.9: Detailed MIRL results on the balanced and unbalanced panel data for log(NAB)

	Balanced Panel									Unbalanced Panel								
	BIC			AIC			C_p			BIC			AIC			C_p		
Variable	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$
Costs																		
Administrative Exp. / Assets	0.603	−0.014	0.023	0.344	−0.016	0.031	0.382	−0.018	0.031	0.822	−0.029	0.022	0.618	−0.045	0.050	0.660	−0.043	0.049
Cost per Borrower	0.251	−0.002	0.014	0.555	0.030	0.105	0.564	0.039	0.110	0.883	−0.052	0.035	0.731	−0.067	0.071	0.721	−0.060	0.065
Cost per Loan	0.997	−0.103	0.033	0.891	−0.132	0.116	0.911	−0.143	0.121	0.406	−0.009	0.031	0.293	0.003*	0.062	0.286	−0.003°	0.058
Fin. Exp. Fund. Liab. / Assets	0.015	0.000°	0.000	0.034	0.000°	0.006	0.024	0.000	0.007	0.115	−0.001	0.005	0.644	−0.027	0.039	0.660	−0.027	0.040
Financial Expense / Assets	0.009	0.000°	0.000	0.153	−0.001**	0.013	0.143	0.000°	0.015	0.030	0.000°	0.002	0.517	0.019	0.039	0.493	0.018	0.038
Operating Expense / Assets	0.352	−0.006	0.018	0.362	−0.030	0.059	0.374	−0.029	0.056	0.043	0.000°	0.005	0.574	−0.041	0.097	0.546	−0.045	0.098
Operating Expense / GLP	0.164	−0.001	0.007	0.194	−0.010	0.031	0.224	−0.009	0.031	0.497	−0.013	0.022	0.811	0.094	0.174	0.818	0.093	0.169
Financial Performance																		
Operational Self-Sufficiency	0.894	0.043	0.039	0.565	0.035	0.038	0.599	0.035	0.037	0.061	0.000**	0.003	0.003	0.000°	0.000	0.005	0.000°	0.001
Profit Margin	0.772	0.032	0.034	0.285	0.012	0.029	0.277	0.011	0.028	0.567	0.010	0.017	0.366	0.008	0.016	0.367	0.007	0.016
Yield on GLP (Real)	0.086	0.000	0.006	0.565	0.029	0.039	0.577	0.028	0.039	0.019	0.000°	0.002	0.114	0.002	0.009	0.133	0.002	0.010
Financing Structure																		
Assets	0.128	0.003	0.013	0.546	0.040**	0.464	0.509	0.026**	0.379	0.252	−0.002°	0.055	0.726	0.103	0.267	0.693	0.099	0.253
Capital / Assets	0.263	−0.012	0.045	0.058	−0.001°	0.097	0.070	0.007	0.107	0.804	−0.037	0.031	0.594	−0.046	0.060	0.568	−0.044	0.060
Deposits / Loans	0.305	0.003	0.014	0.100	0.001°	0.020	0.059	0.001°	0.018	0.113	−0.001	0.009	0.750	−0.059	0.113	0.751	−0.056	0.110
Deposits / Total Assets	0.992	0.113	0.031	0.833	0.109	0.063	0.848	0.109	0.060	0.954	0.062	0.022	0.636	0.051	0.043	0.623	0.049	0.043
Donated Equity / Equity	0.745	−0.021	0.022	0.551	−0.026	0.029	0.553	−0.027	0.028	0.508	−0.007	0.012	0.335	−0.008	0.013	0.349	−0.009	0.013
Equity	0.578	0.013	0.031	0.894	0.140	0.135	0.892	0.140	0.123	0.383	−0.010	0.033	0.519	−0.038	0.063	0.535	−0.038	0.061
Liabilities	0.711	0.022	0.029	0.924	0.373	0.421	0.935	0.385	0.372	0.940	0.061	0.057	0.750	0.127	0.218	0.743	0.132	0.214
Liabilities / Assets	0.958	0.152	0.049	0.873	0.148	0.111	0.900	0.159	0.118	0.739	0.030	0.030	0.269	0.014	0.055	0.292	0.016	0.055
Paid in Capital / Equity	0.788	−0.026	0.028	0.603	−0.028	0.028	0.626	−0.029	0.028	0.016	0.000*	0.000	0.221	−0.001	0.007	0.207	−0.001**	0.006
Operational Aspects																		
Av. Loan Size: Female	0.522	−0.010	0.021	0.273	−0.009	0.023	0.288	−0.009	0.022	0.263	−0.005	0.036	0.542	−0.078	0.254	0.527	−0.081	0.232
Av. Loan Size: Large Corp.	0.254	−0.002	0.012	0.575	−0.034	0.041	0.567	−0.032	0.040	0.065	0.000°	0.006	0.373	−0.011	0.105	0.381	−0.015	0.108
Av. Loan Size: Legal Entity	0.775	0.023	0.032	0.398	0.017	0.028	0.376	0.017	0.028	0.047	0.000°	0.002	0.344	0.030	0.198	0.329	0.028	0.164

(Table continued on next page)

Table A.9: Detailed MIRL results for log(NAB) *(continued)*

Variable	Balanced Panel									Unbalanced Panel								
	BIC			AIC			C_p			BIC			AIC			C_p		
	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$
Av. Outst. Bal. / GNI p. c.	0.602	−0.017	0.037	0.704	−0.054	0.070	0.715	−0.057	0.073	0.700	−0.025	0.044	0.776	−0.040	0.047	0.775	−0.039	0.038
Client Prot.: Complaint Mech.	0.926	0.044	0.023	0.467	0.028	0.032	0.465	0.027	0.033	0.907	0.046	0.018	0.527	0.028	0.028	0.528	0.029	0.028
Client Prot.: Privacy Data Clause	0.461	−0.007	0.017	0.453	−0.015	0.020	0.488	−0.015	0.020	0.018	0.000°	0.000	0.050	0.000	0.003	0.054	0.000	0.003
Gross Loan Portfolio (GLP)	0.013	0.000°	0.001	0.985	−0.472	0.321	0.979	−0.472	0.294	0.000	0.000°	0.000	0.875	−0.154	0.116	0.891	−0.156	0.120
Income on Pen. Fees / GLP	0.562	−0.010	0.018	0.315	−0.009	0.016	0.358	−0.011	0.018	0.818	−0.027	0.018	0.441	−0.016	0.020	0.415	−0.015	0.020
Net GLP / GLP	0.654	−0.013	0.023	0.534	−0.025	0.029	0.566	−0.027	0.029	0.098	0.000	0.003	0.102	−0.001	0.006	0.092	−0.001	0.006
Portfolio at Risk: > 30 Days	0.111	−0.001	0.006	0.111	-0.001°	0.031	0.105	0.000°	0.028	0.793	−0.034	0.023	0.841	−0.059	0.067	0.849	−0.062	0.070
Portfolio at Risk: > 90 Days	0.248	−0.004	0.011	0.688	−0.052	0.058	0.683	−0.055	0.057	0.039	0.000°	0.005	0.497	0.012	0.058	0.515	0.015	0.060
% Borrowers: Female	0.589	0.012	0.023	0.435	0.018	0.031	0.394	0.016	0.032	0.937	0.057	0.022	0.736	0.069	0.063	0.709	0.065	0.060
% Borrowers: Legal Entity	0.269	−0.002	0.011	0.003	0.000°	0.002	0.007	0.000	0.003	0.837	−0.028	0.018	0.336	−0.013	0.022	0.361	−0.014	0.022
% GLP: Consumption	0.018	0.000°	0.001	0.387	0.006	0.024	0.398	0.006	0.025	0.158	0.001	0.005	0.561	0.017	0.025	0.547	0.017	0.025
% GLP: Household Financing	0.551	−0.010	0.021	0.698	−0.032	0.036	0.714	−0.033	0.036	0.032	0.000	0.003	0.395	−0.014	0.026	0.404	−0.014	0.026
% GLP: Other Household Fin.	0.158	−0.001	0.007	0.182	0.000°	0.011	0.134	−0.001	0.009	0.743	−0.021	0.018	0.108	−0.003	0.010	0.112	−0.003	0.010
% GLP: Renegotiated Loans	0.036	0.000°	0.000	0.608	0.028	0.040	0.633	0.030	0.039	0.018	0.000°	0.001	0.303	0.005	0.011	0.320	0.004	0.011
% GLP: Solidarity Group	0.989	0.081	0.021	0.629	0.052	0.044	0.635	0.052	0.045	0.666	0.016	0.017	0.385	0.012	0.018	0.350	0.011	0.018
Organisational Strategy																		
Has Poverty Target	0.007	0.000°	0.000	0.051	0.000	0.004	0.041	0.000	0.004	0.703	0.017	0.017	0.322	0.011	0.018	0.294	0.010	0.017
Incentives: # Clients	0.971	0.068	0.021	0.585	0.044	0.039	0.595	0.045	0.039	0.897	0.037	0.017	0.496	0.023	0.025	0.514	0.024	0.024
Personnel Structure																		
Av. Salary / GNI p. c.	0.989	0.096	0.027	0.853	0.099	0.049	0.868	0.101	0.047	0.451	0.006	0.016	0.415	0.011	0.017	0.409	0.011	0.018
Borrowers / Office	0.170	−0.002	0.013	0.469	−0.018	0.043	0.510	−0.016	0.044	0.589	0.011	0.023	0.220	0.006	0.019	0.205	0.007	0.023
Borrowers / Staff Member	1.000	0.261	0.031	0.971	0.271	0.068	0.964	0.270	0.073	1.000	0.282	0.020	0.992	0.304	0.036	0.991	0.303	0.037
Personnel / Office	0.959	0.059	0.021	0.751	0.062	0.044	0.744	0.061	0.044	0.822	0.031	0.023	0.460	0.022	0.026	0.472	0.022	0.027
Personnel Expense / Assets	0.022	0.000°	0.000	0.287	-0.003^{**}	0.037	0.257	−0.003	0.033	0.000	0.000°	0.000	0.674	0.052	0.081	0.702	0.055	0.085
Personnel Expense / GLP	0.038	0.000^{**}	0.001	0.259	−0.004	0.027	0.241	−0.004	0.027	0.546	−0.013	0.021	0.871	−0.098	0.111	0.871	−0.097	0.107
# Personnel	1.000	0.309	0.024	0.974	0.294	0.056	0.986	0.297	0.045	1.000	0.310	0.020	0.989	0.332	0.041	0.995	0.334	0.032
% Board Members: Female	0.789	−0.025	0.025	0.385	−0.015	0.022	0.358	−0.014	0.022	0.565	−0.009	0.014	0.314	−0.009	0.015	0.340	−0.009	0.015

(Table continued on next page)

Table A.9: Detailed MIRL results for log(NAB) *(continued)*

	Balanced Panel									Unbalanced Panel								
	BIC			**AIC**			**C_p**			**BIC**			**AIC**			**C_p**		
Variable	$\hat{\pi}$	b_{MIRL}	$\text{sd}_{b_{\text{MIRL}}}$	$\hat{\pi}$	b_{MIRL}	$\text{sd}_{b_{\text{MIRL}}}$	$\hat{\pi}$	b_{MIRL}	$\text{sd}_{b_{\text{MIRL}}}$	$\hat{\pi}$	b_{MIRL}	$\text{sd}_{b_{\text{MIRL}}}$	$\hat{\pi}$	b_{MIRL}	$\text{sd}_{b_{\text{MIRL}}}$	$\hat{\pi}$	b_{MIRL}	$\text{sd}_{b_{\text{MIRL}}}$
% Managers: Female	0.268	−0.003	0.009	0.001	0.000°	0.000	0.009	0.000	0.002	0.750	−0.022	0.017	0.315	−0.012	0.021	0.297	−0.011	0.020
% Staff: Board Members	1.000	−0.316	0.030	0.992	−0.332	0.045	0.988	−0.330	0.049	1.000	−0.354	0.015	0.998	−0.374	0.027	0.996	−0.372	0.032
% Staff: Female	0.431	−0.006	0.013	0.324	−0.009	0.019	0.337	−0.009	0.020	0.915	−0.048	0.021	0.568	−0.036	0.035	0.607	−0.039	0.035
% Staff: Loan Officers	0.993	0.096	0.025	0.777	0.076	0.050	0.754	0.074	0.051	0.973	0.067	0.018	0.670	0.044	0.034	0.681	0.045	0.034
% Staff: Managers	0.990	−0.083	0.028	0.670	−0.055	0.044	0.655	−0.054	0.045	1.000	−0.147	0.014	0.884	−0.133	0.050	0.897	−0.135	0.048
Services																		
Offers Compulsory Insurance	0.620	0.012	0.018	0.297	0.011	0.021	0.273	0.010	0.020	0.857	0.035	0.017	0.436	0.023	0.027	0.469	0.025	0.028
Offers Health Services	0.902	0.040	0.026	0.512	0.026	0.029	0.499	0.026	0.029	0.820	0.030	0.017	0.481	0.022	0.025	0.473	0.022	0.024
Offers Other Fin. Services	0.497	0.006	0.014	0.280	0.010	0.021	0.293	0.009	0.019	0.894	0.039	0.017	0.504	0.030	0.032	0.500	0.030	0.032
Offers Savings Products	0.749	0.020	0.021	0.462	0.027	0.036	0.510	0.029	0.036	0.550	0.011	0.015	0.242	0.011	0.022	0.245	0.011	0.022
Offers Voluntary Insurance	0.979	0.067	0.020	0.508	0.038	0.039	0.498	0.038	0.040	0.760	0.023	0.016	0.357	0.015	0.022	0.356	0.015	0.022
Offers Women's Empow. Serv.	0.149	0.001	0.005	0.165	0.002	0.010	0.150	0.003	0.010	0.681	0.018	0.017	0.316	0.012	0.019	0.307	0.012	0.019
Optimal threshold $\hat{\pi}^*$	0.711			0.462			0.465			0.666			0.395			0.404		
# variables in the final model	23			32			33			29			36			36		
Overall $R^2_{\text{av.}}$ (adjusted $\bar{R}^2_{\text{av.}}$)	0.799 (0.795)			0.798 (0.793)			0.800 (0.795)			0.766 (0.765)			0.757 (0.754)			0.756 (0.754)		

Note: The table contains the empirical selection probabilities and MIRL coefficients of all variables that were selected to at least one of the six final models. sd_{MIRL} denotes the empirical standard deviation of the initial estimates that the MIRL estimates are computed from. All effects are significant at the 1% level based on bootstrap confidence intervals, unless otherwise stated: ** (*) denotes that the respective effect is significant at the 5% (10%) level, ° implies that it is not significant at the 10% level. Grey font symbolises that the respective variable is not included in the final model. $R^2_{\text{av.}}$ ($\bar{R}^2_{\text{av.}}$) denotes the average (adjusted) R^2 across all 10 imputed data sets.

Table A.10: Detailed MIRL results on the balanced and unbalanced panel data for –log(ALBG)

	Balanced Panel									Unbalanced Panel								
	BIC			**AIC**			**C_p**			**BIC**			**AIC**			**C_p**		
Variable	$\hat{\pi}$	b_{MIRL}	$sd_{b_{\text{MIRL}}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{\text{MIRL}}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{\text{MIRL}}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{\text{MIRL}}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{\text{MIRL}}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{\text{MIRL}}}$
Costs																		
Cost per Borrower	0.904	−0.089	0.061	0.859	−0.116	0.064	0.830	−0.114	0.067	0.965	−0.124	0.149	0.944	−0.244	0.225	0.952	−0.242	0.218
Cost per Loan	0.605	−0.025	0.055	0.052	−0.002	0.032	0.039	0.001°	0.032	0.311	0.040	0.147	0.797	0.138	0.226	0.787	0.137	0.220
Fin. Exp. Fund. Liab. / Assets	0.639	0.050	0.109	0.988	0.231	0.091	0.990	0.236	0.091	0.370	0.028	0.068	0.905	0.129	0.082	0.893	0.128	0.082
Financial Expense / Assets	0.491	−0.045	0.107	0.981	−0.222	0.090	0.982	−0.227	0.092	0.671	−0.045	0.070	0.898	−0.138	0.082	0.909	−0.137	0.082
Operating Expense / Assets	0.490	0.021	0.054	0.642	0.046	0.130	0.636	0.046	0.132	0.604	0.032	0.089	0.370	−0.023	0.097	0.329	−0.019	0.091
Operating Expense / GLP	0.653	0.036	0.064	0.556	0.064	0.126	0.567	0.065	0.129	0.920	0.113	0.108	0.983	0.400	0.279	0.994	0.400	0.273
Financial Performance																		
Financial Revenue / Assets	0.547	0.014	0.050	0.195	0.002°	0.047	0.145	0.002*	0.039	0.742	0.030	0.057	0.049	0.001°	0.018	0.064	0.001°	0.019
Operational Self-Sufficiency	0.874	−0.047	0.034	0.599	−0.040	0.037	0.640	−0.043	0.037	0.812	−0.042	0.027	0.504	−0.026	0.029	0.500	−0.025	0.029
Profit Margin	0.486	−0.016	0.047	0.823	−0.097	0.077	0.818	−0.095	0.077	0.420	−0.010	0.023	0.385	−0.010	0.026	0.409	−0.010	0.026
Return on Assets	0.692	0.042	0.065	0.917	0.145	0.075	0.928	0.147	0.072	0.538	0.034	0.056	0.718	0.071	0.061	0.719	0.071	0.063
Financing Structure																		
Assets	0.040	0.005	0.048	0.482	0.084	0.263	0.459	0.084	0.233	0.006	0.000°	0.006	0.110	0.002*	0.051	0.122	0.005	0.057
Deposits / Loans	0.134	0.000**	0.007	0.611	0.029	0.052	0.642	0.031	0.053	0.128	−0.008	0.050	0.887	−0.190	0.215	0.899	−0.193	0.234
Deposits / Total Assets	0.036	0.000	0.003	0.549	−0.030	0.054	0.514	−0.031	0.054	0.458	−0.015	0.022	0.137	−0.006	0.020	0.130	−0.006	0.019
Equity	0.000	0.000°	0.000	0.350	0.025	0.062	0.303	0.023	0.056	0.033	0.000*	0.006	0.408	0.014	0.041	0.418	0.016	0.039
Liabilities	0.115	0.000°	0.031	0.678	−0.042	0.223	0.662	−0.044	0.190	0.644	−0.030	0.044	0.229	−0.009	0.049	0.201	−0.010	0.049
Liabilities / Assets	0.709	−0.034	0.033	0.543	−0.032	0.083	0.520	−0.030	0.088	0.692	−0.031	0.027	0.406	−0.020	0.045	0.426	−0.020	0.047
Operational Aspects																		
Av. Loan Size: Household Fin.	0.659	−0.021	0.022	0.496	−0.020	0.027	0.458	−0.019	0.028	0.190	0.015**	0.218	0.120	0.013*	0.243	0.110	0.006°	0.175
Av. Loan Size: Male	0.611	−0.021	0.029	0.341	−0.019	0.034	0.358	−0.019	0.035	0.232	0.031	0.257	0.487	0.035	0.324	0.511	0.051	0.328
Av. Loan Size: Microenterpr.	0.572	−0.016	0.023	0.368	−0.017	0.029	0.349	−0.014	0.027	0.856	−0.674	1.231	0.928	−1.197	1.560	0.941	−1.188	1.546
Av. Loan Size: Rural	0.723	−0.026	0.033	0.464	−0.021	0.035	0.478	−0.020	0.034	0.351	−0.007	0.017	0.233	−0.006	0.018	0.230	−0.006	0.018
Av. Loan Size: Sm. & Med. Ent.	0.584	−0.012	0.016	0.434	−0.014	0.020	0.411	−0.014	0.020	0.320	0.001°	0.149	0.156	−0.009	0.128	0.165	−0.003°	0.042
Av. Loan Size: Solidarity Group	0.514	−0.009	0.016	0.421	−0.011	0.018	0.409	−0.011	0.018	0.107	−0.001	0.005	0.162	−0.002	0.008	0.153	−0.003	0.011
Av. Loan Size: Urban	0.927	−0.067	0.034	0.641	−0.055	0.050	0.651	−0.056	0.050	0.199	0.002°	0.307	0.360	0.062	0.552	0.349	0.024°	0.620

(Table continued on next page)

Table A.10: Detailed MIRL results for –log(ALBG) *(continued)*

	Balanced Panel									Unbalanced Panel								
	BIC			AIC			C_p			BIC			AIC			C_p		
Variable	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$
Client Prot.: Decl. Bal. Method	0.748	−0.026	0.020	0.380	−0.013	0.018	0.407	−0.013	0.018	0.851	−0.051	0.024	0.477	−0.033	0.036	0.527	−0.036	0.036
GLP / Total Assets	0.736	−0.027	0.038	0.620	−0.040	0.043	0.651	−0.041	0.041	0.849	−0.050	0.030	0.594	−0.039	0.038	0.544	−0.036	0.039
Gross Loan Portfolio (GLP)	0.981	−0.112	0.066	0.849	−0.178	0.191	0.853	−0.175	0.188	0.768	−0.048	0.049	0.734	−0.097	0.079	0.772	−0.101	0.076
Interest Income on GLP / GLP	0.990	0.114	0.035	0.783	0.075	0.051	0.759	0.071	0.050	0.994	0.164	0.045	0.826	0.143	0.077	0.799	0.139	0.081
Portfolio at Risk: > 30 Days	0.044	0.000°	0.012	0.565	0.024	0.100	0.570	0.018	0.066	0.046	0.000°	0.004	0.446	0.006	0.030	0.413	0.006	0.029
Retained Earnings / Equity	0.335	0.006	0.014	0.370	0.012	0.018	0.359	0.011	0.018	0.157	0.001	0.005	0.416	0.004	0.017	0.384	0.004	0.017
Tax Expense / Assets	0.496	−0.008	0.016	0.212	−0.003	0.011	0.235	−0.003	0.012	0.031	0.000°	0.001	0.002	0.000°	0.000	0.007	0.000°	0.002
# Active Borrowers	0.077	−0.001	0.009	0.652	−0.065	0.096	0.659	−0.066	0.095	0.000	0.000°	0.000	0.047	0.001°	0.019	0.050	0.001	0.016
# Loans Outstanding	0.080	0.000	0.004	0.508	−0.022	0.082	0.541	−0.026	0.080	0.056	0.000°	0.005	0.557	−0.009	0.043	0.563	−0.011	0.043
% Borrowers: Female	0.366	0.010	0.028	0.177	0.006	0.034	0.133	0.003	0.032	0.945	0.117	0.061	0.692	0.099	0.072	0.696	0.102	0.074
% Borrowers: Male	0.757	−0.037	0.035	0.636	−0.039	0.045	0.613	−0.039	0.046	0.391	−0.001°	0.052	0.085	−0.004	0.028	0.088	−0.004	0.029
% GLP: >3 Months Delinquent	0.287	−0.004	0.012	0.570	−0.021	0.031	0.533	−0.018	0.028	0.009	0.000*	0.001	0.108	−0.001	0.009	0.091	−0.001	0.011
% GLP: Consumption	0.048	0.000°	0.002	0.347	0.003	0.015	0.319	0.002	0.014	0.051	0.000°	0.004	0.131	0.000°	0.009	0.136	0.000°	0.009
% GLP: Individual	0.679	−0.027	0.027	0.016	0.000°	0.011	0.020	−0.001	0.006	0.707	−0.039	0.039	0.044	−0.003	0.017	0.026	−0.002	0.013
% GLP: Renegotiated Loans	0.474	−0.011	0.017	0.406	−0.015	0.027	0.419	−0.015	0.025	0.180	−0.002	0.006	0.175	−0.003	0.008	0.184	−0.002	0.010
% GLP: Solidarity Group	0.365	0.007	0.017	0.575	0.028	0.030	0.564	0.028	0.028	0.710	0.031	0.035	0.603	0.048	0.043	0.592	0.047	0.043
Organisational Strategy																		
Goals: Econ. Improvement	0.407	−0.007	0.015	0.460	−0.011	0.018	0.448	−0.012	0.019	0.044	0.000	0.003	0.045	0.000	0.003	0.042	0.000	0.003
Goals: Education Opport.	0.095	0.000	0.004	0.335	0.004	0.012	0.371	0.004	0.012	0.023	0.000	0.001	0.040	0.000	0.003	0.035	0.000	0.002
Goals: Health Infrastructure	0.599	−0.018	0.020	0.452	−0.015	0.020	0.425	−0.015	0.020	0.202	−0.003	0.009	0.170	−0.004	0.009	0.170	−0.004	0.010
Personnel Structure																		
Av. Salary / GNI p. c.	1.000	−0.438	0.020	1.000	−0.456	0.021	0.999	−0.455	0.025	1.000	−0.387	0.039	0.993	−0.407	0.057	0.994	−0.407	0.056
Borrowers / Loan Officer	0.260	0.003	0.020	0.520	0.016	0.049	0.544	0.017	0.050	0.495	0.012	0.023	0.319	0.007	0.016	0.321	0.007	0.018
Borrowers / Office	0.807	0.030	0.047	0.653	0.033	0.045	0.678	0.038	0.048	0.940	0.079	0.053	0.777	0.065	0.061	0.793	0.068	0.062
Borrowers / Staff Member	1.000	0.325	0.050	0.995	0.361	0.068	0.998	0.362	0.067	1.000	0.233	0.038	0.955	0.261	0.071	0.952	0.259	0.071
Personnel / Office	0.608	−0.014	0.029	0.479	−0.015	0.031	0.514	−0.019	0.034	0.912	−0.068	0.040	0.683	−0.055	0.052	0.697	−0.057	0.052
Personnel Expense / Assets	0.996	0.145	0.078	0.935	0.174	0.127	0.935	0.171	0.128	0.967	0.147	0.087	0.976	0.265	0.115	0.975	0.263	0.114

(Table continued on next page)

Table A.10: Detailed MIRL results for –log(ALBG) *(continued)*

Variable	Balanced Panel									Unbalanced Panel								
	BIC			AIC			C_p			BIC			AIC			C_p		
	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$	$\hat{\pi}$	b_{MIRL}	$sd_{b_{MIRL}}$
Personnel Expense / GLP	0.984	0.118	0.077	0.839	0.098	0.126	0.844	0.101	0.126	0.493	−0.010	0.097	0.887	−0.161	0.157	0.891	−0.162	0.154
# Personnel	0.097	0.000	0.007	0.843	0.070	0.060	0.865	0.075	0.058	0.227	0.004	0.013	0.552	0.025	0.034	0.579	0.026	0.036
% Staff: Female	0.460	0.010	0.018	0.330	0.008	0.015	0.374	0.009	0.016	0.123	−0.001	0.009	0.302	−0.002	0.021	0.320	−0.003	0.024
% Staff: Loan Officers	0.905	0.056	0.027	0.637	0.045	0.043	0.645	0.046	0.043	0.882	0.059	0.028	0.496	0.040	0.044	0.481	0.038	0.043
% Staff: Managers	0.802	−0.033	0.027	0.494	−0.020	0.024	0.511	−0.021	0.024	0.033	0.000°	0.002	0.074	0.000^{**}	0.005	0.059	0.000°	0.005
Services																		
Offers Compulsory Insurance	0.915	0.049	0.023	0.507	0.025	0.028	0.547	0.027	0.028	0.908	0.063	0.023	0.592	0.041	0.036	0.610	0.042	0.036
Offers Other Fin. Services	0.905	−0.052	0.025	0.540	−0.027	0.028	0.550	−0.028	0.029	0.903	−0.067	0.026	0.594	−0.048	0.043	0.591	−0.047	0.042
Offers Savings Products	0.529	−0.011	0.017	0.406	−0.016	0.024	0.434	−0.018	0.025	0.640	−0.026	0.025	0.377	−0.020	0.029	0.395	−0.021	0.029
Optimal threshold $\hat{\pi}^*$	0.490			0.341			0.349			0.882			0.944			0.952		
# variables in the final model	36			49			49			12			5			5		
Overall $R^2_{av.}$ (adjusted $\bar{R}^2_{av.}$)	0.854 (0.849)			0.860 (0.854)			0.860 (0.854)			0.597 (0.595)			0.423 (0.422)			0.423 (0.422)		

Note: The table contains the empirical selection probabilities and MIRL coefficients of all variables that were selected to at least one of the six final models. sd_{MIRL} denotes the empirical standard deviation of the initial estimates that the MIRL estimates are computed from. All effects are significant at the 1% level based on bootstrap confidence intervals, unless otherwise stated: ** (*) denotes that the respective effect is significant at the 5% (10%) level, $^{\circ}$ implies that it is not significant at the 10% level. Grey font symbolises that the respective variable is not included in the final model. $R^2_{av.}$ ($\bar{R}^2_{av.}$) denotes the average (adjusted) R^2 across all 10 imputed data sets.

A.5 Detailed Two-Step Lasso Results

Table A.11: Detailed two-step lasso results on the balanced and unbalanced panel data for OSS

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
Costs																		
Cost per Loan	−0.065	0.018	−0.051**	0.020	−0.052	0.020	—	—	−0.055**	0.025	−0.055**	0.025	—	—	—	—	—	—
Financial Expense/Assets	−0.091	0.017	−0.104	0.018	−0.102	0.017	−0.060	0.020	−0.072	0.023	−0.072	0.023	—	—	—	—	—	—
Interest Exp. on Borr./Borr.	—	—	—	—	—	—	—	—	−0.009°	0.018	−0.009°	0.018	—	—	—	—	—	—
Operating Expense/Assets	—	—	—	—	—	—	0.003°	0.047	—	—	—	—	—	—	—	—	—	—
Financial Performance																		
Financial Revenue/Assets	—	—	0.064**	0.027	0.060**	0.027	—	—	—	—	—	—	—	—	—	—	—	—
Profit Margin	0.367	0.028	0.407	0.030	0.401	0.029	0.717	0.022	0.792	0.043	0.792	0.043	0.053	0.019	0.059	0.019	0.059	0.019
Return on Assets	0.158	0.027	0.143	0.027	0.144	0.027	—	—	−0.101**	0.040	−0.101**	0.040	0.163	0.019	0.148	0.019	0.148	0.019
Return on Equity	—	—	−0.029**	0.012	−0.033	0.012	—	—	—	—	—	—	—	—	—	—	—	—
Yield on GLP (Real)	—	—	—	—	—	—	—	—	−0.030°	0.028	−0.030°	0.028	—	—	—	—	—	—
Financing Structure																		
Capital/Assets	—	—	—	—	—	—	0.145	0.023	0.123°	0.135	0.123°	0.135	—	—	—	—	—	—
Debt/Equity	—	—	—	—	—	—	—	—	−0.030°	0.022	−0.030°	0.022	—	—	—	—	—	—
Deposits/Loans	—	—	—	—	—	—	—	—	0.159	0.051	0.159	0.051	—	—	—	—	—	—
Deposits/Total Assets	—	—	—	—	—	—	—	—	−0.176	0.056	−0.176	0.056	—	—	0.036**	0.018	0.036**	0.018
Liabilities/Assets	—	—	—	—	—	—	—	—	−0.041°	0.133	−0.041°	0.133	—	—	−0.041**	0.017	−0.041**	0.017
Paid in Capital/Equity	—	—	—	—	—	—	—	—	0.041*	0.022	0.041*	0.022	—	—	—	—	—	—
Subord. Debt/Equity	—	—	0.026*	0.015	0.026*	0.015	—	—	0.020°	0.019	0.020°	0.019	—	—	—	—	—	—
% Assets: Non-Earn. Liq. Ass.	—	—	0.030*	0.016	0.027*	0.016	0.045**	0.018	0.028°	0.020	0.028°	0.020	−0.049	0.016	−0.054	0.016	−0.054	0.016
Operational Aspects																		
Av. Loan Bal./GNI p. c.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.026*	0.016	0.026*	0.016
Av. Loan Size: Consumption	—	—	0.017°	0.013	—	—	—	—	—	—	—	—	—	—	—	—	—	—

(Table continued on next page)

Table A.11: Detailed two-step lasso results for OSS *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C$_p$**		**BIC**		**AIC**		**C$_p$**		**BIC**		**AIC**		**C$_p$**	
Variable	b	sd$_b$	b	sd$_b$	b	sd$_b$	b	sd$_b$	b	sd$_b$	b	sd$_b$	b	sd$_b$	b	sd$_b$	b	sd$_b$
Av. Loan Size: Ext. Customers	−0.030	0.011	−0.019°	0.014	−0.018°	0.014	—	—	−0.028°	0.023	−0.028°	0.023	—	—	—	—	—	—
Av. Loan Size: Female	—	—	—	—	—	—	—	—	0.028°	0.022	0.028°	0.022	—	—	0.029*	0.016	0.029*	0.016
Av. Loan Size: Large Corp.	—	—	—	—	—	—	—	—	0.013°	0.018	0.013°	0.018	—	—	—	—	—	—
Av. Loan Size: Mortgage, Housing	—	—	—	—	—	—	—	—	0.025°	0.019	0.025°	0.019	—	—	—	—	—	—
Av. Loan Size: Solidarity Group	—	—	—	—	—	—	—	—	−0.013°	0.019	−0.013°	0.019	—	—	—	—	—	—
Av. Loan Size: Urban	−0.048	0.017	−0.051	0.018	−0.047	0.016	—	—	0.027°	0.025	0.027°	0.025	—	—	—	—	—	—
Av. Loan Size: Village (SHG)	—	—	—	—	—	—	—	—	−0.023°	0.019	−0.023°	0.019	—	—	—	—	—	—
Av. Outst. Bal./GNI p. c.	—	—	—	—	—	—	—	—	0.050**	0.024	0.050**	0.024	—	—	—	—	—	—
Borrower Retention Rate	−0.035	0.011	−0.039	0.011	−0.038	0.011	—	—	—	—	—	—	—	—	—	—	—	—
Client Prot.: Clear Debt Coll.	—	—	0.030**	0.014	0.029**	0.014	—	—	−0.042**	0.019	−0.042**	0.019	—	—	—	—	—	—
Client Prot.: Complaint Mech.	—	—	—	—	—	—	—	—	−0.012°	0.020	−0.012°	0.020	—	—	—	—	—	—
Client Prot.: Full Disclosure	—	—	—	—	—	—	—	—	0.012°	0.018	0.012°	0.018	—	—	—	—	—	—
Client Prot.: Over-Indebt. Prev.	0.044	0.014	0.043	0.014	0.046	0.014	—	—	−0.039**	0.019	−0.039**	0.019	—	—	—	—	—	—
Client Prot.: Privacy Data Clause	—	—	—	—	—	—	—	—	0.024°	0.019	0.024°	0.019	—	—	—	—	—	—
GLP/Total Assets	—	—	−0.036	0.013	−0.037	0.013	—	—	−0.033*	0.020	−0.033*	0.020	—	—	—	—	—	—
Gross Loan Portfolio (GLP)	—	—	—	—	—	—	—	—	−0.051*	0.027	−0.051*	0.027	—	—	—	—	—	—
Imp. Loss Allow./GLP	—	—	—	—	—	—	—	—	−0.017°	0.023	−0.017°	0.023	—	—	—	—	—	—
Loan Impairm. Prov./Assets	—	—	0.040**	0.016	0.048	0.015	0.096	0.020	0.103	0.022	0.103	0.022	−0.070	0.016	−0.060	0.018	−0.060	0.018
Measures Client Poverty	—	—	−0.029**	0.015	−0.031**	0.015	—	—	−0.022°	0.019	−0.022°	0.019	—	—	—	—	—	—
Portfolio at Risk: > 30 Days	—	—	−0.053°	0.059	−0.061°	0.058	—	—	−0.024°	0.028	−0.024°	0.028	—	—	—	—	—	—
Portfolio at Risk: > 90 Days	—	—	0.000°	0.057	0.006°	0.057	—	—	—	—	—	—	—	—	—	—	—	—
Retained Earnings/Equity	—	—	−0.020°	0.013	—	—	—	—	−0.046**	0.022	−0.046**	0.022	—	—	—	—	—	—
Tax Expense/Assets	0.074	0.013	0.079	0.014	0.083	0.014	—	—	0.042**	0.020	0.042**	0.020	0.095	0.016	0.093	0.016	0.093	0.016
Write-Off Ratio	—	—	0.012°	0.014	—	—	—	—	—	—	—	—	—	—	−0.009°	0.018	−0.009°	0.018
# Loans Outstanding	—	—	0.077**	0.031	—	—	—	—	—	—	—	—	—	—	—	—	—	—
# New Borrowers/NAB	—	—	—	—	—	—	—	—	−0.015°	0.018	−0.015°	0.018	—	—	—	—	—	—
% Borrowers: Ext.Customers	—	—	0.000°	0.014	−0.001°	0.014	—	—	—	—	—	—	—	—	—	—	—	—
% Borrowers: Female	—	—	—	—	—	—	—	—	−0.058**	0.026	−0.058**	0.026	—	—	—	—	—	—

(Table continued on next page)

Table A.11: Detailed two-step lasso results for OSS *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
% Borrowers: Legal Entity	—	—	0.018°	0.012	0.017°	0.012	—	—	—	—	—	—	—	—	—	—	—	—
% Borrowers: Man. & Staff	—	—	-0.014°	0.016	-0.017°	0.016	—	—	-0.013°	0.023	-0.013°	0.023	—	—	—	—	—	—
% Borrowers: Rural	—	—	—	—	—	—	—	—	0.048^{**}	0.019	0.048^{**}	0.019	—	—	—	—	—	—
% GLP: >3 Months Delinquent	—	—	—	—	—	—	—	—	-0.018°	0.027	-0.018°	0.027	—	—	—	—	—	—
% GLP: Enterprise Finance	—	—	-0.012°	0.014	—	—	—	—	-0.041°	0.025	-0.041°	0.025	—	—	—	—	—	—
% GLP: Household Financing	—	—	—	—	—	—	—	—	-0.071	0.027	-0.071	0.027	—	—	—	—	—	—
% GLP: Large Corporations	—	—	—	—	—	—	—	—	-0.022°	0.019	-0.022°	0.019	—	—	—	—	—	—
% GLP: Mortgage, Housing	—	—	—	—	—	—	—	—	-0.033^{*}	0.020	-0.033^{*}	0.020	—	—	—	—	—	—
% GLP: Small & Med. Enterpr.	—	—	—	—	—	—	—	—	0.013°	0.018	0.013°	0.018	—	—	0.026^{*}	0.016	0.026^{*}	0.016
% GLP: Solidarity Group	—	—	—	—	—	—	—	—	-0.054^{**}	0.023	-0.054^{**}	0.023	—	—	—	—	—	—
% GLP: Village Banking (SHG)	—	—	—	—	—	—	—	—	-0.099	0.022	-0.099	0.022	—	—	—	—	—	—
Organisational Strategy																		
Board: Social Training	—	—	—	—	—	—	-0.051	0.018	-0.030°	0.019	-0.030°	0.019	—	—	—	—	—	—
Has Poverty Target	—	—	0.021°	0.015	—	—	—	—	0.018°	0.019	0.018°	0.019	—	—	—	—	—	—
Incentives: Data-Collect. Qual.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	-0.040	0.016	-0.040	0.016
Incentives: Interaction Qual.	—	—	-0.032^{**}	0.015	-0.033^{**}	0.015	—	—	—	—	—	—	—	—	—	—	—	—
Incentives: Portf. Qual.	—	—	—	—	—	—	—	—	-0.037^{**}	0.018	-0.037^{**}	0.018	—	—	—	—	—	—
Target Market: Urban	—	—	—	—	—	—	—	—	0.021°	0.018	0.021°	0.018	—	—	—	—	—	—
Target Market: Youth	—	—	—	—	—	—	—	—	0.038^{**}	0.018	0.038^{**}	0.018	—	—	—	—	—	—
Personnel Structure																		
Av. Salary/GNI p. c.	—	—	0.061	0.022	0.060	0.022	—	—	—	—	—	—	—	—	-0.028^{*}	0.016	-0.028^{*}	0.016
Borrowers/Office	—	—	—	—	—	—	—	—	-0.015°	0.021	-0.015°	0.021	—	—	—	—	—	—
Borrowers/Staff Member	—	—	—	—	—	—	—	—	0.062^{**}	0.029	0.062^{**}	0.029	—	—	0.047	0.017	0.047	0.017
Personnel Expense/Assets	—	—	—	—	—	—	-0.086^{*}	0.046	-0.051°	0.031	-0.051°	0.031	—	—	—	—	—	—
# Personnel	—	—	—	—	—	—	—	—	0.033°	0.023	0.033°	0.023	—	—	—	—	—	—
% Board Members: Female	—	—	—	—	—	—	—	—	-0.025°	0.019	-0.025°	0.019	—	—	—	—	—	—
% Loan Officers: Female	—	—	—	—	—	—	—	—	0.025°	0.030	0.025°	0.030	—	—	—	—	—	—

(Table continued on next page)

Table A.11: Detailed two-step lasso results for OSS *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
% Staff: Board Members	−0.089	0.017	−0.086	0.017	−0.090	0.017	—	—	−0.087	0.020	−0.087	0.020	—	—	—	—	—	—
% Staff: Employed > 1 Year	—	—	0.029**	0.011	0.028**	0.011	—	—	—	—	—	—	—	—	—	—	—	—
% Staff: Female	—	—	−0.039**	0.018	−0.042**	0.018	—	—	0.052*	0.030	0.052*	0.030	—	—	—	—	—	—
% Staff: Leaving During Period	—	—	—	—	—	—	—	—	0.032*	0.018	0.032*	0.018	—	—	—	—	—	—
% Staff: Loan Officers	—	—	0.030°	0.022	—	—	—	—	—	—	—	—	−0.052	0.016	−0.062	0.017	−0.062	0.017
% Staff: Managers	−0.080	0.015	−0.086	0.015	−0.087	0.015	0.105	0.018	0.101	0.019	0.101	0.019	—	—	0.021°	0.016	0.021°	0.016
Services																		
Offers Enterprise Services	—	—	−0.022°	0.018	—	—	—	—	—	—	—	—	—	—	0.032**	0.016	0.032**	0.016
Offers Health Services	—	—	—	—	—	—	—	—	0.047**	0.020	0.047**	0.020	—	—	—	—	—	—
Offers Non-Inc. Gen. Loans Also	—	—	—	—	—	—	—	—	0.017°	0.020	0.017°	0.020	—	—	—	—	—	—
Offers Other Fin. Services	—	—	—	—	—	—	−0.053	0.019	−0.047**	0.022	−0.047**	0.022	—	—	—	—	—	—
Offers Savings Products	—	—	—	—	—	—	—	—	−0.021°	0.027	−0.021°	0.027	—	—	—	—	—	—
Offers Voluntary Insurance	—	—	0.020°	0.016	0.021°	0.016	—	—	—	—	—	—	—	—	—	—	—	—
Offers Women's Empow. Serv.	—	—	—	—	—	—	—	—	0.042**	0.021	0.042**	0.021	—	—	—	—	—	—
# variables in the final model	11		37		29		10		67		67		6		17		17	
Overall $R^2_{av.}$ (adjusted $\bar{R}^2_{av.}$)	0.882 (0.857)		0.891 (0.865)		0.889 (0.864)		0.604 (0.601)		0.658 (0.639)		0.658 (0.639)		0.073 (0.071)		0.082 (0.078)		0.082 (0.078)	
Within $R^2_{av.}$ (adjusted $\bar{R}^2_{av.}$)	0.544 (0.448)		0.579 (0.477)		0.573 (0.473)		—		—		—		—		—		—	

Note: The table contains the two-step lasso coefficients of all variables that were selected to at least one of the nine final models. sd_b denotes the OLS standard error of the estimate. All effects are significant at the 1% level based on standard confidence intervals under normality, unless otherwise stated: ** (*) denotes that the respective effect is significant at the 5% (10%) level, ° implies that it is not significant at the 10% level. "—" symbolises that the respective variable is not selected by lasso. $R^2_{av.}$ ($\bar{R}^2_{av.}$) denotes the average (adjusted) R^2 across all 10 imputed data sets.

Table A.12: Detailed two-step lasso results on the balanced and unbalanced panel data for log(NAB)

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		$\mathbf{C_p}$		**BIC**		**AIC**		$\mathbf{C_p}$		**BIC**		**AIC**		$\mathbf{C_p}$	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
Costs																		
Administrative Exp./Assets	—	—	—	—	—	—	—	—	−0.065	0.021	−0.063	0.021	−0.032	0.012	−0.061	0.020	−0.061	0.020
Cost per Borrower	−0.050	0.017	-0.042^{**}	0.017	-0.042^{**}	0.017	—	—	—	—	—	—	−0.039	0.010	−0.035	0.010	−0.035	0.010
Cost per Loan	—	—	—	—	—	—	−0.131	0.018	−0.113	0.019	−0.113	0.019	—	—	—	—	—	—
Fin. Exp. Fund. Liab./Assets	—	—	—	—	—	—	—	—	—	—	—	—	—	—	−0.083	0.021	−0.083	0.021
Financial Expense/Assets	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.086	0.023	0.086	0.023
Operating Expense/GLP	—	—	—	—	—	—	—	—	—	—	—	—	−0.047	0.010	—	—	—	—
Financial Performance																		
Financial Revenue/Assets	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.031°	0.024	0.031°	0.024
Operational Self-Sufficiency	—	—	—	—	—	—	—	—	-0.030°	0.019	-0.029°	0.019	—	—	—	—	—	—
Profit Margin	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.007°	0.012	0.007°	0.012
Return on Assets	0.009°	0.010	0.022^{**}	0.011	0.022^{**}	0.011	0.065	0.015	0.037^{*}	0.022	0.037^{*}	0.022	0.017°	0.012	0.004°	0.022	0.004°	0.022
Return on Equity	0.012^{*}	0.007	0.010°	0.007	0.011°	0.007	—	—	—	—	—	—	—	—	—	—	—	—
Yield on GLP (Real)	—	—	—	—	—	—	—	—	0.053	0.021	0.054	0.021	—	—	0.007°	0.016	0.007°	0.016
Financing Structure																		
Assets	—	—	—	—	—	—	—	—	-0.369°	0.844	-0.415°	0.842	—	—	—	—	—	—
Borrowings/Liabilities	—	—	—	—	—	—	—	—	0.018°	0.013	0.018°	0.013	—	—	0.008°	0.009	0.008°	0.009
Capital/Assets	—	—	−0.045	0.016	−0.048	0.016	-0.074°	0.108	0.029°	0.109	0.032°	0.109	−0.052	0.016	−0.046	0.016	−0.046	0.016
Debt/Equity	—	—	—	—	—	—	—	—	0.021°	0.018	0.022°	0.018	—	—	—	—	—	—
Deposits/Loans	—	—	—	—	—	—	—	—	—	—	—	—	—	—	−0.028	0.010	−0.028	0.010
Deposits/Total Assets	—	—	—	—	—	—	0.097	0.019	0.137	0.022	0.140	0.022	0.076	0.014	0.073	0.015	0.073	0.015
Donated Equity/Equity	—	—	—	—	—	—	—	—	-0.043^{*}	0.023	-0.043^{*}	0.023	—	—	-0.013°	0.009	-0.013°	0.009
Equity	—	—	—	—	—	—	—	—	0.108°	0.167	0.114°	0.167	—	—	−0.056	0.019	−0.056	0.019
Liabilities	—	—	—	—	—	—	0.032^{*}	0.020	0.334°	0.694	0.356°	0.693	0.062	0.011	0.100	0.017	0.100	0.017
Liabilities/Assets	—	—	—	—	—	—	0.124°	0.107	0.205^{*}	0.109	0.207^{*}	0.109	0.035^{**}	0.017	0.033^{*}	0.017	0.033^{*}	0.017
Paid in Capital/Equity	—	—	—	—	—	—	—	—	−0.066	0.020	−0.064	0.020	—	—	—	—	—	—

(Table continued on next page)

Table A.12: Detailed two-step lasso results for log(NAB) *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
% Assets: Non-Earn. Liq. Ass.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.019*	0.010	0.019*	0.010
Operational Aspects																		
Av. Loan Bal./GNI p. c.	−0.080	0.018	−0.097	0.017	−0.097	0.017	—	—	−0.033°	0.054	—	—	−0.025°	0.020	−0.020°	0.020	−0.020°	0.020
Av. Loan Size: Consumption	—	—	—	—	—	—	—	—	−0.012°	0.016	−0.013°	0.016	—	—	−0.026	0.009	−0.026	0.009
Av. Loan Size: Ext. Customers	—	—	—	—	—	—	—	—	0.019°	0.015	0.018°	0.015	—	—	—	—	—	—
Av. Loan Size: Female	—	—	—	—	—	—	−0.038**	0.016	−0.027°	0.017	−0.029*	0.017	—	—	—	—	—	—
Av. Loan Size: Individual	—	—	—	—	—	—	—	—	−0.053	0.014	−0.053	0.014	—	—	—	—	—	—
Av. Loan Size: Legal Entity	—	—	—	—	—	—	—	—	0.028*	0.015	0.032**	0.015	—	—	—	—	—	—
Av. Loan Size: Mortgage, Housing	—	—	—	—	—	—	—	—	0.028*	0.015	0.031**	0.015	—	—	0.024**	0.010	0.024**	0.010
Av. Loan Size: Other Househ. Fin.	—	—	—	—	—	—	—	—	0.013°	0.014	—	—	—	—	—	—	—	—
Av. Loan Size: Rural	−0.008°	0.010	−0.001°	0.010	0.000°	0.010	—	—	—	—	—	—	—	—	—	—	—	—
Av. Loan Size: Solidarity Group	0.017**	0.008	0.012°	0.008	0.012°	0.008	—	—	0.022°	0.014	0.021°	0.014	—	—	—	—	—	—
Av. Loan Size: Urban	—	—	—	—	—	—	—	—	−0.001°	0.020	−0.003°	0.020	—	—	—	—	—	—
Av. Loan Size: Village (SHG)	—	—	0.014°	0.009	0.014°	0.009	—	—	—	—	—	—	—	—	0.009°	0.009	0.009°	0.009
Av. Outst. Bal./GNI p. c.	—	—	—	—	—	—	—	—	−0.031°	0.052	−0.050**	0.021	−0.031°	0.020	−0.039*	0.021	−0.039*	0.021
Borrower Retention Rate	0.016**	0.007	0.027	0.009	0.026	0.009	0.045	0.014	0.034**	0.014	0.035**	0.014	0.027	0.009	0.027	0.009	0.027	0.009
Client Prot.: Clear Debt Coll.	−0.022**	0.009	−0.021**	0.009	−0.023	0.009	—	—	0.011°	0.015	—	—	—	—	0.004°	0.010	0.004°	0.010
Client Prot.: Complaint Mech.	—	—	—	—	—	—	0.053	0.015	0.061	0.015	0.064	0.015	0.064	0.010	0.058	0.010	0.058	0.010
Client Prot.: Decl. Bal. Method	—	—	—	—	—	—	—	—	—	—	—	—	—	—	−0.015°	0.009	−0.015°	0.009
Client Prot.: Full Disclosure	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.006°	0.009	0.006°	0.009
Client Prot.: Over-Indebt. Prev.	—	—	—	—	—	—	—	—	0.012°	0.015	0.014°	0.015	—	—	0.011°	0.010	0.011°	0.010
Client Prot.: Privacy Data Clause	—	—	—	—	—	—	—	—	−0.046	0.014	−0.042	0.014	—	—	—	—	—	—
Client Prot.: Rob. Payment Eval.	—	—	−0.004°	0.008	—	—	—	—	0.016°	0.015	—	—	—	—	0.011°	0.009	0.011°	0.009
Deposit Accounts/Staff	0.067	0.013	0.067	0.013	0.069	0.013	—	—	—	—	—	—	—	—	0.009°	0.009	0.009°	0.009
GLP/Total Assets	—	—	—	—	—	—	—	—	—	—	—	—	—	—	−0.016°	0.010	−0.016°	0.010
Imp. Loss Allow./GLP	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.017°	0.011	0.017°	0.011
Income on Pen. Fees/GLP	—	—	—	—	—	—	—	—	−0.026*	0.014	−0.027*	0.014	−0.037	0.009	−0.038	0.009	−0.038	0.009
Loan Impairm. Prov./Assets	—	—	0.034	0.008	0.034	0.008	—	—	—	—	—	—	—	—	0.022°	0.014	0.022°	0.014

(Table continued on next page)

Table A.12: Detailed two-step lasso results for log(NAB) *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
Measures Client Poverty	—	—	—	—	—	—	0.037**	0.015	0.040	0.015	0.041	0.015	0.020**	0.010	0.019*	0.010	0.019*	0.010
Net GLP/GLP	—	—	—	—	—	—	—	—	−0.021°	0.016	−0.023°	0.016	—	—	—	—	—	—
Portfolio at Risk: > 30 Days	−0.008°	0.013	−0.014°	0.019	−0.015°	0.019	−0.018°	0.015	−0.031**	0.016	−0.034**	0.016	−0.051	0.010	−0.063	0.012	−0.063	0.012
Retained Earnings/Equity	—	—	0.006°	0.008	0.006°	0.008	—	—	0.000°	0.026	0.002°	0.026	—	—	—	—	—	—
Risk Coverage	—	—	—	—	—	—	—	—	0.040	0.014	0.039	0.014	—	—	0.015°	0.009	0.015°	0.009
Tax Expense/Assets	0.007°	0.008	0.013°	0.008	0.013*	0.008	—	—	0.037**	0.014	0.038	0.014	0.033	0.009	0.035	0.010	0.035	0.010
Write-Off Ratio	—	—	—	—	—	—	—	—	0.012°	0.016	0.014°	0.016	0.058	0.009	0.050	0.010	0.050	0.010
# New Borrowers/NAB	—	—	0.024	0.009	0.023**	0.009	—	—	—	—	—	—	—	—	0.020**	0.009	0.020**	0.009
% Borrowers: Ext.Customers	—	—	0.009°	0.007	0.009°	0.007	—	—	0.022°	0.015	0.022°	0.015	—	—	0.014°	0.009	0.014°	0.009
% Borrowers: Female	—	—	—	—	—	—	0.015°	0.019	—	—	—	—	0.082	0.012	0.030°	0.023	0.030°	0.023
% Borrowers: Legal Entity	—	—	0.011°	0.007	0.010°	0.007	—	—	—	—	—	—	−0.036	0.010	−0.045	0.011	−0.045	0.011
% Borrowers: Male	—	—	—	—	—	—	—	—	−0.017°	0.019	−0.017°	0.019	—	—	−0.047**	0.021	−0.047**	0.021
% Borrowers: Rural	—	—	—	—	—	—	—	—	−0.023°	0.016	−0.024°	0.016	—	—	—	—	—	—
% Borrowers: Urban	−0.053	0.012	−0.053	0.012	−0.054	0.012	—	—	—	—	—	—	−0.026	0.009	−0.039	0.010	−0.039	0.010
% GLP: 1-3 Months Delinquent	—	—	—	—	—	—	—	—	0.028**	0.014	0.028**	0.014	—	—	—	—	—	—
% GLP: >3 Months Delinquent	—	—	−0.012°	0.012	−0.012°	0.012	—	—	—	—	—	—	—	—	—	—	—	—
% GLP: Enterprise Finance	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.010°	0.011	0.010°	0.011
% GLP: Household Financing	—	—	—	—	—	—	—	—	−0.040*	0.021	−0.038*	0.021	—	—	—	—	—	—
% GLP: Individual	—	—	—	—	—	—	—	—	−0.012°	0.026	−0.010°	0.026	—	—	0.005°	0.019	0.005°	0.019
% GLP: Large Corporations	—	—	−0.009°	0.008	−0.009°	0.008	—	—	—	—	—	—	—	—	—	—	—	—
% GLP: Microenterprise	−0.035	0.010	−0.034	0.010	−0.035	0.010	—	—	—	—	—	—	—	—	—	—	—	—
% GLP: Mortgage, Housing	—	—	—	—	—	—	—	—	0.036**	0.016	0.033**	0.016	—	—	—	—	—	—
% GLP: Other Household Fin.	—	—	—	—	—	—	—	—	−0.011°	0.018	−0.007°	0.017	−0.043	0.009	−0.035	0.010	−0.035	0.010
% GLP: Renegotiated Loans	−0.034	0.012	−0.026*	0.013	−0.024*	0.013	—	—	—	—	—	—	—	—	0.022**	0.010	0.022**	0.010
% GLP: Small & Med. Enterpr.	—	—	—	—	—	—	—	—	0.009°	0.014	0.013°	0.014	—	—	−0.015°	0.009	−0.015°	0.009
% GLP: Solidarity Group	—	—	0.039	0.014	0.039	0.014	0.093	0.016	0.104	0.024	0.106	0.024	0.032	0.010	0.039**	0.019	0.039**	0.019
% GLP: Village Banking (SHG)	—	—	−0.046	0.013	−0.047	0.013	—	—	—	—	—	—	—	—	0.012°	0.014	0.012°	0.014

(Table continued on next page)

Table A.12: Detailed two-step lasso results for log(NAB) *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
Organisational Strategy																		
Board: Social Training	−0.026	0.008	−0.029	0.008	−0.030	0.008	—	—	−0.023°	0.014	−0.023°	0.014	0.020**	0.009	0.022**	0.009	0.022**	0.009
Goals: Econ. Improvement	−0.038	0.009	−0.036	0.009	−0.035	0.009	−0.042	0.014	−0.014°	0.020	−0.016°	0.020	—	—	−0.022**	0.011	−0.022**	0.011
Goals: Education Opport.	—	—	—	—	—	—	—	—	−0.022°	0.016	−0.022°	0.016	—	—	—	—	—	—
Goals: Health Infrastructure	0.040	0.010	0.037	0.010	0.038	0.010	—	—	—	—	—	—	—	—	—	—	—	—
Goals: Women's Empow.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.015°	0.010	0.015°	0.010
Has Poverty Target	—	—	0.011°	0.009	0.012°	0.009	—	—	—	—	—	—	0.031	0.010	0.027	0.010	0.027	0.010
Incentives: Data-Collect. Qual.	—	—	—	—	—	—	—	—	−0.040**	0.016	−0.040**	0.016	—	—	0.010°	0.010	0.010°	0.010
Incentives: Interaction Qual.	—	—	—	—	—	—	—	—	−0.031*	0.016	−0.028*	0.016	—	—	−0.033	0.010	−0.033	0.010
Incentives: Portf. Qual.	0.043	0.009	0.033**	0.013	0.030**	0.013	—	—	0.053**	0.021	0.052**	0.021	—	—	0.024*	0.012	0.024*	0.012
Incentives: # Clients	—	—	0.009°	0.013	0.013°	0.013	0.076	0.014	0.065	0.021	0.065	0.021	0.064	0.009	0.053	0.012	0.053	0.012
Target Market: Rural	—	—	—	—	—	—	—	—	−0.045**	0.021	−0.045**	0.021	—	—	−0.019*	0.012	−0.019*	0.012
Target Market: Urban	—	—	—	—	—	—	—	—	0.033*	0.019	0.033*	0.019	—	—	0.024**	0.011	0.024**	0.011
Target Market: Youth	0.034	0.010	0.032	0.010	0.030	0.010	—	—	−0.033**	0.015	−0.034**	0.015	—	—	−0.013°	0.010	−0.013°	0.010
Personnel Structure																		
Av. Salary/GNI p. c.	—	—	—	—	—	—	0.086	0.015	0.115	0.017	0.113	0.017	—	—	0.024**	0.010	0.024**	0.010
Borrowers/Loan Officer	—	—	—	—	—	—	—	—	0.008°	0.013	0.008°	0.014	—	—	0.015*	0.009	0.015*	0.009
Borrowers/Office	—	—	—	—	—	—	—	—	−0.052	0.020	—	—	—	—	—	—	—	—
Borrowers/Staff Member	0.299	0.019	0.287	0.019	0.286	0.019	0.255	0.020	0.268	0.023	0.251	0.022	0.278	0.011	0.282	0.012	0.282	0.012
Depositors/Staff	—	—	—	—	—	—	—	—	−0.041**	0.019	−0.043**	0.018	—	—	—	—	—	—
Personnel/Office	0.034	0.010	0.029	0.010	0.030	0.010	0.072	0.015	0.101	0.019	0.072	0.016	0.074	0.010	0.069	0.010	0.069	0.010
Personnel Expense/GLP	—	—	—	—	—	—	—	—	—	—	—	—	—	—	−0.037**	0.016	−0.037**	0.016
# Personnel	0.296	0.024	0.277	0.023	0.278	0.023	0.307	0.018	0.274	0.019	0.281	0.019	0.288	0.011	0.300	0.012	0.300	0.012
% Board Members: Female	—	—	—	—	—	—	—	—	−0.041	0.015	−0.043	0.015	—	—	−0.014°	0.010	−0.014°	0.010
% Loan Officers: Female	—	—	—	—	—	—	—	—	−0.041**	0.017	−0.039**	0.017	—	—	—	—	—	—
% Managers: Female	—	—	—	—	—	—	—	—	0.001°	0.016	0.000°	0.016	−0.039	0.011	−0.034	0.011	−0.034	0.011
% Staff: Board Members	−0.035	0.010	−0.029	0.010	−0.029	0.010	−0.236	0.015	−0.207	0.016	−0.208	0.016	−0.226	0.010	−0.219	0.010	−0.219	0.010
% Staff: Employed > 1 Year	—	—	—	—	—	—	—	—	−0.011°	0.014	−0.013°	0.014	—	—	−0.015*	0.009	−0.015*	0.009

(Table continued on next page)

Table A.12: Detailed two-step lasso results for log(NAB) *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
% Staff: Female	—	—	−0.019*	0.011	−0.020*	0.011	—	—	—	—	—	—	−0.045	0.012	−0.043	0.012	−0.043	0.012
% Staff: Loan Officers	—	—	—	—	—	—	0.106	0.018	0.090	0.019	0.090	0.019	0.086	0.011	0.082	0.011	0.082	0.011
% Staff: Managers	—	—	—	—	—	—	−0.087	0.015	−0.055	0.015	−0.058	0.015	−0.157	0.010	−0.145	0.010	−0.145	0.010
Services																		
Offers Compulsory Insurance	—	—	—	—	—	—	—	—	—	—	—	—	0.029	0.010	0.029	0.010	0.029	0.010
Offers Education Services	—	—	0.017°	0.011	0.017°	0.011	—	—	0.013°	0.016	0.014°	0.016	—	—	0.002°	0.011	0.002°	0.011
Offers Enterprise Services	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.005°	0.010	0.005°	0.010
Offers Health Services	—	—	—	—	—	—	0.054	0.015	0.054	0.016	0.052	0.016	0.044	0.010	0.047	0.010	0.047	0.010
Offers Non-Inc. Gen. Loans Also	—	—	—	—	—	—	—	—	0.030*	0.016	0.029*	0.016	—	—	0.020**	0.010	0.020**	0.010
Offers Other Fin. Services	—	—	0.028*	0.015	—	—	0.051	0.017	0.066	0.017	0.065	0.017	0.081	0.011	0.078	0.011	0.078	0.011
Offers Savings Products	—	—	—	—	—	—	—	—	—	—	—	—	0.005°	0.013	0.002°	0.013	0.002°	0.013
Offers Voluntary Insurance	—	—	—	—	—	—	0.073	0.015	0.062	0.015	0.064	0.015	0.011°	0.009	0.012°	0.009	0.012°	0.009
Offers Women's Empow. Serv.	—	—	—	—	—	—	—	—	0.013°	0.017	0.013°	0.017	0.021**	0.010	0.016°	0.012	0.016°	0.012
# variables in the final model	23		40		38		25		80		79		39		85		85	
Overall $R^2_{av.}$ (adjusted $\bar{R}^2_{av.}$)	0.956 (0.946)		0.959 (0.949)		0.959 (0.949)		0.762 (0.757)		0.797 (0.783)		0.795 (0.782)		0.697 (0.694)		0.709 (0.703)		0.709 (0.703)	
Within $R^2_{av.}$ (adjusted $\bar{R}^2_{av.}$)	0.400 (0.265)		0.438 (0.300)		0.436 (0.299)		—		—		—		—		—		—	

Note: The table contains the two-step lasso coefficients of all variables that were selected to at least one of the nine final models. sd_b denotes the OLS standard error of the estimate. All effects are significant at the 1% level based on standard confidence intervals under normality, unless otherwise stated: ** (*) denotes that the respective effect is significant at the 5% (10%) level, ° implies that it is not significant at the 10% level. "—" symbolises that the respective variable is not selected by lasso. $R^2_{av.}$ ($\bar{R}^2_{av.}$) denotes the average (adjusted) R^2 across all 10 imputed data sets.

Table A.13: Detailed two-step lasso results on the balanced and unbalanced panel data for –log(ALBG)

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		$\mathbf{C_p}$		**BIC**		**AIC**		$\mathbf{C_p}$		**BIC**		**AIC**		$\mathbf{C_p}$	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
Costs																		
Administrative Exp./Assets	—	—	−0.001°	0.013	−0.001°	0.013	—	—	−0.075**	0.031	—	—	—	—	−0.112	0.028	−0.112	0.028
Cost per Borrower	−0.098	0.018	−0.116	0.018	−0.116	0.018	−0.115	0.018	−0.168	0.038	−0.118	0.018	−0.071	0.011	−0.395	0.046	−0.395	0.046
Cost per Loan	—	—	—	—	—	—	—	—	0.059°	0.038	—	—	—	—	0.307	0.047	0.307	0.047
Fin. Exp. Fund. Liab./Assets	—	—	0.149	0.025	0.149	0.025	—	—	0.224	0.038	0.232	0.038	—	—	0.163	0.024	0.163	0.024
Financial Expense/Assets	—	—	−0.164	0.028	−0.164	0.028	—	—	−0.217	0.040	−0.216	0.039	−0.047	0.011	−0.163	0.023	−0.163	0.023
Interest Exp. on Borr./Borr.	—	—	0.005°	0.008	0.005°	0.008	—	—	0.006°	0.012	0.006°	0.012	—	—	0.019**	0.009	0.019**	0.009
Operating Expense/Assets	—	—	—	—	—	—	—	—	0.119*	0.067	—	—	0.090	0.030	0.130*	0.074	0.130*	0.074
Operating Expense/GLP	—	—	—	—	—	—	0.132	0.034	0.193	0.033	0.195	0.027	0.088	0.018	0.631	0.100	0.631	0.100
Financial Performance																		
Financial Revenue/Assets	—	—	—	—	—	—	0.036°	0.026	—	—	—	—	0.079	0.019	—	—	—	—
Operational Self-Sufficiency	—	—	—	—	—	—	—	—	−0.048	0.018	−0.043**	0.018	−0.046	0.010	−0.045	0.010	−0.045	0.010
Profit Margin	−0.047	0.011	−0.056	0.013	−0.056	0.013	—	—	−0.124	0.031	−0.115	0.031	—	—	−0.019°	0.013	−0.019°	0.013
Return on Assets	—	—	—	—	—	—	—	—	0.176	0.029	0.176	0.028	—	—	0.100	0.019	0.100	0.019
Return on Equity	0.026	0.008	0.029	0.008	0.029	0.008	—	—	0.010°	0.013	0.006°	0.012	—	—	0.005°	0.010	0.005°	0.010
Yield on GLP (Real)	—	—	—	—	—	—	—	—	—	—	—	—	0.000°	0.020	0.015°	0.020	0.015°	0.020
Financing Structure																		
Assets	—	—	—	—	—	—	—	—	−3.270	0.714	—	—	—	—	—	—	—	—
Borrowings/Liabilities	—	—	—	—	—	—	—	—	—	—	—	—	—	—	−0.009°	0.009	−0.009°	0.009
Capital/Assets	—	—	—	—	—	—	—	—	0.100°	0.093	0.010°	0.092	—	—	−0.008°	0.017	−0.008°	0.017
Debt/Equity	—	—	—	—	—	—	−0.022*	0.013	−0.006°	0.015	−0.009°	0.015	—	—	0.012°	0.016	0.012°	0.016
Deposits/Loans	—	—	—	—	—	—	—	—	0.035°	0.035	—	—	—	—	−0.353	0.063	−0.353	0.063
Deposits/Total Assets	—	—	—	—	—	—	—	—	−0.058°	0.040	—	—	−0.054	0.012	−0.027*	0.016	−0.027*	0.016
Donated Equity/Equity	—	—	—	—	—	—	—	—	0.021°	0.016	—	—	—	—	−0.007°	0.009	−0.007°	0.009
Equity	—	—	—	—	—	—	—	—	0.660	0.142	—	—	—	—	0.038*	0.023	0.038*	0.023
Liabilities	—	—	—	—	—	—	−0.126	0.015	2.534	0.581	−0.015°	0.063	−0.069	0.011	−0.011°	0.044	−0.011°	0.044
Liabilities/Assets	−0.060	0.016	−0.066	0.018	−0.066	0.018	−0.036**	0.014	0.079°	0.093	−0.015°	0.091	−0.026**	0.012	−0.028°	0.018	−0.028°	0.018

(Table continued on next page)

Table A.13: Detailed two-step lasso results for –log(ALBG) *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
Subord. Debt/Equity	—	—	0.011°	0.010	0.011°	0.010	—	—	−0.014°	0.012	−0.015°	0.012	—	—	0.005°	0.010	0.005°	0.010
% Assets: Non-Earn. Liq. Ass.	0.021*	0.011	0.034	0.011	0.034	0.011	—	—	—	—	—	—	—	—	0.004°	0.010	0.004°	0.010
Operational Aspects																		
Av. Loan Size: Consumption	—	—	—	—	—	—	—	—	0.023°	0.021	—	—	—	—	−0.304**	0.126	−0.304**	0.126
Av. Loan Size: Enterpr. Fin.	—	—	−0.005°	0.008	−0.005°	0.008	—	—	—	—	—	—	—	—	−0.037°	0.046	−0.037°	0.046
Av. Loan Size: Ext. Customers	—	—	—	—	—	—	—	—	0.011°	0.015	—	—	—	—	—	—	—	—
Av. Loan Size: Female	−0.013°	0.011	−0.014°	0.011	−0.014°	0.011	—	—	−0.011°	0.017	−0.011°	0.017	—	—	−1.151	0.227	−1.151	0.227
Av. Loan Size: Household Fin.	—	—	—	—	—	—	−0.023*	0.014	−0.039*	0.021	−0.019°	0.014	—	—	—	—	—	—
Av. Loan Size: Large Corp.	—	—	—	—	—	—	0.053	0.012	0.049	0.013	0.054	0.012	—	—	0.218°	0.139	0.218°	0.139
Av. Loan Size: Legal Entity	—	—	—	—	—	—	—	—	0.020°	0.013	—	—	—	—	—	—	—	—
Av. Loan Size: Male	−0.023*	0.013	−0.017°	0.013	−0.017°	0.013	−0.013°	0.018	−0.012°	0.021	−0.007°	0.020	—	—	1.277	0.491	1.277	0.491
Av. Loan Size: Man. & Staff	—	—	−0.011°	0.008	−0.011°	0.008	—	—	0.008°	0.012	—	—	—	—	2.569	0.622	2.569	0.622
Av. Loan Size: Microenterpr.	−0.047	0.014	−0.043	0.014	−0.043	0.014	−0.023°	0.016	−0.017°	0.018	−0.025°	0.017	—	—	−2.839	0.267	−2.839	0.267
Av. Loan Size: Mortgage, Housing	—	—	—	—	—	—	—	—	0.036	0.013	0.030**	0.013	−0.018*	0.010	−0.012°	0.010	−0.012°	0.010
Av. Loan Size: Other Househ. Fin.	—	—	0.005°	0.007	0.005°	0.007	—	—	−0.005°	0.012	—	—	—	—	0.260°	0.496	0.260°	0.496
Av. Loan Size: Rural	−0.003°	0.011	−0.001°	0.011	−0.001°	0.011	—	—	−0.012°	0.017	—	—	—	—	−0.023°	0.017	−0.023°	0.017
Av. Loan Size: Sm. & Med. Enterpr.	—	—	—	—	—	—	—	—	−0.024*	0.013	−0.013°	0.012	—	—	—	—	—	—
Av. Loan Size: Solidarity Group	—	—	−0.008°	0.008	−0.008°	0.008	—	—	−0.024**	0.012	−0.023*	0.012	—	—	−0.009°	0.009	−0.009°	0.009
Av. Loan Size: Urban	−0.021*	0.012	−0.015°	0.012	−0.015°	0.012	−0.064	0.017	−0.051	0.018	−0.052	0.017	—	—	—	—	—	—
Av. Loan Size: Village (SHG)	—	—	—	—	—	—	—	—	0.010°	0.012	—	—	—	—	0.007°	0.009	0.007°	0.009
Borrower Retention Rate	—	—	0.013*	0.008	0.013*	0.008	—	—	0.031**	0.015	0.027*	0.015	—	—	−0.012°	0.010	−0.012°	0.010
Client Prot.: Clear Debt Coll.	—	—	—	—	—	—	—	—	0.010°	0.013	—	—	—	—	−0.001°	0.011	−0.001°	0.011
Client Prot.: Complaint Mech.	—	—	—	—	—	—	—	—	−0.002°	0.013	—	—	—	—	0.010°	0.011	0.010°	0.011
Client Prot.: Decl. Bal. Method	—	—	—	—	—	—	−0.025**	0.012	−0.016°	0.013	−0.027**	0.012	−0.048	0.010	−0.048	0.010	−0.048	0.010
Client Prot.: Full Disclosure	—	—	−0.007°	0.007	−0.007°	0.007	—	—	−0.005°	0.012	—	—	—	—	−0.011°	0.010	−0.011°	0.010
Client Prot.: Over-Indebt. Prev.	—	—	—	—	—	—	—	—	0.025**	0.013	0.027**	0.012	—	—	−0.023**	0.010	−0.023**	0.010
Client Prot.: Privacy Data Clause	—	—	—	—	—	—	—	—	−0.010°	0.012	−0.011°	0.012	—	—	0.009°	0.010	0.009°	0.010
Client Prot.: Rob. Payment Eval.	—	—	—	—	—	—	—	—	−0.015°	0.012	−0.012°	0.012	—	—	0.002°	0.010	0.002°	0.010

(Table continued on next page)

Table A.13: Detailed two-step lasso results for –log(ALBG) *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
Deposit Accounts/Staff	—	—	—	—	—	—	—	—	0.017°	0.016	0.011°	0.015	—	—	−0.091**	0.036	−0.091**	0.036
GLP/Total Assets	—	—	−0.016*	0.009	−0.016*	0.009	—	—	−0.036**	0.016	−0.034**	0.015	−0.092	0.011	−0.099	0.011	−0.099	0.011
Gross Loan Portfolio (GLP)	—	—	—	—	—	—	—	—	−0.025°	0.103	−0.141**	0.064	—	—	−0.100*	0.053	−0.100*	0.053
Imp. Loss Allow./GLP	—	—	—	—	—	—	—	—	—	—	—	—	—	—	−0.002°	0.012	−0.002°	0.012
Income on Pen. Fees/GLP	—	—	0.018**	0.008	0.018**	0.008	—	—	—	—	—	—	—	—	0.008°	0.009	0.008°	0.009
Interest Income on GLP/GLP	0.063	0.015	0.063	0.015	0.063	0.015	0.148	0.020	0.124	0.021	0.121	0.020	0.174	0.017	0.122	0.019	0.122	0.019
Loan Impairm. Prov./Assets	—	—	−0.031	0.010	−0.031	0.010	—	—	—	—	—	—	—	—	−0.020°	0.013	−0.020°	0.013
Loan Loss Rate	—	—	—	—	—	—	—	—	0.011°	0.013	0.021°	0.013	—	—	−0.010°	0.009	−0.010°	0.009
Measures Client Poverty	—	—	−0.010°	0.010	−0.010°	0.010	—	—	0.009°	0.013	0.013°	0.013	—	—	0.013°	0.010	0.013°	0.010
Net GLP/GLP	—	—	0.037	0.012	0.037	0.012	—	—	0.008°	0.015	0.011°	0.015	—	—	−0.022**	0.010	−0.022**	0.010
Portfolio at Risk: > 30 Days	—	—	—	—	—	—	—	—	0.103*	0.060	—	—	—	—	0.173	0.065	0.173	0.065
Portfolio at Risk: > 90 Days	—	—	—	—	—	—	—	—	−0.120**	0.058	−0.030°	0.022	—	—	−0.206	0.066	−0.206	0.066
Retained Earnings/Equity	—	—	−0.011°	0.009	−0.011°	0.009	—	—	0.033*	0.018	0.018°	0.014	—	—	0.020°	0.015	0.020°	0.015
Risk Coverage	—	—	−0.010°	0.008	−0.010°	0.008	—	—	—	—	—	—	—	—	−0.009°	0.009	−0.009°	0.009
Tax Expense/Assets	—	—	—	—	—	—	−0.035	0.013	−0.011°	0.014	−0.015°	0.013	—	—	−0.002°	0.010	−0.002°	0.010
Write-Off Ratio	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.004°	0.011	0.004°	0.011
# Active Borrowers	—	—	−0.171	0.034	−0.171	0.034	—	—	−0.152	0.051	—	—	—	—	−0.058°	0.036	−0.058°	0.036
# Loans Outstanding	—	—	—	—	—	—	—	—	0.041°	0.035	—	—	—	—	−0.015°	0.031	−0.015°	0.031
# New Borrowers/NAB	—	—	—	—	—	—	—	—	0.014°	0.015	0.013°	0.015	—	—	−0.010°	0.009	−0.010°	0.009
% Borrowers: Ext.Customers	—	—	0.002°	0.008	0.002°	0.008	—	—	0.015°	0.015	0.015°	0.015	—	—	0.004°	0.011	0.004°	0.011
% Borrowers: Female	0.061**	0.027	0.051*	0.026	0.051*	0.026	0.053	0.017	0.059	0.018	0.048°	0.035	0.096	0.013	0.021°	0.025	0.021°	0.025
% Borrowers: Legal Entity	—	—	—	—	—	—	—	—	0.002°	0.013	—	—	−0.053	0.010	−0.037	0.011	−0.037	0.011
% Borrowers: Male	−0.036*	0.021	−0.028°	0.021	−0.028°	0.021	—	—	—	—	0.002°	0.032	—	—	−0.049**	0.022	−0.049**	0.022
% Borrowers: Man. & Staff	—	—	—	—	—	—	—	—	−0.016°	0.018	−0.007°	0.014	—	—	−0.011°	0.011	−0.011°	0.011
% Borrowers: Rural	—	—	−0.049	0.012	−0.049	0.012	—	—	−0.013°	0.037	−0.019°	0.013	—	—	—	—	—	—
% Borrowers: Urban	—	—	—	—	—	—	—	—	0.014°	0.037	—	—	—	—	0.008°	0.011	0.008°	0.011
% GLP: 1-3 Months Delinquent	—	—	0.011°	0.007	0.011°	0.007	—	—	−0.003°	0.013	0.013°	0.012	—	—	−0.042*	0.024	−0.042*	0.024
% GLP: >3 Months Delinquent	0.033	0.009	0.035	0.009	0.035	0.009	—	—	—	—	—	—	—	—	0.031°	0.029	0.031°	0.029

(Table continued on next page)

Table A.13: Detailed two-step lasso results for –log(ALBG) *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
% GLP: Consumption	—	—	—	—	—	—	—	—	0.017°	0.029	0.044**	0.017	—	—	—	—	—	—
% GLP: Enterprise Finance	—	—	—	—	—	—	0.020°	0.017	0.040*	0.021	0.053	0.019	—	—	0.055	0.019	0.055	0.019
% GLP: Household Financing	—	—	—	—	—	—	—	—	0.063°	0.039	—	—	—	—	0.058	0.021	0.058	0.021
% GLP: Individual	—	—	—	—	—	—	−0.037**	0.015	−0.041**	0.016	−0.010°	0.021	−0.048	0.015	−0.021°	0.021	−0.021°	0.021
% GLP: Large Corporations	—	—	−0.010°	0.008	−0.010°	0.008	—	—	—	—	−0.009°	0.013	—	—	0.005°	0.011	0.005°	0.011
% GLP: Microenterprise	—	—	—	—	—	—	0.013°	0.018	0.076**	0.033	0.020°	0.019	0.009°	0.013	0.003°	0.026	0.003°	0.026
% GLP: Mortgage, Housing	—	—	—	—	—	—	—	—	−0.014°	0.019	0.002°	0.013	−0.039	0.011	−0.046	0.011	−0.046	0.011
% GLP: Other Household Fin.	—	—	0.013*	0.008	0.013*	0.008	—	—	0.020°	0.021	0.035	0.013	—	—	−0.011°	0.012	−0.011°	0.012
% GLP: Renegotiated Loans	—	—	−0.021**	0.009	−0.021**	0.009	—	—	—	—	0.004°	0.020	—	—	0.015°	0.019	0.015°	0.019
% GLP: Small & Med. Enterpr.	—	—	—	—	—	—	—	—	0.042**	0.021	—	—	−0.047	0.012	−0.058	0.017	−0.058	0.017
% GLP: Solidarity Group	—	—	—	—	—	—	—	—	—	—	0.032°	0.020	0.028*	0.014	0.040**	0.020	0.040**	0.020
% GLP: Village Banking (SHG)	0.051	0.014	0.047	0.013	0.047	0.013	—	—	−0.039**	0.015	—	—	—	—	0.019°	0.015	0.019°	0.015
Organisational Strategy																		
Board: Social Training	—	—	0.013°	0.009	0.013°	0.009	—	—	−0.003°	0.012	—	—	—	—	0.001°	0.010	0.001°	0.010
Goals: Econ. Improvement	—	—	—	—	—	—	—	—	−0.009°	0.018	—	—	—	—	0.004°	0.013	0.004°	0.013
Goals: Education Opport.	—	—	—	—	—	—	—	—	0.032**	0.015	0.033**	0.015	—	—	−0.011°	0.011	−0.011°	0.011
Goals: Health Infrastructure	—	—	−0.019*	0.010	−0.019*	0.010	−0.027**	0.012	−0.034**	0.015	−0.043	0.014	—	—	−0.007°	0.011	−0.007°	0.011
Goals: Women's Empow.	—	—	—	—	—	—	—	—	0.028*	0.015	0.021°	0.013	—	—	−0.012°	0.012	−0.012°	0.012
Has Poverty Target	—	—	—	—	—	—	—	—	—	—	—	—	0.029	0.010	0.007°	0.011	0.007°	0.011
Incentives: Data-Collect. Qual.	—	—	—	—	—	—	—	—	−0.014°	0.013	−0.017°	0.012	—	—	−0.016°	0.011	−0.016°	0.011
Incentives: Interaction Qual.	0.020*	0.010	0.017*	0.010	0.017*	0.010	—	—	−0.010°	0.014	—	—	—	—	0.016°	0.011	0.016°	0.011
Incentives: Portf. Qual.	—	—	—	—	—	—	—	—	−0.027°	0.018	—	—	—	—	−0.016°	0.013	−0.016°	0.013
Incentives: # Clients	—	—	—	—	—	—	—	—	0.022°	0.018	0.007°	0.012	—	—	0.027**	0.013	0.027**	0.013
Target Market: Rural	—	—	—	—	—	—	—	—	0.021°	0.018	—	—	—	—	0.024**	0.012	0.024**	0.012
Target Market: Urban	—	—	—	—	—	—	—	—	−0.011°	0.017	—	—	—	—	−0.011°	0.012	−0.011°	0.012
Target Market: Women	—	—	−0.017*	0.010	−0.017*	0.010	—	—	—	—	—	—	—	—	0.005°	0.013	0.005°	0.013
Target Market: Youth	—	—	—	—	—	—	—	—	−0.014°	0.013	−0.011°	0.013	—	—	−0.009°	0.010	−0.009°	0.010

(Table continued on next page)

Table A.13: Detailed two-step lasso results for –log(ALBG) *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
Personnel Structure																		
Av. Salary/GNI p. c.	−0.098	0.015	−0.097	0.015	−0.097	0.015	−0.415	0.013	−0.422	0.013	−0.418	0.013	−0.326	0.010	−0.326	0.010	−0.326	0.010
Borrowers/Loan Officer	—	—	—	—	—	—	—	—	0.006°	0.011	—	—	—	—	0.029	0.009	0.029	0.009
Borrowers/Office	—	—	—	—	—	—	—	—	0.023°	0.014	0.019°	0.014	—	—	0.062	0.014	0.062	0.014
Borrowers/Staff Member	0.208	0.020	0.265	0.021	0.265	0.021	0.359	0.017	0.399	0.022	0.368	0.021	0.274	0.012	0.260	0.014	0.260	0.014
Depositors/Staff	—	—	−0.029**	0.014	−0.029**	0.014	—	—	—	—	—	—	—	—	0.090**	0.035	0.090**	0.035
Personnel/Office	—	—	—	—	—	—	—	—	—	—	—	—	−0.022**	0.010	−0.060	0.013	−0.060	0.013
Personnel Expense/Assets	0.035°	0.030	0.055*	0.031	0.055*	0.031	0.104	0.037	0.106	0.040	0.155	0.022	0.077**	0.034	0.261	0.061	0.261	0.061
Personnel Expense/GLP	0.073**	0.029	0.031°	0.031	0.031°	0.031	0.058°	0.045	—	—	—	—	0.000°	0.025	−0.334	0.060	−0.334	0.060
# Personnel	—	—	—	—	—	—	—	—	0.095	0.032	—	—	—	—	0.074	0.021	0.074	0.021
% Board Members: Female	—	—	—	—	—	—	—	—	−0.009°	0.012	—	—	—	—	0.016°	0.010	0.016°	0.010
% Loan Officers: Female	—	—	—	—	—	—	—	—	—	—	—	—	0.054	0.010	0.041	0.014	0.041	0.014
% Managers: Female	—	—	0.012°	0.009	0.012°	0.009	—	—	0.024*	0.014	0.024*	0.014	—	—	0.016°	0.012	0.016°	0.012
% Staff: Board Members	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.004°	0.010	0.004°	0.010
% Staff: Employed > 1 Year	—	—	—	—	—	—	—	—	−0.006°	0.012	—	—	—	—	−0.009°	0.009	−0.009°	0.009
% Staff: Female	—	—	—	—	—	—	0.039	0.012	0.022°	0.015	0.016°	0.015	—	—	0.006°	0.017	0.006°	0.017
% Staff: Leaving During Period	—	—	0.002°	0.008	0.002°	0.008	—	—	—	—	—	—	—	—	0.013°	0.010	0.013°	0.010
% Staff: Loan Officers	—	—	—	—	—	—	0.072	0.015	0.046	0.016	0.063	0.015	0.078	0.011	0.065	0.012	0.065	0.012
% Staff: Managers	−0.041	0.010	−0.042	0.009	−0.042	0.009	−0.059	0.012	−0.047	0.013	−0.042	0.013	—	—	−0.004°	0.010	−0.004°	0.010
Services																		
Offers Compulsory Insurance	—	—	0.023**	0.012	0.023**	0.012	0.067	0.013	0.055	0.013	0.055	0.013	0.070	0.010	0.067	0.010	0.067	0.010
Offers Education Services	—	—	—	—	—	—	—	—	—	—	—	—	—	—	−0.020*	0.011	−0.020*	0.011
Offers Enterprise Services	—	—	−0.022*	0.012	−0.022*	0.012	—	—	−0.016°	0.013	−0.020°	0.012	—	—	0.016°	0.011	0.016°	0.011
Offers Health Services	—	—	—	—	—	—	—	—	−0.007°	0.013	—	—	0.034	0.011	0.028	0.011	0.028	0.011
Offers Non-Inc. Gen. Loans Also	—	—	—	—	—	—	−0.009°	0.013	−0.004°	0.014	−0.009°	0.014	−0.027	0.010	−0.028	0.011	−0.028	0.011
Offers Other Fin. Services	—	—	—	—	—	—	−0.086	0.014	−0.057	0.016	−0.068	0.015	−0.091	0.011	−0.084	0.011	−0.084	0.011
Offers Savings Products	—	—	—	—	—	—	—	—	−0.024°	0.018	−0.030*	0.016	—	—	−0.034**	0.014	−0.034**	0.014
Offers Voluntary Insurance	—	—	—	—	—	—	—	—	0.021°	0.013	0.018°	0.013	—	—	0.025**	0.010	0.025**	0.010

(Table continued on next page)

Table A.13: Detailed two-step lasso results for –log(ALBG) *(continued)*

	Balanced Panel, Fixed Effects						Balanced Panel, Pooled OLS						Unbalanced Panel, Pooled OLS					
	BIC		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**		**BIC**		**AIC**		**C_p**	
Variable	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b	b	sd_b
Offers Women's Empow. Serv.	—	—	0.003°	0.011	0.003°	0.011	—	—	−0.015°	0.015	—	—	0.035	0.011	0.037	0.012	0.037	0.012
# variables in the final model	21		55		55		29		108		74		34		123		123	
Overall $R^2_{\text{av.}}$ (adjusted $\bar{R}^2_{\text{av.}}$)	0.950 (0.939)		0.956 (0.944)		0.956 (0.944)		0.835 (0.831)		0.862 (0.850)		0.855 (0.846)		0.650 (0.647)		0.693 (0.683)		0.693 (0.683)	
Within $R^2_{\text{av.}}$ (adjusted $\bar{R}^2_{\text{av.}}$)	0.315 (0.162)		0.395 (0.235)		0.395 (0.235)		—		—		—		—		—		—	

Note: The table contains the two-step lasso coefficients of all variables that were selected to at least one of the nine final models. sd_b denotes the OLS standard error of the estimate. All effects are significant at the 1% level based on standard confidence intervals under normality, unless otherwise stated: ** (*) denotes that the respective effect is significant at the 5% (10%) level, ° implies that it is not significant at the 10% level. "—" symbolises that the respective variable is not selected by lasso. $R^2_{\text{av.}}$ ($\bar{R}^2_{\text{av.}}$) denotes the average (adjusted) R^2 across all 10 imputed data sets.

B Appendix to Chapter 3: Comparing Forecast Performance on Large Panel Data with Unknown Clustering Structure

B.1 Proof of Consistency

In the following, we show that $|\hat{\sigma}_{\bar{d}}^2 - \sigma_{\bar{d}}^2| = o_P(1)$. Let $\check{\sigma}_{\bar{d}}^2 = \frac{1}{N}\sum_{i,j} s_{ij}$. By the triangle inequality, we have

$$|\hat{\sigma}_{\bar{d}}^2 - \sigma_{\bar{d}}^2| \leq |\hat{\sigma}_{\bar{d}}^2 - \check{\sigma}_{\bar{d}}^2| + |\check{\sigma}_{\bar{d}}^2 - \sigma_{\bar{d}}^2|.$$

We conduct the proof by first showing that $\max_{ij} |\hat{s}_{ij} - s_{ij}| = O_P(\omega_{NT})$, where $\omega_{NT} = L\sqrt{\frac{\log[(L+1)N^2T]}{T}}$ (**Part I**). With that, we can show that $|\hat{\sigma}_{\bar{d}}^2 - \check{\sigma}_{\bar{d}}^2| = o_P(1)$ (**Part II**). Finally, we show that $|\check{\sigma}_{\bar{d}}^2 - \sigma_{\bar{d}}^2| = o(1)$ (**Part III**) which concludes the proof.

Part I: Show that $\max_{ij} |\hat{s}_{ij} - s_{ij}| = O_P(\omega_{NT})$.

Recall that $z_{h,ij,t} = (d_{it}d_{j,t-h} - \mathbb{E}d_{it}d_{j,t-h})\,\omega(h,L)\mathbb{1}\{t \geq h\}$; $L = 4\,(T/100)^{\frac{2}{9}}$. Set $\alpha_{NT} = \omega_{NT}/L$, define $v^2 = \max_{0\leq h\leq L}\max_{ij}\sup_{t>0}\left(\mathrm{Var}(z_{h,ij,t}) + 2\sum_{t'>t}|\mathrm{Cov}(z_{h,ij,t}, z_{h,ij,t'})|\right)$ and let $\log N = O(\log T)$. By Assumption 2 (i), $0 < v^2 < \infty$. Under Assumptions 2 (ii) and (iii), we can apply Theorem 2 from Merlevède et al. (2009) and obtain the following, where C is a positive constant that depends on C_3:

$$\begin{aligned}
P\left(\max_{0\leq h\leq L}\max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T} z_{h,ij,t}\right| > \alpha_{NT}\right) &\leq (L+1)N^2 \max_{0\leq h\leq L}\max_{ij} P\left(\left|\sum_{t=1}^{T} z_{h,ij,t}\right| > T\alpha_{NT}\right) \\
&\leq (L+1)N^2 \exp\left\{-\frac{C(T\alpha_{NT})^2}{v^2T + C_4^2 + C_4T\alpha_{NT}(\log T)^2}\right\} \\
&= \exp\left\{\log[(L+1)N^2] - \frac{C\log[(L+1)N^2T]}{v^2 + \frac{1}{T}C_4^2 + C_4\sqrt{\frac{\log[(L+1)N^2T]}{T}}(\log T)^2}\right\} \\
&\to 0 \text{ for } T \to \infty.
\end{aligned}$$

Consequently, $\max_h \max_{ij} |\frac{1}{T}\sum_{t=1}^T z_{h,ij,t}| = O_P\left(\sqrt{\frac{\log[(L+1)N^2T]}{T}}\right)$, and

$$\begin{aligned}
\max_{ij} |\hat{s}_{ij} - s_{ij}| &\leq \max_{ij} |\frac{1}{T}\sum_{t=1}^T d_{it}d_{jt} - \mathbb{E}d_{it}d_{jt}| \\
&\quad + 2\max_{ij} |\frac{1}{T}\sum_{h=1}^L \omega(h,L) \sum_{t=h+1}^T d_{it}d_{j,t-h} - \mathbb{E}d_{it}d_{j,t-h}| \\
&\leq \max_{ij} |\frac{1}{T}\sum_{t=1}^T d_{it}d_{jt} - \mathbb{E}d_{it}d_{jt}| \\
&\quad + 2L \max_{1\leq h\leq L} \max_{ij} |\frac{1}{T}\sum_{t=1}^T \omega(h,L)\mathbb{1}\{t \geq h\} \left(d_{it}d_{j,t-h} - \mathbb{E}d_{it}d_{j,t-h}\right)| \\
&= O_P\left(\sqrt{\frac{\log[(L+1)N^2T]}{T}}\right) + LO_P\left(\sqrt{\frac{\log[(L+1)N^2T]}{T}}\right) \\
&= O_P\left(L\sqrt{\frac{\log[(L+1)N^2T]}{T}}\right) = O_P(\omega_{NT}).
\end{aligned}$$

Part II: Show that $|\hat{\sigma}_{\bar{d}}^2 - \check{\sigma}_{\bar{d}}^2| = o_P(1)$.

Note that $\hat{\sigma}_{\bar{d}}^2 = \frac{1}{N}\sum_{(i,j)\in S_l} \hat{s}_{ij}$ and $\check{\sigma}_{\bar{d}}^2 = \frac{1}{N}\sum_{ij} s_{ij}$. Hence,

$$|\hat{\sigma}_{\bar{d}}^2 - \check{\sigma}_{\bar{d}}^2| \leq \frac{1}{N}\sum_{(i,j)\notin \hat{S}_l} |\hat{s}_{ij} - s_{ij}| + \frac{1}{N}\sum_{(i,j)\in \hat{S}_l} |\hat{s}_{ij} - s_{ij}|.$$

Recall that $\rho_{ij,h} = \max_{t\leq T} [|\mathbb{E}d_{it}d_{j,t-j}| + |\mathbb{E}d_{i,t-h}d_{jt}|]$. In the following, we need:

- $\boldsymbol{\max_{ij} |\hat{s}_{ij} - s_{ij}| = O_P(\omega_{NT})}$. Proof: Part I.
- $\boldsymbol{\lambda_{ij} = M\omega_{NT}\sqrt{s_{ii}s_{jj}} + o_P(\omega_{NT})}$. Proof:

$$\begin{aligned}
\lambda_{ij} &= M\omega_{NT}\sqrt{\hat{s}_{ii}\hat{s}_{jj}} \\
&= M\omega_{NT}\sqrt{(s_{ii} + \Delta_{ii})(s_{jj} + \Delta_{jj})},
\end{aligned}$$

where $\Delta_{ii} = \hat{s}_{ii} - s_{ii},\ \Delta_{jj} = \hat{s}_{jj} - s_{jj}$. By Assumptions 2 (i) and 3 (ii), $0 < C_6 < s_{ii} \leq C_2$. With $\max_{ij} |\hat{s}_{ij} - s_{ij}| = O_P(\omega_{NT})$, we have $\Delta_{ii}, \Delta_{jj} = O_P(\omega_{NT})$. Hence,

$$\frac{(s_{ii} + \Delta_{ii})(s_{jj} + \Delta_{jj})}{s_{ii}s_{jj}} = 1 + \frac{\Delta_{ii}}{s_{ii}} + \frac{\Delta_{jj}}{s_{jj}} + O_P(\omega_{NT}^2) = 1 + O_P(\omega_{NT}).$$

Using $\sqrt{1+u} = 1 + \frac{1}{2}u + o(u)$ for $u = O_P(\omega_{NT})$, we obtain

$$\sqrt{\hat{s}_{ii}\hat{s}_{jj}} = \sqrt{s_{ii}s_{jj}}\,(1 + o_P(1)).$$

Therefore,

$$\begin{aligned}\lambda_{ij} &= M\omega_{NT}\sqrt{s_{ii}s_{jj}}\,(1 + o_P(1)) \\ &= M\omega_{NT}\sqrt{s_{ii}s_{jj}} + o_P(\omega_{NT}).\end{aligned}$$

- $\boldsymbol{\lambda_{ij} - \max_{ij}|\hat{s}_{ij} - s_{ij}| > 0}$ **with probability approaching one**. Proof: By Assumption 3 (ii) $s_{ii} > C_6 > 0$. From $\lambda_{ij} = M\omega_{NT}\sqrt{s_{ii}s_{jj}} + o_P(\omega_{NT})$, we have $\lambda_{ij}/\omega_{NT} \xrightarrow{P} M\sqrt{s_{ii}s_{jj}} \geq MC_6$. Since $\max_{ij}|\hat{s}_{ij} - s_{ij}| = O_P(\omega_{NT})$, for any $\varepsilon > 0$, there exists $K > 0$ such that $P(\max_{ij}|\hat{s}_{ij} - s_{ij}| \leq K\omega_{NT}) > 1 - \varepsilon$ for large T. Choose M sufficiently large so that $MC_6 > K$. Then,

$$P\Big(\lambda_{ij} - \max_{ij}|\hat{s}_{ij} - s_{ij}| > 0\Big) \geq P[(MC_6 - K)\omega_{NT} + o_P(\omega_{NT}) > 0] \to 1.$$

- $\boldsymbol{|s_{ij}| \leq \sum_{h=0}^{L}\rho_{ij,h}}$. Proof:

$$\begin{aligned}|s_{ij}| &= |\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}d_{it}d_{jt} + \frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}[\mathbb{E}d_{it}d_{j,t-h} + \mathbb{E}d_{i,t-h}d_{jt}]\,| \\ &\leq \rho_{ij,h} + \frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}\rho_{ij,h} \\ &\leq \sum_{h=0}^{L}\rho_{ij,h}.\end{aligned}$$

Note that by Assumption 2 (i), $\sqrt{s_{ii}s_{jj}} \leq C_2$. With that, the above bullet points and denoting the indicator function for event A by $\mathbb{1}\{A\}$, we obtain for $q \in (0,1)$ and $K > 0$

$$\begin{aligned}\frac{1}{N}\sum_{(i,j)\notin\hat{S}_l}|\hat{s}_{ij} - s_{ij}| &= \frac{1}{N}\sum_{(i,j)\notin\hat{S}_l}|s_{ij}| \\ &\leq \frac{1}{N}\sum_{ij}|s_{ij}|\,\mathbb{1}\{|\hat{s}_{ij}| < \lambda_{ij}\} \\ &\leq \frac{1}{N}\sum_{ij}|s_{ij}|\,\mathbb{1}\{|s_{ij}| \leq |\hat{s}_{ij}| + |\hat{s}_{ij} - s_{ij}|,\, |\hat{s}_{ij}| < \lambda_{ij}\}\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{N}\sum_{ij}|s_{ij}|\,\mathbb{1}\{|s_{ij}|\leq|\lambda_{ij}|+\max_{ij}|\hat{s}_{ij}-s_{ij}|\}\\
&\leq \frac{1}{N}\sum_{ij}|s_{ij}|\,\frac{(|\lambda_{ij}|+\max_{ij}|\hat{s}_{ij}-s_{ij}|)^{1-q}}{|s_{ij}|^{1-q}}\,\mathbb{1}\{|s_{ij}|\leq|\lambda_{ij}|+\max_{ij}|\hat{s}_{ij}-s_{ij}|\}\\
&\leq \frac{1}{N}\sum_{ij}|s_{ij}|^{q}\left[M\omega_{NT}\sqrt{s_{ii}s_{jj}}+K\omega_{NT}+o_P(\omega_{NT})\right]^{1-q}\\
&\leq \frac{1}{N}\sum_{ij}|s_{ij}|^{q}\,(M\sqrt{s_{ii}s_{jj}}+K)^{1-q}\omega_{NT}^{1-q}+o_P(\omega_{NT}^{1-q})\\
&\leq (MC_2+K)^{1-q}\omega_{NT}^{1-q}\,\frac{1}{N}\sum_{ij}|s_{ij}|^{q}+o_P(\omega_{NT}^{1-q})\\
&\leq (MC_2+K)^{1-q}\omega_{NT}^{1-q}\,\frac{1}{N}\sum_{ij}\left(\sum_{h=0}^{L}\rho_{ij,h}\right)^{q}+o_P(\omega_{NT}^{1-q})\\
&\leq (MC_2+K)^{1-q}\,\omega_{NT}^{1-q}\max_{i}\sum_{j}\left(\sum_{h=0}^{L}\rho_{ij,h}\right)^{q}+o_P(\omega_{NT}^{1-q})\\
&= o(1)+o_P(\omega_{NT}^{1-q})=o_P(1),
\end{aligned}
$$

where the last equation follows from Assumption 2 (iv). Further,

$$
\begin{aligned}
\frac{1}{N}\sum_{(i,j)\in\hat{S}_l}|\hat{s}_{ij}-s_{ij}| &\leq \frac{1}{N}\sum_{(i,j)\in\hat{S}_l}\max_{ij}|\hat{s}_{ij}-s_{ij}|\\
&\leq \frac{1}{N}\sum_{ij}\max_{ij}|\hat{s}_{ij}-s_{ij}|\,\mathbb{1}\{|\hat{s}_{ij}|\geq\lambda_{ij}\}\\
&\quad+\frac{1}{N}\sum_{i}\max_{ij}|\hat{s}_{ij}-s_{ij}|\,\mathbb{1}\{|\hat{s}_{ii}|<\lambda_{ii}\}\\
&\leq \frac{1}{N}\sum_{ij}\max_{ij}|\hat{s}_{ij}-s_{ij}|\,\mathbb{1}\{|s_{ij}|\geq|\hat{s}_{ij}|-|\hat{s}_{ij}-s_{ij}|,|\hat{s}_{ij}|\geq\lambda_{ij}\}\\
&\quad+\frac{1}{N}\sum_{i}\max_{ij}|\hat{s}_{ij}-s_{ij}|\,\mathbb{1}\{|\hat{s}_{ii}|<M\omega_{NT}|\hat{s}_{ii}|\}\\
&\leq \frac{1}{N}\sum_{ij}\max_{ij}|\hat{s}_{ij}-s_{ij}|\,\mathbb{1}\{|s_{ij}|\geq\lambda_{ij}-\max_{ij}|\hat{s}_{ij}-s_{ij}|\}\\
&\quad+\frac{1}{N}\sum_{i}\max_{ij}|\hat{s}_{ij}-s_{ij}|\,\mathbb{1}\{1<M\omega_{NT}\}\\
&\leq \frac{1}{N}\sum_{ij}\max_{ij}|\hat{s}_{ij}-s_{ij}|\,\frac{|s_{ij}|^{q}}{(\lambda_{ij}-\max_{ij}|\hat{s}_{ij}-s_{ij}|)^{q}}\\
&\quad\times\mathbb{1}\{|s_{ij}|\geq|\lambda_{ij}|-\max_{ij}|\hat{s}_{ij}-s_{ij}|\}+o(\omega_{NT})
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{1}{N}\sum_{ij}|s_{ij}|^q \frac{\max_{ij}|\hat{s}_{ij}-s_{ij}|}{(\lambda_{ij}-\max_{ij}|\hat{s}_{ij}-s_{ij}|)^q}+o(\omega_{NT})\\
&\leq \frac{1}{N}\sum_{ij}\left(\sum_{h=0}^{L}\rho_{ij,h}\right)^q \frac{K\omega_{NT}+o_P(\omega_{NT})}{\left[M\omega_{NT}\sqrt{s_{ii}s_{jj}}-K\omega_{NT}+o_P(\omega_{NT})\right]^q}+o(\omega_{NT})\\
&\leq \frac{K}{(MC_6-K)^q}\,\omega_{NT}^{1-q}\max_i\sum_j\left(\sum_{h=0}^{L}\rho_{ij,h}\right)^q+o_P(\omega_{NT}^{1-q})\\
&= o(1)+o_P(\omega_{NT}^{1-q})=o_P(1)
\end{aligned}
$$

and consequently

$$
|\hat{\sigma}_{\bar{d}}^2-\check{\sigma}_{\bar{d}}^2|\leq \frac{1}{N}\sum_{(i,j)\notin\hat{S}_l}|\hat{s}_{ij}-s_{ij}|+\frac{1}{N}\sum_{(i,j)\in\hat{S}_l}|\hat{s}_{ij}-s_{ij}|=o_P(1).
$$

Part III: Show that $|\check{\boldsymbol{\sigma}}_{\bar{d}}^{\mathbf{2}}-\boldsymbol{\sigma}_{\bar{d}}^{\mathbf{2}}|=\boldsymbol{o}(\mathbf{1})$.

First, note that with $\mathbf{1}$ representing an $N\times 1$ all-ones vector, we have

$$
\begin{aligned}
|\mathbb{E}\mathbf{1}'d_td_{t-h}'\mathbf{1}+\mathbb{E}\mathbf{1}'d_{t-h}d_t'\mathbf{1}| &= \|\mathbb{E}\mathbf{1}'d_td_{t-h}'\mathbf{1}+\mathbb{E}\mathbf{1}'d_{t-h}d_t'\mathbf{1}\|\\
&\leq \|\mathbf{1}\|\|\mathbb{E}d_td_{t-h}'\|\|\mathbf{1}\|+\|\mathbf{1}\|\|\mathbb{E}d_{t-h}d_t'\|\|\mathbf{1}\|\leq N\gamma_{NT}(h).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
|\check{\sigma}_{\bar{d}}^2-\sigma_{\bar{d}}^2| &\leq |\frac{1}{NT}\sum_{h=1}^{T-1}\sum_{t=h+1}^{T}\left[\mathbb{E}\mathbf{1}'d_td_{t-h}'\mathbf{1}+\mathbb{E}\mathbf{1}'d_{t-h}d_t'\mathbf{1}\right]|\\
&\quad+|\frac{1}{NT}\sum_{h=1}^{L}(1-\omega(h,L))\sum_{t=h+1}^{T}\left[\mathbb{E}\mathbf{1}'d_td_{t-h}'\mathbf{1}+\mathbb{E}\mathbf{1}'d_{t-h}d_t'\mathbf{1}\right]|\\
&\leq \frac{1}{T}\sum_{h=1}^{T-1}\sum_{t=h+1}^{T}\gamma_{NT}(h)+\frac{1}{T}\sum_{h=1}^{L}(1-\omega(h,L))\sum_{t=h+1}^{T}\gamma_{NT}(h)\\
&\leq \sum_{h>L}\gamma_{NT}(h)+\sum_{h=1}^{L}|1-\omega(h,L)|\,\gamma_{NT}(h)=o(1).
\end{aligned}
$$

The first term goes to zero because $L\to\infty$ as $T\to\infty$ and the second term goes to zero due to $|1-\omega(h,L)|\to 0$ as $L\to\infty$ (Assumption 3 (i)) and $\gamma_{NT}(h)$ being summable over h (Assumption 2 (i)).

B.2 Descriptive Statistics

Table B.1: Descriptive statistics

Variable	Mean	SD	Min	Q1	Median	Q3	Max	N	%NA
Sovereign CDS Spread Changes									
AT	-0.006	1.471	-15.700	-0.500	0.000	0.500	13.033	22752	0.112
BE	-0.004	1.573	-14.117	-0.592	0.000	0.583	13.250	22752	0.116
DE	-0.000	0.930	-8.750	-0.300	0.000	0.292	10.000	22752	0.109
ES	-0.000	2.124	-17.062	-0.842	0.000	0.833	22.200	22752	0.070
FI	-0.004	1.374	-10.000	-0.375	0.000	0.375	9.500	22752	0.159
FR	-0.001	1.137	-8.500	-0.400	0.000	0.393	13.000	22752	0.103
IE	-0.015	2.968	-54.167	-1.250	0.000	1.242	35.000	22752	0.090
IT	-0.011	2.063	-22.250	-0.783	0.000	0.750	23.000	22752	0.067
NL	-0.001	1.398	-14.417	-0.445	0.000	0.432	12.000	22752	0.145
PT	0.003	4.891	-77.468	-1.393	0.000	1.492	82.990	22752	0.098
UK	-0.004	1.290	-9.500	-0.375	0.000	0.375	14.750	22752	0.122
Intraday Variables									
$r_{d,t}^{\text{EUSTX50}}$	-0.000	0.003	-0.032	-0.001	0.000	0.001	0.020	21330	0.002
$\text{RV}_{d,t}^{\text{EUSTX50}}$	0.002	0.001	0.000	0.001	0.002	0.003	0.018	21330	0.002
Daily Variables									
FS_d	0.335	0.228	-0.375	0.136	0.321	0.492	1.006	1422	0.002
Slope_d	0.887	0.530	0.104	0.451	0.762	1.290	2.075	1422	0.002
$\Delta \ln FX_d^{\text{GBP/EUR}}$	0.000	0.005	-0.022	-0.003	0.000	0.003	0.031	1422	0.000
$\Delta \ln FX_d^{\text{CHF/EUR}}$	0.000	0.005	-0.084	-0.001	0.000	0.002	0.031	1422	0.000
$\Delta \ln FX_d^{\text{NOK/EUR}}$	0.000	0.005	-0.033	-0.003	0.000	0.003	0.022	1422	0.000
ΔVIX_d	-0.032	1.732	-12.940	-0.750	-0.110	0.530	16.000	1422	0.028
ΔEVZ_d	-0.018	0.494	-3.800	-0.250	-0.030	0.190	3.290	1422	0.028
$\Delta\text{iTraxx}_d^{\text{Corp}}$	0.017	0.501	-2.843	-0.236	0.015	0.288	3.973	1422	0.020

Note: Summary statistics of the completed data set, see Table 3.1 for variable definitions. $Q1$ and $Q3$ denote the first (third) quartile. N denotes the total number of observations; *%NA* gives the proportion of imputed values. Data cover January 2009 to December 2014. The number of observations differs between CDS spread changes and the intraday regressors because CDS data are available from 09:30 to 17:00 (used as regressors and 1-step-ahead target variable), whereas intraday variables are included only from 09:30 to 16:30 (used as regressors only).

B.3 Detailed Results

Table B.2: Pairwise Diebold-Mariano test results for equal predictive accuracy of 30-minute sovereign CDS forecasts

			Hard-Thresholding					Zero-Lag			Newey-West			Driscoll-Kraay		
Model m_1	**Model m_2**	$\bar{d}$	M	$\hat{\sigma}_{\bar{d},NT}$	J_{NT}	p	p^{BF}	$\hat{\sigma}^{ZL}_{\bar{d},NT}$	J^{ZL}_{NT}	p_{ZL}	$\hat{\sigma}^{NW}_{\bar{d},NT}$	J^{NW}_{NT}	p_{NW}	$\hat{\sigma}^{DK}_{\bar{d},NT}$	J^{DK}_{NT}	p_{DK}
Last-value	GARCH	8.880	0.00	0.601	14.769	0.000	0.000	0.292	30.404	0.000	0.555	15.986	0.000	0.601	14.769	0.000
Last-value	Linear	9.619	0.00	0.597	16.107	0.000	0.000	0.283	33.940	0.000	0.538	17.873	0.000	0.597	16.107	0.000
Last-value	Linear^{R2}	9.641	0.00	0.600	16.079	0.000	0.000	0.285	33.875	0.000	0.541	17.836	0.000	0.600	16.079	0.000
Last-value	Linear^{R3}	9.637	0.00	0.600	16.064	0.000	0.000	0.285	33.826	0.000	0.541	17.818	0.000	0.600	16.064	0.000
Last-value	PVAR-X	9.803	0.00	0.638	15.366	0.000	0.000	0.310	31.652	0.000	0.584	16.800	0.000	0.638	15.366	0.000
Last-value	PVAR-X^{R2}	9.813	0.00	0.638	15.369	0.000	0.000	0.310	31.675	0.000	0.584	16.808	0.000	0.638	15.369	0.000
Last-value	PVAR-X^{R3}	9.808	0.00	0.638	15.364	0.000	0.000	0.310	31.651	0.000	0.584	16.802	0.000	0.638	15.364	0.000
Last-value	RF	8.313	0.00	0.575	14.467	0.000	0.000	0.289	28.809	0.000	0.529	15.725	0.000	0.575	14.467	0.000
Last-value	RF^{R2}	9.347	0.00	0.597	15.660	0.000	0.000	0.288	32.426	0.000	0.541	17.273	0.000	0.597	15.660	0.000
Last-value	RF^{R3}	9.642	0.00	0.608	15.856	0.000	0.000	0.292	33.022	0.000	0.550	17.543	0.000	0.608	15.856	0.000
Last-value	RF^{R4}	9.774	0.00	0.624	15.673	0.000	0.000	0.299	32.695	0.000	0.566	17.260	0.000	0.624	15.673	0.000
Last-value	XGB	9.325	0.00	0.599	15.559	0.000	0.000	0.292	31.946	0.000	0.544	17.126	0.000	0.599	15.559	0.000
Last-value	XGB^{R2}	9.346	0.00	0.598	15.628	0.000	0.000	0.290	32.272	0.000	0.542	17.233	0.000	0.598	15.628	0.000
Last-value	XGB^{R3}	9.462	0.00	0.610	15.518	0.000	0.000	0.297	31.888	0.000	0.554	17.085	0.000	0.610	15.518	0.000
Last-value	XGB^{R4}	9.579	0.00	0.632	15.157	0.000	0.000	0.309	31.039	0.000	0.578	16.569	0.000	0.632	15.157	0.000
GARCH	Linear	0.739	0.00	0.053	13.987	0.000	0.000	0.030	24.674	0.000	0.043	17.304	0.000	0.053	13.987	0.000
GARCH	Linear^{R2}	0.761	0.00	0.052	14.516	0.000	0.000	0.030	25.547	0.000	0.042	18.225	0.000	0.052	14.516	0.000
GARCH	Linear^{R3}	0.757	0.00	0.052	14.450	0.000	0.000	0.030	25.283	0.000	0.042	18.156	0.000	0.052	14.450	0.000
GARCH	PVAR-X	0.923	0.00	0.057	16.068	0.000	0.000	0.034	26.923	0.000	0.044	20.787	0.000	0.057	16.068	0.000
GARCH	PVAR-X^{R2}	0.933	0.00	0.058	16.098	0.000	0.000	0.035	27.036	0.000	0.045	20.856	0.000	0.058	16.098	0.000
GARCH	PVAR-X^{R3}	0.928	0.00	0.058	15.985	0.000	0.000	0.035	26.622	0.000	0.045	20.659	0.000	0.058	15.985	0.000
GARCH	RF	-0.567	**0.10**	0.069	-8.195	0.000	0.000	0.057	-9.967	0.000	0.062	-9.107	0.000	0.069	-8.210	0.000
GARCH	RF^{R2}	0.467	0.00	0.054	8.623	0.000	0.000	0.038	12.281	0.000	0.047	10.012	0.000	0.054	8.623	0.000
GARCH	RF^{R3}	0.762	0.00	0.049	15.463	0.000	0.000	0.031	24.322	0.000	0.037	20.390	0.000	0.049	15.463	0.000
GARCH	RF^{R4}	0.894	0.00	0.050	17.904	0.000	0.000	0.030	29.895	0.000	0.035	25.435	0.000	0.050	17.904	0.000
GARCH	XGB	0.445	**0.03**	0.053	8.417	0.000	0.000	0.042	10.494	0.000	0.045	9.973	0.000	0.053	8.417	0.000

(Table continued on next page)

Table B.2: Pairwise Diebold-Mariano test results for equal predictive accuracy of 30-minute sovereign CDS forecasts *(continued)*

			Hard-Thresholding					**Zero-Lag**			**Newey-West**			**Driscoll-Kraay**		
Model m_1	**Model m_2**	$\bar{d}$	M	$\hat{\sigma}_{\bar{d},NT}$	J_{NT}	p	p^{BF}	$\hat{\sigma}^{ZL}_{\bar{d},NT}$	J^{ZL}_{NT}	p_{ZL}	$\hat{\sigma}^{NW}_{\bar{d},NT}$	J^{NW}_{NT}	p_{NW}	$\hat{\sigma}^{DK}_{\bar{d},NT}$	J^{DK}_{NT}	p_{DK}
GARCH	XGB^{R2}	0.466	0.00	0.057	8.218	0.000	0.000	0.043	10.842	0.000	0.049	9.582	0.000	0.057	8.218	0.000
GARCH	XGB^{R3}	0.582	0.00	0.055	10.531	0.000	0.000	0.041	14.205	0.000	0.046	12.695	0.000	0.055	10.531	0.000
GARCH	XGB^{R4}	0.699	0.00	0.058	11.972	0.000	0.000	0.042	16.819	0.000	0.047	14.838	0.000	0.058	11.972	0.000
Linear	$Linear^{R2}$	0.021	0.00	0.005	3.905	0.000	**0.011**	0.004	5.616	0.000	0.004	4.885	0.000	0.005	3.905	0.000
Linear	$Linear^{R3}$	0.017	0.00	0.006	2.801	0.005	**0.612**	0.004	3.999	0.000	0.005	3.537	0.000	0.006	2.801	0.005
Linear	PVAR-X	0.184	**0.31**	0.052	3.561	0.000	**0.044**	0.033	5.565	0.000	0.051	3.603	0.000	0.051	3.615	0.000
Linear	$PVAR\text{-}X^{R2}$	0.194	**0.26**	0.052	3.723	0.000	**0.024**	0.033	5.829	0.000	0.051	3.772	0.000	0.051	3.771	0.000
Linear	$PVAR\text{-}X^{R3}$	0.189	**0.25**	0.052	3.627	0.000	**0.034**	0.033	5.656	0.000	0.051	3.674	0.000	0.051	3.673	0.000
Linear	RF	-1.306	0.00	0.062	-21.207	0.000	0.000	0.051	-25.511	0.000	0.052	-24.894	0.000	0.062	-21.207	0.000
Linear	RF^{R2}	-0.272	**0.02**	0.025	-10.947	0.000	0.000	0.025	-10.955	0.000	0.023	-11.778	0.000	0.025	-10.947	0.000
Linear	RF^{R3}	0.023	**0.06**	0.021	1.121	0.262	**1.000**	0.020	1.133	0.257	0.019	1.217	0.224	0.021	1.121	0.262
Linear	RF^{R4}	0.155	0.00	0.036	4.249	0.000	0.003	0.026	5.947	0.000	0.035	4.423	0.000	0.036	4.249	0.000
Linear	XGB	-0.294	**0.03**	0.035	-8.407	0.000	0.000	0.036	-8.155	0.000	0.034	-8.606	0.000	0.035	-8.418	0.000
Linear	XGB^{R2}	-0.273	**0.03**	0.034	-8.156	0.000	0.000	0.033	-8.381	0.000	0.032	-8.670	0.000	0.034	-8.157	0.000
Linear	XGB^{R3}	-0.157	0.00	0.035	-4.497	0.000	0.001	0.033	-4.789	0.000	0.033	-4.789	0.000	0.035	-4.497	0.000
Linear	XGB^{R4}	-0.041	**0.06**	0.053	-0.765	0.444	**1.000**	0.040	-1.013	0.311	0.051	-0.791	0.429	0.053	-0.767	0.443
$Linear^{R2}$	$Linear^{R3}$	-0.004	0.00	0.002	-1.601	0.109	**1.000**	0.002	-1.990	0.047	0.002	-2.252	0.024	0.002	-1.601	0.109
$Linear^{R2}$	PVAR-X	0.162	**0.35**	0.049	3.293	0.001	**0.119**	0.032	5.100	0.000	0.049	3.330	0.001	0.048	3.351	0.001
$Linear^{R2}$	$PVAR\text{-}X^{R2}$	0.172	**0.33**	0.049	3.481	0.000	**0.060**	0.032	5.419	0.000	0.049	3.521	0.000	0.049	3.541	0.000
$Linear^{R2}$	$PVAR\text{-}X^{R3}$	0.167	**0.34**	0.049	3.381	0.001	**0.087**	0.032	5.240	0.000	0.049	3.419	0.001	0.049	3.438	0.001
$Linear^{R2}$	RF	-1.328	0.00	0.062	-21.268	0.000	0.000	0.051	-25.990	0.000	0.053	-25.120	0.000	0.062	-21.268	0.000
$Linear^{R2}$	RF^{R2}	-0.293	**0.10**	0.024	-12.009	0.000	0.000	0.024	-12.153	0.000	0.022	-13.052	0.000	0.024	-12.014	0.000
$Linear^{R2}$	RF^{R3}	0.002	0.00	0.019	0.089	0.929	**1.000**	0.019	0.085	0.932	0.017	0.097	0.923	0.019	0.089	0.929
$Linear^{R2}$	RF^{R4}	0.134	**0.10**	0.034	3.938	0.000	0.010	0.025	5.362	0.000	0.033	4.078	0.000	0.034	3.946	0.000
$Linear^{R2}$	XGB	-0.316	**0.05**	0.035	-8.986	0.000	0.000	0.036	-8.814	0.000	0.034	-9.327	0.000	0.035	-9.000	0.000
$Linear^{R2}$	XGB^{R2}	-0.295	**0.05**	0.033	-8.987	0.000	0.000	0.032	-9.286	0.000	0.031	-9.616	0.000	0.033	-8.989	0.000
$Linear^{R2}$	XGB^{R3}	-0.179	**0.01**	0.034	-5.252	0.000	0.000	0.032	-5.581	0.000	0.032	-5.651	0.000	0.034	-5.252	0.000
$Linear^{R2}$	XGB^{R4}	-0.062	**0.06**	0.051	-1.224	0.221	**1.000**	0.039	-1.596	0.111	0.049	-1.264	0.206	0.051	-1.226	0.220

(Table continued on next page)

Table B.2: Pairwise Diebold-Mariano test results for equal predictive accuracy of 30-minute sovereign CDS forecasts *(continued)*

			Hard-Thresholding					**Zero-Lag**			**Newey-West**			**Driscoll-Kraay**		
Model m_1	**Model m_2**	$\bar{d}$	M	$\hat{\sigma}_{\bar{d},NT}$	J_{NT}	p	p^{BF}	$\hat{\sigma}^{ZL}_{\bar{d},NT}$	J^{ZL}_{NT}	p_{ZL}	$\hat{\sigma}^{NW}_{\bar{d},NT}$	J^{NW}_{NT}	p_{NW}	$\hat{\sigma}^{DK}_{\bar{d},NT}$	J^{DK}_{NT}	p_{DK}
Linear$^{\text{R3}}$	PVAR-X	0.166	**0.34**	0.049	3.397	0.001	**0.082**	0.032	5.273	0.000	0.048	3.436	0.001	0.048	3.456	0.001
Linear$^{\text{R3}}$	PVAR-X$^{\text{R2}}$	0.176	**0.33**	0.049	3.587	0.000	**0.040**	0.032	5.595	0.000	0.049	3.629	0.000	0.048	3.647	0.000
Linear$^{\text{R3}}$	PVAR-X$^{\text{R3}}$	0.171	**0.33**	0.049	3.489	0.000	**0.058**	0.032	5.429	0.000	0.049	3.529	0.000	0.048	3.550	0.000
Linear$^{\text{R3}}$	RF	-1.324	0.00	0.062	-21.190	0.000	0.000	0.051	-25.995	0.000	0.053	-25.003	0.000	0.062	-21.190	0.000
Linear$^{\text{R3}}$	RF$^{\text{R2}}$	-0.289	**0.09**	0.025	-11.712	0.000	0.000	0.024	-12.059	0.000	0.023	-12.809	0.000	0.025	-11.716	0.000
Linear$^{\text{R3}}$	RF$^{\text{R3}}$	0.006	0.00	0.018	0.309	0.758	**1.000**	0.019	0.297	0.766	0.017	0.335	0.738	0.018	0.309	0.758
Linear$^{\text{R3}}$	RF$^{\text{R4}}$	0.138	**0.10**	0.034	4.090	0.000	**0.005**	0.025	5.596	0.000	0.032	4.246	0.000	0.034	4.101	0.000
Linear$^{\text{R3}}$	XGB	-0.312	**0.04**	0.035	-8.854	0.000	0.000	0.036	-8.752	0.000	0.034	-9.202	0.000	0.035	-8.857	0.000
Linear$^{\text{R3}}$	XGB$^{\text{R2}}$	-0.291	0.00	0.033	-8.853	0.000	0.000	0.032	-9.203	0.000	0.031	-9.472	0.000	0.033	-8.853	0.000
Linear$^{\text{R3}}$	XGB$^{\text{R3}}$	-0.175	**0.01**	0.034	-5.162	0.000	0.000	0.032	-5.524	0.000	0.032	-5.543	0.000	0.034	-5.162	0.000
Linear$^{\text{R3}}$	XGB$^{\text{R4}}$	-0.058	**0.07**	0.050	-1.153	0.249	**1.000**	0.039	-1.504	0.133	0.049	-1.190	0.234	0.050	-1.154	0.249
PVAR-X	PVAR-X$^{\text{R2}}$	0.010	0.00	0.004	2.750	0.006	**0.715**	0.002	3.937	0.000	0.003	3.736	0.000	0.004	2.750	0.006
PVAR-X	PVAR-X$^{\text{R3}}$	0.005	0.00	0.004	1.189	0.235	**1.000**	0.003	1.633	0.102	0.003	1.599	0.110	0.004	1.189	0.235
PVAR-X	RF	-1.490	0.00	0.086	-17.385	0.000	0.000	0.058	-25.899	0.000	0.076	-19.634	0.000	0.086	-17.385	0.000
PVAR-X	RF$^{\text{R2}}$	-0.456	**0.03**	0.056	-8.206	0.000	0.000	0.037	-12.430	0.000	0.053	-8.564	0.000	0.056	-8.206	0.000
PVAR-X	RF$^{\text{R3}}$	-0.161	**0.02**	0.045	-3.610	0.000	**0.037**	0.030	-5.424	0.000	0.043	-3.781	0.000	0.045	-3.611	0.000
PVAR-X	RF$^{\text{R4}}$	-0.029	**0.14**	0.023	-1.232	0.218	**1.000**	0.018	-1.565	0.118	0.023	-1.276	0.202	0.023	-1.238	0.216
PVAR-X	XGB	-0.478	0.00	0.056	-8.594	0.000	0.000	0.041	-11.785	0.000	0.053	-9.065	0.000	0.056	-8.594	0.000
PVAR-X	XGB$^{\text{R2}}$	-0.457	**0.01**	0.059	-7.809	0.000	0.000	0.040	-11.352	0.000	0.055	-8.285	0.000	0.059	-7.809	0.000
PVAR-X	XGB$^{\text{R3}}$	-0.341	**0.01**	0.051	-6.685	0.000	0.000	0.035	-9.738	0.000	0.046	-7.346	0.000	0.051	-6.685	0.000
PVAR-X	XGB$^{\text{R4}}$	-0.225	0.00	0.029	-7.649	0.000	0.000	0.025	-9.166	0.000	0.025	-9.149	0.000	0.029	-7.649	0.000
PVAR-X$^{\text{R2}}$	PVAR-X$^{\text{R3}}$	-0.005	0.00	0.002	-2.826	0.005	**0.565**	0.002	-2.982	0.003	0.001	-3.417	0.001	0.002	-2.826	0.005
PVAR-X$^{\text{R2}}$	RF	-1.500	0.00	0.086	-17.406	0.000	0.000	0.058	-26.043	0.000	0.076	-19.698	0.000	0.086	-17.406	0.000
PVAR-X$^{\text{R2}}$	RF$^{\text{R2}}$	-0.466	**0.02**	0.056	-8.350	0.000	0.000	0.037	-12.736	0.000	0.053	-8.732	0.000	0.056	-8.350	0.000
PVAR-X$^{\text{R2}}$	RF$^{\text{R3}}$	-0.171	**0.02**	0.045	-3.825	0.000	**0.016**	0.030	-5.770	0.000	0.043	-4.006	0.000	0.045	-3.827	0.000
PVAR-X$^{\text{R2}}$	RF$^{\text{R4}}$	-0.039	**0.16**	0.023	-1.653	0.098	**1.000**	0.018	-2.100	0.036	0.023	-1.707	0.088	0.023	-1.665	0.096
PVAR-X$^{\text{R2}}$	XGB	-0.488	0.00	0.056	-8.688	0.000	0.000	0.041	-11.996	0.000	0.053	-9.201	0.000	0.056	-8.688	0.000
PVAR-X$^{\text{R2}}$	XGB$^{\text{R2}}$	-0.467	**0.01**	0.059	-7.970	0.000	0.000	0.040	-11.674	0.000	0.055	-8.466	0.000	0.059	-7.970	0.000

(Table continued on next page)

Table B.2: Pairwise Diebold-Mariano test results for equal predictive accuracy of 30-minute sovereign CDS forecasts *(continued)*

			Hard-Thresholding					Zero-Lag			Newey-West			Driscoll-Kraay		
Model m_1	**Model m_2**	$\bar{d}$	M	$\hat{\sigma}_{\bar{d},NT}$	J_{NT}	p	p^{BF}	$\hat{\sigma}^{ZL}_{\bar{d},NT}$	J^{ZL}_{NT}	p_{ZL}	$\hat{\sigma}^{NW}_{\bar{d},NT}$	J^{NW}_{NT}	p_{NW}	$\hat{\sigma}^{DK}_{\bar{d},NT}$	J^{DK}_{NT}	p_{DK}
PVAR-X^{R2}	XGB^{R3}	-0.351	0.00	0.051	-6.853	0.000	0.000	0.035	-10.050	0.000	0.047	-7.545	0.000	0.051	-6.853	0.000
PVAR-X^{R2}	XGB^{R4}	-0.234	0.00	0.029	-8.088	0.000	0.000	0.024	-9.669	0.000	0.024	-9.666	0.000	0.029	-8.088	0.000
PVAR-X^{R3}	RF	-1.495	0.00	0.086	-17.357	0.000	0.000	0.058	-25.963	0.000	0.076	-19.628	0.000	0.086	-17.357	0.000
PVAR-X^{R3}	RF^{R2}	-0.461	0.00	0.056	-8.257	0.000	0.000	0.037	-12.593	0.000	0.053	-8.637	0.000	0.056	-8.257	0.000
PVAR-X^{R3}	RF^{R3}	-0.166	**0.02**	0.045	-3.718	0.000	**0.024**	0.030	-5.615	0.000	0.043	-3.889	0.000	0.045	-3.720	0.000
PVAR-X^{R3}	RF^{R4}	-0.034	**0.15**	0.023	-1.438	0.150	**1.000**	0.018	-1.830	0.067	0.023	-1.486	0.137	0.023	-1.448	0.148
PVAR-X^{R3}	XGB	-0.483	0.00	0.056	-8.600	0.000	0.000	0.041	-11.884	0.000	0.053	-9.106	0.000	0.056	-8.600	0.000
PVAR-X^{R3}	XGB^{R2}	-0.462	**0.01**	0.059	-7.891	0.000	0.000	0.040	-11.551	0.000	0.055	-8.377	0.000	0.059	-7.892	0.000
PVAR-X^{R3}	XGB^{R3}	-0.346	0.00	0.051	-6.767	0.000	0.000	0.035	-9.943	0.000	0.047	-7.439	0.000	0.051	-6.767	0.000
PVAR-X^{R3}	XGB^{R4}	-0.229	0.00	0.029	-7.917	0.000	0.000	0.024	-9.481	0.000	0.024	-9.462	0.000	0.029	-7.917	0.000
RF	RF^{R2}	1.034	0.00	0.056	18.344	0.000	0.000	0.047	22.141	0.000	0.050	20.662	0.000	0.056	18.344	0.000
RF	RF^{R3}	1.329	0.00	0.063	20.998	0.000	0.000	0.046	28.739	0.000	0.053	25.219	0.000	0.063	20.998	0.000
RF	RF^{R4}	1.461	0.00	0.075	19.481	0.000	0.000	0.052	27.923	0.000	0.064	22.928	0.000	0.075	19.481	0.000
RF	XGB	1.012	0.00	0.049	20.726	0.000	0.000	0.037	27.459	0.000	0.042	24.148	0.000	0.049	20.726	0.000
RF	XGB^{R2}	1.033	0.00	0.054	19.017	0.000	0.000	0.045	23.170	0.000	0.047	21.982	0.000	0.054	19.017	0.000
RF	XGB^{R3}	1.149	0.00	0.063	18.252	0.000	0.000	0.047	24.455	0.000	0.054	21.232	0.000	0.063	18.252	0.000
RF	XGB^{R4}	1.266	0.00	0.079	15.937	0.000	0.000	0.059	21.536	0.000	0.072	17.675	0.000	0.079	15.937	0.000
RF^{R2}	RF^{R3}	0.295	0.00	0.023	12.824	0.000	0.000	0.018	16.462	0.000	0.020	14.396	0.000	0.023	12.824	0.000
RF^{R2}	RF^{R4}	0.427	**0.02**	0.041	10.291	0.000	0.000	0.030	14.269	0.000	0.038	11.116	0.000	0.041	10.292	0.000
RF^{R2}	XGB	-0.022	**0.11**	0.035	-0.640	0.522	**1.000**	0.033	-0.688	0.492	0.035	-0.643	0.520	0.035	-0.642	0.521
RF^{R2}	XGB^{R2}	-0.001	**0.02**	0.024	-0.055	0.956	**1.000**	0.022	-0.059	0.953	0.023	-0.057	0.955	0.024	-0.055	0.956
RF^{R2}	XGB^{R3}	0.115	**0.02**	0.031	3.755	0.000	**0.021**	0.025	4.565	0.000	0.028	4.054	0.000	0.030	3.756	0.000
RF^{R2}	XGB^{R4}	0.231	**0.10**	0.052	4.435	0.000	0.001	0.040	5.797	0.000	0.050	4.602	0.000	0.052	4.437	0.000
RF^{R3}	RF^{R4}	0.132	**0.01**	0.029	4.537	0.000	0.001	0.022	5.953	0.000	0.027	4.974	0.000	0.029	4.537	0.000
RF^{R3}	XGB	-0.317	**0.01**	0.034	-9.386	0.000	0.000	0.031	-10.364	0.000	0.032	-9.977	0.000	0.034	-9.389	0.000
RF^{R3}	XGB^{R2}	-0.296	0.00	0.030	-9.718	0.000	0.000	0.027	-11.178	0.000	0.028	-10.553	0.000	0.030	-9.718	0.000
RF^{R3}	XGB^{R3}	-0.181	0.00	0.025	-7.323	0.000	0.000	0.020	-8.862	0.000	0.023	-7.994	0.000	0.025	-7.323	0.000

(Table continued on next page)

Table B.2: Pairwise Diebold-Mariano test results for equal predictive accuracy of 30-minute sovereign CDS forecasts *(continued)*

			Hard-Thresholding					**Zero-Lag**			**Newey-West**			**Driscoll-Kraay**		
Model m_1	**Model m_2**	$\bar{d}$	M	$\hat{\sigma}_{\bar{d},NT}$	J_{NT}	p	p^{BF}	$\hat{\sigma}^{ZL}_{\bar{d},NT}$	J^{ZL}_{NT}	p_{ZL}	$\hat{\sigma}^{NW}_{\bar{d},NT}$	J^{NW}_{NT}	p_{NW}	$\hat{\sigma}^{DK}_{\bar{d},NT}$	J^{DK}_{NT}	p_{DK}
$\mathrm{RF}^{\mathrm{R3}}$	$\mathrm{XGB}^{\mathrm{R4}}$	-0.064	**0.02**	0.043	-1.465	0.143	**1.000**	0.034	-1.861	0.063	0.041	-1.537	0.124	0.043	-1.466	0.143
$\mathrm{RF}^{\mathrm{R4}}$	XGB	-0.449	**0.02**	0.045	-10.045	0.000	0.000	0.037	-12.277	0.000	0.041	-10.885	0.000	0.045	-10.050	0.000
$\mathrm{RF}^{\mathrm{R4}}$	$\mathrm{XGB}^{\mathrm{R2}}$	-0.428	**0.01**	0.046	-9.314	0.000	0.000	0.035	-12.147	0.000	0.043	-10.055	0.000	0.046	-9.314	0.000
$\mathrm{RF}^{\mathrm{R4}}$	$\mathrm{XGB}^{\mathrm{R3}}$	-0.312	0.00	0.041	-7.692	0.000	0.000	0.031	-9.944	0.000	0.036	-8.738	0.000	0.041	-7.692	0.000
$\mathrm{RF}^{\mathrm{R4}}$	$\mathrm{XGB}^{\mathrm{R4}}$	-0.196	**0.11**	0.027	-7.130	0.000	0.000	0.023	-8.513	0.000	0.025	-7.803	0.000	0.027	-7.147	0.000
XGB	$\mathrm{XGB}^{\mathrm{R2}}$	0.021	**0.03**	0.026	0.823	0.411	**1.000**	0.025	0.852	0.394	0.024	0.862	0.388	0.026	0.823	0.411
XGB	$\mathrm{XGB}^{\mathrm{R3}}$	0.137	**0.01**	0.033	4.143	0.000	0.004	0.028	4.975	0.000	0.031	4.478	0.000	0.033	4.143	0.000
XGB	$\mathrm{XGB}^{\mathrm{R4}}$	0.254	0.00	0.050	5.060	0.000	0.000	0.041	6.201	0.000	0.048	5.239	0.000	0.050	5.060	0.000
$\mathrm{XGB}^{\mathrm{R2}}$	$\mathrm{XGB}^{\mathrm{R3}}$	0.116	**0.03**	0.027	4.336	0.000	0.002	0.023	4.934	0.000	0.025	4.647	0.000	0.027	4.340	0.000
$\mathrm{XGB}^{\mathrm{R2}}$	$\mathrm{XGB}^{\mathrm{R4}}$	0.233	**0.02**	0.051	4.564	0.000	0.001	0.040	5.841	0.000	0.050	4.692	0.000	0.051	4.564	0.000
$\mathrm{XGB}^{\mathrm{R3}}$	$\mathrm{XGB}^{\mathrm{R4}}$	0.117	**0.01**	0.044	2.638	0.008	**1.000**	0.034	3.386	0.001	0.042	2.801	0.005	0.044	2.638	0.008

Note: Pairwise Diebold-Mariano tests for 30-minute sovereign CDS forecasts (11 countries, January 2011–December 2014, i.e., $T = 14,205$) for each ordered pair (m_1, m_2), with mean loss differential $\bar{d}$, hard-thresholded standard error estimate $\hat{\sigma}_{\bar{d},NT}/\sqrt{NT}$ ($L = 4(T/100)^{2/9} = 12$, M determined via cross-validation over $M \in \{0, 0.01, \ldots, 0.40\}$), corresponding statistic J_{NT}, and raw and Bonferroni-adjusted p-values p and $p^{\mathrm{BF}} = 120 \times p$. Positive $\bar{d}$ indicates that m_1 has a higher MSPE than m_2. For comparison, we also report the results for zero-lag ($L = 0$, $M = \infty$), Newey-West ($L = 12$; $M = \infty$; NW), and Driscoll-Kraay ($L = 12$; $M = 0$; DK) standard errors. All models, see Table B.2, produce conditional-mean forecasts estimated with 24-month rolling windows and the regressor sets R2–R4 defined in Section 3.3.2. Where no regressor set is indicated, forecasts are based on R1. Grey rows mark the model pairs from Table 3.4 emphasised in the discussion. Bold values indicate $M > 0$ or $p^{\mathrm{BF}} > 0.01$.

C Appendix to Chapter 4: Direction Augmentation in the Evaluation of Armed Conflict Predictions

In this appendix, we provide proofs for the results on the optimal point forecasts for the TADDA specification illustrated in Chapter 4 as well as corresponding results for two further variations of the score. To distinguish them clearly, we will denote the variant discussed in the main text by $\text{TADDA1}_\epsilon^{\text{L1}}$. We will further address the same score based on an L2 rather than L1 distance, denoted by $\text{TADDA1}_\epsilon^{\text{L2}}$. Lastly, we consider an alternative handling of the tolerance region proposed by Vesco et al. (2022) under L1 distance. Following the convention from Vesco et al. (2022), we denote the latter by $\text{TADDA2}_\epsilon^{\text{L1}}$. As mentioned in Section 4.4, we will use $\pi_- := \text{Pr}_F(Y < -\epsilon)$ and $\pi_+ := \text{Pr}_F(Y > \epsilon)$.

C.1 Derivation of the Optimal Point Forecast for $\text{TADDA1}_\epsilon^{\text{L1}}$

Result: The optimal point forecast for

$$\text{TADDA1}_\epsilon^{\text{L1}}(\hat{y}, y) = |\hat{y} - y| + a_\epsilon(\hat{y}, y),$$

where $a_\epsilon(\hat{y}, y)$ is defined as in Equation 4.2 is given by

$$\hat{y}_{\text{OPF}} = \begin{cases} F^{-1}\{0.5 \times (1 + \pi_+)\} & \text{if } \pi_- \geq 0.5 \times (1 + \pi_+) \\ -\epsilon & \text{if } 0.5 < \pi_- < 0.5 \times (1 + \pi_+) \\ F^{-1}(0.5) \quad = m & \text{if } \pi_- \leq 0.5 \text{ and } \pi_+ \leq 0.5 \\ \epsilon & \text{if } 0.5 < \pi_+ \leq 0.5 \times (1 + \pi_-) \\ F^{-1}\{0.5 \times (1 - \pi_-)\} & \text{if } \pi_+ > 0.5 \times (1 + \pi_-). \end{cases}$$

Note that the definition of the direction augmentation term is equivalent to

$$a_\epsilon(\hat{y}, y) = \begin{cases} |\hat{y} - \epsilon| & \text{if} \quad \hat{y} > \epsilon \text{ and } y < -\epsilon \\ |\hat{y} + \epsilon| & \text{if} \quad \hat{y} < -\epsilon \text{ and } y > \epsilon \\ 0 & \text{otherwise,} \end{cases} \tag{C.1}$$

which we will use in the following.

Proof: The expectation of $\text{TADDA1}_\epsilon^{\text{L1}}(\hat{y}, Y)$ under the predictive distribution F reads

$$\mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L1}}(\hat{y}, Y)] = \underbrace{\mathbb{E}_F|\hat{y} - Y|}_{\text{(a)}} \quad + \quad \underbrace{\mathbb{E}_F[a_\epsilon(\hat{y}, Y)]}_{\text{(b)}}.$$

Both terms (a) and (b) are obviously non-negative, which translates directly to their expectation. Term (a) is minimised by $\hat{y} = m$ and increases monotonically to either side of m. Further, term (b) is zero for $\hat{y} \in [-\epsilon, \epsilon]$ and non-negative for $\hat{y} \notin [-\epsilon, \epsilon]$. Note that it is zero in particular for $\hat{y} = \pm\epsilon$. We show the result via proof by cases.

Case 1 $(-\epsilon \leq m \leq \epsilon)$: Due to the above-mentioned characteristics of terms (a) and (b), the optimal point forecast is given by $\hat{y}_{\text{OPF}} = m$.

Case 2 $(m > \epsilon)$: For all $\hat{y} < \epsilon$, term (a) satisfies

$$\mathbb{E}_F|\epsilon - Y| \leq \mathbb{E}_F|\hat{y} - Y|.$$

Additionally, we know that no $\hat{y} < \epsilon$ can achieve a smaller value of term (b) than $\hat{y} = \epsilon$. Combined, we have that

$$\mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L1}}(\epsilon, Y)] \leq \mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L1}}(\hat{y}, Y)] \quad \text{for all } \hat{y} < \epsilon.$$

We can thus restrict our search for the optimal point forecast to $\hat{y} \geq \epsilon$. On this segment of the real line, the expected score is given by (see equation C.1)

$$\begin{aligned} \mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L1}}(\hat{y}, Y)] &= \mathbb{E}_F|\hat{y} - Y| \quad + \quad \pi_- \times \mathbb{E}_F|\hat{y} - \epsilon| \\ &\propto \mathbb{E}_F|\hat{y} - Z|, \end{aligned}$$

where

$$Z = \begin{cases} Y & \text{with probability } \frac{1}{1+\pi_-} \\ \epsilon & \text{with probability } \frac{\pi_-}{1+\pi_-}. \end{cases}$$

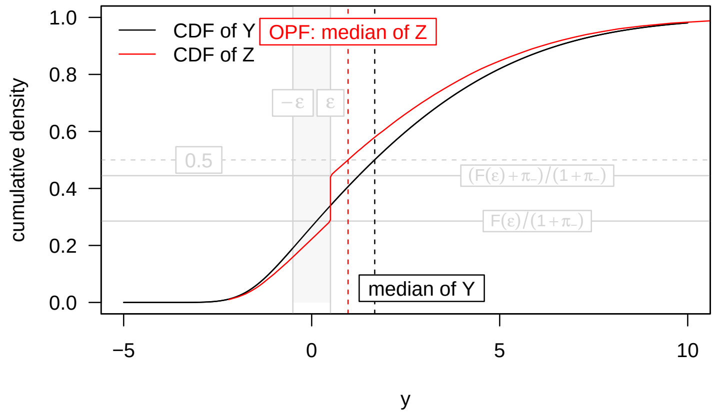

Figure C.1: Illustration of the CDFs F and G for Case 2 of $\text{TADDA1}_\epsilon^{\text{L1}}$, where $\epsilon > 0$ and $m > \epsilon$

The optimal choice for $\hat{y}$ is consequently the median m_Z of Z. The CDF of Z is given by

$$G(z) = \begin{cases} \frac{1}{1+\pi_-} \times F(z) & \text{for} \quad z < \epsilon \\ \frac{1}{1+\pi_-} \times F(z) + \frac{\pi_-}{1+\pi_-} & \text{for} \quad z \geq \epsilon, \end{cases} \tag{C.2}$$

see Figure C.1 for an illustration. Since $m > \epsilon$ by assumption of Case 2, we know that $F(\epsilon) < 0.5$. The median m_Z of Z is obtained by setting $G(m_Z) = 0.5$ in Equation C.2 and solving for m_Z. This yields

$$\hat{y}_{\text{OPF}} = m_Z = \begin{cases} \epsilon & \text{if } \frac{F(\epsilon)+\pi_-}{1+\pi_-} \geq 0.5 \\ F^{-1}\{0.5 \times (1-\pi_-)\} & \text{otherwise} . \end{cases}$$

We can substitute $F(\epsilon) = 1 - \pi_+$ in the condition, which leads to

$$\pi_+ \leq 0.5 \times (1 + \pi_-).$$

In conjunction with $\pi_+ > 0.5$ (by assumption of Case 2), this gives the stated part of the result.

Case 3 $(m < -\epsilon)$: Analogously to Case 2, we have that

$$\mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L1}}(-\epsilon, Y)] \leq \mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L1}}(\hat{y}, Y)] \quad \text{for all } \hat{y} > -\epsilon.$$

Therefore, we can restrict our search for the optimal point forecast to $\hat{y} \leq -\epsilon$. For this segment of the real line, we obtain the expected score via

$$\mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L1}}(\hat{y}, Y)] = \mathbb{E}_F|\hat{y} - Y| \quad + \quad \pi_+ \times \mathbb{E}_F|\hat{y} - (-\epsilon)|$$

$$\propto \mathbb{E}_F|\hat{y} - Z|,$$

where

$$Z = \begin{cases} Y & \text{with probability } \frac{1}{1+\pi_+} \\ -\epsilon & \text{with probability } \frac{\pi_+}{1+\pi_+}. \end{cases}$$

Again, the optimal choice for $\hat{y}$ is the median m_Z of Z. The CDF of Z is given by

$$G(z) = \begin{cases} \frac{1}{1+\pi_+} \times F(z) & \text{for} \quad z < -\epsilon \\ \frac{1}{1+\pi_+} \times F(z) + \frac{\pi_+}{1+\pi_+} & \text{for} \quad z \geq -\epsilon \end{cases} \tag{C.3}$$

and visualised in Figure C.2. Setting $G(m_Z) = 0.5$ in Equation C.3 and solving for m_Z leads to

$$\hat{y}_{\text{OPF}} = m_Z = \begin{cases} F^{-1}\{0.5 \times (1+\pi_+)\} & \text{if } \frac{\Pr_F(Y<-\epsilon)}{1+\pi_+} \geq 0.5 \\ -\epsilon & \text{otherwise.} \end{cases}$$

We can rewrite the condition as

$$\frac{\pi_-}{1+\pi_+} \geq 0.5 \iff \pi_- \geq 0.5 \times (1+\pi_+).$$

This concludes the proof.

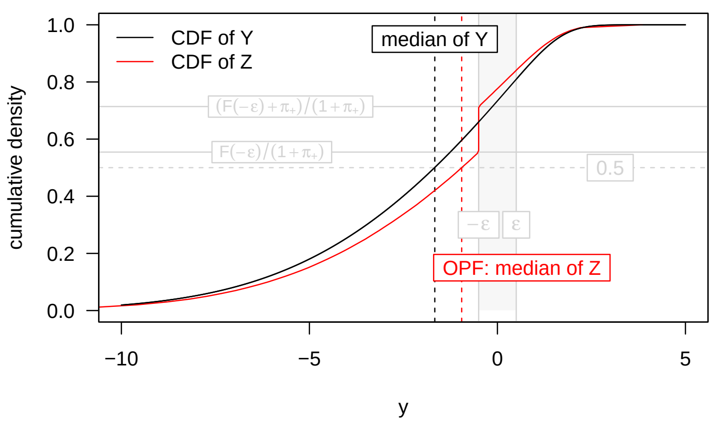

Figure C.2: Illustration of the CDFs F and G for Case 3 of TADDA1_{ϵ}^{L1}, where $\epsilon > 0$ and $m < -\epsilon$

C.2 Derivation of the Optimal Point Forecast for TADDA1$_\epsilon^{\mathbf{L2}}$

Result: The optimal point forecast for

$$\text{TADDA1}_\epsilon^{\text{L2}}(\hat{y}, y) = (\hat{y} - y)^2 + a_\epsilon(\hat{y}, y),$$

where $a_\epsilon(\hat{y}, y)$ is defined as

$$a_\epsilon(\hat{y}, y) = \begin{cases} (\hat{y} - \epsilon)^2 & \text{if} \quad \hat{y} > \epsilon \text{ and } y < -\epsilon \\ (\hat{y} + \epsilon)^2 & \text{if} \quad \hat{y} < -\epsilon \text{ and } y > \epsilon \\ 0 & \text{otherwise,} \end{cases}$$

is given by

$$\hat{y}_{\text{OPF}} = \begin{cases} \mu \times \frac{1}{1+\pi_+} - \epsilon \times \frac{\pi_+}{1+\pi_+} & \text{if } \mu < -\epsilon \\ \mu & \text{if } -\epsilon \leq \mu \leq \epsilon \\ \mu \times \frac{1}{1+\pi_-} + \epsilon \times \frac{\pi_-}{1+\pi_-} & \text{if } \mu > \epsilon. \end{cases}$$

Proof: The expectation of TADDA1$_\epsilon^{\text{L2}}(\hat{y}, Y)$ reads

$$\mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L2}}(\hat{y}, Y)] = \underbrace{\mathbb{E}_F[(\hat{y} - Y)^2]}_{\text{(a)}} \quad + \quad \underbrace{\mathbb{E}_F[a_\epsilon(\hat{y}, Y)]}_{\text{(b)}}.$$

As in the L1 case, terms (a) and (b) are non-negative. Term (a) is minimised by the mean μ of F, i.e., $\hat{y} = \mu$, and increases monotonically to each side of μ. Further, term (b) is zero for $\hat{y} \in [-\epsilon, \epsilon]$ and non-negative for $\hat{y} \notin [-\epsilon, \epsilon]$. Note that it is zero in particular for $\hat{y} = \pm\epsilon$. Again, we show the result via proof by cases.

Case 1 $(-\epsilon \leq \mu \leq \epsilon)$: Due to the above-mentioned characteristics of terms (a) and (b), the optimal point forecast is given by $\hat{y}_{\text{OPF}} = \mu$.

Case 2 $(\mu > \epsilon)$: For all $\hat{y} < \epsilon$, term (a) satisfies

$$\mathbb{E}_F[(\epsilon - Y)^2] \leq \mathbb{E}_F[(\hat{y} - Y)^2],$$

and we know that no $\hat{y} < \epsilon$ can achieve a smaller expected value of term (b) than $\hat{y} = \epsilon$. Combined, we have that

$$\mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L2}}(\epsilon, Y)] \leq \mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L2}}(\hat{y}, Y)] \text{ for all } \hat{y} < \epsilon.$$

We can thus restrict our search for the optimal point forecast to $\hat{y} \geq \epsilon$. On this segment of the real line, the expected score is given by

$$\begin{aligned}\mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L2}}(\hat{y}, Y)] &= \mathbb{E}_F[(\hat{y} - Y)^2] \quad + \quad \pi_- \times \mathbb{E}_F[(\hat{y} - \epsilon)^2] \\ &\propto \mathbb{E}_F[(\hat{y} - Z)^2],\end{aligned}$$

where

$$Z = \begin{cases} Y & \text{with probability } \frac{1}{1+\pi_-} \\ \epsilon & \text{with probability } \frac{\pi_-}{1+\pi_-}. \end{cases}$$

The optimal choice for $\hat{y}$ is consequently the mean μ_Z of Z, i.e.,

$$\hat{y}_{\text{OPF}} = \mu_Z = \mathbb{E}_F[Z] = \mu \times \frac{1}{1 + \pi_-} + \epsilon \times \frac{\pi_-}{1 + \pi_-}.$$

Case 3 $(\mu < -\epsilon)$: Analogously to Case 2, we have that

$$\mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L2}}(-\epsilon, Y)] \leq \mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L2}}(\hat{y}, Y)] \text{ for all } \hat{y} > -\epsilon.$$

Therefore, restricting our search for the optimal point forecast to $\hat{y} \leq -\epsilon$, we have that

$$\begin{aligned}\mathbb{E}_F[\text{TADDA1}_\epsilon^{\text{L2}}(\hat{y}, Y)] &= \mathbb{E}_F[(\hat{y} - Y)^2] \quad + \quad \pi_+ \times \mathbb{E}_F[(\hat{y} - (-\epsilon))^2] \\ &\propto \mathbb{E}_F[(\hat{y} - Z)^2],\end{aligned}$$

where

$$Z = \begin{cases} Y & \text{with probability } \frac{1}{1+\pi_+} \\ -\epsilon & \text{with probability } \frac{\pi_+}{1+\pi_+}. \end{cases}$$

Again, the optimal choice for $\hat{y}$ is the mean μ_Z of Z, which is given by

$$\hat{y}_{\text{OPF}} = \mu_Z = \mathbb{E}_F[Z] = \mu \times \frac{1}{1 + \pi_+} - \epsilon \times \frac{\pi_+}{1 + \pi_+}.$$

C.3 Derivation of the Optimal Point Forecast for $\text{TADDA2}_\epsilon^{\text{L1}}$

Result: The optimal point forecast for

$$\text{TADDA2}_\epsilon^{\text{L1}}(\hat{y}, y) = |\hat{y} - y| + a_\epsilon(\hat{y}, y),$$

where $a_\epsilon(\hat{y}, Y)$ is defined as

$$a_\epsilon(\hat{y}, y) = \begin{cases} |\hat{y} - \epsilon| & \text{if } \{\hat{y} \leq \epsilon \text{ and } y > \epsilon\} \text{ or } \{\hat{y} > \epsilon \text{ and } y \in [-\epsilon, \epsilon]\} \\ |\hat{y} + \epsilon| & \text{if } \{\hat{y} \geq -\epsilon \text{ and } y < -\epsilon\} \text{ or } \{\hat{y} < -\epsilon \text{ and } y \in [-\epsilon, \epsilon]\} \\ 0 & \text{otherwise,} \end{cases}$$

is given by

$$\hat{y}_{\mathrm{OPF}} = \begin{cases} F^{-1}\{0.5 \times (2 - \pi_-)\} & \text{if } \pi_- \geq \frac{2}{3} \\ -\epsilon & \text{if } m < -\epsilon \text{ and } \pi_- < \frac{2}{3} \\ & \quad \text{or } -\epsilon \leq m \leq \epsilon \text{ and } \pi_- \geq \frac{1+\pi_+ - 2\mathrm{Pr}_F(Y=-\epsilon)}{3} \\ F^{-1}\{0.5 \times (1 - \pi_- + \pi_+)\} & \text{if } -\epsilon \leq m \leq \epsilon \text{ and } \pi_- < \frac{1+\pi_+ - 2\mathrm{Pr}_F(Y=-\epsilon)}{3} \text{ and } \pi_+ < \frac{1+\pi_-}{3} \\ \epsilon & \text{if } -\epsilon \leq m \leq \epsilon \text{ and } \pi_+ \geq \frac{1+\pi_-}{3} \\ & \quad \text{or } m > \epsilon \text{ and } \pi_+ \leq \frac{2}{3} \\ F^{-1}(0.5 \times \pi_+) & \text{if } \pi_+ > \frac{2}{3}. \end{cases}$$

A graphical display illustrating the definition of the score is provided in Figure C.3.
Proof: The expectation of TADDA2$_\epsilon^{\mathrm{L1}}(\hat{y}, Y)$ is given by

$$\mathbb{E}_F[\mathrm{TADDA2}_\epsilon^{\mathrm{L1}}(\hat{y}, Y)] = \underbrace{\mathbb{E}_F|\hat{y} - Y|}_{\text{(a)}} \quad + \quad \underbrace{\mathbb{E}_F[a_\epsilon(\hat{y}, Y)]}_{\text{(b)}}.$$

We again prove the result by cases.

Case 1 $(-\epsilon \leq m \leq \epsilon)$: For $\hat{y} = \epsilon$, term (b) is

$$\mathbb{E}_F[a_\epsilon(\epsilon, Y)] = \pi_- \times 2\epsilon,$$

while for $\hat{y} > \epsilon$, we get

$$\mathbb{E}_F[a_\epsilon(\hat{y}, Y)] = \pi_- \times \underbrace{|\hat{y} + \epsilon|}_{\geq 2\epsilon} + \underbrace{(1 - \pi_- - \pi_+) \times |\hat{y} - \epsilon|}_{\geq 0}.$$

For $\hat{y} = -\epsilon$, term (b) is

$$\mathbb{E}_F[a_\epsilon(-\epsilon, Y)] = \pi_+ \times 2\epsilon,$$

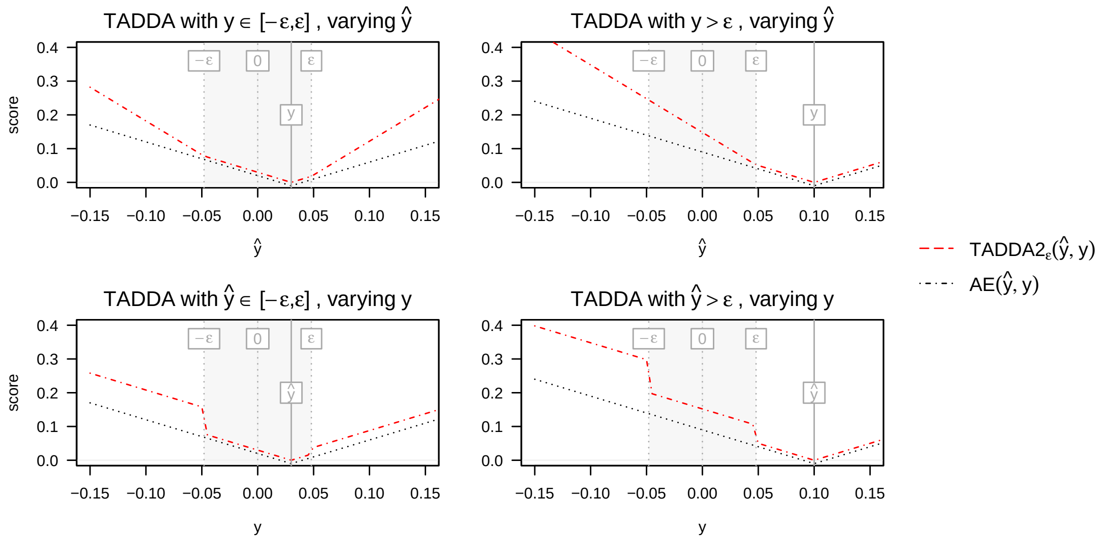

Figure C.3: Illustration of the absolute error (AE) and TADDA2$_\epsilon$ with $\epsilon = 0.048$

Note: Top: as a function of $\hat{y}$ with fixed observation y. Bottom: as a function of y with fixed prediction $\hat{y}$. In each row we show one example where $\hat{y}$ or y, respectively, are inside $[-\epsilon, \epsilon]$ and one where it is outside. The lines for the two scores are slightly shifted to avoid overplotting.

while for $\hat{y} < -\epsilon$, we have

$$\mathbb{E}_F[a_\epsilon(\hat{y}, Y)] = \pi_+ \times \underbrace{|\hat{y} - \epsilon|}_{\geq 2\epsilon} + \underbrace{(1 - \pi_- - \pi_+) \times |\hat{y} + \epsilon|}_{\geq 0}.$$

We can thus conclude that term (b) is smaller for $\hat{y} = \pm\epsilon$ than for any $\hat{y} < -\epsilon$ or $\hat{y} > \epsilon$ and restrict our search for the optimal point forecast to $\hat{y} \in [-\epsilon, \epsilon]$. For this segment of the real line, we obtain

$$\begin{aligned}\mathbb{E}_F[\text{TADDA2}_\epsilon^{\text{L1}}(\hat{y}, Y)] &= \mathbb{E}_F|\hat{y} - Y| \quad + \quad \pi_- \times \mathbb{E}_F|\hat{y} + \epsilon| \quad + \quad \pi_+ \times \mathbb{E}_F|\hat{y} - \epsilon| \\ &\propto \mathbb{E}_F|\hat{y} - Z|,\end{aligned}$$

where

$$Z = \begin{cases} Y & \text{with probability } \frac{1}{1+\pi_-+\pi_+} \\ -\epsilon & \text{with probability } \frac{\pi_-}{1+\pi_-+\pi_+} \\ \epsilon & \text{with probability } \frac{\pi_+}{1+\pi_-+\pi_+}, \end{cases}$$

The term $\mathbb{E}_F|\hat{y} - Z|$ is minimised by the median m_Z of Z. The CDF of Z is given by

$$G(z) = \begin{cases} \frac{1}{1+\pi_-+\pi_+} \times F(z) & \text{for } z < -\epsilon \\ \frac{1}{1+\pi_-+\pi_+} \times F(z) + \frac{\pi_-}{1+\pi_-+\pi_+} & \text{for } -\epsilon \leq z < \epsilon \\ \frac{1}{1+\pi_-+\pi_+} \times F(z) + \frac{\pi_-+\pi_+}{1+\pi_-+\pi_+} & \text{for } z \geq \epsilon. \end{cases}$$

Setting $F^{-1}(m_Z) = 0.5$, it can be shown that

$$\hat{y}_{\text{OPF}} = m_Z = \begin{cases} -\epsilon & \text{if } \pi_- \geq \frac{1+\pi_+-2\text{Pr}_F(Y=-\epsilon)}{3} \\ F^{-1}\{0.5 \times (1-\pi_- + \pi_+)\} & \text{if } \pi_- < \frac{1+\pi_+-2\text{Pr}_F(Y=-\epsilon)}{3} \text{ and } \pi_+ < \frac{1+\pi_-}{3} \\ \epsilon & \text{if } \pi_+ \geq \frac{1+\pi_-}{3}. \end{cases}$$

Note that in many practically relevant cases $-2\text{Pr}_F(Y = -\epsilon) = 0$ holds, so that this term can be omitted in the above equation.

Case 2 $(m > \epsilon)$: For $\hat{y} = \epsilon$, term (b) is

$$\mathbb{E}_F[a_\epsilon(\epsilon, Y)] = \pi_- \times 2\epsilon,$$

while for $-\epsilon \leq \hat{y} < \epsilon$, it is

$$\begin{aligned} \mathbb{E}_F[a_\epsilon(\hat{y}, Y)] &= \pi_- \times (\hat{y} + \epsilon) + \pi_+ \times (\epsilon - \hat{y}) \\ &= (\pi_- + \pi_+) \times \epsilon + (\pi_- - \pi_+) \times \hat{y} \\ &= 2 \times \pi_- \times \epsilon + (\pi_+ - \pi_-) \times \epsilon + (\pi_- - \pi_+) \times \hat{y} \\ &= \pi_- \times 2\epsilon + \underbrace{(\pi_+ - \pi_-)}_{<0} \times \underbrace{(\epsilon - \hat{y})}_{>0} \\ &> \pi_- \times 2\epsilon. \end{aligned}$$

For $\hat{y} < -\epsilon$, we have

$$\begin{aligned} \mathbb{E}_F[a_\epsilon(\hat{y}, Y)] &= (1 - \pi_- - \pi_+) \times (-\epsilon - \hat{y}) + \pi_+ \times (\epsilon - \hat{y}) \\ &= (-1 + \pi_- + 2\pi_+) \times \epsilon + (-1 + \pi_-) \times \hat{y} \\ &= 2 \underbrace{\pi_+}_{>\pi_-} \times \epsilon + \underbrace{(-1 + \pi_-)}_{<0} \times \underbrace{(\hat{y} + \epsilon)}_{<0} \\ &> \pi_- \times 2\epsilon. \end{aligned}$$

We can thus conclude that term (b) is smaller for $\hat{y} = \epsilon$ than for any $\hat{y} < \epsilon$. As we assumed that term (a) is minimised by $m > \epsilon$, we can restrict our search for the optimal point forecast to $\hat{y} \geq \epsilon$ and we have per definition

$$\begin{aligned}\mathbb{E}_F[\text{TADDA2}_\epsilon^{\text{L1}}(\hat{y}, Y)] &= \mathbb{E}_F|\hat{y} - Y| \quad + \quad \pi_- \times \mathbb{E}_F|\hat{y} + \epsilon| \quad + \quad (1 - \pi_- - \pi_+) \times \mathbb{E}_F|\hat{y} - \epsilon| \\ &\propto \mathbb{E}_F|\hat{y} - Z|,\end{aligned}$$

where

$$Z = \begin{cases} Y & \text{with probability } \frac{1}{2-\pi_+} \\ -\epsilon & \text{with probability } \frac{\pi_-}{2-\pi_+} \\ \epsilon & \text{with probability } \frac{1-\pi_--\pi_+}{2-\pi_+}. \end{cases}$$

$\mathbb{E}_F|\hat{y} - Z|$ is minimised by the median m_Z of Z under F. The CDF G of Z can be expressed through F as

$$G(z) = \begin{cases} \frac{1}{2-\pi_+} \times F(z) & \text{if } z < -\epsilon \\ \frac{1}{2-\pi_+} \times F(z) + \frac{\pi_-}{2-\pi_+} & \text{if } -\epsilon \leq z < \epsilon \\ \frac{1}{2-\pi_+} \times F(z) + \frac{1-\pi_+}{2-\pi_+} & \text{if } z \geq \epsilon. \end{cases} \tag{C.4}$$

We obtain the median m_Z by setting $G(m_Z) = 0.5$. Using

$$\frac{1}{2 - \pi_+} \times \underbrace{F(\epsilon)}_{=1-\pi_+} + \frac{1 - \pi_+}{2 - \pi_+} \geq 0.5$$

and some simple algebra, we find that the median is ϵ whenever $\pi_+ \leq \frac{2}{3}$. This leads to the distinction

$$\hat{y}_{\text{OPF}} = m_Z = \begin{cases} \epsilon & \text{if } \pi_+ \leq \frac{2}{3} \\ F^{-1}(0.5 \times \pi_+) & \text{otherwise.} \end{cases}$$

Case 3 $(m < -\epsilon)$: Following essentially the same arguments as in Case 2, it can be shown that term (b) is smaller for $\hat{y} = -\epsilon$ than for any $\hat{y} > -\epsilon$. As we know that term (a) is minimised by $m < -\epsilon$, we can restrict our search for the optimal point forecast to $\hat{y} \leq -\epsilon$. For this part of the real line, we have

$$\begin{aligned}\mathbb{E}_F[\text{TADDA2}_\epsilon^{\text{L1}}(\hat{y}, Y)] &= \mathbb{E}_F|\hat{y} - Y| \quad + \quad (1 - \pi_- - \pi_+) \times \mathbb{E}_F|\hat{y} + \epsilon| \quad + \quad \pi_+ \times \mathbb{E}_F|\hat{y} - \epsilon| \\ &\propto \mathbb{E}_F|\hat{y} - Z|,\end{aligned}$$

where

$$Z = \begin{cases} Y & \text{with probability } \frac{1}{2-\pi_-} \\ -\epsilon & \text{with probability } \frac{1-\pi_- -\pi_+}{2-\pi_-} \\ \epsilon & \text{with probability } \frac{\pi_+}{2-\pi_-}. \end{cases}$$

The expectation $\mathbb{E}_F|\hat{y} - Z|$ is minimised by the median m_Z of Z. The CDF of Z is given by

$$G(z) = \begin{cases} \frac{1}{2-\pi_-} \times F(z) & \text{if } z < -\epsilon \\ \frac{1}{2-\pi_-} \times F(z) + \frac{1-\pi_- -\pi_+}{2-\pi_-} & \text{if } -\epsilon \leq z < \epsilon \\ \frac{1}{2-\pi_-} \times F(z) + \frac{1-\pi_-}{2-\pi_-} & \text{if } z \geq \epsilon. \end{cases}$$

From $m < -\epsilon$, we know that $\pi_- > 0.5$. Using $G(m_Z) = 0.5$ and $\hat{y} \leq -\epsilon$, we get

$$\hat{y}_{\text{OPF}} = m_Z = \begin{cases} -\epsilon & \text{if } \pi_- < \frac{2}{3} \\ F^{-1}\{0.5 \times (2 - \pi_-)\} & \text{otherwise.} \end{cases}$$

This concludes the proof.

D Appendix to Chapter 5: Contribution to the 2023/24 VIEWS Prediction Competition

D.1 Modelling Details

D.1.1 Negative Binomial Distribution

We use the negative binomial distribution to model the predictive distribution of the s-step-ahead number of fatalities $Y_{i,t+s}$ of country i issued at time t. The NB is characterised by two parameters, r and p. Following the parameterisation from Lindén and Mäntyniemi (2011), $Y \sim NB(r,p)$ implies that

$$P_{\mathrm{NB}}(Y = y \,|\, r, p) = \binom{y + r - 1}{y} p^r (1-p)^y,$$

where the parameters can be expressed in terms of the expected value μ and variance σ^2 of Y, i.e.

$$r = \frac{\mu^2}{\sigma^2 - \mu} \tag{D.1}$$

and

$$p = \frac{\mu}{\sigma^2}. \tag{D.2}$$

As shown by Bliss and Fisher (1953), we can estimate μ and σ^2 via the mean and the empirical variance of Y and use Equations D.1 and D.2 to obtain estimates for r and p. In our case, we estimate $\mu_{i,t+s}$ and $\sigma^2_{i,t+s}$ from the previous w observations $\{y_{i,t-w+1}, ..., y_{i,t}\}$ of $Y_{i,t}$ via

$$\hat{\mu}^w_{i,t} = \frac{1}{w} \sum_{l=t-w+1}^{t} y_{i,l}$$

and

$$\hat{\sigma}^{2,w}_{i,t} = \frac{1}{w} \sum_{l=t-w+1}^{t} (y_{i,l} - \hat{\mu}^w_{i,t})^2.$$

Plugging these into Equations D.1 and D.2 leaves us with a fully specified probability distribution for $Y_{i,t+s}$. We determine w in a data-driven way; see Section 5.3.1.

D.1.2 Hurdle Model

In contrast to the NB approach, the hurdle model contains an additional Bernoulli variable $Z_{i,t+s}$ whose parameter $\pi_{i,t+s}$ denotes the probability of a non-zero number of fatalities for country i at time $t+s$. In that way, zero fatalities are accounted for separately and positive numbers of fatalities are modelled by a truncated negative binomial distribution. Omitting the indices $i, t+s$, the probability mass function of the hurdle model is given by

$$P_{\mathrm{H}}(Y=y) = \begin{cases} P_{\mathrm{TNB}}(Y=y \mid r,p)\cdot\pi, & y>0, \\ 1-\pi, & y=0, \end{cases} \tag{D.3}$$

see Porter and White (2012), p. 111. Here, P_{TNB} denotes the probability mass function of the truncated negative binomial distribution. The cumulative distribution function of the TNB is

$$P_{\mathrm{TNB}}(Y \leq y \mid r,p) = \frac{P_{\mathrm{NB}}(Y \leq y \mid r,p) - P_{\mathrm{NB}}(Y=0 \mid r,p)}{1-P_{\mathrm{NB}}(Y=0 \mid r,p)}.$$

We estimate $\pi_{i,t+s}$ via $\hat{\pi}_{i,t+s} = \frac{1}{w}\sum_{l=t-w+1}^{t} \mathbf{1}\{y_{i,l}>0\}$, the relative occurrence of positive fatality counts in a window of w past observations. The parameters of the TNB are calculated as described in Section D.1.1, where only positive values in the last w numbers of fatalities are used for estimation of $\mu_{i,t+s}$ and $\sigma^2_{i,t+s}$.

D.1.3 Neural Networks

The hidden layers of the neural networks each utilise the ReLu activation function $f(x) = \max(0,x)$. We choose the number of neurons per hidden layer n_j to be the same for all h hidden layers $j = 1, ..., h$ and the total number of neurons in the hidden layers $N_h = \sum_{j=1}^{h} n_j$ to lie in the interval $[\min(n_{\mathrm{input}}, n_{\mathrm{output}}), \max(n_{\mathrm{input}}, n_{\mathrm{output}})]$, where n_{input} (n_{output}) denotes the number of input (output) neurons.

As mentioned in Section 5.3.3, hyperparameter optimisation is performed through random search. In each iteration, the parameter combination to be examined is randomly sampled from the distributions shown in Table D.1. For each country, the chosen set of hyperparameters is that with the minimum CRPS on the validation dataset.

The number of hidden layers h is set to a maximum of 6 to mitigate overfitting and the batch size b is kept relatively small due to limited data availability. Adam is used as the optimisation algorithm (Kingma and Ba, 2017) with a minimum learning rate of 0.001. In our case, smaller learning rates tend to produce worse results. The feature subset f_{set} is randomly drawn from five possible sets; see Section 5.2, where ged_{sb} denotes fatality data only and *all* includes all features. We consider lead time $s = 8$ in the tuning process.

Table D.1: Sampling distributions for neural-network hyperparameters in the random search

Parameter	**Parameter Space**
w	$U_d(1, 12)$
h, b	$U_d(1, 6)$
N_h	$U_d([\min(n_{\text{input}}, n_{\text{output}}), \max(n_{\text{input}}, n_{\text{output}})])$
l	$U(0.001, 0.15)$
d	$U(0.1, 0.5) \times Ber(0.5)$
e	$U_d(3, 40)$
f_{set}	$U(\{\textit{conflicthistory}, \textit{vdem}, \textit{wdi}, \textit{ged}_{\text{sb}}, \textit{all}\})$

Note: U denotes the continuous uniform distribution, U_d the discrete uniform distribution, and Ber the Bernoulli distribution. The feature subsets *conflicthistory*, *vdem*, *wdi*, $\textit{ged}_{\text{sb}}$, and *all* are defined in Sections 5.2 and D.1.3.

E Appendix to Chapter 6: Challenges in Evaluating Conflict Fatality Forecasts from an Onset Perspective

E.1 Supplementary Figures

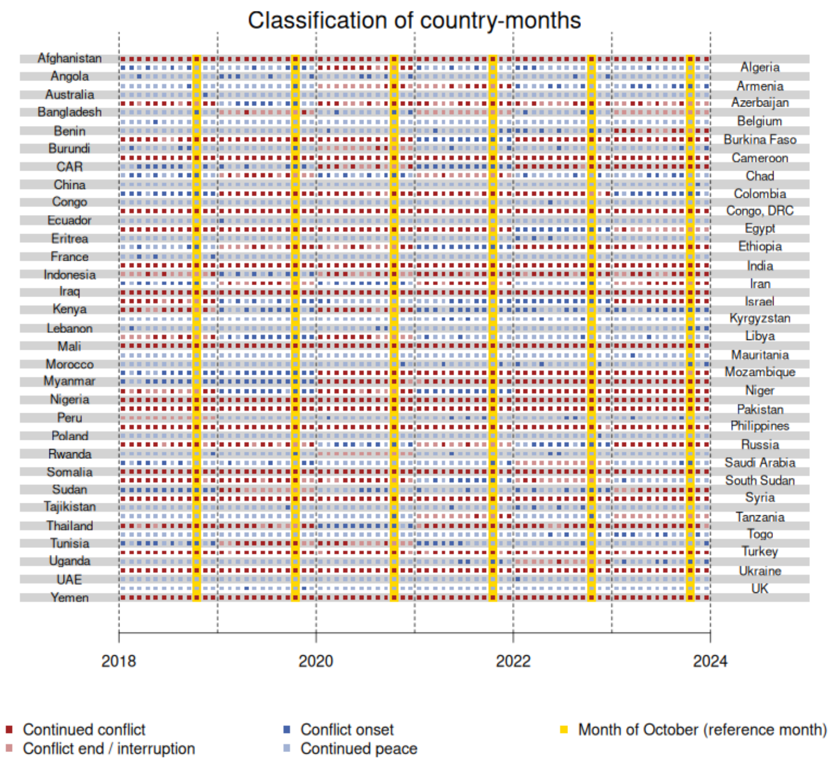

Figure E.1: Classification of monthly conflict situations according to the definition in Equation 6.1 for all countries with at least one conflict between 2018 and 2023

E.2 Mathematical Background

The link between the Brier score, CRPS and Brier score is well-known, but for clarity we re-state it in the notation of the manuscript. Keep in mind that we defined

$$y^{(a)} = \mathbf{1}(x > a)$$

and

$$p_a = \mathrm{Prob}_F(x > a) = \mathrm{Prob}_F(y^{(a)} = 1).$$

Now Equation 6.6 results from re-writing

$$\begin{aligned}
\mathrm{CRPS}(F, x) &= \int_{-\infty}^{\infty} [F(a) - \mathbf{1}(a \geq x)]^2 \mathrm{d}a \\
&= \int_{-\infty}^{\infty} [(1 - p_a) - (1 - y^{(a)})]^2 \mathrm{d}a \\
&= \int_{-\infty}^{\infty} \mathrm{BS}[(1 - p_a), (1 - y^{(a)})] \mathrm{d}a \\
&= \int_{-\infty}^{\infty} \mathrm{BS}(p_a, y^{(a)}) \mathrm{d}a,
\end{aligned}$$

where the last step holds because the Brier score is invariant to whether it is applied to an event $y^{(a)}$ or its converse $1 - y^{(a)}$. For count-valued outcomes, where $x \in \{0, 1, 2, \dots\}$ and F is a step function with jumps only at those same values, Equation 6.7 arises from

$$\begin{aligned}
\mathrm{CRPS}(F, x) &= \int_{-\infty}^{\infty} [F(a) - \mathbf{1}(a \geq x)]^2 \mathrm{d}a \\
&= \sum_{i=0}^{\infty} \int_{i}^{i+1} [F(a) - \mathbf{1}(a \geq x)]^2 \mathrm{d}a \\
&= \sum_{i=0}^{\infty} [F(i) - \mathbf{1}(i \geq x)]^2 \\
&= \sum_{i=0}^{\infty} \mathrm{BS}[(1 - p_i), (1 - y^{(i)})] \\
&= \sum_{i=0}^{\infty} \mathrm{BS}(p_i, y^{(i)}) = \mathrm{RPS}(F, x).
\end{aligned}$$

We note that this equality only holds if F is indeed a suitable step function. In practice this is not the case for all VIEWS submissions, some of which contained real-valued samples. This is the reason why we had to round all samples prior to evaluation, see Section 6.2.1.

Bibliography

Akgun, O., A. Pirotte, G. Urga, and Z. Yang (2024): “Equal Predictive Ability Tests Based on Panel Data with Applications to OECD and IMF Forecasts,” *International Journal of Forecasting*, 40, 202–228.

Allen, S., D. Ginsbourger, and J. Ziegel (2023): “Evaluating Forecasts for High-Impact Events Using Transformed Kernel Scores,” *SIAM/ASA Journal on Uncertainty Quantification*, 11, 906–940.

Andrews, D. W. K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817–858.

Armendáriz, B. and A. Szafarz (2011): “On Mission Drift in Microfinance Institutions,” *Handbook of Microfinance*.

Assefa, E., N. Hermes, and A. Meesters (2013): “Competition and the Performance of Microfinance Institutions,” *Applied Financial Economics*, 23, 767–782.

Athey, S. and G. W. Imbens (2017): “The State of Applied Econometrics: Causality and Policy Evaluation,” *Journal of Economic Perspectives*, 31, 3–32.

Awaworyi Churchill, S. (2020): “Microfinance Financial Sustainability and Outreach: Is There a Trade-Off?” *Empirical Economics*, 59, 1329–1350.

Ayayi, A. G. and M. Sene (2010): “What Drives Microfinance Institution’s Financial Sustainability,” *The Journal of Developing Areas*, 303–324.

Azur, M. J., E. A. Stuart, C. Frangakis, and P. J. Leaf (2011): “Multiple Imputation By Chained Equations: What Is It and How Does It Work?” *International Journal of Methods in Psychiatric Research*, 20, 40–49.

Azzalini, A. (2013): *The Skew-Normal and Related Families*, Institute of Mathematical Statistics Monographs, Cambridge University Press.

Bai, J., S. H. Choi, and Y. Liao (2024): "Standard Errors for Panel Data Models with Unknown Clusters," *Journal of Econometrics*, 240, 105004.

Baltagi, B. H. (2008): "Econometric Analysis of Panel Data," *Rohn Wiley.*

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): "The Diffusion of Microfinance," *Science*, 341.

Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan (2015): "The Miracle of Microfinance? Evidence from a Randomized Evaluation," *American Economic Journal: Applied Economics*, 7, 22–53.

Banerjee, A. V. and E. Duflo (2025): *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, Hachette UK.

Bartlett, J. W., S. R. Seaman, I. R. White, J. R. Carpenter, and Alzheimer's Disease Neuroimaging Initiative* (2015): "Multiple Imputation of Covariates by Fully Conditional Specification: Accommodating the Substantive Model," *Statistical Methods in Medical Research*, 24, 462–487.

Basharat, B., M. Hudon, and A. Nawaz (2015): "Does Efficiency Lead to Lower Prices? A New Perspective from Microfinance Interest Rates," *Strategic Change*, 24, 49–66.

Bassem, B. S. (2012): "Social and Financial Performance of Microfinance Institutions: Is There a Trade-off?" *Journal of Economics and International Dinance*, 4, 92.

Bazzi, S., R. A. Blair, C. Blattman, O. Dube, M. Gudgeon, and R. Peck (2022): "The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia," *Review of Economics and Statistics*, 104, 764–779.

Becker, F., F. Krüger, and M. Schienle (2025): "Simple Macroeconomic Forecast Distributions for the G7 Economies," *The Annals of Applied Statistics*, 19, 2878–2897.

Belloni, A. and V. Chernozhukov (2013): "Least Squares After Model Selection in High-Dimensional Sparse Models," *Bernoulli*, 19, 521–547.

Belloni, A., V. Chernozhukov, C. Hansen, and D. Kozbur (2016): "Inference in High-Dimensional Panel Models with an Application to Gun Control," *Journal of Business & Economic Statistics*, 34, 590–605.

Berge, L. I. O., K. Bjorvatn, and B. Tungodden (2015): "Human and Financial Capital for Microenterprise Development: Evidence from a Field and Lab Experiment," *Management Science*, 61, 707–722.

Bergstra, J. and Y. Bengio (2012): "Random Search for Hyper-Parameter Optimization," *The Journal of Machine Learning Research*, 13, 281–305.

Bester, C. A., T. G. Conley, and C. B. Hansen (2011): "Inference with Dependent Data Using Cluster Covariance Estimators," *Journal of Econometrics*, 165, 137–151.

Bickel, P. J. and E. Levina (2008): "Regularized Estimation of Large Covariance Matrices," *The Annals of Statistics*, 36, 199–227.

Bliss, C. I. and R. A. Fisher (1953): "Fitting the Negative Binomial Distribution to Biological Data," *Biometrics*, 9, 176–200.

Bodentien, T. and L. Rüter (2024): "Forecasting Monthly Fatalities via a Negative Binomial Distribution and Comparison with a Hurdle Model and Neural Networks," Tech. rep., The 2023/2024 VIEWS Prediction Challenge, https://viewsforecasting.org/wp-content/uploads/Bodentien_VIEWSPredictionChallenge2023.pdf.

Bogan, V. L. (2012): "Capital Structure and Sustainability: An Empirical Study of Microfinance Institutions," *Review of Economics and Statistics*, 94, 1045–1058.

Boisbunon, A., S. Canu, D. Fourdrinier, W. Strawderman, and M. T. Wells (2013): "AIC, C_p and Estimators of Loss for Elliptically Symmetric Distributions," *arXiv preprint arXiv:1308.2766*.

Bosse, N. I., S. Abbott, A. Cori, E. van Leeuwen, J. Bracher, and S. Funk (2023): "Scoring Epidemiological Forecasts on Transformed Scales," *PLOS Computational Biology*, 19, 1–23.

Bracher, J., E. L. Ray, T. Gneiting, and N. G. Reich (2021): "Evaluating Epidemic Forecasts in an Interval Format," *PLOS Computational Biology*, 17, 1–15.

Bracher, J., L. Rüter, F. Krüger, S. Lerch, and M. Schienle (2023): "Direction Augmentation in the Evaluation of Armed Conflict Predictions," *International Interactions*, 49, 989–1004.

Brandt, P. T. (2024): "Bayesian Density Forecasts for Views," Tech. rep., The 2023/2024 VIEWS Prediction Challenge, https://viewsforecasting.org/wp-content/uploads/Brandt_VIEWSPredictionChallenge2023.pdf.

Brandt, P. T., V. D'Orazio, L. Khan, Y.-F. Li, J. Osorio, and M. Sianan (2022): "Conflict Forecasting with Event Data and Spatio-Temporal Graph Convolutional Networks," *International Interactions*, 48, 800–822.

Brandt, P. T., J. R. Freeman, and P. A. Schrodt (2014): "Evaluating Forecasts of Political Conflict Dynamics," *International Journal of Forecasting*, 30, 944–962.

Brau, J. C. and G. M. Woller (2004): "Microfinance: A Comprehensive Review of the Existing Literature," *Journal of Entrepreneurial Finance*, 9, 1–28.

Brehmer, J. R., K. Kraus, T. Gneiting, M. Herrmann, and W. Marzocchi (2024): "Enhancing the Statistical Evaluation of Earthquake Forecasts–An Application to Italy," *Seismological Research Letters*, 96, 1966–1988.

Buera, F. J., J. P. Kaboski, and Y. Shin (2021): "The Macroeconomics of Microfinance," *Review of Economic Studies*, 88, 126–161.

Buse, R. and M. Schienle (2019): "Measuring Connectedness of Euro Area Sovereign Risk," *International Journal of Forecasting*, 35, 25–44.

Buse, R., M. Schienle, and J. Urban (2022): "Assessing the Impact of Policy and Regulation Interventions in European Sovereign Credit Risk Networks: What Worked Best?" *Journal of International Economics*, 139, 103673.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference with Multiway Clustering," *Journal of Business Economic Statistics*, 29, 238–249.

Cederman, L.-E. and N. B. Weidmann (2017): "Predicting Armed Conflict: Time to Adjust Our Expectations?" *Science*, 355, 474–476.

Chen, K. and T. J. Vogelsang (2024): "Fixed-b Asymptotics for Panel Models with Two-Way Clustering," *Journal of Econometrics*, 244, 105831.

Cheston, S. and L. Kuhn (2002): "Empowering Women through Microfinance," in *Pathways Out of Poverty: Innovations in Microfinance for the Poorest Families*, ed. by S. Daley-Harris, Kumarian Press.

Chiang, H. D., B. E. Hansen, and Y. Sasaki (2024): "Standard Errors for Two-Way Clustering with Serially Correlated Time Effects," *Review of Economics and Statistics*, 1–40.

Collier, P. and N. Sambanis (2002): "Understanding Civil War: A New Agenda," *The Journal of Conflict Resolution*, 46, 3–12.

Cragg, J. G. (1971): "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica*, 39, 829–844.

Creal, D. and J. Kim (2024): "Bayesian Estimation of Cluster Covariance Matrices of Unknown Form," *Journal of Econometrics*, 241, 105725.

Cull, R., A. Demirgüç-Kunt, and J. Morduch (2007): "Financial Performance and Outreach: A Global Analysis of Leading Microbanks," *The Economic Journal*, 117, 107–133.

Davezies, L., X. D'Haultfoeuille, and Y. Guyonvarch (2025): "Analytic Inference with Two-Way Clustering," *arXiv preprint arXiv:2506.20749*.

Davies, S., G. Engström, T. Pettersson, and M. Öberg (2024): "Organized Violence 1989–2023, and the Prevalence of Organized Crime Groups," *Journal of Peace Research*, 61, 673–693.

De'Ath, G. and K. E. Fabricius (2000): "Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis," *Ecology*, 81, 3178–3192.

Dernoncourt, D., B. Hanczar, and J.-D. Zucker (2014): "Analysis of Feature Selection Stability on High Dimension and Small Sample Data," *Computational Statistics & Data Analysis*, 71, 681–693.

Diebold, F. and R. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business Economic Statistics*, 13, 253–263.

Dimitriadis, T., T. Gneiting, A. I. Jordan, and P. Vogel (2024): "Evaluating Probabilistic Classifiers: The Triptych," *International Journal of Forecasting*, 40, 1101–1122.

Driscoll, J. C. and A. C. Kraay (1998): "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *Review of Economics and Statistics*, 80, 549–560.

Efron, B. and B. Narasimhan (2020): "The Automatic Construction of Bootstrap Confidence Intervals," *Journal of Computational and Graphical Statistics*, 29, 608–619.

Ehm, W., T. Gneiting, A. Jordan, and F. Krüger (2016): "Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations and Forecast Rankings," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 78, 505–562.

Fall, F. S., H. Tchakoute Tchuigoua, A. Vanhems, and L. Simar (2023): "Investigating the Unobserved Heterogeneity Effect on Outreach to Women: Lessons from Microfinance Institutions," *Annals of Operations Research*, 328, 1365–1386.

Fan, J., Y. Liao, and M. Mincheva (2013): "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society Series B*, 75, 603–680.

Field, E. and R. Pande (2008): "Repayment Frequency and Default in Microfinance: Evidence from India," *Journal of the European Economic Association*, 6, 501–509.

Field, E., R. Pande, J. Papp, and N. Rigol (2013): "Does the Classic Microfinance Model Discourage Entrepreneurship Among the Poor? Experimental Evidence from India," *American Economic Review*, 103, 2196–2226.

Friedman, J., T. Hastie, and R. Tibshirani (2010): "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, 33, 1–22.

Gao, J., B. Peng, and Y. Yan (2024): "Higher-Order Expansions and Inference for Panel Data Models," *Journal of the American Statistical Association*, 119, 2760–2771.

Gneiting, T. (2011): "Making and Evaluating Point Forecasts," *Journal of the American Statistical Association*, 746–762.

Gneiting, T. and M. Katzfuss (2014): "Probabilistic Forecasting," *Annual Review of Statistics and Its Application*, 1, 125–151.

Gneiting, T. and A. E. Raftery (2007): "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

Goldstein, H., J. R. Carpenter, and W. J. Browne (2014): "Fitting Multilevel Multivariate Models with Missing Data in Responses and Covariates That May Include Interactions and Non-Linear Terms," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 553–564.

Gonçalves, S. (2011): "The Moving Blocks Bootstrap for Panel Linear Regression Models with Individual Fixed Effects," *Econometric Theory*, 27, 1048–1082.

Gyntelberg, J., P. Hoerdahl, K. Ters, and J. Urban (2013): "Intraday Dynamics of Euro Area Sovereign CDS and Bonds," Tech. Rep. 423, Bank for International Settlements.

Hartarska, V. and D. Nadolnyak (2007): "Do Regulated Microfinance Institutions Achieve Better Sustainability and Outreach? Cross-Country Evidence," *Applied Economics*, 39, 1207–1222.

Hautsch, N. (2004): *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models*, Springer.

Hegre, H., M. Allansson, M. Basedau, M. Colaresi, M. Croicu, H. Fjelde, F. Hoyles, L. Hultman, S. Högbladh, R. Jansen, et al. (2019): "ViEWS: A Political Violence Early-Warning System," *Journal of Peace Research*, 56, 155–174.

Hegre, H., C. Bell, M. Colaresi, M. Croicu, F. Hoyles, R. Jansen, M. R. Leis, A. Lindqvist-McGowan, D. Randahl, E. G. Rød, and P. Vesco (2021): "ViEWS2020: Revising and Evaluating the ViEWS Political Violence Early-Warning System," *Journal of Peace Research*, 58, 599–611.

Hegre, H., N. W. Metternich, H. M. Nygård, and J. Wucherpfennig (2017): "Introduction: Forecasting in Peace Research," *Journal of Peace Research*, 54, 113–124.

Hegre, H., P. Vesco, and M. Colaresi (2022): "Lessons from an Escalation Prediction Competition," *International Interactions*, 48, 521–554.

Hegre, H., P. Vesco, M. Colaresi, and J. Vestby (2023): "Invitation to the 2023/24 VIEWS Prediction Competition," Tech. rep., Peace Research Institute Oslo (PRIO) and

Uppsala University, https://viewsforecasting.org/wp-content/uploads/VIEWS_2023.24_Prediction_Competition_Invitation.pdf.

Hegre, H., P. Vesco, M. Colaresi, J. Vestby, A. Timlick, N. S. Kazmi, A. Lindqvist-McGowan, F. Becker, M. Binetti, T. Bodentien, T. Bohne, P. T. Brandt, T. Chadefaux, S. Drauz, C. Dworschak, V. D'Orazio, H. Frank, C. Fritz, K. S. Gleditsch, S. Häffner, M. Hofer, F. L. Klebe, L. Macis, A. Malaga, M. Mehrl, N. W. Metternich, D. Mittermaier, D. Muchlinski, H. Mueller, C. Oswald, P. Pisano, D. Randahl, C. Rauh, L. Rüter, T. Schincariol, B. Seimon, E. Siletti, M. Tagliapietra, C. Thornhill, J. Vegelius, and J. Walterskirchen (2025): "The 2023/24 VIEWS Prediction Challenge: Predicting the Number of Fatalities in Armed Conflict, with Uncertainty," *Journal of Peace Research*, 62, 2070–2087.

Hermes, N. and M. Hudon (2019): "Determinants of the Performance of Microfinance Institutions: A Systematic Review," *Contemporary Topics in Finance: A Collection of Literature Surveys*, 297–330.

Hermes, N. and R. Lensink (2007): "The Empirics of Microfinance: What Do We Know?" *Economic Journal*, 117, 1–10.

Hermes, N., R. Lensink, and A. Meesters (2011): "Outreach and Efficiency of Microfinance Institutions," *World Development*, 39, 938–948.

Hewamalage, H., K. Ackermann, and C. Bergmeir (2023): "Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices," *Data Mining and Knowledge Discovery*, 37, 788–832.

Hidalgo, J. and M. Schafgans (2021): "Inference Without Smoothing for Large Panels with Cross-Sectional and Temporal Dependence," *Journal of Econometrics*, 223, 125–160.

Hong, T., P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman (2016): "Probabilistic Energy Forecasting: Global Energy Forecasting Competition 2014 and Beyond," *International Journal of Forecasting*, 32, 896–913.

Huddleston, D., F. Liu, and L. Stentoft (2023): "Intraday Market Predictability: A Machine Learning Approach," *Journal of Financial Econometrics*, 21, 485–527.

Hulme, D. (2000): "Is Microdebt Good for Poor People? A Note on the Dark Side of Microfinance," *Small Enterprise Development*, 11, 26–28.

Ibragimov, R. and U. K. Müller (2010): "t-Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business Economic Statistics*, 28, 453–468.

——— (2016): "Inference with Few Heterogeneous Clusters," *Review of Economics and Statistics*, 98, 83–96.

Imai, K. S., R. Gaiha, G. Thapa, and S. K. Annim (2012): "Microfinance and Poverty–A Macro Perspective," *World Development*, 40, 1675–1689.

Joseph, V. R. (2022): "Optimal Ratio for Data Splitting," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15, 531–538.

Khandker, S. R. (2005): "Microfinance and Poverty: Evidence Using Panel Data from Bangladesh," *World Bank Economic Review*, 19, 263–286.

Kim, M. S. and Y. Sun (2013): "Heteroskedasticity and Spatiotemporal Dependence Robust Inference for Linear Panel Models with Fixed Effects," *Journal of Econometrics*, 177, 85–108.

Kinde, B. A. (2012): "Financial Sustainability of Microfinance Institutions (MFIs) in Ethiopia," *European Journal of Business and Management*, 4, 1–10.

Kingma, D. P. and J. Ba (2017): "Adam: A Method for Stochastic Optimization," .

Klemperer, P. (1999): "Auction Theory: A Guide to the Literature," *Journal of Economic Surveys*, 13, 227–286.

——— (2002): "What Really Matters in Auction Design," *Journal of Economic Perspectives*, 16, 169–189.

Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

Kolassa, S. (2020): "Why the "Best" Point Forecast Depends on the Error or Accuracy Measure," *International Journal of Forecasting*, 36, 208–211.

Kyereboah-Coleman, A. and K. A. Osei (2008): "Outreach and Profitability of Microfinance Institutions: The Role of Governance," *Journal of Economic Studies*, 35, 236–248.

Lerch, S., T. L. Thorarinsdottir, F. Ravazzolo, and T. Gneiting (2017): "Forecaster's Dilemma: Extreme Events and Forecast Evaluation," *Statistical Science*, 106–127.

Lindén, A. and S. Mäntyniemi (2011): "Using the Negative Binomial Distribution to Model Overdispersion in Ecological Count Data," *Ecology*, 92, 1414–1421.

Little, R. J. and D. B. Rubin (2019): *Statistical Analysis with Missing Data*, vol. 793, John Wiley & Sons.

Liu, Y., Y. Wang, Y. Feng, and M. M. Wall (2016): "Variable Selection and Prediction with Incomplete High-Dimensional Data," *Annals of Applied Statistics*, 10, 418–450.

Macis, L., M. Tagliapietra, E. Siletti, and P. Pisano (2024): "Predicting Fatalities with Pre-Trained Temporal Transformers: A Time Series Regression Approach," Tech. rep., The 2023/2024 VIEWS Prediction Challenge, https://viewsforecasting.org/wp-content/uploads/Unito_VIEWSPredictionChallenge2023.pdf.

Mallows, C. L. (2000): "Some Comments on C_p," *Technometrics*, 42, 87–94.

Manski, C. F. (2013): *Public Policy in an Uncertain World: Analysis and Decisions*, Harvard University Press.

McIntosh, C. (2008): "Estimating Treatment Effects from Spatial Policy Experiments: An Application to Ugandan Microfinance," *The Review of Economics and Statistics*, 90, 15–28.

McIntosh, C. and B. Wydick (2005): "Competition and Microfinance," *Journal of Development Economics*, 78, 271–298.

Meinshausen, N. and P. Bühlmann (2010): "Stability Selection," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 72, 417–473.

Menzel, K. (2021): "Bootstrap with Cluster-Dependence in Two or More Dimensions," *Econometrica*, 89, 2143–2188.

Merlevède, F., M. Peligrad, and E. Rio (2009): "Bernstein Inequality and Moderate Deviations Under Strong Mixing Conditions," in *High Dimensional Probability V: The Luminy Volume*, Institute of Mathematical Statistics, vol. 5, 273–293.

Moghimbeigi, A., M. R. Eshraghian, K. Mohammad, and B. Mcardle (2008): "Multilevel Zero-Inflated Negative Binomial Regression Modeling for Over-Dispersed Count Data with Extra Zeros," *Journal of Applied Statistics*, 35, 1193–1202.

Morduch, J. (1999): "The Microfinance Promise," *Journal of Economic Literature*, 37, 1569–1614.

Muchlinski, D. and C. Thornhill (2024): "A Zero-Inflated Poisson Generalized Additive Model for Forecasting Conflict Fatalities," Tech. rep., The 2023/2024 VIEWS Prediction Challenge, https://viewsforecasting.org/wp-content/uploads/Muchlinski_VIEWSPredictionChallenge2023.pdf.

Mueller, H. and C. Rauh (2022a): "The Hard Problem of Prediction for Conflict Prevention," *Journal of the European Economic Association*, 20, 2440–2467.

——— (2022b): "Using Past Violence and Current News to Predict Changes in Violence," *International Interactions*, 48, 579–596.

Málaga, A., H. Mueller, C. Rauh, and B. Seimon (2024): "Predicting Fatalities Using Newspaper Text," Tech. rep., The 2023/2024 VIEWS Prediction Challenge, https://viewsforecasting.org/wp-content/uploads/ConflictForecast_VIEWSPredictionChallenge2023.pdf.

Newey, W. and K. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Newey, W. K. and K. D. West (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *The Review of Economic Studies*, 61, 631–653.

Oswald, C. and D. Ohrenhofer (2022): "Click, Click Boom: Using Wikipedia Data to Predict Changes in Battle-Related Deaths," *International Interactions*, 48, 678–696.

Pesaran, M. H. and T. Yamagata (2008): "Testing Slope Homogeneity in Large Panels," *Journal of Econometrics*, 142, 50–93.

Petersen, M. A. (2009): "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches," *The Review of Financial Studies*, 22, 435–480.

Pettersson, T. and M. Öberg (2020): "Organized Violence, 1989–2019," *Journal of Peace Research*, 57, 597–613.

Phillips, P. C. B. and H. R. Moon (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

Porter, M. D. and G. White (2012): "Self-Exciting Hurdle Models for Terrorist Activity," *The Annals of Applied Statistics*, 6, 106–124.

Qu, R., A. Timmermann, and Y. Zhu (2024): "Comparing Forecasting Performance with Panel Data," *International Journal of Forecasting*, 40, 918–941.

Quayes, S. (2012): "Depth of Outreach and Financial Sustainability of Microfinance Institutions," *Applied Economics*, 44, 3421–3433.

——— (2015): "Outreach and Performance of Microfinance Institutions: A Panel Analysis," *Applied Economics*, 47, 1909–1925.

Quayes, S. and G. Joseph (2022): "Determinants of Social Outreach of Microfinance Institutions," in *Capitalism: An Unsustainable Future?*, Routledge, 210–220.

R Core Team (2021): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Raghunathan, T., J. Lepkowski, J. V. Hoewyk, and P. Solenberger (2001): "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models," *Survey Methodology*, 27, 85–96.

Randahl, D. and J. Vegelius (2024): "Forecasting Densities of Fatalities from State-Based Conflicts Using Observed Markov Models," *arXiv preprint arXiv:2410.12374*.

Roberts, P. W. (2013): "The Profit Orientation of Microfinance Institutions and Effective Interest Rates," *World Development*, 41, 120–131.

Rød, E. G., T. Gåsste, and H. Hegre (2024): "A Review and Comparison of Conflict Early Warning Systems," *International Journal of Forecasting*, 40, 96–112.

Rüter, L. and M. Schienle (2025): "Model Determination for High-Dimensional Longitudinal Data with Missing Observations: An Application to Microfinance Data," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, qnae144.

Schafer, J. L. and J. W. Graham (2002): "Missing Data: Our View of the State of the Art," *Psychological Methods*, 7, 147–177.

Schincariol, T., H. Frank, and T. Chadefaux (2024): "Temporal Patterns in Conflict Prediction: An Improved Shape-Based Approach," Tech. rep., The 2023/2024 VIEWS Prediction Challenge, https://viewsforecasting.org/wp-content/uploads/Pace_VIEWSPredictionChallenge2023.pdf.

Sela, R. J. and J. S. Simonoff (2012): "RE-EM Trees: A Data Mining Approach for Longitudinal and Clustered Data," *Machine Learning*, 86, 169–207.

Shah, R. D. and R. J. Samworth (2013): "Variable Selection with Error Control: Another Look at Stability Selection," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75, 55–80.

Strøm, R. Ø., B. D'Espallier, and R. Mersland (2014): "Female Leadership, Performance, and Governance in Microfinance Institutions," *Journal of Banking & Finance*, 42, 60–75.

Swain, R. B. and F. Y. Wallentin (2009): "Does Microfinance Empower Women? Evidence from Self-Help Groups in India," *International Review of Applied Economics*, 23, 541–556.

Taylor, A. D. (2002): "The Manipulability of Voting Systems," *The American Mathematical Monthly*, 109, 321–337.

Ters, K. and J. Urban (2018): "Intraday Dynamics of Credit Risk Contagion Before and During the Euro Area Sovereign Debt Crisis," *International Review of Economics Finance*, 54, 123–142.

Tollefsen, A. F., H. Strand, and H. Buhaug (2012): "PRIO-GRID: A Unified Spatial Data Structure," *Journal of Peace Research*, 49, 363–374.

van Buuren, S. (2007): "Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification," *Statistical Methods in Medical Research*, 16, 219–242.

van Buuren, S. and K. Groothuis-Oudshoorn (2011): "mice: Multivariate Imputation by Chained Equations in R," *Journal of Statistical Software*, 45, 1–67.

van Rooyen, C., R. Stewart, and T. de Wet (2012): "The Impact of Microfinance in Sub-Saharan Africa: A Systematic Review of the Evidence," *World Development*, 40, 2249–2262.

Varian, H. R. (2014): "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, 28, 3–28.

Ver Hoef, J. M. and P. L. Boveng (2007): "Quasi-Poisson Vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data?" *Ecology*, 88, 2766–2772.

Vesco, P., H. Hegre, M. Colaresi, R. B. Jansen, A. Lo, G. Reisch, and N. B. Weidmann (2022): "United They Stand: Findings from an Escalation Prediction Competition," *International Interactions*, 48, 860–896.

Vitart, F., A. W. Robertson, A. Spring, F. Pinault, R. Roškar, W. Cao, S. Bech, A. Bienkowski, N. Caltabiano, E. De Coning, B. Denis, A. Dirkson, J. Dramsch, P. Dueben, J. Gierschendorf, H. S. Kim, K. Nowak, D. Landry, L. Lledó, L. Palma, S. Rasp, and S. Zhou (2022): "Outcomes of the WMO Prize Challenge to Improve Subseasonal to Seasonal Predictions Using Artificial Intelligence," *Bulletin of the American Meteorological Society*, 103, E2878–E2886.

Vogelsang, T. J. (2012): "Heteroskedasticity, Autocorrelation, and Spatial Correlation Robust Inference in Linear Panel Models with Fixed Effects," *Journal of Econometrics*, 166, 303–319.

Wang, S., B. Nan, S. Rosset, and J. Zhu (2011): "Random Lasso," *Annals of Applied Statistics*, 5, 468–485.

White, H. (2014): *Asymptotic Theory for Econometricians*, Academic Press.

White, I. R., P. Royston, and A. M. Wood (2011): "Multiple Imputation Using Chained Equations: Issues and Guidance for Practice," *Statistics in Medicine*, 30, 377–399.

Wooldridge, J. M. (2012): *Introductory Econometrics: A Modern Approach*, 5 ed.

Yap, L. (2025): "Asymptotic Theory for Two-Way Clustering," *Journal of Econometrics*, 249, 106001.

Yunus, M. (2009): *Creating a World Without Poverty: Social Business and the Future of Capitalism*, Public Affairs.

Zeckhauser, R. (1973): "Voting Systems, Honest Preferences and Pareto Optimality," *American Political Science Review*, 67, 934–946.

Zhao, Y. and Q. Long (2017): "Variable Selection in the Presence of Missing Data: Imputation-Based Methods," *Wiley Interdisciplinary Reviews: Computational Statistics*, 9, e1402.

Zheng, X. and W.-Y. Loh (1995): "Consistent Variable Selection in Linear Models," *Journal of the American Statistical Association*, 90, 151–156.

Zhou, H., M. R. Elliott, and T. E. Raghunathan (2016): "A Two-Step Semiparametric Method to Accommodate Sampling Weights in Multiple Imputation," *Biometrics*, 72, 242–252.

Zhou, J., H. Li, and W. Zhong (2021): "A Modified Diebold-Mariano Test for Equal Forecast Accuracy with Clustered Dependence," *Economics Letters*, 207, 110029.

Declaration of the Use of Generative Artificial Intelligence Tools

Documentation of generative artificial intelligence (AI) usage and assessment of its impact

Section	**Role of Generative AI**	**Impact**
Introduction	Assistance with phrasing of an early draft, followed by substantial revision by the author	Medium
	Light language refinements, consistency in British English and correction of typographical errors	Low
Chapter 2	Consistency in British English and correction of typographical errors	Low
Chapter 3	Assistance with phrasing of an early draft of the introduction paragraph, followed by substantial revision by the author	Medium
	Suggestions for restructuring of the code to improve readability and efficiency	Low
	LaTeX formatting of Tables 3.1–3.4, B.1 and B.2 based on variable descriptions, model information and R output	High
	Generation of Figure 3.1 with R from existing estimation results	High
	Assistance with phrasing of an early draft of the results section, followed by substantial revision by the author	Medium
	Light language refinements, consistency in British English and correction of typographical errors	Low
Chapter 4	Consistency in British English and correction of typographical errors	Low
Chapter 5	Consistency in British English and correction of typographical errors	Low
Chapter 6	Assistance with phrasing of an early draft of the introduction paragraph, followed by substantial revision by the author	Medium
	Light language refinements, consistency in British English and correction of typographical errors	Low

Note: Generative AI tools used under author supervision: DeepL and ChatGPT 5.2 (OpenAI).