Essays on the use of MIDAS regressions in banking and finance

DISSERTATION

of the University of St. Gallen, School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by

Alexander Kostrov

from

Russia

Approved on the application of

Prof. Francesco Audrino, PhD

and

Prof. Eric Ghysels, PhD

Dissertation no. 5121

Difo-Druck GmbH, Untersiemau 2021

Essays on the use of MIDAS regressions in banking and finance

DISSERTATION

of the University of St. Gallen, School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by

Alexander Kostrov

from

Russia

Approved on the application of

Prof. Francesco Audrino, PhD

and

Prof. Eric Ghysels, PhD
Prof. Dr. Reto Föllmi

Dissertation no. 5121

Difo-Druck GmbH, Untersiemau 2021

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science, hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, May 28, 2021

The President:

Prof. Dr. Bernhard Ehrenzeller

Acknowledgements

Working on a PhD thesis is a life-changing experience. I am deeply grateful to my thesis advisor Prof. Francesco Audrino and co-advisor Prof. Juan-Pablo Ortega for the advice and mentoring that they provided to me in doing research and communications with students, as well as outside the classroom. Their inspiring teaching methods helped me a lot in challenging situations. I also thank Prof. Reto Föllmi for his comments on my research proposal and participation in the thesis committee.

I express my gratitude and special appreciation for Prof. Eric Ghysels for the numerous discussions, valuable feedback, and guidance during my ongoing research visit at the University of North Carolina at Chapel Hill. Despite the unusual circumstances of the year 2020, I had unrestricted opportunities to learn and move forward with my dissertation. This would have been impossible without financial support from the Swiss National Science Foundation.

I am indebted to Prof. Anastasija Tetereva for her contribution to this work and to Prof. Lyudmila Grigoryeva for sharing the dataset with me. I am thankful to the former and current PhD students Alexander, Daniele, Dominik, Jeannine, Mikhail, Philipp, Robert, Thomas, and Yauhen. Moreover, I am grateful to all the participants of the PEF Literature and PhD Seminars at the University of St. Gallen for their suggestions and exchange of ideas. Additionally, many thanks to Margit and Mirela for the administrative help throughout my PhD studies.

Last, but not the least, I would like to thank my parents, wife Arina, and daughter Anna for their trust and familial comfort.

Alexander Kostrov

Chapel Hill, 2021

Summary

Economic time series are available at different frequencies due to their origin and data collection techniques. A mixed data sampling (MIDAS) regression is mainly a forecasting tool designed to harness mixed-frequency data. This dissertation proposes a computationally efficient estimation approach for MIDAS models and demonstrates its advantages in applications involving Big Data, including those related to banking and finance.

In the first essay, the logit model is generalized by introducing the MIDAS logit model. It is applied to predict U.S. bank failures. The proposed model yields a strongly significant improvement in classification accuracy both in statistical and economic terms. Some of the largest bank failures previously misclassified are now correctly predicted. Furthermore, an algorithm is suggested to mitigate the class-imbalance problem that is pervasive across many datasets.

The second essay is also related to the investigation of the banks' financial stability. The objective of this paper is to forecast bank closures costs in the United States. It is found that the primary federal regulators supervise banks inconsistently which affects the formation of costs. The accuracy of forecasting is low for banks supervised by some regulators. This evidence questions the quality of supervision and implies that the off-site monitoring might be blinded.

In the third essay, MIDAS regressions are applied to predict the realized covariance matrices. A covariance matrix forecast is important in risk-management and portfolio applications. A study is performed across nearly all stocks in Dow Jones and it is discovered that the MIDAS model has superior forecasting accuracy as compared to the HAR model. The economic value of the proposed method is evaluated using portfolio management applications in terms of the investor's utility and annualized returns.

The fourth essay aims at developing a computationally efficient approach to estimating MIDAS regressions and demonstrates its advantages in applications involving Big Data, in which estimation has previously not been feasible. I revise an optimization procedure for a MIDAS-NLS estimator and reach a major enhancement in its performance. In order to facilitate the use of my approach, that is the MIDAS-NLS with revised optimization, the "MIDAS Analytic" toolbox was released. Documentation for the toolbox is provided in Chapter 5.

Zusammenfassung

Aufgrund ihrer unterschiedlichen Quellen und Erfassungsmethoden variieren die Frequenzen ökonomischer Zeitreihen stark. Die MIDAS-Regression (Mixed Data Sampling) ist eine populäre Prognosemethode zur Auswertung von Daten mit unterschiedlichen Frequenzen. Diese Arbeit stellt eine rechnerisch effiziente Schätzungsmethode für MIDAS-Modelle vor und demonstriert deren Vorteile in Big Data-Anwendungen, u. a. im Banken- und Finanzwesen.

Der erste Aufsatz erweitert das Logit-Modell durch Einführung des MIDAS-Logit-Modells. Es wird angewendet, um amerikanische Bankausfälle vorherzusagen. Das eingeführte Modell erhöht die statistische und ökonomische Klassifizierungsgenauigkeit in signifikanter Weise. Ausserdem wird eine neue Methode zum Ausgleich des Klassenungleichgewichts vorgestellt.

Der zweite Aufsatz beschäftigt sich weiter mit der Untersuchung der finanziellen Stabilität von Banken. Ziel dieser Forschung ist es, die Kosten von Schliessungen amerikanischer Banken vorherzusagen. Es wird festgestellt, dass die primären Bundesaufsichtsbehörden die Banken uneinheitlich überwachen, was sich auf die Bildung der Kosten auswirkt. Darüber hinaus haben einige Agenturen eine geringe Vorhersagekraft. Dieser Nachweis stellt die Qualität der Überwachung in Frage und impliziert, dass Regulierungsbehörden möglicherweise getäuscht werden können.

Der dritte Aufsatz konzentriert sich auf die Verwendung von MIDAS-Regressionen zur Prognose von realisierten Kovarianzmatrizen. Diese Prognose ist ein wichtiger Bestandteil des Asset Pricings und des Portfoliomanagements. Die vorgelegte Studie weist auf eine höhere Vorhersagekraft des Midas-Modells im Vergleich zum konventionellen HAR-Modell hin. Der ökonomische Wert der vorgestellten Methode wird anhand von Portfoliomanagement-Anwendungen bewertet.

Der letzte Teil zielt auf die Entwicklung einer rechnerisch effizienten Schätzung von MIDAS Regressionen ab und demonstriert deren Vorteile für Big Data-Anwendungen. Ein vorgeschlagenes überarbeitetes Optimierungsverfahren zur Berechnung von MIDAS-NLS Schätzern führt zu einer wesentlichen Leistungssteigerung in Bezug auf Rechenaufwand und Genauigkeit. Um die Anwendung dieses Ansatzes zu vereinfachen, wurde diese Methode als Toolbox „MIDAS Analytic“ dokumentiert und implementiert.

Contents

List of Figures vii

List of Tables viii

1 Predicting U.S. bank failures with MIDAS logit models (published in the JFQA) 4

1.1 Introduction . . . 5
1.2 Methodology . . . 9
1.2.1 Simple logit model . . . 9
1.2.2 MIDAS weighting scheme . . . 12
1.2.3 MIDAS in a logistic regression framework . . . 16
1.2.4 Reweighting of observations in the log-likelihood function . . . 17
1.2.5 Classification accuracy when classes are imbalanced . . . 17
1.3 Enhancing the default probability model for U.S. banks . . . 19
1.3.1 Bank-level statistics . . . 20
1.3.2 Forecasting procedure design . . . 21
1.3.3 Empirical results . . . 24
1.4 Conclusion . . . 28
Appendix 1.A Analytical Gradient for MIDAS Logit . . . 32
Appendix 1.B Simulation study for the MIDAS logit model . . . 32

2 Inconsistent supervision when banks hide losses: Is it a consequence of regulatory forbearance or misreporting? 34

2.1 Introduction . . . 35
2.2 Data and model . . . 38
2.2.1 Bank failure cases . . . 39
2.2.2 Empirical predictors of bank failure cost . . . 40
2.3 Modeling and forecasting procedure . . . 41
2.4 Estimation results . . . 42
2.4.1 The Lasso model: Full sample and subsamples . . . 42
2.4.2 Out-of-sample forecasting results . . . 45

2.4.3 Fraud or forbearance: Circumstantial evidence . . . 45
2.5 Conclusion . . . 50
Appendix 2.A Candidate predictors of bank failure cost . . . 51
Appendix 2.B Lasso estimation using all observations . . . 53

3 Forecasting realized covariances: a MIDAS approach 55
3.1 Introduction . . . 56
3.2 Estimating the realized covariance matrix from high-frequency data . . . 58
3.3 Modeling the realized covariances . . . 60
3.3.1 HAR model for the realized covariances . . . 60
3.3.2 MIDAS model for the realized covariances . . . 63
3.4 Multiperiod forecasts and forecast evaluation . . . 65
3.4.1 Direct and iterated forecasts . . . 65
3.4.2 Bias correction . . . 66
3.4.3 Forecasting errors . . . 67
3.5 Empirical study . . . 69
3.5.1 Data . . . 69
3.5.2 Forecasting results . . . 69
3.5.3 Portfolio management application . . . 76
3.6 Conclusion . . . 81
Appendix 3.A List of 29 companies . . . 83
Appendix 3.B Estimated models and coefficients . . . 84

4 Estimating MIDAS regressions via MIDAS-NLS with revised optimization 87
4.1 Introduction . . . 88
4.2 Reconsidering the MIDAS-NLS estimation . . . 90
4.2.1 MIDAS regressions . . . 90
4.2.2 Revised approach to MIDAS-NLS . . . 93
4.3 Simulation study . . . 94
4.3.1 Data generating process . . . 94
4.3.2 Performance measures . . . 95
4.3.3 Findings . . . 96
4.3.4 Importance of fast estimation . . . 100
4.4 Volatility forecasting application . . . 101
4.5 Conclusion . . . 103
Appendix 4.A Measure of the computational efficiency . . . 104
Appendix 4.B Sensitivity analysis for profiling . . . 105
Appendix 4.C Three-parameter Beta case . . . 106

5 "MIDAS Analytic" extension to the MIDAS MATLAB Toolbox (software) 107
5.1 Classical univariate MIDAS model 109
5.2 Multivariate MIDAS model 114
5.3 MIDAS logit model 119
5.4 MIDAS VAR model (mixed frequency VAR) 122
5.5 Notations 128
Appendix 5.A Analytic derivatives for the classical univariate MIDAS model . 130
Appendix 5.B Analytic derivatives for the multivariate MIDAS model 143
Appendix 5.C Analytic derivatives for the MIDAS logit model 145
Appendix 5.D Analytic derivatives for the MIDAS VAR model 151

Bibliography 156

Curriculum vitae 165

List of Figures

1.1 Trend in the number of U.S. banks . . . 6
1.2 MIDAS weighting profiles . . . 14
1.3 Estimated MIDAS weighting profiles . . . 15
1.4 Temporal design of MIDAS logit . . . 16
1.5 Risk group in terms of a confusion matrix . . . 19
1.6 Risk group and area under risk group . . . 20
1.7 Forecasting procedure design for U.S. bank failures . . . 23
1.8 Comparison of ROC curves . . . 26

2.1 Formation of negative net worth in banks . . . 36
2.2 NCAR1 as an early warning signal of bank financial distress . . . 48

3.1 MIDAS weighting schemes driven by θ_1 and θ_2 . . . 64
3.2 Four losses to compare the covariance matrix forecasts . . . 72
3.3 Forecasting accuracy for elements of the covariance matrix . . . 74
3.4 Efficient frontiers for the crisis and non-crisis periods . . . 78
3.A.1 Estimated coefficients in the HAR model . . . 84
3.A.2 Averaged coefficients in the MIDAS model: Rolling window of 200 days . . . 86

4.1 Beta MIDAS weights with zero-last-lag ($\theta_3 = 0$) . . . 92
4.2 Relation between a grid step and the computational complexity . . . 100
4.A.1 Sensitivity analysis for the Beta polynomial . . . 105

List of Tables

1.1 A confusion (classification) matrix and its elements . . . 10
1.2 Instance confusion matrix . . . 18
1.3 Descriptive statistics for Cole's factors . . . 21
1.4 Estimated coefficients for "Simple" and "MIDAS" logit models . . . 24
1.5 Classification accuracy results . . . 25
1.6 Statistical significance: t-test . . . 27
1.7 Confusion matrix for McNemar's test . . . 28
1.8 Statistical significance: McNemar's test . . . 29
1.9 Statistical significance: Wilcoxon's test . . . 30
1.A.1 Simulation study results . . . 33

2.1 The number and types of bank failures in the United States . . . 38
2.2 Descriptive statistics for 524 U.S. bank closures in our sample, 2004–2018 . . . 39
2.3 Out-of-sample forecasting with expanding window design using Lasso . . . 46
2.4 Path to failure: Descriptive statistics by quarter . . . 47
2.5 Counterfactual bank closure cost . . . 49
2.A.1 Summary statistics for candidate predictors of bank failure cost . . . 51
2.A.2 Estimation of the Lasso model using bank failures in the period 2004–2018 53

3.1 Empirical bias correction and forecasting accuracy . . . 70
3.2 Forecasting the realized covariance matrices in 2003–2017. . . . 77
3.3 Performance of GMV portfolio investment . . . 79
3.4 Economic value of switching to the MIDAS model, annualized Δ. . . . 81
3.A.1 29 Dow Jones stocks which are used in the empirical study . . . 83
3.A.2 Descriptive statistics for estimated coefficients . . . 85

4.1 Beta lag polynomials . . . 91
4.2 MIDAS-NLS with revised optimization: Optimization by θ_2 . . . 97
4.3 MIDAS-NLS with revised optimization: Optimization by θ_2 and θ_3 . . . 98
4.4 MIDAS-NLS with revised optimization: Optimization by θ_1 and θ_2 . . . 99
4.5 Out-of-sample RV forecasting, MSE loss function . . . 102

4.6 Out-of-sample RV forecasting, QLike loss function 102
4.A.1 Computational efficiency of MIDAS-NLS with revised optimization 104
4.A.2 MIDAS-NLS with revised optimization: Optimization by θ_1, θ_2, and θ_3 . . . 106

Preface

This dissertation titled "Essays on the use of MIDAS regressions in banking and finance" comprises four essays that address prominent forecasting tasks in banking and finance. Aiming to complete the tasks, I take into account the availability of time series sampled at different frequencies, as it can improve the accuracy of predictions. In recent years, econometric models that deal with mixed frequency data have attracted considerable attention. In particular, regression models in which a lower frequency variable of interest is conditioned on the relevant information observed at higher frequencies (mixed data sampling or MIDAS regressions) have gained prominence in academia as well as in the private sector. Notably, MIDAS regressions reveal superior performance in many practical applications due to their parsimony and flexibility, achieving particularly sound results in the fields of Macroeconomics and Financial Econometrics. In this dissertation, most of the suggested models rely on the MIDAS approach.

Chapter 1 is based on a published paper by Audrino, Kostrov and Ortega (2019), written in collaboration with my thesis advisor and co-advisor. The MIDAS approach was applied to study the financial stability of banks. Moreover, the logit model was generalized by introducing a novel MIDAS logit model to improve prediction accuracy in U.S. bank failures. Further, an algorithm has been suggested in order to mitigate the class-imbalance problem that is pervasive across many datasets. The proposed MIDAS logit model, which is applied to the forecasting of U.S. bank failures in the period 2004–2016, yields a strongly significant improvement in classification accuracy, both in statistical and economic terms. Constructing flow financial variables (measured over a period of time) using the MIDAS scheme relaxes the limitation of conventional annual aggregation schemes. The approach that this dissertation proposes is general and seems promising in many applications in addition to forecasting bank failures. Moreover, the discussion regarding the classification accuracy evaluation in the case when one class of observations is severely undersampled will contribute to enhancing existing practices in financial and medical studies.

Chapter 2 is also focused on banks' financial stability and rests on some ideas mentioned in Karminsky and Kostrov (2017), Kostrov and Mamonov (2019), and Chapter 1. This part of my dissertation makes a two-fold contribution. First, it contributes to the knowledge base in the literature about the cost of bank failures in the United States. Using a sample of 524 bank closures during the period from 2004 to 2018, I consider the role played by the primary federal regulators in the formation of costs. So, banks supervised by different agencies failed with unequal closure costs, which were forecast by distinct groups of predictors. This result supports the finding of Agarwal et al. (2014) that U.S. banking regulators implement identical rules differently, i.e., there are inconsistencies in supervision. Second, I attempt to reconsider the phenomenon of regulatory forbearance, according to which delayed regulator's response to bank distress leads to more costly bank failures. It was discovered that financial reports of banks supervised by the two primary federal regulators do not help to predict bank resolution costs. Poor forecasting accuracy calls into question the quality of supervision and indicates low informativeness of the off-site monitoring. For these banks, regulatory forbearance seems to be a minor issue, whereas fraudulent reporting is a major one.

Chapter 3 is motivated by the empirical evidence that MIDAS regressions are instrumental to improve volatility predictions (Ghysels et al., 2006, 2019). Considering a gap in the literature, my joint work with Anastasija Tetereva explains how to improve forecasts of realized covariance matrices using a MIDAS approach. This topic is highly important in risk management, asset pricing, and for contagion analysis in financial markets. A challenge in predicting a covariance matrix is to guarantee the positive semidefiniteness of the forecast matrix. To comply with this requirement, the matrix logarithm transformation was applied in the spirit of Bauer and Vorkink (2011). A total of 29 stocks in the Dow Jones Industrial Average index were used to show that the MIDAS approach is more accurate than the established HAR model, and that it holds good for long and short forecasting horizons. According to Chiriac and Voev (2011) and Symitsi et al. (2018), the HAR model is a valid benchmark in this forecasting set-up. The value of using the MIDAS model was measured in a portfolio application and in terms of investor's utility. Switching to the MIDAS model generates up to 7% of annualized returns depending on the forecasting horizon. A strong result is that the model performs very well when most needed, in crisis. During the Great Recession, an investor would avoid any losses. More than that, the forecasting error is 40% smaller in the turbulent period as compared to the HAR model. A cornerstone of this paper is an efficient estimation procedure because the MIDAS regressions are estimated thousands of times.

Chapter 4 builds upon estimation principles that are formulated for MIDAS regressions in Chapters 1 and 3. The advantages of a MIDAS regression are offered at the expense of computational complexity. Estimation is challenging because a non-linear optimization problem has to be solved, and there is empirical evidence that existing methods are computationally demanding and may produce inaccurate solutions. Despite its importance, the literature on solving this computational burden is rather scarce. MIDAS regressions are typically estimated via nonlinear least squares (NLS) with nice asymptotic properties stated by Andreou et al. (2010). However, Ghysels and Qian (2019) report that the known numerical pitfalls of the MIDAS-NLS estimator make it ill-behaved in some applications. That is why they propose a profiling estimator that would be helpful in such cases. For instance, a profiling estimator is attractive in the context of regression-based volatility forecasting. However, their method inherits certain disadvantages attributed to the grid-search origin of profiling (the estimation is time-consuming). In their recent study, Bollerslev et al. (2018) applied a MIDAS regression to predict volatility. They explained that the rolling window estimation design is not feasible. The same pattern of high computational complexity is evident in multistep forecasting. In this chapter, the numerical drawbacks of the original NLS estimation are mitigated. In MIDAS-NLS with revised optimization, the computational problem to be solved at the time of estimation and the corresponding computations has been vectorized. Our optimization technique relies on analytic derivatives. To promote estimating MIDAS regressions via MIDAS-NLS with revised optimization, I have released the "MIDAS Analytic extension" to the MIDAS MATLAB Toolbox (Ghysels, 2013), which currently supports several MIDAS models. It is particularly beneficial in situations involving Big Data. Software documentation and key derivations are provided in Chapter 5.

Chapter 1

Predicting U.S. bank failures with MIDAS logit models (published in the JFQA)

Francesco Audrino [1], Alexander Kostrov [2], and Juan-Pablo Ortega [3]

[1]University of St. Gallen, Faculty of Mathematics and Statistics, francesco.audrino@unisg.ch
[2]University of St. Gallen, Faculty of Mathematics and Statistics, alexander.kostrov@unisg.ch
[3]University of St. Gallen, Faculty of Mathematics and Statistics; Centre National de la Recherche Scientifique (CNRS), juan-pablo.ortega@unisg.ch

Abstract

We propose a new approach based on a generalization of the logit model to improve prediction accuracy in U.S. bank failures. Mixed-data sampling (MIDAS) is introduced in the context of a logistic regression. We also mitigate the class-imbalance problem in data and adjust the classification accuracy evaluation. In applying the suggested model to the period from 2004 to 2016, we show that it correctly classifies significantly more bank failure cases than the classic logit model, in particular for long-term forecasting horizons. Some of the largest recent bank failures in the United States that had been previously misclassified are now correctly predicted.

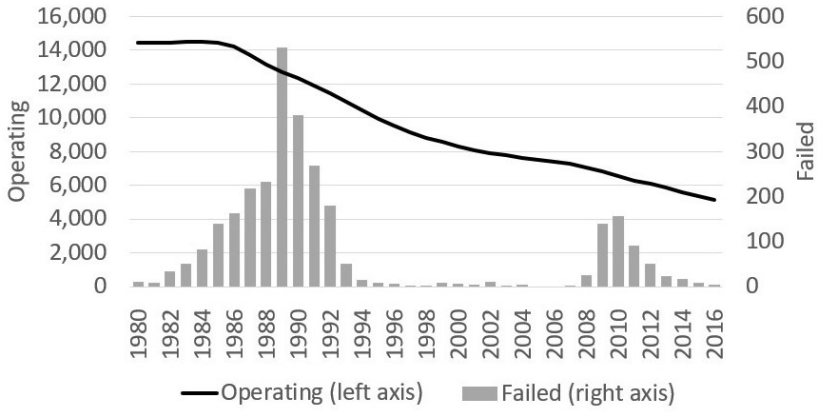
1.1 Introduction

Both regulators and bank counterparties are interested in spotting vulnerable banks and in having accurate models in order to forecast failures of U.S. banks several periods in advance. A huge amount of money could potentially be saved if failures were correctly predicted. Consequently, there is a vast literature studying the determining factors of failures in the U.S. banking sector and whether bank failures are largely predictable. These papers focus on bank failures during two recent banking crises. From 1985 to 1992, almost 2,500 banks left the market due to failure. More recently, about 400 U.S. banks collapsed from 2009 to 2011 (see Figure 1.1). Despite consolidation and waves of failures, about 5,000 banks still operate in the U.S. banking sector, and some players remain comparatively weak. The official Federal Deposit Insurance Corporation (FDIC) confidential Problem Bank List contained 123 and 95 banks at the end of 2016 and 2017, respectively.

One of the most relevant results shown in literature is empirical evidence that the main factors driving bank failures in the U.S. banking system have been quite stable since 1980. In fact, it is well documented that the relevant explanatory variables for bank failures from 1987 to 1992 do a very good job of explaining U.S. bank failures during the recent banking crisis (see, e.g., Aubuchon and Wheelock (2010); Cole and White (2012); Cole et al. (2020),

Figure 1.1 Trend in the number of U.S. banks

Figure 1.1 plots the number of operating and failed commercial banks in the United States, 1980–2016, annually. Data are provided by the Federal Deposit Insurance Corporation (FDIC).



and Mayes and Stremmel (2013)). In particular, Mayes and Stremmel (2013) assume in their research design that the two U.S. banking crises share many similar features. They aggregate failures across these crises and treat them as a group, which requires some homogeneity assumptions. There seems to be a consensus in the banking literature on the set of baseline predictors in the default probability models for U.S. banks. Selected variables characterize a bank with respect to capital, assets, management, earnings, liquidity, and sensitivity to market risk (CAMELS). The continuing use of CAMELS proxies in the literature also reflects the stability of the factors used to forecast banking failures.

Logistic regressions are a well-established, powerful approach to classifying binary outcomes. In applications, there is a specific setting when the model is used to identify a few unhealthy units in a large set of observations, an issue known as the class-imbalance problem. In addition to financial research, data samples with severe class imbalance are typical in medical studies with a few sick people in a mainly healthy population: Li et al. (2010) consider diabetes and liver disorders; Mazurowski et al. (2008), Malof et al. (2012), and Miller et al. (2014) investigate breast cancer cases. In numerous financial papers with class imbalance in their data, weak business units can be distinguished from others, for instance, among banks (Demyanyk and Hasan (2010)) or companies (Agarwal and Taffler (2008)).

It is therefore not surprising that the logistic regression is still the key benchmark model used by both the academic community and in the private sector to identify future bank failures (see, among others, Kolari et al. (2002), Mayes and Stremmel (2013), or Cole et al. (2020)). The main question we raise in this study is whether the logit model can be extended

to more accurately forecast U.S. bank failures. This would not only provide significant improvements in statistical terms but would also result in concrete economic advantages for the whole banking system. In fact, for the benchmark logistic model, bank assets of $103.4 billion are liquidated between 2010 and 2016 because of misclassified failures in forecasting bank failures 8 quarters ahead. In our empirical analysis, we show that by adopting some modifications in a logistic regression framework, up to $4.59 billion (i.e., 9.1% of the total assets of failed banks predicted by the benchmark model) could be saved thanks to an improved, correct identification of failures as much as 2 years in advance. The stability of the most relevant factors driving bank failures over time can be seen as an indication that the improvements in classifying historical bank failures we obtain by modifying the simple logit model are likely to remain in the future.

The new methodological changes that we adopt to improve the classification accuracy of the benchmark logistic regression model can be summarized as follows. First, we introduce a mixed-data sampling (MIDAS) weighting scheme (Ghysels et al. (2007)) for constructing the most relevant explanatory variable(s). This approach is instrumental in constructing more accurate flow predictors. By definition, flow variables are measured with reference to the period of time (not at a point in time). Conventionally, most flow variables in finance are annual, with an equal contribution of the different quarters (e.g., annual return on assets (ROA) or change in bad loans). At the expense of estimating a few extra parameters, the MIDAS weighting scheme relaxes both constraints: relevant periods are automatically selected and the weight given to each quarter is different. Since both constraints are not strongly motivated, MIDAS weighting could significantly improve the fit of the model to the data and its forecasting power. The individual weights obtained in the MIDAS weighting scheme characterize the relationship profile between a dependent variable and past values of an independent covariate. This novel approach is applicable far beyond forecasting bank failures, the central application of the present study.

Second, we address the issue of classification accuracy for the logit model with a severe imbalance of classes in the data. We suggest assessing the classification accuracy of a model using a "risk group" concept instead of standard classification accuracy indicators, which are misleading in this setup.

Third, we explain the use of the MIDAS for the logit model. Since Ghysels et al. (2004) introduced the MIDAS regression as a parsimonious model designed for mixed-frequency data sets, a burgeoning literature on MIDAS has implemented this approach in a large array of practical examples coming from academic research and industrial applications. A growing literature uses MIDAS in the context of volatility and correlation analysis for stock returns.

Ghysels et al. (2005) apply the MIDAS approach to predict realized volatility. Engle et al. (2013) suggest the generalized autoregressive conditional heteroskedasticity-MIDAS model. Dynamic conditional correlation-MIDAS and constant conditional correlation-MIDAS are found to outperform standard benchmark models in Audrino (2014) and Colacito et al. (2011). MIDAS is also instrumental in macroeconomic nowcasting (see, e.g., Foroni and Marcellino (2014) and the Factor-Augmented MIDAS in Marcellino and Schumacher (2010)) and in other general forecasting contexts (see, among others, Ferrara et al. (2014), the Markov-switching MIDAS in Guérin and Marcellino (2013), and the MIDAS-Generalized Autoregressive Score in Gorgi et al. (2019)). Ghysels, Plazzi and Valkanov (2016) rely on MIDAS for the modeling of conditional quantiles. Foroni et al. (2015) introduce MIDAS regressions with unrestricted linear lag polynomials (U-MIDAS), which are expected to be particularly useful when the differences in sampling frequencies of explanatory and explained variables are small. Semiparametric and nonparametric MIDAS regressions are presented in Chen and Ghysels (2011) and Breitung and Roling (2015), respectively. To the best of our knowledge, MIDAS has never been combined with the logit model. Freitag (2016) explicitly claims that a MIDAS logit regression has yet to be introduced. The combined model is nonlinear and optimization requires some effort.

Finally, we implement reweighting in the log-likelihood function to remedy the imbalance of classes in the data. An optimal weight of the rare-class observations (i.e., bank failure cases) in the log-likelihood function is selected by cross-validation. We call the resulting model MIDAS logit with reweighting of observations.

We investigate the accuracy of the new model in a realistic, correctly specified simulation setting generated to mimic as closely as possible the U.S. bank failure data. Results of our simulations support the use of the new methodology based on the MIDAS weighting scheme. In fact, although the data-generating process is nonlinear, the MIDAS logit model is able to recognize the correct values of the parameters and the correct MIDAS weighting structure.

Following the literature, in our empirical analysis, we employ the extended set of CAMELS factors used in Cole and White (2012). The introduction of the MIDAS weighting scheme significantly enhances the out-of-sample classification accuracy of the standard logit model, in particular for long-term forecasting from 6 to 8 quarters ahead. The difference is not large or significant for shorter forecasting horizons. The reweighting of the observations to accommodate the class imbalance problem brings further improvements. In economic terms, the most striking result is the following: Through the use of the new MIDAS logit model, some of the largest recent U.S. bank failures, such as the Eurobank ($2.5 billionn in assets) and Charter Bank ($1.2 billionn in assets) failures in 2010, can be correctly classified

as many as 8 quarters in advance. This is not the case if one follows the predictions of the benchmark logit model. As a consequence, the economic value of correct predictions of bank failures measured in terms of banks' total asset value increases by approximately 3%–9% (depending on the forecasting horizon) due to the use of the MIDAS logit model.

The logit model allows for a straightforward enhancement by introducing extra predictors that are added to a standard set of CAMELS explanatory variables. Numerous papers have investigated the impact of a new factor on the probability of U.S. bank failures (DeYoung and Torna (2013), Jin et al. (2013)). MIDAS weighting is a convenient tool to augment the logit model with additional flow explanatory variables.

The content of the paper can be summarized as follows: In Section 1.2, we introduce the new MIDAS logit model. Section 1.3 summarizes the main empirical findings using the MIDAS logit model in the prediction of U.S. bank failures. Section 1.4 concludes.

1.2 Methodology

1.2.1 Simple logit model

Logit and probit models have been instrumental in predicting U.S. bank failures. Demyanyk and Hasan (2010) review the prediction methods that have historically been used for bank failures. They stress that discriminant analysis was previously the leading technique to forecast bank failures. More recently, it was replaced by maximum likelihood-based methods and machine learning, which allow for more general distributional assumptions.

Although many machine learning methods have been recently introduced to forecast bank performance (Demyanyk and Hasan (2010), Gogas et al. (2018), and Iturriaga and Sanz (2015)), the parametric logit and probit models are still very common and competitive in explaining and predicting bank failures. This is partially due to the availability of well-developed inference techniques for these methods. Using a training data sample, the logit and probit models forecast the probability of a bank failure, which is mapped into the predicted outcomes.

Mayes and Stremmel (2013) estimate the logit model to explain bank failures in the United States for the period from 1992 to 2012. Mayes and Stremmel (Table 1) show that the logit and probit models are very popular methods to explain bank failures before the recent crisis in the United States and globally. They remain common in recent research on the U.S. banking sector. Kolari et al. (2002) employ the logit model and the trait recognition model to study failures among large U.S. banks in the period from 1989 to 1992. Although

Table 1.1 A confusion (classification) matrix and its elements

Table 1.1 describes the construction of a confusion matrix. The number of false positive (FP), false negative (FN), true positive (TP), and true negative (TN) outcomes of classification are reported.

	Actual Outcome	
Predicted Outcome	Operating State ($y = 0$)	Failure State ($y = 1$)
Operating state ($\hat{y} = 0$)	No. of TN	No. of FP (Type I errors)
Failure state ($\hat{y} = 1$)	No. of FN (Type II errors)	No. of TP

the logit model showed inferior out-of-sample performance, both models were successful in classifying in-sample observations (with accuracy rates over 95%). Cole et al. (2020) compare a time-varying hazard model with a static probit model to predict bank failures in the United States for the years 2009 and 2010. They discover that a probit model dominates when the information set is limited to financial data available at the time of prediction. The logit model is superior at longer forecasting horizons (over 1 year ahead) and in the classification of failure type observations. Using the probit model, Cole and White (2012) find that the variable "commercial real estate investments" contains predictive information to explain U.S. bank failures. Kerstein and Kozberg (2013) utilize regulatory enforcement actions as a predictor variable to improve the forecasting performance of the probit model. DeYoung and Torna (2013) examine the structure of incomes from nontraditional banking activities to forecast bank distress in a logistic regression framework. Berger and Bouwman (2013) provide evidence of the effect of capital on banks of different sizes. Lu and Whidbee (2013) identify the causes of bank failures with a logit regression, focusing on the impact of bank age and charter type on solvency. Jin et al. (2011) apply the probit model to exhibit the importance of auditor type and its core specialization to explain bank failures. Using a logistic regression, Jin et al. (2013) provide empirical evidence that banks complying with Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA) internal control are financially more stable.

In the standard logit model, we use a matrix Z_t of explanatory variables measured at time t to forecast bank performance h quarters ahead:

$$\boldsymbol{y}_{t+h} = \Lambda(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta}) + \boldsymbol{\epsilon}_t, \tag{1.1}$$

where $\boldsymbol{y}_{t+h} = (y_{1,t+h}, \ldots, y_{k,t+h})^T$ is a vector of dummy variables denoting whether we observe the failure of an existing bank $i,\ i = 1, \ldots, k$, at period $t+h$; that is,

$$y_{i,t+h} = \begin{cases} 1, & \text{if bank } i \text{ fails in period } t+h, \\ 0, & \text{otherwise.} \end{cases}$$

$\Lambda(\cdot) = \exp(\cdot)/(\boldsymbol{\iota} + \exp(\cdot))$ denotes the vector of logistic function outcomes (with a slight abuse of notation); α is a scalar, $\boldsymbol{\beta}$ is a vector of coefficients, and $\boldsymbol{\epsilon}_t$ denotes a vector of errors. The parameters α and $\boldsymbol{\beta}$ are selected to maximize the (log-)likelihood of observing the failure outcomes $\boldsymbol{y}$:

$$\text{LL}(\alpha, \boldsymbol{\beta}) = \boldsymbol{y}_{t+h}^T(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta}) - \boldsymbol{\iota}^T \ln(\boldsymbol{\iota} + \exp(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta})). \tag{1.2}$$

Once the parameters are estimated, the fitted probabilities can be obtained as

$$\hat{\boldsymbol{p}}_{t+h} = \text{prob}(\boldsymbol{y}_{t+h} = \boldsymbol{\iota} | Z_t) = \Lambda(\hat{\alpha}\boldsymbol{\iota} + Z_t\hat{\boldsymbol{\beta}}). \tag{1.3}$$

The output of the binary choice logit model is a vector of predicted probabilities $\hat{\boldsymbol{p}}_{t+h}$ for the forecasting horizon h of interest. In order to evaluate the accuracy of the classification of the model into the two classes, the probabilities have to be mapped into the binary predicted outcomes (0 or 1). A conventional mapping assigns class "1" to an object i if the predicted probability $\hat{p}_{i,t+h}$ lies above a threshold value μ (for $i = 1, \ldots, k$); that is,

$$\hat{y}_{i,t+h} = \begin{cases} 1, & \text{if } \hat{p}_{i,t+h} > \mu, \\ 0, & \text{otherwise.} \end{cases} \tag{1.4}$$

In practice, the value μ has to be determined by the researcher. In our particular application to bank failures (and crises) classification, typical solutions adopted in the literature are as follows: Mayes and Stremmel (2013) fix μ at some exogenously specified level. In contrast, Demirgüç-Kunt and Detragiache (2000) select μ to optimize a loss function based on classification costs of type I and type II errors. Similarly, Lo Duca and Peltonen (2013) and Sarlin (2013) choose the value of μ that optimizes a loss function based on a usefulness measure in terms of weighted type I and type II error rates, which relates to the gain from using a model compared to naive classification. Finally, Betz et al. (2014) develop the previous approach by introducing bank- and class-specific misclassification costs.

As can be seen from this short literature review, ambiguity in the choice of μ seems to remain, since the costs of type I and type II errors are also unknown. Once μ is fixed, a confusion matrix can be constructed (see Table 1.1) and the classification accuracy indicators can be computed using its elements. The most widely used measures are

$$\text{SPECIFICITY} = \text{TN}/(\text{FP}+\text{TN}),$$
$$\text{ACCURACY} = (\text{TP}+\text{TN})/(\text{TP}+\text{FN}+\text{FP}+\text{TN}), \text{ and}$$
$$\text{SHARE_OF_CORRECTLY_PREDICTED_FAILURES} = \text{TP}/(\text{TP}+\text{FP}).$$

1.2.2 MIDAS weighting scheme

In some applications, we aim at explaining y_t using past values of an explanatory factor x_t^m, where the superscript m means that this variable is sampled at frequency m times higher than the explained variable y_t. For instance, we could consider a relationship between annual gross domestic product (GDP) and monthly statistics on jobless claims ($m = 12$, 12 months in a year). We could include lagged versions of x_t^m as individual regressors to estimate a distributed lag model:

$$\boldsymbol{y}_t = \alpha + \sum_{j=1}^{d} \beta_{t-\frac{j}{m}} \boldsymbol{x}_{t-\frac{j}{m}} + \boldsymbol{e}_t, \tag{1.5}$$

where d is the maximum-order lag parameter. Large d requires many parameters to be estimated. This may lead to an intractable estimation procedure, overfitting, and noisy results.

Alternatively, it is possible to aggregate high-frequency data before estimating a regression equation. Monthly jobless claims could be summed and reported annually. This sort of aggregation assumes that all aggregated periods are equally important to explain the dependent variable, which is not necessarily true. Implicitly, a flat aggregation increases the importance of the selection of the tuning parameter d as well.

MIDAS (mixed-data sampling) is another weighting technique originally developed to explain a low-frequency outcome variable using high-frequency covariates (Ghysels et al. (2004), (2007)). This method allows for an individual weighting of high-frequency increments, reduces the importance of the selection of the maximum-order lag parameter d (which we will also refer to as the MIDAS temporal aggregation period), and keeps the parsimony of a model, hence reducing the danger of overfitting. The MIDAS approach for

the same regression model introduced in equation (1.5) works as follows:

$$\boldsymbol{y}_t = \alpha + \beta \sum_{j=1}^{d} \gamma(j, \theta_1, \theta_2) \boldsymbol{x}_{t-\frac{j}{m}} + \boldsymbol{e}_t, \tag{1.6}$$

where d is the maximum-order lag considered and $\gamma(j, \theta_1, \theta_2) = \exp(\theta_1 j + \theta_2 j^2)$ denotes the exponential Almon (1965) lag function with two parameters, which is common for MIDAS-type models. Dependent on the three parameters β, θ_1, and θ_2, the Almon's scheme is very flexible and allows for different weighting profiles $\gamma(j, \theta_1, \theta_2)$ that characterize the individual weights for the different lags j. MIDAS is parsimonious and has tractable estimation given that it depends only on three parameters regardless of the maximum lag d. Some illustrative weighting profiles are shown in Figure 1.2.

The classical MIDAS approach presented in equation (1.6) employs past values of a high-frequency factor x to predict the low-frequency outcome variable y. In this paper we follow an analog approach.

Let us consider a flow explanatory variable x_t (e.g., quarterly net income normalized by current assets). We want to verify whether a family of lagged versions of x_t $\{x_{t-1}, x_{t-2}, \ldots, x_{t-d}\}$ has some predictive power for the probability of bank failure:

$$x_{t-j} = \frac{\text{QUARTERLY_NET_INCOME}_{t-j}}{\text{ASSETS}_t}, \tag{1.7}$$

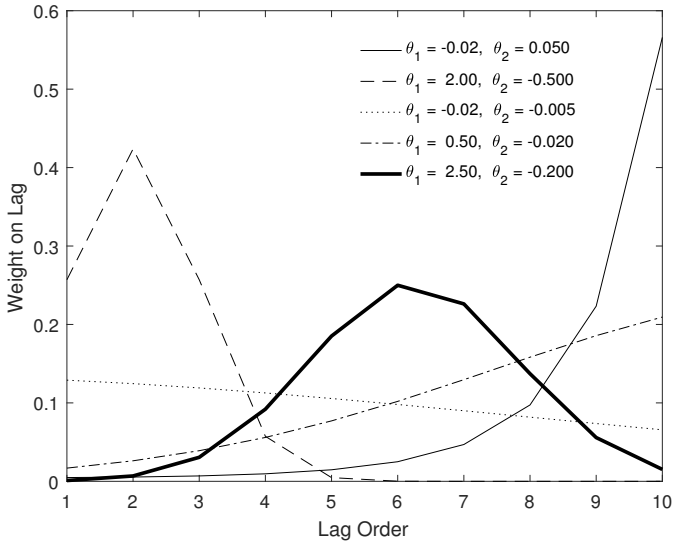
where the maximum-order lag d could be large and unknown. For large d, it is counterproductive to use the past values of x_t as individual regressors because of the high number of parameters involved. Aggregating quarterly variables to annual ones is a standard solution in the financial literature. This procedure generates a new annual variable, x^a, an aggregated version of x:

$$x_t^a = \frac{\sum_{j=1}^{4} \text{QUARTERLY_NET_INCOME}_{t-j}}{\text{ASSETS}_t} \approx \text{ROA}_t. \tag{1.8}$$

Although the annual aggregation is well-established in practice, it has some limitations: It implicitly assumes an equal contribution (weight) of all quarters in the aggregated variable x_t^a and the exogenously given number of quarters (four quarters in the case of annual aggregation). Since both constraints have no true theoretical justification, we could gain performance from relaxing them and improve the fit of the model to the data. Moreover, for the autocorrelated series x_t^a, the use of a flat weighting scheme leads to an aggregation bias (Andreou et al. (2010)), which causes a degradation in the forecasting accuracy. At

Figure 1.2 MIDAS weighting profiles

Figure 1.2 presents mixed-data sampling (MIDAS) weighting profiles driven by θ_1 and θ_2. Weights are normalized to 1.



the limited cost of estimating a few parameters, the MIDAS weighting scheme allows for individual weights for every quarter, which carries in its wake an endogenous and automatic selection of the relevant periods. In other words, the MIDAS logit regression captures the proper mixed-frequency information in a parsimonious way. We therefore construct a MIDAS-weighted variable $\tilde{x}^d$ with a temporal aggregation period d:

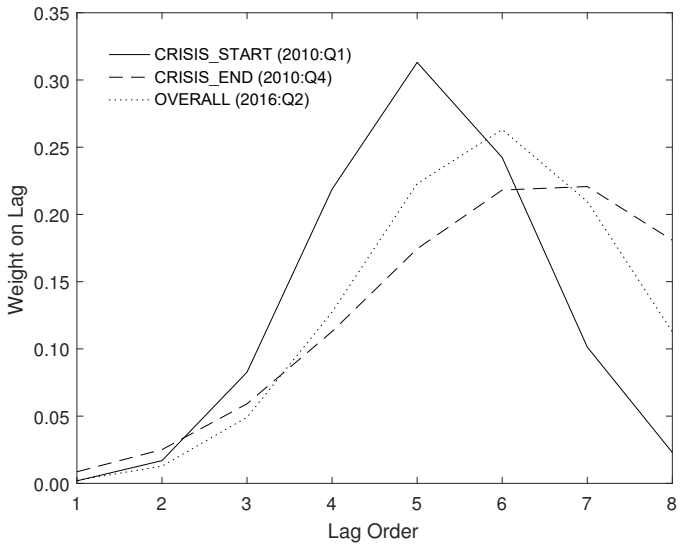
$$\tilde{x}_t^d = \frac{\sum_{j=1}^{d} \gamma_j(\theta_1, \theta_2) \times \text{QUARTERLY_NET_INCOME}_{t-j}}{\text{ASSETS}_t}, \tag{1.9}$$

where the individual weights γ_j characterize the relationship between the dependent variable and the past lagged values of the high-frequency explanatory variable x (quarterly net income in our specific example).

MIDAS weighting can be used in many situations. It is a valuable tool to correct variables based on a profit and loss statement and the corresponding financial ratios such as ROA (i.e., the net income earned during the previous 4 quarters over assets); the cost-to-income ratio (CIR) (i.e., the incurred cost over earned income during the previous 4 quarters; or the net interest margin (NIM) (i.e., the ratio of net interest income earned during the previous 4 quarters to assets). Moreover, MIDAS weighting is also valid for constructing a factor reflecting a change in a stock variable (e.g.„ change in bad loans). It helps to select the optimal weights for the recent quarterly changes and the number of quarters to add.

Figure 1.3 Estimated MIDAS weighting profiles

Figure 1.3 describes the relationship between the probability of bank failure and mixed-data sampling (MIDAS)-weighted earnings normalized by assets. We report the estimated weighting profiles for the selected periods: the first and last crisis quarters and the full sample.



In general, we suggest following these steps to find an optimal change in a stock variable (from a balance sheet):

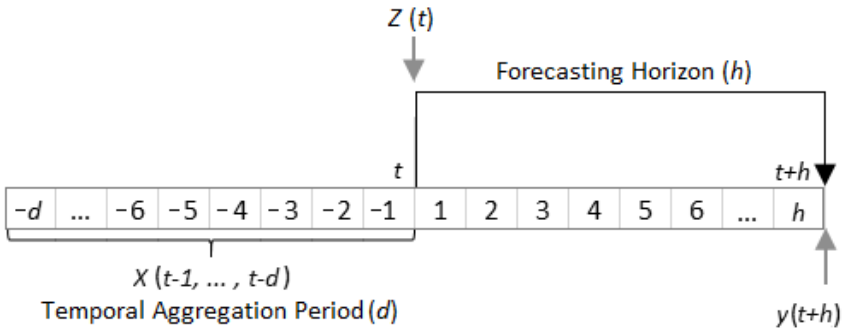
- Assume that the change in x influences y;
- Find the first differences for x; and
- Use a MIDAS-weighted sum of the first differences as a factor in the regression model.

In Figure 1.3, we show the estimated weighting profiles for a ratio of MIDAS-weighted net income to assets for a temporal aggregation period of 8 quarters in the case of predicting U.S. bank failures 2 years in advance in 2010 and 2016.

As can be seen, the estimated weights differ significantly across quarters. Additionally, weights corresponding to lags of more than 1 year seem to be highly crucial to obtain a good forecasting model. The MIDAS weighting profile is quite robust with respect to the estimation periods considered. Our results support the relaxation of the assumptions implicitly assumed by the flat annual aggregation scheme.

Figure 1.4 Temporal design of MIDAS logit

Figure 1.4 explains the temporal design of the mixed-data sampling (MIDAS) logit model (with MIDAS-weighted factor X). Forecasting horizon: h quarters; temporal aggregation period: d quarters.



1.2.3 MIDAS in a logistic regression framework

We have found no attempts to use MIDAS in a logistic regression framework. To the best of our knowledge, only Freitag (2014), (2016) merge the probit model (which is closely related to the logit model) with MIDAS: Weekly credit default swap data are used to predict low-frequency changes in the sovereign ratings of European countries. In our opinion, there are two possible reasons for this. First, the optimization problem that needs to be solved at the time of estimating the model is computationally intensive. Second, studying the relationship between variables sampled at different frequencies is not a standard task for the binary choice model (if we ignore weighting flow variables). As an example, we use MIDAS weighting scheme to correct the annual ROA variable as in equation (1.9). Our modified logit model with MIDAS-weighted part (MIDAS logit) is given by

$$\boldsymbol{y}_{t+h} = \Lambda(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta} + \gamma\tilde{\boldsymbol{x}}_t^d(\theta_1,\theta_2)) + \epsilon_t, \tag{1.10}$$

where we use a MIDAS-weighted factor $\tilde{x}^d$ with a temporal aggregation period d. Figure 1.4 describes the temporal design of our regression. We note that one could generalize model (1.10) by adding multiple MIDAS-weighted variables to it.

The parameters of the MIDAS logit model are estimated by maximizing the corresponding log-likelihood function with respect to the parameters α, $\boldsymbol{\beta}$, γ, and $\boldsymbol{\theta}$:

$$\mathrm{LL}(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\theta}) = \boldsymbol{y}_{t+h}^T(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta} + \gamma\tilde{\boldsymbol{x}}_t^d(\boldsymbol{\theta})) - \boldsymbol{\iota}^T\ln(\boldsymbol{\iota} + \exp(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta} + \gamma\tilde{\boldsymbol{x}}_t^d(\boldsymbol{\theta}))). \tag{1.11}$$

The analytic derivation of a gradient for the target function is provided in Appendix 1.A (for a single MIDAS-weighted covariate).

1.2.4 Reweighting of observations in the log-likelihood function

Class-imbalance is a well-known problem in statistical analysis. Techniques to smooth out the problem using adjusted indicators have been suggested in the literature (Hu and Dong (2014), Longadge et al. (2013), and Menon et al. (2013)). When we receive only a small number of observations for a rare class (failure cases), the log-likelihood maximizer may neglect the rare-class observations and focus only on optimizing the fit for the observations belonging to the large class. Over-sampling the underrepresented class is a standard technique for overcoming the imbalance of classes (e.g., García et al. (2013), Japkowicz et al. (2000)). One can implement this idea by increasing the weight of rare-class observations in the log-likelihood function:

$$\begin{aligned}&\mathrm{LL}(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\theta}) = \\ &\left(\boldsymbol{y}_{t+h}^T \odot \boldsymbol{w}^T\right)(\alpha\boldsymbol{\iota}+Z_t\boldsymbol{\beta}+\gamma\tilde{\boldsymbol{x}}_t^d(\boldsymbol{\theta})) - \left(\boldsymbol{\iota}^T \odot \boldsymbol{w}^T\right)\ln(\boldsymbol{\iota}+\exp(\alpha\boldsymbol{\iota}+Z_t\boldsymbol{\beta}+\gamma\tilde{\boldsymbol{x}}_t^d(\boldsymbol{\theta}))),\end{aligned} \quad (1.12)$$

where $\boldsymbol{w}$ is a vector of weights. By default, $\boldsymbol{w}=\boldsymbol{\iota}$. Reweighting means that for every observation i in the sample,

$$w_i = \begin{cases} m>1, & \text{if } y_i = 1, \\ 1, & \text{otherwise.}\end{cases}$$

The weight multiplier m is selected from a finite set of candidate values to maximize the cross-validated classification accuracy.

1.2.5 Classification accuracy when classes are imbalanced

Evaluation classification accuracy is important for comparing classifiers. Sun et al. (2007) show that standard indicators are not valid when the imbalance of classes is present in the data.

The following example explains the nature of the problem: Consider 7,150 classified units (e.g., banks) with 150 observations for a rare class (e.g., bank failure). Table 1.2 presents a sample confusion matrix. In this simple example, we suspect 1,100 units to belong to class "1" (to fail), but the guess is wrong in 1,000 cases (type II errors). Moreover, 50 banks are not

Table 1.2 Instance confusion matrix

Table 2 is a confusion matrix for an example with 7,150 classified units (150 units belong to a rare class "1").

	Actual Outcome	
Predicted Outcome	Operating State ($y = 0$)	Failure State ($y = 1$)
Operating state ($\hat{y} = 0$)	6,000	50 (Type I errors)
Failure state ($\hat{y} = 1$)	1,000 (Type II errors)	100

suspected but, nevertheless, fail (type I errors). Although the standard measures introduced at the end of Section 1.2.1 indicate a general high accuracy of the classifiers, the number of false negatives is large compared to the sample size:

$$\text{SPECIFICITY} = 6{,}000/(50+6{,}000) = 0.99, \text{ almost perfect,}$$
$$\text{ACCURACY} = (100+6{,}000)/7{,}150 = 0.85, \text{ very good,}$$
$$\text{SHARE_OF_CORRECTLY_PREDICTED_FAILURES} = 100/(100+50) = 0.67, \text{ good.}$$

In both medical and financial studies, a potential user of a model is interested in predicting a "risk group" of observations with the highest probability of failing or being infected. Those units would receive a specific treatment, consideration, or deeper inspection. Once a user defines the size of a risk group (e.g., 5% of units), a threshold μ in equation (1.4) is automatically selected. There is no longer any need to set μ using some advanced approach.

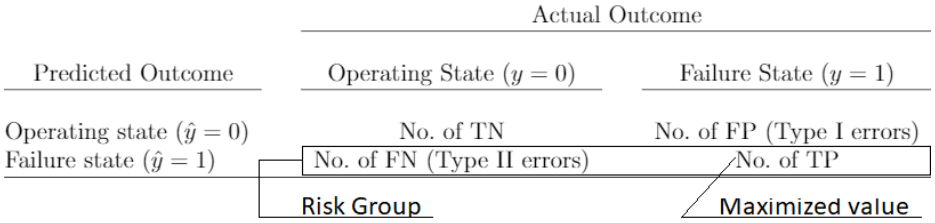
To evaluate the accuracy of the classification, we compute the share (or the number) of actual failures that appear to be in the risk group at a moment of a failure. In terms of a confusion matrix, we maximize the number of true positive outcomes for a risk group of a fixed size (see Figure 1.5).

We define a risk group as the α fraction of observations with the highest predicted probability to appear in class "1." The classification accuracy indicator "risk group of size α" (RG_α) shows the share (or the number) of actual failures captured by a risk group of size α. Cole et al. (2020) and Karminsky and Kostrov (2017) use similar classification accuracy indicators for a 5% risk group.

A second measure of classification accuracy reflects the distribution of failures in a risk group. For a risk group of a selected size α, the detection of more actual failure cases among the observations with the highest predicted probability of failing is a signal of a

Figure 1.5 Risk group in terms of a confusion matrix

In Figure 1.5, there is a risk group of a fixed size: false negative (FN) + true positive (TP) = const. For a model with the best classification accuracy, TP or (equivalently) TP/(FN + TP) is maximized.



better classification performance. Therefore, we compute the area under the curve denoted by S_{shaded}; in Figure 1.6 we show an illustrative example for the case $\alpha = 5\%$. The ratio of S_{shaded} to the full rectangular area measures the accuracy of the classification in a risk group. This indicator, called "area under RG_α" (AURG_α), is close in spirit to the area under the ROC curve (AUC), a popular measure of classification accuracy for binary choice models:

$$\text{AURG}_\alpha = \frac{S_{\text{shaded}}}{100\% \times \alpha}. \tag{1.13}$$

AURG_α takes values between 0 and 1. Higher values of the ratio correspond to better classification performances.

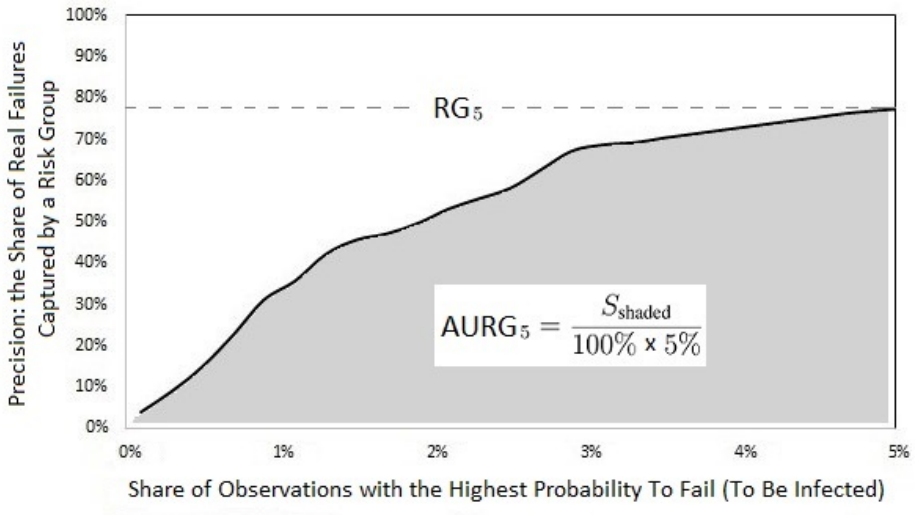
We believe that the suggested measures of classification accuracy RG_α and AURG_α for a risk group size α such as 5% are clearly preferable to alternatives in our setting. In fact, they have the following advantages: First, they are intuitive and straightforward and allow for a sensible interpretation in financial and medical studies. Second, they come with an automatic and interpretable selection of the threshold value μ, where both type I and type II errors are implicitly considered. Third, they are robust to the class-imbalance problem. Finally, they can be computed universally across classification methods.

1.3 Enhancing the default probability model for U.S. banks

In this section, we improve a logit model that predicts U.S. bank failures from 2010 to the second quarter of 2016 (2016:Q2). In the baseline logit model, we employ the extended set of 15 CAMELS predictors suggested by Cole and White (2012) (Cole's factors) to forecast bank performance h quarters ahead. First, we apply MIDAS weighting to correct the annual ROA variable as in equation (1.9) and estimate the MIDAS logit model (1.10). A simulation study contained in Appendix 1.B confirms that MIDAS logit is able to find the correct pattern in data. In line with the literature, we use up to 4 years ($16 = 8 + 8$ quarters) of

Figure 1.6 Risk group and area under risk group

Figure 1.6 displays a curve "share of real failures captured by a risk group – share of observations with the highest probability to fail" for a risk group of 5%. AURG_5 characterizes the distribution of failures in a risk group.



information before a failure event to forecast it in our setting. In a second step, we also use the reweighting of the observations to fix the class-imbalance problem.

1.3.1 Bank-level statistics

To replicate the Cole's factors, we use the bank-specific financial characteristics coming from call reports on FDIC-insured banks from 2004 to 2016:Q2. These reports are disseminated by FDIC in quarterly sets of files with detailed banking statistics. Most bank characteristics can be found in files called "Assets and Liabilities," "Net Loans and Leases," and "Past Due and Nonaccrual Assets." Our sample includes 50 quarters of data for 9,936 unique banks. Table 1.3 contains descriptive statistics for the replicated Cole's factors.

The information on the 525 failure cases in the investigated period is documented in the FDIC's Failed Bank list. A model with Cole's factors is used to predict bank failures for the out-of-sample period from 2010 to the second quarter of 2016.

The common practice in the previous literature is to define extra criteria for bank failures. For example, Cole and White (2012) introduce a "technical failure" defined as a weak financial position:

$$\text{EQUITY} + \text{RESERVES} - 0.5 \times \text{NONPERFORMING_ASSETS} < 0. \tag{1.14}$$

Table 1.3 Descriptive statistics for Cole's factors

Table 1.3 contains descriptive statistics for replicated Cole's factors. The sample for 2004–2016:Q2 contains 374,822 observations without missing values for operating banks and 525 failure cases. Variables are reported as a decimal fraction of total assets (ROA and ln(ASSETS) are exceptions).

			Operating: 374,822		Failed: 525	
	Variable	Explanation	Mean	Std. Dev.	Mean	Std. Dev.
1	EQ	Bank equity capital	0.117	0.074	0.031	0.027
2	LNATRES	Loan loss allowance	0.010	0.007	0.030	0.017
3	ROA	Return on assets (ROA), %	0.798	3.544	-5.702	5.548
4	NPA	Nonperforming assets	0.010	0.012	0.034	0.028
5	SC	Total securities	0.219	0.160	0.101	0.084
6	BRO	Brokered deposits	0.028	0.154	0.100	0.142
7	ln(ASSETS)	Natural logarithm of total assets	12.022	1.378	12.437	1.380
8	CHBAL	Cash and balances due from depository institutions	0.073	0.081	0.099	0.074
9	INTAN	Goodwill and other intangibles	0.005	0.021	0.002	0.008
10	LNRERES	1–4 family residential mortgages	0.197	0.153	0.182	0.143
11	LNREMULT	Real estate multifamily residential mortgages	0.017	0.037	0.031	0.046
12	LNRECONS	Construction and development loans	0.054	0.070	0.146	0.121
13	LNRENRES	Commercial real estate nonresidential mortgages	0.152	0.115	0.228	0.123
14	LNCI	Commercial and industrial loans	0.085	0.073	0.078	0.068
15	LNCON	Loans to individuals	0.043	0.068	0.015	0.023

Similarly, Wheelock and Wilson (2000) register a bank failure when

$$\frac{\text{EQUITY} - \text{GOODWILL}}{\text{ASSETS}} < 2\%. \tag{1.15}$$

Many erroneous predictions occur when weak banks are predicted to fail but go on operating (type II errors). Such modifications artificially improve the accuracy of classification, removing the barrier between failed and weak institutions. For these reasons, we do not use any extra criteria to define a bank failure and focus only on the actual ones.

1.3.2 Forecasting procedure design

When the goal of a given exercise is forecasting, obviously only information available at the time of forecasting must be used. Looking into the future should be avoided since such forecasts are not feasible in practice. Cole and Wu (2020, p. 4) point out such concerns: "In practice, future bank financial data are not available at any given point of time, so regulators must rely upon what data actually are available, without 'peaking' at future data,

as academics have done." Mayes and Stremmel (2013) confirm that the use of unattainable information confers an unfair advantage in a forecasting exercise.

For the logit and probit models, the horizon of forecasting is embedded into a regression design: Shifted factors measured at time t are used to predict bank performance at time $t+h$ (h - shift size and forecasting horizon). Papers that attempt to explain and predict bank failures in the United States have used different forecasting horizons, in most cases of between 1 quarter and 2 years (Bologna (2015), Cole and White (2012), and Kerstein and Kozberg (2013)). In practice, some potential users of the model, such as bank regulators, are concerned about identifying bank failures well in advance. For this reason, Cole et al. (2020) constructed a model with a forecasting horizon of 3 years. According to Jordan et al. (2010), the 4 years of data prior to a financial distress contain meaningful information to predict it.

We report and compare the classification accuracy for U.S. bank failure predictions obtained using the following competing models:

- "Simple": simple logit with 15 Cole's factors in Z:

$$\boldsymbol{y}_{t+h} = \Lambda(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta}) + \epsilon_t.$$

- "MIDAS": simple logit augmented with a MIDAS-weighted predictor that is a ratio of weighted quarterly net income to assets:

$$\boldsymbol{y}_{t+h} = \Lambda(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta} + \gamma\tilde{\boldsymbol{x}}_t^d(\theta_1, \theta_2)) + \epsilon_t.$$

  The "MIDAS" logit model requires the estimation of only three extra parameters $(\gamma, \theta_1, \theta_2)$ relative to the "Simple" logit.

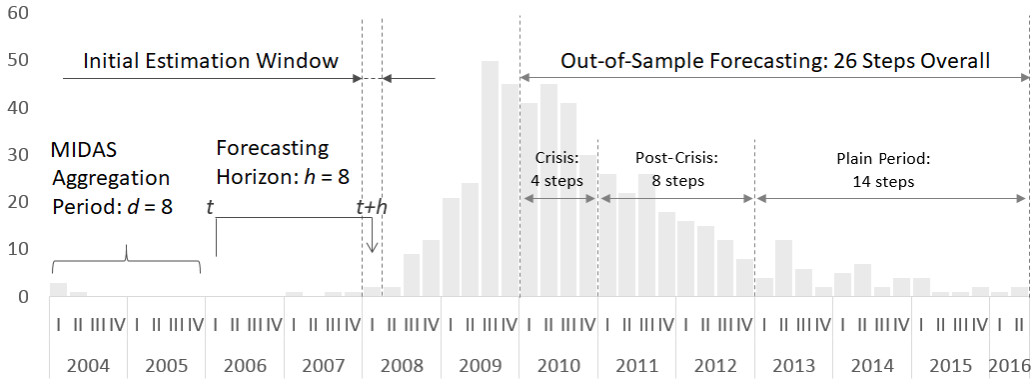
- "MIDAS_REW": the MIDAS logit model in which we also explicitly consider the class-imbalance problem by reweighting the observations in a log-likelihood function.

It is worth mentioning that in settings like ours involving quarterly data, the U-MIDAS model by Foroni et al. (2015) is an attractive option. In that case, instead of the three parameters in the MIDAS-weighted part, one would need to estimate d unrestricted parameters. Unfortunately, we found in our setup that the accuracy of the U-MIDAS logit model is inferior for both simulated and real data. The results of these simulations are available upon request.

In this paper, we apply direct multistep forecasting (out-of-sample) for the period from 2010 to the second quarter of 2016 (Figure 1.7). The forecasting horizon h is taken to be

Figure 1.7 Forecasting procedure design for U.S. bank failures

There are 525 cases of bank failures in the United States in 2004–2016:Q2. Figure 1.7 displays the example for the forecasting horizon h of 8 quarters and temporal aggregation period d of 8 quarters. We expand the estimation window in every step, starting from the first quarter of 2008 as the initial sample.



2, 4, 6, or 8 quarters. The temporal aggregation period d is fixed at 8 quarters. In every step, the parameters of the model are re-estimated. The forecasting procedure ensures no looking into the future. We maximize the log-likelihood function for the MIDAS logit model in MATLAB using the "fminunc" optimizer with a "trust-region" algorithm and the gradient explicitly derived for the target function. The accuracy of the optimization result is confirmed using the profiling estimation procedure by Ghysels and Qian (2019) with a small grid step. The algorithm brings the estimation close to the standard logit by profiling the MIDAS weighting scheme parameters (grid search). The use of profiling does not improve optimization results or forecasting accuracy.

As Figure 1.7 describes, our data sample enforces the use of the first quarter of 2008 as the earliest estimation window available in forecasting bank failures 8 quarters ahead. Failure frequency is very low for prior periods. Consequently, in the first forecasting step, we predict bank failures in the first quarter of 2010. This explains the choice of the out-of-sample forecasting period, which is split into three parts: a crisis period (2010, 4 quarters), a post-crisis period (years 2011 and 2012, 8 quarters), and a plain period (from 2013 to the second quarter of 2016, 14 quarters).

We use 10-fold cross-validation to select the optimal weighting parameter m in the estimation of MIDAS_REW equation (1.12) (i.e., the MIDAS logit model with reweighting of observations): In particular, we split the pooled estimation sample into 10 random, equal-sized, nonintersecting subsamples with 10% $(\frac{1}{10})$ of the observations from each class. We then train the model on nine of the subsamples and use the last one for validation with

Table 1.4 Estimated coefficients for "Simple" and "MIDAS" logit models

In Table 1.4, the full data set is used as the estimation sample. γ, θ_1, and θ_2 are extra coefficients estimated for the mixed-data sampling (MIDAS)-weighted factor. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variables are as defined in Table 1.3.

	Forecasting Horizon							
	$h=2$		$h=4$		$h=6$		$h=8$	
Variable	Simple	MIDAS	Simple	MIDAS	Simple	MIDAS	Simple	MIDAS
EQ	-83.80***	-89.54***	-57.57***	-60.99***	-39.31***	-44.23***	-24.21***	-28.44***
LNATRES	31.99***	15.32***	46.12***	24.11***	47.79***	33.47***	40.17***	20.67***
ROA	0.01***	-0.15***	0.05***	-0.13***	0.05***	-0.07***	0.05***	-0.10***
NPA	13.76***	12.57***	18.89***	18.71***	18.42***	18.13***	19.01***	18.58***
SC	-5.25***	-5.25***	-3.62***	-3.35***	-2.64***	-2.69***	-2.17***	-2.13***
BRO	1.72***	1.60***	0.15**	0.16**	0.14**	0.14**	0.13*	0.13*
ln(ASSETS)	0.065***	0.116***	0.038***	0.056***	0.003	0.009**	0.006*	0.005
CHBAL	-4.25***	-4.76***	-3.90***	-4.29***	-4.03***	-5.69***	-4.78***	-5.19***
INTAN	43.59***	46.37***	31.78***	36.20***	26.56***	31.09***	18.44***	21.70***
LNRERES	-2.56***	-2.78***	-0.97***	-1.02***	0.17	-0.10	0.35*	0.06
LNREMULT	1.69*	2.34**	2.31***	2.78***	3.60***	3.61***	4.14***	4.14***
LNRECONS	3.41***	3.41***	6.16***	6.46***	7.37***	7.42***	7.70***	7.90***
LNRENRES	-2.62***	-2.53***	-1.12***	-0.95***	0.38**	0.33*	1.18***	1.08***
LNCI	-2.35***	-1.96***	-0.69	-0.17	0.66	0.68	0.94**	0.88**
LNCON	-14.03***	-14.46***	-13.95***	-13.37***	-10.95***	-10.79***	-9.80***	-9.67***
Intercept	0.38***	0.05***	-1.87***	-1.79***	-3.68***	-3.03***	-5.11***	-4.38***
γ		6.99		5.03		3.34		-7.32
θ_1		0.09		0.11		-0.08		2.33
θ_2		-0.05		-0.05		-0.001		-0.20

respect to the RG_5 performance measure. Each of the 10 subsamples is used exactly once as a validation set. We finally average the 10 results for the performance measure RG_5 in order to obtain a single cross-validated value. We repeat this procedure for every integer value of m between 1 and 150. Then, we select as the optimal value of m the one that yields the highest cross-validated RG_5. Initially, we started with a wider set of candidates. However, values above 150 were never selected as optimal.

1.3.3 Empirical results

The use of MIDAS does not alter significantly the relevance of the coefficients of the Cole's factors (Table 1.4). The only exception is the coefficient of the ROA variable that is directly affected by the MIDAS weighting scheme of the net income lags in the MIDAS logit model.

Table 1.5 presents the classification accuracy results of the competing models. Naturally, the forecasting power of the models decreases when increasing the forecasting horizon. Not surprisingly, the models exhibit a lower accuracy in crisis times compared to post-crisis and good periods for the economy.

Table 1.5 Classification accuracy results

Table 1.5 reports the out-of-sample subperiods: crisis, post-crisis, and plain times. The number of bank failures to predict is reported for every subperiod and forecasting horizon. Performance measures: RG_5 defined as a share and number (in parentheses) of correctly predicted failures, $AURG_5$ – area under RG_5, and area under the ROC curve (AUC).

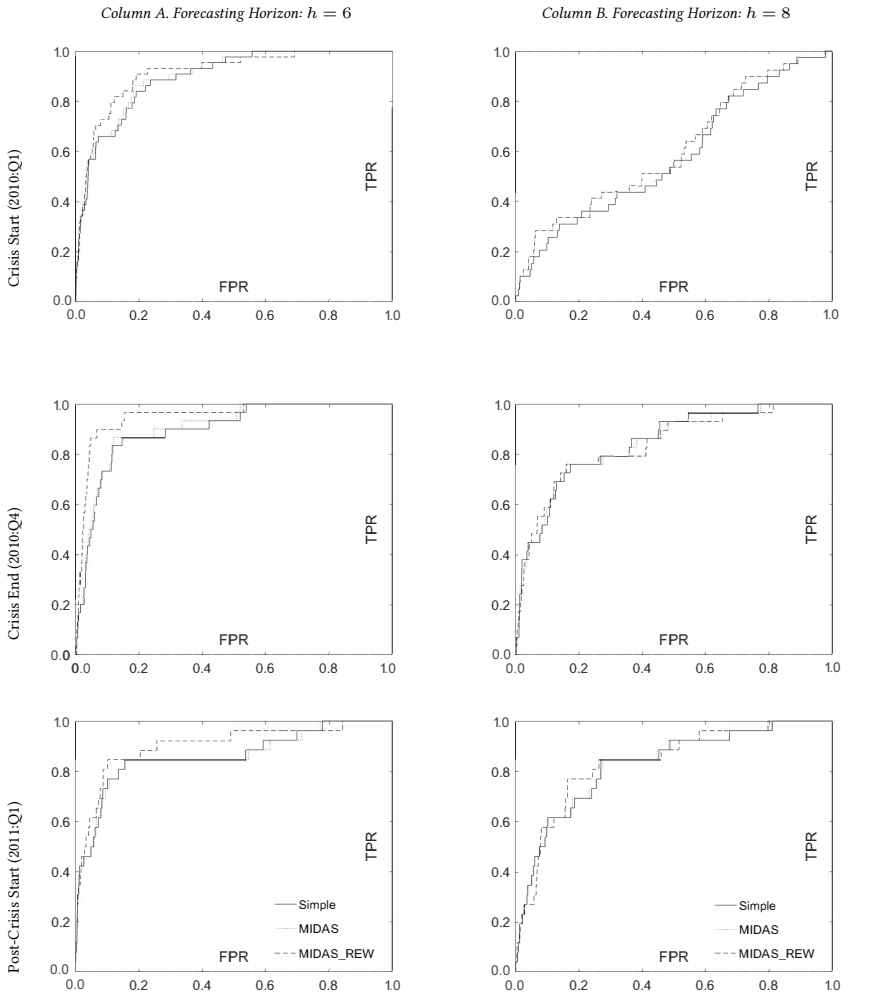
	Overall			Crisis			Post-Crisis			Plain		
	(2010–2016:Q2: 26 quarters)			(2010: 4 quarters)			(2011–2012: 8 quarters)			(2013–2016:Q2: 14 quarters)		
Model	RG_5	$AURG_5$	AUC	RG_5	$AURG_5$	AUC	RG_5	$AURG_5$	AUC	RG_5	$AURG_5$	AUC
Panel A. Forecasting Horizon: $h = 2$												
Failures		353			157			143			53	
Simple	0.972 (343)	0.888	0.991	0.955 (150)	0.852	0.983	0.986 (141)	0.909	0.992	0.981 (52)	0.935	0.993
MIDAS	0.972 (343)	0.893	0.992	0.955 (150)	0.850	0.984	0.986 (141)	0.921	0.994	0.981 (52)	0.943	0.993
MIDAS_REW	0.972 (343)	0.867	0.993	0.955 (150)	0.806	0.982	0.986 (141)	0.905	0.993	0.981 (52)	0.949	0.997
Panel B. Forecasting Horizon: $h = 4$												
Failures		353			157			143			53	
Simple	0.884 (312)	0.683	0.975	0.822 (129)	0.611	0.954	0.930 (133)	0.709	0.976	0.943 (50)	0.826	0.980
MIDAS	0.887 (313)	0.689	0.976	0.828 (130)	0.609	0.955	0.923 (132)	0.719	0.977	0.962 (51)	0.850	0.981
MIDAS_REW	0.887 (313)	0.651	0.980	0.803 (126)	0.553	0.958	0.951 (136)	0.683	0.979	0.962 (51)	0.851	0.988
Panel C. Forecasting Horizon: $h = 6$												
Failures		351			155			143			53	
Simple	0.593 (208)	0.411	0.943	0.497 (77)	0.321	0.888	0.608 (87)	0.413	0.934	0.830 (44)	0.667	0.963
MIDAS	0.604 (212)	0.425	0.945	0.503 (78)	0.331	0.895	0.608 (87)	0.415	0.935	0.887 (47)	0.728	0.965
MIDAS_REW	0.684 (240)	0.453	0.959	0.594 (92)	0.365	0.916	0.699 (100)	0.442	0.957	0.906 (48)	0.743	0.973
Panel D. Forecasting Horizon: $h = 8$												
Failures		349			153			143			53	
Simple	0.424 (148)	0.273	0.879	0.314 (48)	0.188	0.772	0.434 (62)	0.269	0.862	0.717 (38)	0.527	0.919
MIDAS	0.441 (154)	0.281	0.899	0.327 (50)	0.196	0.782	0.448 (64)	0.270	0.869	0.755 (40)	0.558	0.950
MIDAS_REW	0.456 (159)	0.296	0.913	0.32 (49)	0.195	0.782	0.476 (68)	0.298	0.899	0.792 (42)	0.580	0.959

At long forecasting horizons of 6 and 8 quarters, the MIDAS logit model is superior to the Simple logit; reweighting the observations to accommodate the class imbalance problem brings some further improvement. The suggested modifications in the Simple logit model are useful when most needed: in long-term forecasting for crisis and post-crisis times. The same pattern holds for ROC curves (Figure 1.8). At short forecasting horizons, MIDAS weighting neither improves nor deteriorates the forecasting results. Moreover, the MIDAS logit model attains a better distribution of failure cases in a risk group ($AURG_5$). When the bank failure is approaching and the forecasting horizon is short, the historical dynamics of the quarterly net income variables no longer seem to matter. In this case, the most recent Cole's factors already contain all the valuable information.

We use statistical tests to prove whether the discovered differences in the classification performance are significant. Japkowicz and Shah (2011) suggest both parametric and nonparametric tests to compare the performance of two classifiers on a data sample. The parametric t-test verifies whether the difference in the mean classification results is

Figure 1.8 Comparison of ROC curves

Figure 1.8 compares ROC curves in forecasting bank failures 6 and 8 quarters ahead for 3 selected periods: the first and the last crisis periods, the first post-crisis period. Solid, dotted, and dashed lines report the performance of the Simple, mixed-data sampling (MIDAS), and MIDAS_REW models, respectively.



statistically relevant. The statistic reads as

$$t = \frac{\bar{d} - 0}{\frac{\bar{\sigma}_d}{\sqrt{n}}} = \frac{\overline{\mathrm{PM}}(f_1) - \overline{\mathrm{PM}}(f_2)}{\frac{\bar{\sigma}_d}{\sqrt{n}}} \sim t_{n-1}, \tag{1.16}$$

where $\mathrm{PM}(f_i)$ is the performance measure of the classification algorithm f_i, $\bar{d} = \overline{\mathrm{PM}}(f_1) - \overline{\mathrm{PM}}(f_2)$ is the difference in means of the performance measures for the two classifiers f_1

Table 1.6 Statistical significance: t-test

Table 1.6 reports the difference in mean classification results for 26 forecasting steps. Performance measures: RG_5 defined as a share of correctly predicted failures, $AURG_5$ – area under RG_5, and area under the ROC curve (AUC).

Compared Models	Significant Improvement	RG_5		$AURG_5$		AUC	
		t-Stat.	p-Value	t-Stat.	p-Value	t-Stat.	p-Value
Panel A. Forecasting Horizon: $h = 2$							
MIDAS vs. Simple	✓	–	–	3.07	0.006	2.30	0.033
MIDAS_REW vs. MIDAS		0.21	0.386	-1.97	0.060	0.75	0.296
MIDAS_REW vs. Simple		0.21	0.386	-0.27	0.380	1.14	0.205
Panel B. Forecasting Horizon: $h = 4$							
MIDAS vs. Simple		0.88	0.266	2.01	0.057	1.98	0.059
MIDAS_REW vs. MIDAS		0.95	0.250	1.69	0.097	1.35	0.159
MIDAS_REW vs. Simple		1.36	0.157	2.52	0.021	1.48	0.132
Panel C. Forecasting Horizon: $h = 6$							
MIDAS vs. Simple	✓	1.60	0.112	2.65	0.016	0.98	0.242
MIDAS_REW vs. MIDAS	✓	2.54	0.020	1.92	0.066	3.40	0.003
MIDAS_REW vs. Simple	✓	3.17	0.005	2.79	0.012	4.05	0.001
Panel D. Forecasting Horizon: $h = 8$							
MIDAS vs. Simple	✓	2.11	0.046	2.01	0.057	2.05	0.053
MIDAS_REW vs. MIDAS	✓	1.74	0.090	1.69	0.097	3.28	0.004
MIDAS_REW vs. Simple	✓	2.57	0.019	2.52	0.021	2.77	0.012

and f_2, $\bar{\sigma}_d$ stands for the sample standard deviation of mean difference $\bar{d}$, and n denotes the sample size.

Results of the t-tests are summarized in Table 1.6. As expected, the improvements in the forecasting power of the MIDAS logit model over the benchmark Simple logit model are found to be significant for the longer forecasting horizons (6 and 8 quarters). At short forecasting horizons, the number of bank failures in a risk group (RG_5) remains unchanged, although the distribution of cases in a risk group improves slightly.

Nonparametric tests are particularly useful when parametric assumptions are not met. The McNemar's test is performed to compare the classification errors of two classifiers. Since we are mainly interested in classifying bank failures, we construct the McNemar's contingency table for the bank failure cases, as illustrated in Table 1.7.

To test the null hypothesis that two classifiers have the same performance, we compute the χ^2_{Mc} statistic,

$$\chi^2_{\text{Mc}} = \frac{(|c_{01} - c_{10}| - 1)^2}{c_{01} + c_{10}} \sim \chi^2_{1,1-\alpha}, \tag{1.17}$$

Table 1.7 Confusion matrix for McNemar's test

Table 1.7 is a confusion (classification) matrix for McNemar's test and its elements. Predictions for bank failure cases only are divided into four groups.

	Classifier f_2	
Classifier f_1	Incorrect (0)	Correct (1)
Incorrect (0)	c_{00}	c_{01}
Correct (1)	c_{10}	c_{11}

and compare it with the respective critical value at the significance level of interest. Results are shown in Table 1.8. Once again, at long forecasting horizons, the improvement due to the use of the MIDAS logit model is statistically significant.

Moreover, the application of the novel approach to predict bank failures in 2 years increases the assets of correctly classified failed banks by 9.1%. The MIDAS model makes it possible to forecast failure cases for a few large banks previously missed. Two of them, Eurobank and Charter Bank, with asset values of $2.5 billion and $1.2 billion, respectively, are among the largest bank failures in the U.S. banking sector in the current millennium. At the forecasting horizon of 6 quarters, the reweighted MIDAS logit model correctly predicts 40 extra bank failures at the cost of only 8 new errors.

As a final nonparametric test, we investigate the results of the Wilcoxon's test applied to the RG_5, $AURG_5$, and AUC performance measures. The idea behind the test is that a given classifier outperforms an alternative method when most of its records are more accurate than those obtained using the alternative approach; where they are less accurate, the difference should be small.

As shown in Table 1.9, the results of the Wilcoxon's test confirm that, for a long forecasting horizon, the use of the innovations included in the MIDAS logit model significantly improves the classification accuracy of the Simple logit model.

1.4 Conclusion

In this paper, we propose several ways to improve the classification accuracy of a logit model. MIDAS weighting is proposed to construct flow explanatory variables in the regression analysis. Conventionally, firm-specific flow predictors in financial research measure company performance during the previous year, with equal contributions from each quar-

Table 1.8 Statistical significance: McNemar's test

Table 1.8 reports the statistical significance of improvements in the classification results for pairs of models, McNemar's test with statistics χ^2_{Mc}. Classification results for failed banks c_{00}, c_{01}, c_{10}, and c_{11} are defined in Table 1.7. GAIN shows the advantage of using an improved model compared to a benchmark model in terms of assets (ASSETS) of failed banks that are correctly classified.

	MIDAS vs. Simple	MIDAS_REW vs. MIDAS	MIDAS_REW vs. Simple
Panel A. Forecasting Horizon: $h = 2$			
c_{10}	0	1	1
c_{01}	0	1	1
χ^2_{Mc}	—	—	—
Panel B. Forecasting Horizon: $h = 4$			
$c_{1,0}$	2	6	6
$c_{0,1}$	1	6	5
χ^2_{Mc}	—	0.08 (0.773)	0.00 (1.000)
Panel C. Forecasting Horizon: $h = 6$			
$c_{1,0}$	5	36	40
$c_{0,1}$	1	8	8
χ^2_{Mc}	1.50 (0.221)	16.57 (0.000)	20.02 (0.000)
GAIN	—	\$2.48 bn (+2.7%)	\$2.77 bn (+3.1%)
Panel D. Forecasting Horizon: $h = 8$			
$c_{1,0}$	8	14	21
$c_{0,1}$	2	9	10
χ^2_{Mc}	2.50 (0.114)	0.70 (0.404)	3.23 (0.072)
$\text{ASSETS}_{c_{1,1}}$	\$49.40 bn	\$51.91 bn	\$47.39 bn
$\text{ASSETS}_{c_{1,0}}$	\$4.68 bn	\$2.87 bn	\$7.39 bn
$\text{ASSETS}_{c_{0,1}}$	\$0.79 bn	\$2.16 bn	\$2.80 bn
GAIN	\$3.89 bn (+7.7%)	\$0.70 bn (+1.3%)	\$4.59 bn (+9.1%)

Table 1.9 Statistical significance: Wilcoxon's test

The statistical significance of classification results: Wilcoxon's test, pair-wise comparison of the models. There is a significant improvement in classification results when the W-statistic is smaller than the corresponding critical value ($W_{\text{crit},5\%}$).

	RG_5		$AURG_5$		AUC		
	W-Stat.	$W_{\text{crit},5\%}$	W-Stat.	$W_{\text{crit},5\%}$	W-Stat.	$W_{\text{crit},5\%}$	Signific. Improv.
Panel A. Forecasting Horizon: $h = 2$							
Simple vs. MIDAS	—		21	47	71	76	✓
MIDAS_REW vs. MIDAS	—		138.5	53	213	84	
MIDAS_REW vs. Simple	—		107.5	60	116	76	
Panel B. Forecasting Horizon: $h = 4$							
Simple vs. MIDAS	—		50	53	100	84	
MIDAS_REW vs. MIDAS	8	3	181.5	75	161	84	
MIDAS_REW vs. Simple	4	2	153	83	157	84	
Panel C. Forecasting Horizon: $h = 6$							
Simple vs. MIDAS	2	2	17	67	53	84	✓
MIDAS_REW vs. MIDAS	1	10	72.5	67	54	84	✓
MIDAS_REW vs. Simple	2	21	57	67	20	84	✓
Panel D. Forecasting Horizon: $h = 8$							
Simple vs. MIDAS	0	0	86.5	75	27	84	✓
MIDAS_REW vs. MIDAS	8	8	114	91	35	69	✓
MIDAS_REW vs. Simple	4	10	94	110	24	84	✓

ter. Our approach allows for individual weights for past values. The optimal temporal aggregation period can be selected endogenously from the data.

We combine the MIDAS weighting scheme with a logistic regression to improve its forecasting power. MIDAS weighting enhances the forecasting power of an established logit model (Cole and White (2012)) for U.S. bank failures during the period from 2004 to the second quarter of 2016. We augment the reference model with one MIDAS-weighted flow variable (which requires three extra parameters to be estimated). Moreover, we implement a cross-validation procedure to minimize the consequences of class imbalances in the data. Improvements in classification accuracy are found to be statistically significant at the forecasting horizons of 6 and 8 quarters. t-tests, McNemar's tests, and Wilcoxon's tests unanimously support this conclusion. In economic terms, the use of the proposed modifications of the classic logit model enables the correct prediction of a few important bank failures that had been previously misclassified.

We also discuss the problem of measuring the predictive performance of classifiers when the classes of observations are highly unbalanced in the data, a pattern that is typical in many financial and medical studies. Standard accuracy indicators are found to provide generally low information. For this reason, we introduce a risk group approach for classification accuracy evaluation that is based on alternative indicators.

Most of the changes we introduce in this study are not restricted to the considered application of forecasting U.S. bank failures. The MIDAS logit model we propose is very general and can be applied in a variety of situations such as disease diagnosis, solvency evaluation, fraud detection, customer churn prediction, and job-market analysis.

Appendices

Appendix 1.A Analytical Gradient for MIDAS Logit

This section of Audrino, Kostrov and Ortega (2019) has been simplified and extended. The most recent version is provided in Chapter 5 (Sections 5.3, 5.5, and 5.C)

Appendix 1.B Simulation study for the MIDAS logit model

We verify that the MIDAS logit model is able to find the true pattern in a correctly specified simulated data-generating process as follows:

$$\boldsymbol{y}_{\boldsymbol{t+h}}^{\text{sim}} = \Lambda(\alpha^{\text{sim}}\boldsymbol{\iota} + Z_t^{\text{sim}}\boldsymbol{\beta}^{\text{sim}} + \gamma\tilde{\boldsymbol{x}}_t^{d,\text{sim}}(\theta_1^{\text{sim}}, \theta_2^{\text{sim}})) + \epsilon_t.$$

We generate multivariate random normal variables Z^{sim} and X^{sim} to replicate 15 Cole's factors and $d = 8$ past values of a MIDAS-weighted variable. Means and variance-covariance matrices for Z^{sim} and X^{sim} are obtained from the real data application presented in Section 1.3. The coefficients α^{sim}, $\boldsymbol{\beta}^{\text{sim}}$, and $(\gamma^{\text{sim}}, \theta_1^{\text{sim}}, \theta_2^{\text{sim}})$ take the values estimated by the MIDAS logit model when applied to U.S. bank failures with a forecasting horizon of $h = 2$ years (see Table 1.4): $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and $(\hat{\gamma}, \hat{\theta}_1, \hat{\theta}_2)$. Once the coefficients and the covariances are set, we compute the fitted simulated probabilities and map them into the binary responses using a Bernoulli distribution. Notably, the class imbalance problem remains in the simulated data: the share of class "1" observations (bank failures) is about 0.008.

We simulate 1,000 data samples and report the estimation results for Cole's factors (in Z^{sim}) and for the MIDAS part. We consider data samples consisting of 10,000 to 200,000 observations. As Table 1.A.1 shows, the estimated coefficients are close to the true values we used to simulate the data. Moreover, the variance of the different estimators decreases as the sample size increases. Although the model is nonlinear, MIDAS logit is able to identify the correct structure from the data.

Table 1.A.1 Simulation study results

Estimation results for 1,000 simulated data samples. Data samples of 10,000, 50,000 100,000, and 200,000 observations are considered. We report coefficients for 15 Cole's factors and an intercept (Panel A). Panels B and C contain coefficients and weights related to the mixed-data sampling (MIDAS)-weighted factor. DGP stands for data-generating process

		Sim: 10,000		Sim: 50,000		Sim: 100,000		Sim: 200,000	
Variable	DGP Coeff.	Coeff.	Std. Dev.	Coeff.	Std. Dev.	Coeff.	Std. Dev.	Coeff.	Std. Dev.
Panel A. 15 Simulated Cole's Explanatory Variables and an Intercept									
Var1	-28.44	-28.99	2.71	-28.66	1.06	-28.46	0.76	-28.48	0.53
Var2	20.67	21.32	21.45	19.96	9.66	20.91	6.77	20.63	4.90
Var3	-0.10	-0.60	8.30	-0.09	3.43	-0.22	2.53	-0.13	1.73
Var4	18.58	19.14	10.73	18.86	5.08	18.55	3.10	18.70	2.40
Var5	-2.13	-2.19	1.34	-2.14	0.55	-2.15	0.39	-2.12	0.27
Var6	0.13	0.21	0.81	0.13	0.36	0.14	0.27	0.13	0.19
Var7	0.01	0.00	0.11	0.01	0.05	0.01	0.03	0.01	0.02
Var8	-5.19	-5.41	2.02	-5.22	0.86	-5.24	0.59	-5.22	0.42
Var9	21.70	21.84	6.58	21.89	2.93	21.70	2.00	21.65	1.32
Var10	0.06	0.05	1.20	0.06	0.50	0.07	0.36	0.06	0.25
Var11	4.14	4.26	3.94	4.27	1.76	4.16	1.20	4.19	0.89
Var12	7.90	8.00	2.14	7.91	0.88	7.91	0.63	7.90	0.45
Var13	1.08	1.10	1.61	1.11	0.66	1.07	0.45	1.08	0.31
Var14	0.88	1.04	2.21	0.94	0.93	0.89	0.69	0.91	0.50
Var15	-9.67	-9.92	2.32	-9.75	1.04	-9.71	0.67	-9.66	0.48
Intercept	-4.38	-4.53	1.52	-4.42	0.64	-4.39	0.45	-4.39	0.30
Panel B. MIDAS Part									
γ	-7.32	-7.70	6.17	-7.62	2.34	-7.42	1.90	-7.46	1.48
θ_1	2.33	2.44	2.54	2.41	1.08	2.43	0.86	2.35	0.54
θ_2	-0.20	-0.38	1.07	-0.22	0.19	-0.23	0.31	-0.21	0.19
Panel C. Weight of Past Values Assumed by the MIDAS Part (γ, θ_1, and θ_2)									
w_1	0.00	0.08	0.19	0.01	0.06	0.01	0.06	0.01	0.04
w_2	0.01	0.06	0.11	0.02	0.05	0.02	0.05	0.02	0.03
w_3	0.05	0.06	0.08	0.06	0.04	0.06	0.05	0.05	0.03
w_4	0.13	0.10	0.08	0.13	0.04	0.13	0.03	0.13	0.02
w_5	0.22	0.16	0.09	0.21	0.04	0.22	0.03	0.22	0.03
w_6	0.26	0.20	0.11	0.25	0.05	0.26	0.04	0.26	0.03
w_7	0.21	0.18	0.10	0.20	0.05	0.21	0.04	0.21	0.03
w_8	0.11	0.15	0.20	0.11	0.06	0.11	0.03	0.12	0.06

Chapter 2

Inconsistent supervision when banks hide losses: Is it a consequence of regulatory forbearance or misreporting?

Alexander Kostrov [1]

[1]University of St. Gallen, Faculty of Mathematics and Statistics, alexander.kostrov@unisg.ch

Abstract

In this paper, we have attempted to forecast the cost of U.S. bank closures between 2004 and 2018 using financial ratios that are based on banks' accounting reports. Using Lasso with 50 candidate predictors, we discovered significant differences in the accuracy of forecasting and selected predictors among the groups of banks split by the primary federal regulator. Banks supervised by the Federal Reserve (FED) and the Federal Deposit Insurance Corporation (FDIC) are found to fail with large and predictable costs and these cases are associated with regulatory forbearance. This pattern is not valid for banks under the authority of the Office of the Comptroller of the Currency (OCC) and the Office of Thrift Supervision (OTS): Highly likely, their financial statements are fraudulent which creates an opaque regulatory environment, questions the quality of supervision, and limits the benefits of the off-site monitoring.

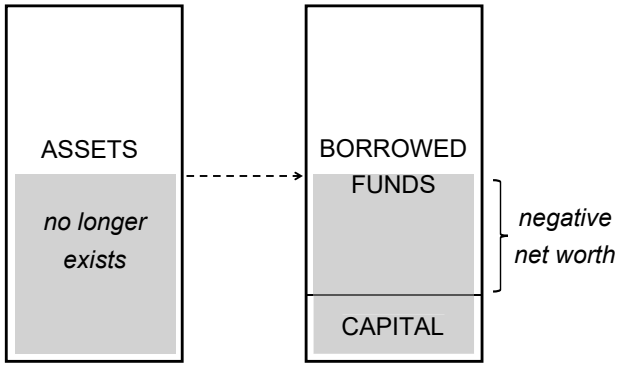
2.1 Introduction

Bank capital has to satisfy official regulatory requirements at all times. However, during periods of financial turbulence, adverse shocks may destruct a substantial part of the assets' value, wipe out bank capital, and initiate a haircut of borrowed funds (Figure 2.1).

McKinsey market review titled "Hidden in plain sight: The hunt for banking capital" was published in the midst of the Great Recession (McKinsey & Company, 2009). This article points out a substantial depletion of bank capital worldwide, underlines "that we hear so little about how to manage capital," and explains the principles of thrifty capital management in banks. However, the lesson has not been learnt. Even ten years later, the problem remains severe: Zombie banks and financial institutions with concealed losses are a threat to the global economy (The Economist, 2019). These banks essentially operate with negative net worth, which is likely to bear significant risks for society. Potential monetary costs are associated with expenses to the government budget and the deposit insurance

Figure 2.1 Formation of negative net worth in banks

According to the fundamental accounting equation, Assets = Borrowed funds + Capital. A considerable loss of assets makes the capital negative. Banks have incentives to hide it from financial authorities in order to stay in business.



system. In turn, possible non-monetary costs originate from decreased availability of bank credit (Kang et al., 2015) and reduced economic activity (Chodorow-Reich, 2014).

In our empirical analysis, we use a sample of 524 failure cases in 2004–2018. We estimate that closed banks report a pre-failure capital-to-assets ratio of $+1.34\%$ on average, whereas the post-failure investigations reveal that losses exceed the capital and expose borrowed funds to a 18.45% reduction in value (Table 2.1). Our results are close to the figures that are reported by Cole and White (2017) and Balla et al. (2015). In general, large losses leading to capital depletion can be treated as a consequence of banks' fraud and regulatory forbearance. The U.S. banking system is subject to a significantly large problem of bank misreporting: James (1991) associated artificially positive levels of capital in weak banks with their attempts to stay in business; Berger and DeYoung (1997) blamed faulty decisions and "bad luck" of bank managers; and Flanagan and Purnanandam (2020) considered the principal-agent problem as a reason for hiding losses and over-reporting net income. At the same time, banking supervisors are prone to forbearance which implies that they do not close distressed banks earlier than they actually do. This might be attributed to political pressure (Brown and Dinç, 2011), the "too-many-to-fail" effect (Acharya and Yorulmazer, 2007), and missing the opportunity to sell weak banks (Bennett and Unal, 2014). A rich body of literature on financial stability has aimed to find the determinants of bank failure and to predict its probability. A review of default probability models can be found in, among others, DeYoung and Torna (2013) and Audrino, Kostrov and Ortega (2019). Fewer papers have studied the costs to society caused by bank failures and attempted to explain (rather

than forecast) the cost of bank closure. Kostrov and Mamonov (2019) surveyed relevant publications.

This paper investigates factors that allow to predict the cost of bank failures. This topic has not been rigorously studied in the prior research. Moreover, Agarwal et al. (2014) have previously shown that U.S. banking regulators implement identical rules differently, which leads to inconsistencies in supervision. Following Agarwal et al. (2014), we have focused on the role of supervisory environment in the formation of losses. We believe that four federal regulators in U.S. banking, namely, the Federal Reserve (FED), the Federal Deposit Insurance Corporation (FDIC), the Office of the Comptroller of the Currency (OCC), and the Office of Thrift Supervision (OTS) supervise banks differently, although they try to synchronize supervision (Labonte, 2020). This hypothesis is supported by our findings and by the empirical evidence that the OTS was considered as a lax regulator and was eventually dissolved in 2011 for poor performance during the Great Recession. In addition, we show that the use of dummy variables for the regulators in the regression analysis as some papers have proposed (e.g., Cole and White, 2017; Lu and Whidbee, 2013; Chernykh and Cole, 2015) is not sufficient to capture the diversity.

There are specific reasons to study whether the cost of bank closure is predictable. First, it is an essential component in the standard credit risk framework, which can be adjusted to evaluate the expected loss of a bank failure ($\$EL$). In order to compute the $\$EL$ value, three components should be known:

$$\$EL = PD * \$EAD * LGD, \tag{2.1}$$

where PD is the probability of default ranging from 0 to 1, which can be derived from the bank's credit rating or is an output of a default probability model; $\$EAD$ stands for the exposure at default and coincides with the amount of bank's borrowed funds; and LGD is the loss given default, that is the portion of borrowed funds which will not be paid back to bank creditors in case of failure:

$$\text{Bank closure cost} = LGD = \frac{\text{Negative net worth}}{\text{Borrowed funds}}. \tag{2.2}$$

Apparently, LGD takes values between 0 and 1. Hereinafter, we use the terms "bank closure cost" and "loss in bank failure" as synonyms to LGD in the context of credit risk modeling.

Second, poor predictability of losses using banks' financial statements would call into question the quality of supervision in banking and would indicate low informativeness of the off-site monitoring by regulators. This is particularly important when turbulent times

Table 2.1 The number and types of bank failures in the United States

Table 2.1 describes the types of U.S. bank failures during the period from 2004 to 2018. Cases are documented in the Failed Bank list which is maintained by the FDIC.

Type	Description	Historically: 548 cases	In our sample: 524 cases
Category 1: Assistance. Charter survives			
A/A	Assistance is provided to an acquirer, who purchases the entire institution.	13	–
Category 2: Failure. Charter is withdrawn, a successor is determined			
PA	The insured and uninsured deposits, certain other liabilities and a portion of the assets are sold to an acquirer.	493	489
PI	The insured deposits only, certain other liabilities and a portion of the assets are sold to an acquirer	15	10
Category 3: Failure. Charter is withdrawn, liquidation			
PO	Payout. The insurer pays the depositors directly and places the assets in a liquidating receivership.	27	25

overlap with periods of delayed reporting by banks (as it happened during the COVID-19 outbreak in 2020): The uncertainty surges and requires corresponding regulatory actions. Moreover, low forecasting accuracy of the bank closure cost would imply the substantial ambiguity of the regulator's environment at the time of decision making. Finally, this evidence would contribute to the ongoing discussion on regulatory forbearance in banking. Therefore, delayed regulator's response to bank distress (i.e., regulatory forbearance) might occur mainly due to bank's misreporting and is not intentional for the financial authorities.

2.2 Data and model

In this section, we first define what is a bank failure, report the sources of data, and proceed to the description of valid predictors of the bank closure costs. Then, we explain the regression framework that we use in order to predict losses associated with bank failures.

Table 2.2 Descriptive statistics for 524 U.S. bank closures in our sample, 2004–2018

		Pre-failure Equity-to-Assets ratio, %				Cost of failure (loss), % of borrowed funds			
	Obs	Mean	SD	min	max	Mean	SD	min	max
Split 1: Primary Federal Regulator									
FED	54	1.48	2.53	-11.75	6.13	18.19	7.81	0	34.36
FDIC	329	1.17	2.91	-13.51	14.96	19.66	8.27	0	42.99
OCC	92	1.94	3.08	-5.26	12.66	16.32	8.07	0.32	35.21
OTS	49	1.25	5.22	-12.13	24.05	14.68	10.59	0	40.77
ANOVA test:		F-stat: 1.47, p-value: 0.221				F-stat: 7.49, p-value: 0.000			
Split 2: Size									
Big	157	1.77	2.70	-6.48	7.97	16.52	9.06	0	39.46
Small	367	1.16	3.37	-13.51	24.05	19.28	8.25	0	42.99
ANOVA test:		F-stat: 4.00, p-value: 0.046				F-stat: 11.57, p-value: 0.001			
Full sample without splitting									
All banks	524	1.34	3.19	-13.51	24.05	18.45	8.59	0	42.99

Note: The ANOVA analysis is used to determine whether any of the means in subsamples is significantly different from others. We are testing the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$, where k is the number of subsamples.

2.2.1 Bank failure cases

Within the period from 2004 to 2018, 548 banks failed in the United States. These cases have been documented in the FDIC's Failed Bank list[2]. Table 2.1 classifies the bank failures and shows that we excluded "Assistance" events because they are treated in a significantly different manner than other failures and removed seven extra cases due to a problem of missing data. Our sample thus includes 60 quarters of data with 524 bank failure cases. The FDIC's Failed Bank list also contains the bank closure cost estimated by the FDIC for each case which we regard as the public cost associated with bank failure. Table 2.2 shows that the banks under FED and FDIC supervision fail with relatively large losses as compared to failed banks with the OCC and the OTS as the primary regulator. Importantly, this difference is not explained by the evidence that big banks fail with relatively smaller losses.

[2]For details about the Failed Bank list (https://banks.data.fdic.gov/explore/failures) we refer the reader to the "Failed Banks Help" (https://banks.data.fdic.gov/explore/failures/help).

2.2.2 Empirical predictors of bank failure cost

Bank-specific financial variables are constructed using the Consolidated Reports of Condition and Income, generally known as Call Reports. FDIC organizes the collection and dissemination of Call Reports on a quarterly basis in the form of Excel files with detailed banking statistics[3]. Table 2.A.1 in Appendix 2.A reports the descriptive statistics for five groups of candidate predictors.

Kostrov and Mamonov (2019) proposed the asset-liability *mismatch* as an explanatory variable for losses in bank failures. Mismatch for bank i at time t is defined as the product of two ratios: Non-monitored deposits to assets (brokered deposits, bro) and commercial & industrial loans to assets (lnci):

$$\text{mismatch}_{i,t} = \text{bro}_{i,t} \times \text{lnci}_{i,t}. \tag{2.3}$$

In practice, asset-liability mismatch may occur due to several reasons. One of them is captive banking: Owners of a non-financial business create a bank to finance their projects at favorable conditions. Since brokered deposits are weakly monitored, captive banks may prefer funding with brokered deposits. Other driving forces that may encourage mismatching include tough competition in product markets and dumping strategies. Kostrov and Mamonov (2019) show that banks with higher mismatch have larger closure costs and justify this finding using a theoretical framework. This result holds in a number of robustness checks.

A list of candidate predictors provided in Table 2.A.1 (Appendix 2.A) is the union of variables from the literature on losses related to bank failures which is enriched by the FDIC Performance and Conditions ratios (Panel 4 of Table 2.A.1) and a few additional variables (Panel 5 of Table 2.A.1) to describe bank's security portfolio. James (1991) analyzed the loss of asset value in failed banks. His study proposed explanatory variables that characterize bank's asset structure. The research following James (1991) developed the analysis in several directions. One example is Schaeck (2008) who considered the differences between high- and low-cost failures and brought forward the set of covariates emphasizing the importance of bank's liability structure (Panel 3 of Table 2.A.1). In another relevant paper, Cole and White (2017) found that the components of bank's non-performing assets and reserves were informative in explaining bank failure costs (Panel 2 of Table 2.A.1).

[3]These data are available for download on the FDIC's website: https://www.fdic.gov/regulations/resources/call/.

2.3 Modeling and forecasting procedure

We apply Lasso as a model selection and prediction device (Tibshirani, 1996) for bank closure costs. A set of 50 possible predictors of losses is considered, which implies a possibility of 50^2 model specifications. Lasso is a computationally efficient method that is designed to select predictors in such multivariate set-ups. Formally, we aim at predicting the closure cost related to bank's i failure at time t using its most recent reported financials:

$$\text{Cost}_{i,t} = \beta_0 + \sum_{k=1}^{50} \beta_k x_{k,i,t-1} + \varepsilon_{i,t}, \tag{2.4}$$

where $\text{Cost}_{i,t}$ is the resolution cost for each failed bank in our sample: The left-hand side of (2.4) is the vector with 524 entries; $x_{k,i,t-1}$ for $k \in \{1, 2, \ldots, 50\}$ are candidate predictors reflecting the bank's financial position at the end of a quarter preceding the failure; $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_5 0]$ is the vector of coefficients that has to be estimated; and $\varepsilon_{i,t}$ is the error term. After standardizing the variables, the coefficients are estimated by solving the following optimization problem with L_1 penalty on the magnitude of the coefficients:

$$\hat{\boldsymbol{\beta}}^{\text{Lasso}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \left[\sum_{i=1}^{524} \left(\text{Cost}_{i,t} - \beta_0 - \sum_{k=1}^{50} \beta_k x_{k,i,t-1} \right)^2 + \lambda \sum_{k=1}^{50} |\beta_k| \right], \tag{2.5}$$

for a positive shrinkage parameter λ with several options for choosing it. In this paper, we selected λ using the leave-one-out cross-validation procedure based on mean squared error (MSE). This choice is motivated by our focus on out-of-sample forecasting in this paper. The Lasso procedure does estimation and variable selection in one step: A sparse solution is produced in which irrelevant coefficients are set exactly to zero.

The forecasting performance of our approach is compared to the model-free forecasting based on the historical average bank closure cost. Another benchmark is the established model suggested by Cole and White (2017):

$$\begin{aligned} \text{Cost}_{i,t} = {} & \beta_0 + \beta_1 \text{size}_{i,t-1} + \beta_2 \text{eq}_{i,t-1} + \beta_3 \text{lnatres}_{i,t-1} + \beta_4 \text{ore}_{i,t-1} + \beta_5 \text{p3asset}_{i,t-1} \\ & + \beta_6 \text{p9asset}_{i,t-1} + \beta_7 \text{naasset}_{i,t-1} + \beta_8 \text{d_OCC} + \beta_9 \text{d_OTS} + \beta_{10} \text{d_FED} + \varepsilon_{i,t}. \end{aligned} \tag{2.6}$$

Description of the explanatory variables in (2.6) are given in Table 2.A.1 (Panel 2) in Appendix 2.A. First, the Lasso model is estimated using the full sample of bank failure cases. Then, using an expanding window regression, we analyze the accuracy of forecasting the bank closure costs. The out-of-sample forecasting horizon is fixed to one quarter. The starting window spans from the first quarter of the year 2004 (2004:Q1) to 2009:Q2 and

includes 58 bank failures. On each of 38 steps of the out-of-sample forecasting (for the quarters from 2009:Q3 to 2018:Q4), we predict the cost of bank closures in the next period. Finally, we report the standard measure of forecasting accuracy, namely, MSE error.

2.4 Estimation results

In this section, we present the estimation results for the Lasso model and compare its forecasting accuracy to the benchmark methods.

2.4.1 The Lasso model: Full sample and subsamples

In the full sample of 524 failure cases, as many as 20 variables survived the variable selection procedure. Cost of failure in subsamples of banks split by the primary federal regulator and bank size is driven by different groups of predictors. Next, we provide the economic interpretation for factors in the final models and comment on the expected correlation with the cost of bank closure.

The description of the predictors is as follows:

- size: *Logarithm of total assets.* Larger banks operate in a stricter regulatory environment and on average fail with smaller losses. We expect a negative correlation of this predictor with bank failure costs.
- mismatch: *Asset-liability mismatch.* Kostrov and Mamonov (2019) discover that higher mismatch leads to costly bank failures. Our expectations are the same.
- lnci: *Share of commercial and industrial loans in assets.* Schaeck (2008) refers to the evidence that banks concentrating on commercial loans are prone to fail with larger losses and confirms this statement.
- TA_growth: *Growth rate of total assets.* High asset value growth rates may reflect attempts of distressed banks to artificially improve the measures of asset quality. Thus, we expect that failures of banks with higher TA_growth are more expensive.
- eqr: *Capital adequacy ratio.* Capital depletion implies that the capital adequacy ratio is small unless there is fraudulent financial reporting. Therefore, we expect a negative correlation of eqr with the cost of bank resolution.
- frepp: *Share of federal funds purchased and repurchase agreements in assets.* We expect that banks with an easy access to cost-efficient borrowing are relatively more healthy.

- lnatres: *Share of loan loss allowance in assets.* Large loan loss reserves reflect the accumulation of risks in a bank, bad risk-management, and aggressive market strategy and lead to considerable losses.

- nclnls: *Share of net charge-offs to loans in assets.* Loans are charged off when a loan repayment is highly unlikely: The credit risk has realized. High charges-offs reported might be the tip of the iceberg of toxic assets.

- bkprem: *Share of bank premises and fixed assets in assets.* Fixed assets do not generate income *per se* and require maintenance. Although some amount of fixed assets is essential for banks, we expect that the effect of bkprem on bank failure cost is positive.

- p3asset and p9asset: *Share of assets past due 30–89 days in assets* and *Share of assets past due 90 or more days in assets*, respectively. These two variables are the components of non-performing assets which are likely to be written off (Cole and White, 2017). We expect that both variables are positively correlated with bank resolution cost.

- ddt: *Share of demand deposits in assets.* This is another source of cheap funding for banks. Although demand deposits can be withdrawn at any time, we expect a negative impact on bank closure cost.

- ts: *Share of total time and savings deposits in assets.* Time and savings deposits are regarded as a reliable (core) funding source in traditional banking. More than that, large ts in a bank confirms high level of clients' trust. However, Schaeck (2008) reports positive correlation of this variable with bank closure cost. Our expectations are mixed.

- scmtgbk and idsccmo: *Share of mortgage-backed securities in assets* and *Share of collaterized mortgage obligations in assets*, respectively. Mortgage-backed securities played a role in triggering the Great Recession. Nevertheless, the value of securities is reported fairly in the balance sheet and should not contribute positively to the cost of bank closure.

- d_OCC and d_OTS: *Indicator variables that take on the value 1 if the OCC and the OTS is the primary federal regulator, respectively.* Banks under OCC and OTS supervision fail with smaller losses (Table 2.2). Thus, both dummy variables should probably have negative effects on the resolution cost.

- intexpy: *Cost of funding earning assets*, nonixay: *Noninterest expense to average assets*. The two FDIC's Performance and Conditions ratios measure the bank's cost efficiency with positive expected contribution to losses.

- nimy: *Net interest margin*, noniiay: *Noninterest income to average assets*, noijy: *Net operating income to assets*. These FDIC's Performance and Conditions ratios characterize the bank's operating efficiency with negative expected impact on bank closure cost.

- iddivnir: *Cash dividends to net income*, roe: *Return on equity*, roeinjr: *Retained earnings to average equity*, rbcrwaj: *Total risk-based capital ratio*. Higher FDIC's Performance and Conditions ratios reflecting the profitability of banking business and dividend payoff might be linked to manager's incentives to hide losses and overreport profit. We expect a positive correlation with bank closure cost.

- nperfv: *Noncurrent assets plus other real estate owned to assets*. The reasoning is the same as for variables p3asset and p9asset with positive expected influence on the bank resolution cost.

- lnlsdepr: *Net loans and leases to deposits*. Our motivation behind the expected positive correlation is that loans funded from non-traditional sources may worsen the bank's risk profile.

It is worth-mentioning that variables appear in the final models without contradictions in the signs of coefficients (Table 2.A.2 in Appendix 2.B). Moreover, these effects agree with results in the literature. A surprising result is that in-sample forecasting errors are relatively larger for banks supervised by the OCC (0.073) and the OTS (0.086) as compared to banks under FED (0.063) and FDIC (0.070) authority. We found this result surprising because the mean size of banks closure cost is much larger for banks regulated by the FED and the FDIC (Table 2.2). In other words, costs associated with impending bank distress seem to be large and predictable for the FED and the FDIC which are conspicuously silent and turn into forbearance, but keep the formation of losses in banks under control. Cost of bank closure is relatively smaller but unforeseen for the OCC and the OTS. This results holds for the estimation in subsamples as well.

The assets-liability *mismatch* variable survived selection in the full sample and in the subsamples of smaller banks and banks under FDIC supervision. As expected, this variable has a positive effect on the cost of bank distress which is consistent with Kostrov and

Mamonov (2019). Motivation for the use of this variable in the regression analysis sounds particularly reasonable for small banks.

2.4.2 Out-of-sample forecasting results

The out-of-sample performance of three method under consideration, namely, Lasso with 50 candidate predictors, OLS regression by Cole and White (2017), and model-free forecasting with a historical average, is compared using 456 bank failures in 2009:Q3–2018:Q4. The expanding window estimation starts with $58 = 524$ (full sample) $- 456$ (out-of-sample) bank failures in 2004Q1–2009:Q2 and aims to predict the cost of bank failure in the next period in each step. The forecasting results are presented for the subsamples of banks taken by the primary federal regulator, by bank size, and for all banks (Table 2.3). Note that the estimation was run sequentially for all bank failures that occurred by the time of making the prediction without splitting them into subgroups. Attempts to forecast for each subsample separately (e.g., predicting resolution cost for banks under FDIC supervision using historical failures of banks under FDIC supervision) leads to a severe reduction in the number of observations and does not improve the forecasting accuracy.

Generally speaking, switching from model-free prediction to the Cole and White (2017) model and then to the Lasso regression improves the precision of forecasting. This statement hold for all banks and does not alter when small and big banks are considered apart. However, the results for subgroups by the primary federal regulator are not so uniform. In particular, the cost of failure for banks under FED and FDIC supervision is predicted much more accurately, but not for financial institutions under OCC and OTS supervision. These results support our hypothesis that the OCC and the OTS supervise their banks differently: In this subsample of bank failures, statistical analysis of financial statements does not improve forecasting errors. Fraudulent financial statements submitted by these banks blind the off-site monitoring and create a vague environment for banking regulators. That is why we suppose that regulatory forbearance for these banks is unlikely and expensive bank failures are explained mainly by the banks' fraud. This idea is indirectly confirmed by the dissolution of the OTS in 2011 due to unsatisfactory supervision during the Great Recession.

2.4.3 Fraud or forbearance: Circumstantial evidence

In this section, using the framework of Cole and White (2017) we show that the OCC and the OTS exercise weaker regulatory forbearance than the FED and the FDIC do. We believe that it is attributed to much higher ambiguity faced by the two former organizations.

Table 2.3 Out-of-sample forecasting with expanding window design using Lasso

Table 2.3 compares the out-of-sample forecasting errors (MSE) of a Lasso regression and the alternative methods. Banks closure costs are predicted for 456 bank failures during the period from 2009:Q3 to 2018:Q4.

MSE	All banks	Primary federal regulator				Size	
		FED	FDIC	OCC	OTS	Small	Big
Panel 1: Model-free prediction (historical average for each subgroup)							
mean	0.0073	0.0063	0.0071	0.0069	0.0101	0.0068	0.0086
SD	(0.0104)	(0.0086)	(0.0102)	(0.0085)	(0.0156)	(0.0102)	(0.0109)
Panel 2: OLS regression by Cole and White (2017)							
mean	0.0067	0.0051	0.0066	0.0068	0.0097	0.0064	0.0075
SD	(0.0105)	(0.0065)	(0.0107)	(0.0097)	(0.0131)	(0.0105)	(0.0105)
p-value: $2-1$	0.034	0.030	0.088	0.459	0.197	0.149	0.039
Panel 3: Lasso regression with 50 candidate predictors							
mean	0.0062	0.0043	0.0056	0.0074	0.0108	0.0058	0.0073
SD	(0.0101)	(0.0069)	(0.0093)	(0.0103)	(0.0156)	(0.0099)	(0.0106)
p-value: $3-2$	0.014	0.150	0.000	0.860	0.866	0.009	0.355
p-value: $3-1$	0.000	0.042	0.000	0.750	0.421	0.002	0.046

Note: Corrected DM-test is used to compare forecasting errors of two models (e_i and e_j). Values i and j are panel indices that we use to define the pair of models under consideration. We are testing the null hypothesis $H_0 : \mathrm{E}(e_i - e_j) = 0$ with an alternative hypothesis $H_1 : \mathrm{E}(e_i - e_j) > 0$, and "p-value: $j-i$" is related to not rejecting H_0. Comparisons that involve the historical averages ("$2-1$" and "$3-1$") are confirmed by McCracken (2007) test for nested models.

Cole and White (2017) proposed the two early warning indicators of bank financial distress, the so-called nonperforming asset coverage ratios, NACR1 and NACR2, which are computed as follows:

$$\text{NACR1} = \text{eq} + \text{lnatres} - 0.2 \times \text{p3asset} - 0.5 \times \text{p9asset} - \text{naasset} - \text{ore}, \tag{2.7}$$

$$\text{NACR2} = \text{eq} + \text{lnatres} - 0.5 \times (\text{p3asset} + \text{p9asset} + \text{naasset} + \text{ore}). \tag{2.8}$$

In the calculation of NACR1 and NACR2, the authors apply different provisioning weights to the components of non-performing assets which are assumed by U.S. banking regulators and the International Monetary Fund, respectively. NARC1 is found to be more conservative than NACR2 since the signal of impending failure, that is, falling below the level of 2%, is received from NACR1 earlier.

A statistically significant deviation of NCAR1 from eqr is registered earlier for banks under FED and FDIC supervision (Table 2.4): 14 quarters before the bank closure versus

Table 2.4 Path to failure: Descriptive statistics by quarter

This table describes 524 failure cases splitting them in two groups: Banks under FED and FDIC supervision and banks under OCC and OTS authority. Risk measures eqr and NACR1 are reported n quarters before bank closure, where $n \in \{1, 2, \ldots, 16\}$. We consider the difference between average eqr and NACR1 for banks within groups and the difference between NACR1 values between groups.

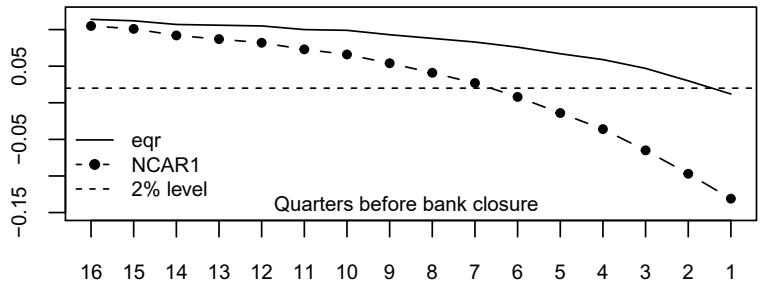
Quarter	Num. of failures	NCAR1		eqr		NCAR1 − eqr		$\text{NCAR1}^{FED}_{FDIC} - \text{NCAR1}^{OCC}_{OTS}$	
		mean	SD	mean	SD	t-stat.	p-val.	t-stat.	p-val.
Panel 1: Banks under OCC and OTS supervision									
1	141	-0.102	0.096	0.017	0.039	-13.56	**0.000**		
2	141	-0.070	0.084	0.033	0.032	-13.75	**0.000**		
3	141	-0.040	0.077	0.050	0.028	-12.99	**0.000**		
4	141	-0.016	0.074	0.064	0.030	-11.84	**0.000**		
5	141	0.001	0.071	0.071	0.031	-10.72	**0.000**		
6	141	0.020	0.071	0.079	0.033	-8.93	**0.000**		
7	141	0.036	0.069	0.084	0.036	-7.38	**0.000**		
8	141	0.050	0.067	0.090	0.041	-6.07	**0.000**		
9	141	0.062	0.079	0.094	0.060	-3.88	**0.000**		
10	141	0.069	0.077	0.097	0.059	-3.39	**0.001**		
11	141	0.075	0.065	0.098	0.045	-3.45	**0.001**		
12	141	0.081	0.073	0.101	0.059	-2.53	0.012		
13	140	0.085	0.061	0.099	0.044	-2.76	0.013		
14	140	0.086	0.063	0.100	0.051	-2.12	0.036		
15	137	0.086	0.064	0.099	0.053	-1.77	0.080		
16	137	0.093	0.080	0.103	0.072	-1.15	0.252		
Panel 2: Banks under FED and FDIC supervision									
1	383	-0.131	0.093	0.012	0.029	-28.75	**0.000**	3.117	**0.002**
2	383	-0.097	0.084	0.030	0.025	-28.38	**0.000**	3.247	**0.001**
3	383	-0.065	0.080	0.047	0.024	-26.36	**0.000**	3.238	**0.001**
4	383	-0.036	0.078	0.059	0.028	-22.41	**0.000**	2.678	**0.008**
5	383	-0.014	0.076	0.067	0.029	-19.42	**0.000**	2.083	0.038
6	383	0.008	0.076	0.076	0.034	-15.86	**0.000**	1.700	0.090
7	383	0.027	0.076	0.083	0.036	-12.98	**0.000**	1.339	0.181
8	383	0.041	0.076	0.088	0.037	-10.93	**0.000**	1.295	0.196
9	383	0.054	0.077	0.093	0.044	-8.67	**0.000**	1.068	0.286
10	383	0.066	0.090	0.099	0.064	-5.90	**0.000**	0.351	0.726
11	381	0.073	0.080	0.100	0.054	-5.59	**0.000**	0.319	0.750
12	381	0.082	0.087	0.105	0.066	-3.94	**0.000**	-0.251	0.802
13	380	0.087	0.080	0.106	0.062	-3.59	**0.000**	-0.805	0.421
14	377	0.092	0.072	0.107	0.056	-3.12	**0.002**	-0.989	0.323
15	376	0.101	0.089	0.112	0.078	-1.88	0.061	-2.048	0.041
16	373	0.105	0.083	0.114	0.074	-1.55	0.122	-1.592	0.112

Note: Following Cole and White (2017), we use t-test to determine if mean NCAR1 deviates significantly from mean eqr a few quarters before bank closure. We are testing the null hypothesis $H_0 : \mu_{NCAR1} = \mu_{eqr}$ with an alternative hypothesis $H_A : \mu_{NCAR1} \neq \mu_{eqr}$. Bold values are related to rejecting H_0 at the 1% significance level.

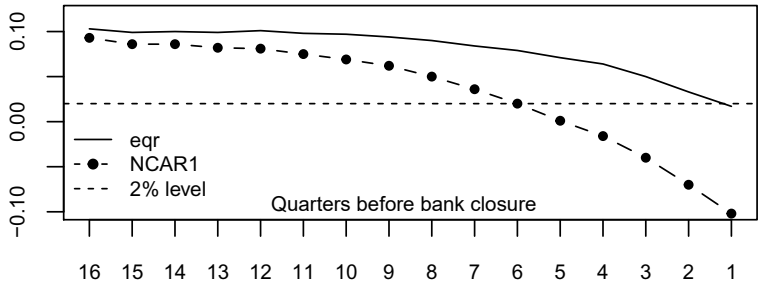
Figure 2.2 NCAR1 as an early warning signal of bank financial distress

Mean NCAR1 significantly deviates from mean eqr 14 quarter before closure for banks under FED and FDIC supervision and 11 quarters for banks under OCC and OTS supervision. This analysis is inspired by Figure 2 in Cole and White (2017). Falling below the 2% level signals bank's weak financial position.

(a) Banks under FED and FDIC supervision.



(b) Banks under OCC and OTS supervision.



11 quarters for banks under OCC and OTS regulation. Moreover, NCAR1 measure for the first subgroups of banks becomes significantly smaller than for banks under OCC and OTS supervision as early as one year prior to bank closure. NCAR1 for the latter subgroup of banks falls below the critical 2% level one quarter later. Overall, it means that we find a stronger and earlier signal of impending financial distress from banks in the responsibility area of the FED and the FDIC. The main result of the study by Cole and White (2017) is that regulator's delay in resolving troubled banks (regulator's forbearance) results in more costly bank failures. We check whether losses would have been smaller if failed banks had been closed earlier. A counterfactual exercise has to be conducted in order to estimate the possible unobserved gain from closing banks before the actual closure time. For each bank failure at time t, where $t \in \{2009{:}Q3, 2018{:}Q4\}$, we find the counterfactual closure cost at time $t-q$, where $q \in \{4, 5, \ldots, 10\}$. It is computed as a forecast made by (2.6) using bank

Table 2.5 Counterfactual bank closure cost

Table 2.5 contains the mean loss value that the society would have incurred if banks had been closed q quarters before the actual failure event. The statistical test shows whether counterfactual losses are significantly smaller than the realized ones.

Quarters before closure	Banks under FED or FDIC supervision				Banks under OCC or OTS supervision			
	mean	SD	t-stat	p-val	mean	SD	t-stat.	p-val.
3	0.189	0.034	-0.063	0.475	0.150	0.030	-0.839	0.201
4	0.186	0.032	-0.655	0.256	0.145	0.028	-1.328	0.093
5	0.180	0.030	-1.873	0.031	0.144	0.030	-1.446	0.075
6	0.176	0.030	-2.602	**0.005**	0.143	0.033	-1.491	0.069
7	0.174	0.029	-3.124	**0.001**	0.137	0.031	-2.111	0.018
8	0.175	0.028	-2.880	**0.002**	0.135	0.030	-2.331	0.011
9	0.174	0.034	-2.855	**0.002**	0.136	0.030	-2.182	0.015
10	0.173	0.034	-2.920	**0.002**	0.136	0.029	-2.195	0.015

Note: Following Cole and White (2017), we use t-test to determine if mean counterfactual bank closure costs are significantly smaller than the actual costs. We are testing the null hypothesis $H_0 : \mu_{\text{counterfactual}} = \mu_{\text{realized}}$ with the alternative hypothesis $H_A : \mu_{\text{counterfactual}} < \mu_{\text{realized}}$. Bold values are related to rejecting H_0 at the 1% significance level.

characteristics at time $t-q-1$:

$$\begin{aligned}\widehat{\text{Cost}}_{i,t,t-q} =& \widehat{\beta_0} + \widehat{\beta_1}\text{size}_{i,t-q-1} + \widehat{\beta_2}\text{eq}_{i,t-q-1} + \widehat{\beta_3}\text{lnatres}_{i,t-q-1} \\ &+ \widehat{\beta_4}\text{ore}_{i,t-q-1} + \widehat{\beta_5}\text{p3asset}_{i,t-q-1} + \widehat{\beta_6}\text{p9asset}_{i,t-q-1} \\ &+ \widehat{\beta_7}\text{naasset}_{i,t-q-1} + \widehat{\beta_8}\text{d_OCC} + \widehat{\beta_9}\text{d_OTS} + \widehat{\beta_{10}}\text{d_FED},\end{aligned} \tag{2.9}$$

where $\widehat{\text{Cost}}_{i,t,t-q}$ is the predicted counterfactual loss, which might have been realized q quarters before the actual bank failure, and the coefficients are estimated using data available at time $t-q-1$. Table 2.5 shows that the average counterfactual bank closure cost tends to grow as the realized closure event approaches. However, it is not significantly smaller than the actual bank closure cost for banks under OCC and OTS supervision. Thus, we conclude that the OCC and the OTS do not practice regulatory forbearance. We believe that an intentional delay of bank closure requires comprehensive control over its financial position which is not feasible for the OCC and the OTS.

This finding might be related to distinctions of banks supervised by the different primary federal regulators as well. For example, the FED is responsible for supervision over the systemically important banks (Labonte, 2020) with potentially high non-monetary costs of failure. Thus, regulatory forbearance can be at least partially justified.

2.5 Conclusion

This paper contributed to the discussion about inconsistent enforcement of similar rules by different regulators in the U.S. banking sector. We examined 524 bank closures during the period 2004–2018 and found that the banks supervised by different federal regulators failed with unequal closure costs and these differences were statistically significant: Banks with the FED and the FDIC as the primary federal regulator lost their capital plus 18.19% and 19.66% of borrowed funds, respectively, whereas banks under the OCC and OTS imposed a smaller haircut on borrowed funds of 16.32% and 14.68%, respectively. We primarily focused on predicting the size of bank closure cost and analyzed whether financial reports contain valuable information to improve the accuracy of forecasting. A good forecasting device is important under the standard risk-management framework and allows to work out guidelines for better supervision in banking. Using Lasso, we selected and interpreted 20 predictors of bank closure costs. The set of selected variables was different for the groups of banks split by the primary federal regulator, which is in line with the inconsistent supervision hypothesis.

A somewhat surprising result was that the analysis of financial statements did not improve the forecasting accuracy of closure cost for banks supervised by the OCC and the OTS as compared to naive predicting with a historical average. Moreover, the mean forecasting error for these banks was almost twice higher than for banks under FED and FDIC authority. In other words, the FED and the FDIC faced large and predictable losses in bank failures whereas the OCC and the OTS supervised in conditions of high uncertainty.

This evidence has two consequences. First, the quality of supervision by the OCC and the OTS is debatable. In particular, the off-site monitoring is blinded and misleading when financial variables weakly reflect bank's vulnerability. This is called fraudulent financial reporting by banks, a joint problem of a regulator, an auditor, and a bank. This situation bears additional risks when the on-site monitoring is limited or impossible for some reasons. Second, as opposed to the OCC and the OTS, the FED and the FDIC are inclined to forbearance in supervision over banks which was confirmed by our analysis inspired by Cole and White (2017).

Appendices

Appendix 2.A Candidate predictors of bank failure cost

Table 2.A.1 describes variables in a universe of candidate predictors of bank failure cost. Variables marked with an asterik* are normalized by total assets. CR in the header stands for "Call report".

Table 2.A.1 Summary statistics for candidate predictors of bank failure cost

		Description	Obs	Mean	SD	Q25	Q75	CR Form
Panel 1: Mismatch (Kostrov and Mamonov, 2019) and asset growth rate (Kang et al., 2015)								
1	mismatch	Product of bro and lnci	524	0.007	0.017	0.000	0.007	–
2	TA_growth	Growth rate of total assets	524	-0.125	0.161	-0.205	-0.072	–
Panel 2: Predictors motivated by Cole and White (2017)								
3	size*	Logarithm of Total assets	524	12.374	1.378	11.502	13.111	Assets and Liabilities
4	eq*	Bank equity capital	524	0.013	0.032	0.002	0.024	
5	lnatres*	Loan loss allowance (reserves)	524	0.033	0.020	0.020	0.041	↑
6	ore*	Other real estate owned	524	0.050	0.048	0.015	0.071	↑
7	p3asset*	Assets past due 30-89 days	524	0.032	0.026	0.014	0.042	Past Due and Nonaccrual Assets
8	p9asset*	Assets past due 90 or more days	524	0.005	0.012	0.000	0.005	
9	naasset*	Assets in nonaccrual status	524	0.110	0.071	0.062	0.140	↑
10	NACR1*	Nonperforming asset coverage ratio 1	524	-0.123	0.095	-0.179	-0.061	–
11	NACR2*	Nonperforming asset coverage ratio 2	524	-0.053	0.059	-0.088	-0.015	–
12	d_OCC	Indicator variable if OCC is the primary federal regulator	524	0.176				–
13	d_FED	Indicator variable if FED is the primary federal regulator	524	0.103				–
14	d_OTS	Indicator variable if OTS is the primary federal regulator	524	0.094				–

Continued on next page

	Description	Obs	Mean	SD	Q25	Q75	CR Form
Panel 3: Predictors motivated by Schaeck (2008)							
15 lnci*	Commercial and industrial loans	524	0.076	0.069	0.030	0.102	Net Loans and Leases
16 frepp*	Federal funds purchased and repurchase agreements	524	0.006	0.021	0.000	0.000	↑
17 lnre*	All real estate loans	524	0.597	0.133	0.523	0.685	↑
18 bro*	Brokered deposits	524	0.088	0.133	0.000	0.114	Total Deposits
19 ddt*	Demand deposits	524	0.075	0.067	0.030	0.099	
20 ts*	Total time and savings deposits	524	0.834	0.099	0.774	0.905	↑
Panel 4: FDIC Performance and Conditions ratios							
21 intincy	Yield on earning assets	524	4.811	0.992	4.207	5.351	Performance and Conditions ratios
22 intexpy	Cost of funding earning assets	524	2.232	1.049	1.426	2.861	
23 nimy	Net interest margin	524	2.579	1.116	1.928	3.305	
24 noniiay	Noninterest income to average assets	524	0.200	1.872	-0.117	0.578	↑
25 nonixay	Noninterest expense to average assets	524	4.483	3.157	3.069	4.824	↑
26 noijy	Net operating income to assets	524	-6.307	6.322	-7.959	-2.797	↑
27 roa	Return on assets (ROA)	524	-6.531	6.458	-8.446	-2.834	↑
28 roaptx	Pretax return on assets	524	-6.547	6.452	-8.348	-2.811	↑
29 roe	Return on Equity (ROE)	524	-254.1	1104	-263.2	-80.90	↑
30 roeinjr	Retained earnings to average equity (ytd only)	524	-254.3	1104	-263.2	-82.82	↑
31 ntlnlsr	Net charge-offs to loans	524	5.244	6.111	1.494	7.041	↑
32 eeffr	Efficiency ratio	524	216.4	1327	107.6	218.1	↑
33 astempm	Assets per employee ($millions)	524	5.428	3.400	3.531	6.211	↑
34 iddivnir	Cash dividends to net income (ytd only)	524	331.4	7572	0.000	0.000	↑
35 lnatresr	Loss allowance to loans	524	4.616	2.703	2.902	5.677	↑
36 nperfv	Noncurrent assets plus other real estate owned to assets	524	16.53	8.790	10.37	21.04	↑
37 nclnlsr	Noncurrent loans to loans	524	16.11	9.573	9.390	20.235	↑
38 lnlsdepr	Net loans and leases to deposits	524	74.20	14.82	65.36	82.32	↑
39 idlncorr	Net loans and leases to core deposits	524	92.81	33.84	72.89	105.7	↑
40 rbc1aaj	Core capital (leverage) ratio	524	1.106	2.972	0.219	2.072	↑
41 rbc1rwaj	Tier 1 risk-based capital ratio	524	1.516	4.916	0.310	2.900	↑
42 rbcrwaj	Total risk-based capital ratio	524	2.426	5.273	0.632	4.231	↑

Continued on next page

	Description	Obs	Mean	SD	Q25	Q75	CR Form
Panel 5: Other candidate predictor							
43 nclnls*	Noncurrent loans and leases	524	0.114	0.072	0.065	0.147	Assets and Liabilities
44 bkprem*	Bank premises and fixed assets	524	0.023	0.020	0.008	0.033	
45 ntlnls*	Net charge-offs	524	0.025	0.027	0.005	0.034	Income and Expense
46 sc*	Total securities	524	0.095	0.083	0.031	0.131	Securities
47 scus*	U.S. Government securities	524	0.074	0.069	0.019	0.110	↑
48 trade*	Trading account assets	524	0.000	0.008	0.000	0.000	↑
49 scmtgbk*	Mortgage-backed securities	524	0.055	0.065	0.002	0.089	↑
50 idsccmo*	Collaterized mortgage obligations	524	0.023	0.045	0.000	0.029	↑
Cost	Cost of failure (negative capital)	524	0.185	0.086	0.123	0.249	–

Appendix 2.B Lasso estimation using all observations

Table 2.A.2 considers the in-sample forecasting accuracy of the Lasso model using the full sample (column "All banks"). We also report accuracy measures for subsamples by the primary federal regulator (rows RMSE_FED, RMSE_FDIC, RMSE_OCC, RMSE_OTS) and by size (rows RMSE_small, RMSE_big). Then, Lasso regressions are estimated separately in subsamples and the results are reported in the corresponding columns.

Table 2.A.2 Estimation of the Lasso model using bank failures in the period 2004–2018

Variable		All banks	Primary federal regulator				Size	
			FED	FDIC	OCC	OTS	Small	Big
intercept		0.150	0.176	0.208	0.152	0.109	0.069	0.170
size	1	-0.009		-0.013			-0.007	
mismatch	3	0.174		0.191			0.186	
lnci	4	0.006						0.089
TA_growth	5	0.040	0.005	0.063				
eqr	6	-0.113				-0.452	-0.001	
frepp	7		-0.856					
lnatres	8	0.215		0.388			0.128	
nclnls	9	0.039		0.010			0.039	
bkprem	11			0.121				
p3asset	13	0.142		0.202	0.412			0.612
p9asset	14	0.655		0.555			0.394	0.413
ddt	17			-0.070				
ts	18	0.092		0.071			0.171	

Continued on next page

			Primary federal regulator				Size	
			FED	FDIC	OCC	OTS	Small	Big
scmtgbk	20	-0.013			-0.013		-0.028	
idsccmo	21	-0.038	-0.002	-0.083	-0.153			
d_OCC	24	-0.007						
d_OTS	26	-0.019						-0.011
intexpy	30	0.019	0.015	0.020	0.001		0.018	0.003
nimy	31		-0.012	-0.003				-0.018
noniiay	32	-0.001		-0.001				
nonixay	33	0.002		0.003			5.47×10^{-4}	
noijy	34					-0.004	-2.14×10^{-4}	
roa	35	-6.07×10^{-5}				-0.002		
roe	37	3.59×10^{-6}		4.99×10^{-6}			3.86×10^{-6}	
roeinjr	38			2.95×10^{-11}			1.51×10^{-8}	
iddivnir	42	4.01×10^{-7}		4.57×10^{-7}				3.06×10^{-7}
nperfv	44			0.001				
lnlsdepr	46						5.02×10^{-6}	
rbcrwaj	50							-2.91×10^{-4}
N obs		524	54	329	92	49	367	157
N vars		20	5	19	4	3	14	8
R_squared		0.303	0.437	0.317	0.131	0.469	0.277	0.295
RMSE		0.072						
RMSE_FED		0.063	0.058					
RMSE_FDIC		0.070		0.068				
RMSE_OCC		0.073			0.075			
RMSE_OTS		0.086				0.076		
RMSE_small		0.069					0.070	
RMSE_big		0.077						0.076

Chapter 3

Forecasting realized covariances: a MIDAS approach

Alexander Kostrov [1], **Anastasija Tetereva** [2]

[1]University of St. Gallen, Faculty of Mathematics and Statistics, alexander.kostrov@unisg.ch
[2]Erasmus University Rotterdam, Chair of Econometrics, tetereva@ese.eur.nl

Abstract

This paper considers a mixed data sampling (MIDAS) approach to forecast realized covariance matrices. The parameters are estimated using a revised computationally efficient MIDAS-NLS procedure. We discover a considerable advantage in out-of-sample forecasting accuracy for one-day-ahead and multiperiod forecasting as compared to the established heterogeneous autoregressive (HAR) model, particularly during the crisis. This improvement is attributed to the smoothness and flexibility of the MIDAS model. The economic value of the proposed method is justified by portfolio management applications in terms of the investor's utility and annualized returns.

3.1 Introduction

The modeling and accurate forecasting of multivariate financial time series is one of the main objectives of financial econometrics. A large number of models have been developed to predict the covariance matrix of financial assets, starting with pioneering studies on the autoregressive conditional heteroskedasticity (ARCH) model by Engle (1982) and the generalized autoregressive conditional heteroskedasticity (GARCH) model by Bollerslev (1986). There is a lack of strong evidence that numerous extensions of the latter model provide better out-of-sample forecasting properties than GARCH$(1,1)$ model (Hansen and Lunde (2005), Francq and Zakoian (2011)). Although ARCH and GARCH models are flexible enough to capture the dynamic nature of the second conditional moment of financial time series, they measure daily volatility based on squared daily returns. This drawback leads to noisy estimation and poor out-of-sample forecasting properties. Moreover, these models are difficult to estimate in higher dimensions as they suffer from the curse of dimensionality.

Andersen et al. (2001) address this weakness by introducing the concept of realized volatility. Daily realized volatility is computed from intraday data and serves as an accurate measure of the conditional variance of daily observations; it treats an inherently latent variable like an observed variable. Recent studies in the field of financial econometrics have

shown that the daily realized variance computed from intraday returns is a good proxy for the conditional variance of daily returns. In addition to the excellent out-of-sample forecasting properties, in contrast with the GARCH type of processes, no specific model is needed to estimate it. In the recent decade, the availability of high-frequency data has resulted in growing attention being given to the concept of realized variance and covariance.

Although there is a significant body of the literature on the estimation and forecasting of the realized covariance matrices, most methods perform well in forecasting the realized variances while failing to accurately predict the realized covariances. This can be attributed to the highly volatile nature of the realized covariances and the presence of multiple change points. Accurate forecasting of covariances plays an important role in a variety of applications such as investment portfolio management, Value-at-Risk forecasting, and the analysis of contagion in financial markets (Engle (2009), Vortelinos (2016)).

Many early studies on the forecasting of the realized covariance matrices focus on the heterogeneous autoregressive model by Corsi (2009) and its extensions. In contrast to the periods of low volatility, during which the HAR model demonstrates a satisfactory prediction accuracy, it fails to deliver accurate forecasts during the periods of financial turbulence. However, many extensions are suggested for the HAR model. However, Symitsi et al. (2018) compare 12 popular covariance forecasting models and show that HAR is still very difficult to outperform.

In our paper, we take advantage of another approach, namely the MIDAS regression. Although the MIDAS regression has been extensively discussed in the literature in the context of modeling the realized volatility, the topic of realized covariance forecasting has not been studied widely. Moreover, this model has not been used for the dynamic modeling of realized covariance matrices in a high-dimensional setting. This may be due to the absence of computationally efficient and accurate estimation procedures. In this paper, we contribute to the literature by adopting the methodology of Audrino, Kostrov and Ortega (2019) to revise a MIDAS-NLS estimator for the MIDAS regression. Our procedure allows the estimation of rolling window MIDAS regressions for the realized covariances at lower computational costs. In contrast to its competitor, the HAR model, the MIDAS model allows for smooth weighting of the past values. This smoothness and flexibility of MIDAS results in a superior forecasting performance.

Our empirical analysis is based on 29 constituents of the Dow Jones Industrial Average (DJIA) index for the period from January 2001 to December 2017. In comparison with the well-established HAR model by Corsi (2009), the MIDAS approach provides superior forecasting performance, which is evaluated using four common loss functions – specifically,

mean absolute error (MAE), mean squared error (MSE), asymmetric loss function (Asym), and quasi-likelihood (QLike). We consider one-day-ahead and multiperiod predictions for both models. In the latter case, we compare direct and iterated forecasting designs for the MIDAS approach and for the benchmark HAR model. We find that HAR direct and MIDAS iterated approaches are most accurate. The superior forecasting performance of the MIDAS model is especially pronounced during the crisis time and is statistically significant, according to the corrected Diebold-Mariano test by Harvey et al. (1997), the fluctuation test by Giacomini and Rossi (2010), and the Model Confidence Set (MCS) approach by Hansen et al. (2011). Improvement is observed for both, shorter and longer time horizons. These conclusions are justified by considering the economic gain from investing in the global minimum variance portfolio (GMV) in terms of utility and returns: During the crisis period, switching from the GVM portfolio suggested by the HAR model to that suggested by the MIDAS is worth 1.2–6.6 percentage points of annualized returns depending on the investment horizon.

This paper is organized as follows. Section 3.2 explains the use of high-frequency data for the realized covariance matrix estimation. In Section 3.3, we describe the HAR model and introduce the MIDAS approach for one-day-ahead forecasting, and Section 3.4 focuses on multiperiod predictions and aspects such as forecasting accuracy evaluation and bias correction. Section 3.5 describes the data and the main findings and lays out the economic significance of the introduced approach. The final section concludes the paper.

3.2 Estimating the realized covariance matrix from high-frequency data

The theoretical considerations of the realized variance model are based on the assumption that the logarithm of the instantaneous price X_t follows the standard continuous time process

$$\mathrm{d}X_t = \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W_t,$$

where $\mu(t)$ is a non-random càdlàg finite variation process referred to as trend and $\sigma(t)$ is the time-varying càdlàg volatility function which is assumed to be independent of a standard Brownian motion process W_t. As shown by Andersen et al. (2001), the realized

variance over a time interval of one day

$$RV_t^d = \sum_{j=0}^{M-1} r_{t-j\cdot\Delta}^2$$

converges in probability to the integral of the instantaneous variance over the one-day interval $[t-1d, t]$

$$IV_t^d = \int_{t-1d}^{t} \sigma_s^2(s)\mathrm{d}s,$$

where $\Delta = 1d/M$, $r_{t-j\cdot\Delta} = X_{t-j\cdot\Delta} - X_{t-(j+1)\cdot\Delta}$ are the intraday returns sampled at time interval Δ and M denotes the number of intraday observations. A full trading day is represented by the time interval $1d$. However, the theoretical properties of the above-mentioned estimator are valid only when one assumes the absence of jumps and microstructure noise. The properties and further extensions of this estimator have been rigorously studied by Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004).

In order to estimate the realized covariance matrix, some methodological adjustments are required. First, the estimates of the matrix of interest are not necessarily positive semidefinite due to the market microstructure noise. Second, non-synchronous trading induces bias in the estimates. Some estimators, e.g., the multivariate realized kernel by Barndorff-Nielsen et al. (2009), are able to deliver valid positive semidefinite estimators. However, this estimator uses refresh-time sampling to synchronize the data. As a result of such a sampling scheme, the number of observations may be reduced heavily because the resulting observation frequency is determined by the least liquid asset. Hautsch et al. (2012) point out that the loss of data can exceed 90%. Other estimators such as Aït-Sahalia et al. (2010) and Lunde et al. (2016) manage to handle the data more efficiently by using the synchronized pairs of price processes. However, these methods fail to guarantee the positive semidefiniteness of the estimator due to separate estimation of the off-diagonal elements. Corsi et al. (2015) address the missing observation problem in the realized covariance estimation using Kalman smoother and expectation maximization algorithm. That approach preserves positive semidefiniteness of the estimated covariance matrices.

In this paper, we employ the CholCov estimator by Boudt et al. (2017) that uses the Cholesky factorization of the covariance matrix. This estimator is positive semidefinite by construction and addresses the heterogeneity present in the trading intensity by estimating the parameters sequentially, according to the decreasing liquidity. The fact that different elements of the matrix are estimated using different time-grids increases the efficiency of

the estimator through the use of more observations. For details on the sequential estimation procedure of the CholCov we refer to Boudt et al. (2017).

Once the realized covariance matrix for n assets is estimated, it needs to be modeled and predicted. We aim to forecast the realized covariance matrix for day $t+z$ conditioned on the information available at time t , where z is the forecasting horizon. The elements of the estimated realized covariance matrix H_t are defined as $H_{ij,t}$, $i,j \in \{1,\dots,n\}$ and $t \in \{1,\dots,T\}$; T denotes the number of trading days in the sample.

3.3 Modeling the realized covariances

Empirical evidence shows that financial time series of volatility exhibit very strong persistence. Moreover, multiscaling and clustering of volatility are typical for financial data. It is important for the model to be able to capture these stylized facts.

Several methods address these issues in the literature. For example, Gouriéroux et al. (2009) suggest the use of a flexible dynamic Wishart autoregressive process for the entire covariance matrix modeling. Vector autoregressive fractionally integrated moving average (VARFIMA) model has been applied to the Cholesky factors of the realized covariance matrix by Chiriac and Voev (2011). However, both specifications lack a clear economic interpretation. Moreover, the ARFIMA model is difficult to extend to high-dimensional settings without assuming some underlying factor model that might be too restrictive in practice. For this reason, a large number of existing studies rely on the HAR model suggested by Corsi (2009). Although this model is intuitive and easy to estimate via ordinary least squares (OLS), it successfully captures the observed pattern of the realized volatility. The HAR model is able to capture stylized facts typical for financial time series including volatility persistence. This model is used as a benchmark to evaluate the performance of the model proposed in this paper. In the next subsection, we describe the HAR model in greater details.

3.3.1 HAR model for the realized covariances

Corsi (2009) introduced the HAR model, which is an additive model with heterogeneous components related to the realized volatility literature. The author motivated his model using the heterogeneous market hypothesis formulated by Müller et al. (1993). This model assumes that each market agent forms expectations for the next period volatility based on different time scales. The AR(1) part of the model characterizes the current realized

volatility that is experienced at the same time scale. The hierarchical component mimics the expectations for longer horizons. In other words, the market participants can be classified into three heterogeneous groups of agents according to their trading horizons, i.e., daily, weekly, and monthly. Each group creates its own component of latent volatility. This cascade of time scales in the volatilities motivates the HAR equation:

$$\log h_{t+1}^{d} = c + \beta^{d} \log h_{t}^{d} + \beta^{w} \log h_{t}^{w} + \beta^{m} \log h_{t}^{m} + \varepsilon_{t+1}, \tag{3.1}$$

where $\boldsymbol{\beta} = (c, \beta^{d}, \beta^{w}, \beta^{m})$ is a parameter vector and the weekly and monthly averages of volatility are defined as $\log h_{t}^{w} = \frac{1}{5}\sum_{q=1}^{5} \log h_{t-q+1}^{d}$ and $\log h_{t}^{m} = \frac{1}{22}\sum_{q=1}^{22} \log h_{t-q+1}^{d}$, $t \in \{22, \dots, T\}$. The logarithmic scale for volatilities is used to get (approximately) the Gaussian distribution and to ensure a valid forecasting of the realized volatility.

Whereas the HAR approach (3.1) is originally designed for the univariate volatility modeling, its extension to the multivariate setting has been developed in the literature. Following Bauer and Vorkink (2011), we make a matrix-logarithmic transformation of the covariance matrices for each $t = \{1, \dots, T\}$ and apply the univariate HAR to forecast each individual element of the matrix. In other words,

$$\mathrm{RC}_{ij,t+1}^{d} = c_{ij} + \beta_{ij}^{d} \mathrm{RC}_{ij,t}^{d} + \beta_{ij}^{w} \mathrm{RC}_{ij,t}^{w} + \beta_{ij}^{m} \mathrm{RC}_{ij,t}^{m} + \varepsilon_{ij,t+1}, \tag{3.2}$$

where $\mathrm{RC}_{ij,t}^{d} = \log(H_{t}^{d})_{ij}$, $i, j \in \{1 \dots n\}$ is an element of a covariance matrix after the matrix-logarithm transformation, $\beta_{ij} = (c_{ij}, \beta_{ij}^{d}, \beta_{ij}^{w}, \beta_{ij}^{m})$ is a parameter vector, weekly and monthly averages are defined as $\mathrm{RC}_{ij,t}^{w} = \frac{1}{5}\sum_{q=1}^{5} \mathrm{RC}_{ij,t-q+1}^{d}$ and $\mathrm{RC}_{ij}^{m} = \frac{1}{22}\sum_{q=1}^{22} \mathrm{RC}_{ij,t-q+1}^{d}$, $t \in \{22, \dots, T\}$. The coefficients are estimated using OLS. Finally, the inverse transformation, the matrix exponential, is used to ensure a positive-definite forecast.

It is worth mentioning that the nonlinear logarithm transformation results in a prediction bias which arises due to the fact that the forecast is made in the transformed space instead of the original one. Details on the bias correction will be discussed in Section 3.4.2. Another way to guarantee the validity of the forecast is to model the Cholesky factorization of the covariance matrix. This approach was suggested by Chiriac and Voev (2011). As was empirically justified in Fengler and Okhrin (2016), both methods produce similar results.

The HAR model can be considered as a restricted version of a more general regression, which is known as the autoregressive model AR(p) in the time series literature or the

U-MIDAS model in the MIDAS literature (Foroni et al., 2015):

$$\mathrm{RC}^{d}_{ij,t+1} = c_{ij} + \sum_{q=1}^{p} \beta_{ij,t-q+1}\, \mathrm{RC}^{d}_{ij,t-q+1} + \varepsilon_{ij,t+1}, \tag{3.3}$$

where p is the maximum lag order parameter which equals 22, $t \in \{p, \ldots, T\}$. Notice that a large p requires many parameters to be estimated which makes the estimation procedure computationally intensive and may lead to overfitting and noisy forecasts.

Alternatively, it is possible to reorganize daily data before estimating a regression equation, similar to the HAR model (3.2), with daily, weekly, and monthly horizons. This equation can be rewritten in terms of daily realized covariances:

$$\begin{aligned}
\mathrm{RC}^{d}_{ij,t+1} =& c_{ij} + \beta^{d}_{ij}\, \mathrm{RC}^{d}_{ij,t} + \beta^{w}_{ij}\, \mathrm{RC}^{w}_{ij,t} + \beta^{m}_{ij}\, \mathrm{RC}^{m}_{ij,t} + \varepsilon_{ij,t+1} \\
=& c_{ij} + \beta^{d}_{ij}\, \mathrm{RC}^{d}_{ij,t} + \frac{1}{5}\beta^{w}_{ij} \sum_{q=1}^{5} \mathrm{RC}^{d}_{ij,t-q+1} + \frac{1}{22}\beta^{m}_{ij} \sum_{q=1}^{22} \mathrm{RC}^{d}_{ij,t-q+1} + \varepsilon_{ij,t+1} \\
=& c_{ij} + (\beta^{d}_{ij} + \frac{1}{5}\beta^{w}_{ij} + \frac{1}{22}\beta^{m}_{ij})\, \mathrm{RC}^{d}_{ij,t} + (\frac{1}{5}\beta^{w}_{ij} + \frac{1}{22}\beta^{m}_{ij}) \sum_{q=2}^{5} \mathrm{RC}^{d}_{ij,t-q+1} \\
&+ \frac{1}{22}\beta^{m}_{ij} \sum_{q=6}^{22} \mathrm{RC}^{d}_{ij,t-q+1} + \varepsilon_{ij,t+1}.
\end{aligned} \tag{3.4}$$

This aggregation assumes that days 2–5 and 6–22 prior to time t equally impact the dependent variable in (3.4), which is not always supported by empirical evidence. The use of this flat weighting scheme may lead to an aggregation bias (Andreou et al., 2010) which may deteriorate the forecasting accuracy.

Despite many instances of the successful application of the HAR model, its validity continues to be debated. Audrino and Knaus (2016) have statistically verified the performance of the HAR model by applying the adaptive Lasso procedure to the AR model with more than 22 lags. Using a portfolio consisting of 19 stocks, the authors observe that during the financial crisis only shorter lags are selected by the Lasso. In contrast, in the beginning and at the end of the financial turbulence, further lags survive the selection. However, the lags which are inconsistent with the HAR model appear to be statistically insignificant.

A broad comparison of the HAR model to a rich variety of its extensions within a volatility forecasting exercise is presented in Audrino, Huang and Okhrin (2019). Among many others, the competitor models such as AR(p) with the order selected by AIC, the AR-Lasso by Audrino and Knaus (2016), the HARQ model by Bollerslev et al. (2016), and the HAR-S-RV-J model by Chen and Ghysels (2010) are compared based on the root-mean-

square error criterion. Audrino, Huang and Okhrin (2019) conclude that the time scales of the individual components of the HAR model depend on the market conditions. However, the authors fail to find a model that brings a statistically significant improvement as compared to HAR in the out-of-sample considerations. The conclusion that no model is uniformly better than all other competitors is supported by the construction of the MCS approach by Hansen et al. (2011). Moreover, Symitsi et al. (2018) report that the HAR model has the most prominent performance in covariance forecasting as compared to 11 other common models. Therefore, HAR remains a valid benchmark for new models in terms of the realized volatility and covariance.

3.3.2 MIDAS model for the realized covariances

The MIDAS weighting scheme

The HAR model is not the only way to predict the covariance matrix based on its past realizations. Mixed data sampling regression is originally introduced to involve high-frequent covariates into explaining a low-frequent dependent variable. We refer to Ghysels et al. (2004) and Ghysels et al. (2007) for more details. This method assigns smooth individual weights to past values, decreases the importance of maximum lag order selection in regression analysis, and keeps the model parsimonious. The MIDAS approach for the regression model introduced in equation (3.3) works as follows:

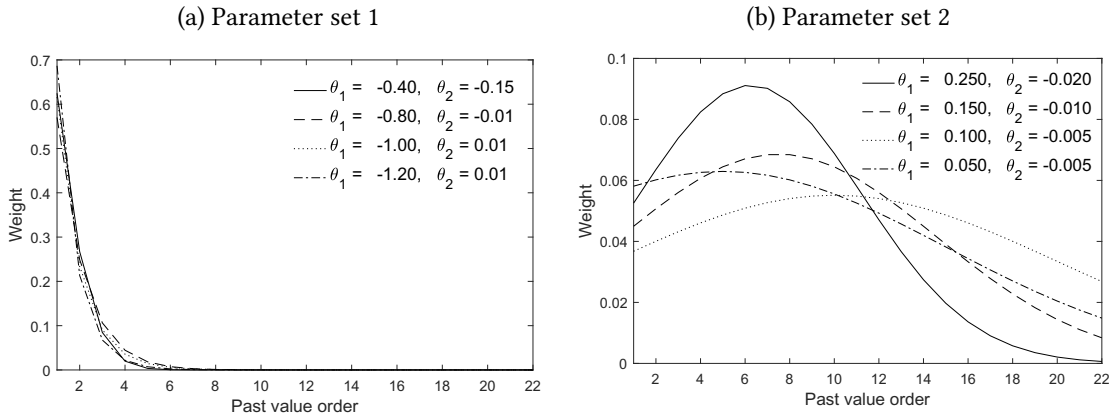
$$\mathrm{RC}^{d}_{ij,t+1} = \beta_{ij} \sum_{q=1}^{p} \gamma(q, \theta_{1,ij}, \theta_{2,ij})\, \mathrm{RC}^{d}_{ij,t-q+1} + \varepsilon_{ij,t+1}, \tag{3.5}$$

where p is the maximum lag order considered (MIDAS temporal aggregation period), $t \in \{p, \ldots, T\}$. The individual weights of the past values are assigned through $\gamma(q, \theta_1, \theta_2) = \exp(\theta_1 q + \theta_2 q^2)$, $q \in \{1, \ldots, p\}$ denoting the exponential Almon (1965) lag function with two parameters, which is common for the MIDAS-type models. It is rather flexible and unifies many weighting profiles $\gamma(\theta_1, \theta_2)$. Some feasible schemes are shown in Figure 3.1, weights sum up to 1. The MIDAS model (3.5) is driven by three parameters regardless of the maximum lag p. This model automatically adjusts the weights of high-order lags in HAR; therefore, it is able to show a superior forecasting behavior. In fact, the HAR model can be considered as a particular case of MIDAS with "day-week-month" step functions.

MIDAS volatility and covariance forecasts

Since Ghysels et al. (2004) introduced the MIDAS regression, this approach has been

Figure 3.1 MIDAS weighting schemes driven by θ_1 and θ_2



implemented in a large array of applications. There is a burgeoning literature that applies MIDAS in the context of volatility and covariance analysis.

Engle et al. (2013) suggest the Generalized Autoregressive Conditional Heteroskedasticity-MIDAS (GARCH-MIDAS). Dynamic Conditional Correlation-MIDAS and Constant Conditional Correlation-MIDAS exhibit a superior performance compared to the standard benchmarks in Audrino (2014) and Colacito et al. (2011). Ghysels et al. (2005) benefit from the MIDAS approach in predicting the realized volatility. Ghysels et al. (2006) show how to harness daily or intraday data in order to facilitate forecasting volatility using MIDAS. Golosnoy et al. (2012) and Anatolyev and Kobotaev (2018) find that the MIDAS-augmented versions of the conditional autoregressive Wishart model and the conditional threshold autoregressive Wishart model tend to have a good out-of-sample performance in forecasting the realized covariance matrix. Becker et al. (2010) employ Cholesky decomposition-MIDAS model to predict realized covariance matrix.

The recent study by Bollerslev et al. (2018) considers the MIDAS approach in the volatility forecasting context. The authors discuss the prediction of realized volatilities across global assets by using the MIDAS regression. The use of the MIDAS model in volatility forecasting is motivated by its ability to "smooth" the distributed lag polynomials in (3.3) for the purpose of modeling the dynamic dependencies. In the MIDAS regression, the authors use common Beta polynomials governed by two parameters, θ_1 and θ_2. The first parameter is set to one: $\theta_1 = 1$. This choice is well-established in the MIDAS literature (Ghysels et al. (2006), Becker et al. (2010)) and relates to a slowly decaying pattern that is common for volatility. The subsequent MIDAS model estimation is based on the efficient profiling approach by

Ghysels and Qian (2019) which applies a grid search for θ_2. However, the rolling window analysis along with profiling remains computationally challenging (Bollerslev et al. (2018)).

The current paper applies an analytical gradient-based method to estimate the MIDAS model. Adopting the methodology of Audrino, Kostrov and Ortega (2019), we explicitly derive the gradient of the objective function in the MIDAS optimization problem and combine it with a trust-region optimization algorithm. The derivation is standard and can be obtained from the authors upon request. Our approach has two peculiarities. First, we estimate both parameters θ_1 and θ_2: A purely data-driven choice is made. Second, the estimation procedure is computationally efficient. The advantage in the model estimation time allows to carry out a real out-of-sample prediction exercise for high-dimensional covariance matrices by re-estimating the MIDAS model in a rolling estimation window. The computational burden has previously been an obstacle in applications (Ghysels and Qian (2019)). Kostrov (2020) contains a simulation study that illustrates advantages of the revised optimization procedure in terms of precision and computation time. It is shown that in contrast to existing estimators, the revisited optimization procedure is fast, guarantees convergence to the true solution, and therefore is computationally feasible for high-dimensional covariance matrices.

3.4 Multiperiod forecasts and forecast evaluation

3.4.1 Direct and iterated forecasts

As has been mentioned before, the main objective of the current paper is to suggest an accurate method for the forecasting of realized covariance matrix. In contrast to the majority of studies that solely consider one-day-ahead forecasts, this paper involves the analysis of longer time horizons. It is a known fact that multiperiod predictions are of crucial importance in portfolio applications since many market players have investment horizons that last longer than one day. For example, the Basel Committee regulations require accurate Value-at-Risk forecasts at 0.99 level for ten-day-ahead forecasts. For that reason, we report the results for forecasting horizons equal to 1, 5, 10, and 22 days. Further on, we will use $\widehat{H}_{t+z|t}$ to denote the forecast for day $t+z$ and $\widehat{H}_{(t+1):(t+z)|t}$ to denote the multiperiod forecast (for the whole period) from day $t+1$ till day $t+z$. Adding up individual forecasts (iterated method) is not the only way to obtain multiperiod forecasts. Direct approach is also widely discussed in the literature. For example, Clements et al. (2008) discover that direct modeling using HAR leads to the most accurate volatility forecasts for currencies.

This study compares both of them for each of the considered models. A z-days-ahead iterated forecast is defined as

$$\widehat{H}^{\text{iterated}}_{(t+1):(t+z)|t} = \sum_{k=1}^{z} \widehat{H}^{d}_{t+k|t}. \tag{3.6}$$

Prediction for day $t+k$, $\widehat{H}_{t+k|t}$, is obtained by forecasting elements of the logarithmic transformation of the covariance matrix $RC_{t+k} = \log(H_{t+k})$ using the HAR model (3.2) or the MIDAS regression (3.5) and making the inverse transformation $\widehat{H}_{t+k|t} = \exp\left(\widehat{RC}_{t+k|t}\right)$, $k = \{1, \dots, z\}$. Note that the procedure involves information available at time t only and that predictions for days $t+1, \dots, t+z$ are iterated from the one-step-ahead model without reestimation of coefficients. To sum up, multiperiod forecast $\widehat{H}^{\text{iterated}}_{(t+1):(t+z)|t}$ is computed as a sum of z daily forecasts for days $t+1, \dots, t+z$.

In contrast, multiperiod z-day-ahead direct forecasts are obtained from

$$\text{RC}^{\text{direct}}_{ij,(t+1):(t+z)} = c_{ij} + \beta^{d}_{ij}\,\text{RC}^{d}_{ij,t} + \beta^{w}_{ij}\,\text{RC}^{w}_{ij,t} + \beta^{m}_{ij}\,\text{RC}^{m}_{ij,t} + \varepsilon_{ij,t+1} \tag{3.7}$$

for the HAR model and from

$$\text{RC}^{\text{direct}}_{ij,(t+1):(t+z)} = \beta_{ij} \sum_{q=1}^{p} \gamma(q, \theta_{1,ij}, \theta_{2,ij})\,\text{RC}^{d}_{ij,t-q+1} + \varepsilon_{ij,t+1} \tag{3.8}$$

for the MIDAS model. In other words, a direct prediction is obtained in one step using an aggregated covariance matrix as a dependent variable in equations (3.7) and (3.8). Empirical studies that compare direct and iterated forecasts provide contradicting evidence in favour of one model or another. MIDAS model's ability to directly predict multiperiod volatility for long horizons is documented in Ghysels et al. (2019). In Chiriac and Voev (2011), an iterated approach to HAR delivers better performance than a direct one in forecasting multiperiod covariance matrices. The current study considers both approaches and selects the most accurate for both models.

3.4.2 Bias correction

An additional challenge that needs to be addressed when forecasting the realized covariance matrices is the bias that arises due to the fact that forecasts are made for the logarithmic transformation of the covariance matrix and not for the original matrix. In other words, unbiased forecasts for the elements of matrix-logarithm do not necessarily result in unbiased forecasts for the original matrix. Therefore, the bias caused by the transformation should be

accounted for. Under certain assumptions, the expression for the bias can be theoretically derived and then adjusted for. We refer the reader to Bauer and Vorkink (2011) for the theoretical bias correction term for matrix-logarithm transformations. Given that the theoretical correction depends on the true parameters of the data generating process that are unknown and is valid under quite restrictive assumptions, including Gaussian distribution of the error terms, data-driven procedure is recommended in practical applications. In this paper, we use the bias correction procedure used in Chiriac and Voev (2011), who decompose a covariance matrix into a correlation matrix and a vector of volatilities, correct bias of volatility forecasts leaving correlations unchanged, and reconstruct the covariance matrix by combining the bias-corrected volatilities and the original predicted correlations. In order to correct the covariance matrix forecast at time $t+1$, bias correction terms $bc_i = \text{median}\left(\frac{\sqrt{H_{ii,k}}}{\sqrt{\widehat{H}_{ii,k}}}\right)$, $i \in \{1,\dots,n\}$, $k \in \{t-u, t-u+1, \dots, t\}$ are calculated and volatilities are rescaled correspondingly. Formally, u is a hyperparameter that needs to be selected. In practice, it is suggested to choose a u value equal to the estimation window. In this paper, we correct for the bias for both one-day-ahead and multiperiod forecasts. Further on, we illustrate improvement in the forecasting accuracy when the above-mentioned bias is taken into account and corrected for.

3.4.3 Forecasting errors

In this paper, we evaluate the predictive ability of competing models based on the four loss functions most commonly used in the literature:
mean squared error

$$\text{MSE}\left(H_{(t+1):(t+z)}, \widehat{H}_{(t+1):(t+z)|t}\right) = \frac{1}{\frac{n(n-1)}{2}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left(H_{ij,(t+1):(t+z)} - \widehat{H}_{ij,(t+1):(t+z)|t}\right)^2, \tag{3.9}$$

mean absolute error

$$\text{MAE}\left(H_{(t+1):(t+z)}, \widehat{H}_{(t+1):(t+z)|t}\right) = \frac{1}{\frac{n(n-1)}{2}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left|H_{ij,(t+1):(t+z)} - \widehat{H}_{ij,(t+1):(t+z)|t}\right|, \tag{3.10}$$

asymmetric loss function

$$\begin{aligned}\text{Asym}\left(H_{(t+1):(t+z)}, \widehat{H}_{(t+1):(t+z)|t}\right) = \frac{1}{b(b-1)} \text{tr}\left(H^b_{(t+1):(t+z)} - \widehat{H}^b_{(t+1):(t+z)|t}\right) \\ - \frac{1}{b-1} \text{tr}\left(\widehat{H}^{b-1}_{(t+1):(t+z)|t} \left(H_{(t+1):(t+z)} - \widehat{H}_{(t+1):(t+z)|t}\right)\right),\end{aligned} \tag{3.11}$$

and quasi-likelihood

$$\begin{aligned}\text{QLike}\left(H_{(t+1):(t+z)}, \widehat{H}_{(t+1):(t+z)|t}\right) = \operatorname{tr}\left(\widehat{H}^{-1}_{(t+1):(t+z)|t} H_{(t+1):(t+z)}\right) \\ - \log|\widehat{H}^{-1}_{(t+1):(t+z)|t} H_{(t+1):(t+z)}| - n,\end{aligned} \tag{3.12}$$

where a constant $b \geq 3$ in (3.11) defines the degree of asymmetry and is set equal to 3 in this paper which results in a mild degree of asymmetry, and a constant n in (3.12) stands for the number of assets and equals 29. The mean error in (3.9) and (3.10) is calculated for the lower triangular elements of a realized covariance matrix H_t including entries in the main diagonal in order to avoid the use of the same error terms twice in a loss function. It is important to note, that for these loss functions, errors for all elements of a matrix are calculated independently of each other. This is not the case for the other two measures, where all elements are included in the loss function and all elements are not assessed individually. As mentioned in Patton (2011), the MSE and MAE loss functions are common in the volatility forecasting literature. MAE is less sensitive to extreme observations while MSE is robust to imperfect volatility estimation. It is important to note, that loss functions in (3.12) and (3.11) are calculated in relative terms and penalize underprediction and overprediction in different ways. In contrast to the QLike loss function (3.12), which puts a bigger penalty on underprediction, asymmetric loss function (3.11) penalizes overprediction more heavily, while both of them have comparable degrees of asymmetry. As was shown in Patton (2011), MSE and QLike are two preferable measures for comparing competing forecasts of the realized volatilities. This is motivated by the fact that these measures are robust to noise in the estimate of the true conditional volatility, unless this volatility proxy is unbiased. However, these measures sometimes give preference to different forecasts. Moreover, it is important to note that the QLike loss function can be seen as Kullback–Leibler divergence (up to a constant which does not influence ordering of losses) between two multivariate Gaussian distributions with the same mean vectors and different covariance matrices. The behavior of this measure in cases when normality assumption is heavily violated has not been theoretically developed in the literature. Therefore, all above-mentioned measures have their strengths and weaknesses, and none of them are theoretically proven to be superior to others. In this paper, the four most commonly used measures are considered to have the full picture about the performance of the proposed model. For more details on the measures of forecasting accuracy see Amendola and Storti (2015), Becker et al. (2015), and Laurent et al. (2013).

3.5 Empirical study

3.5.1 Data

We consider the tick-by-tick prices of assets that contain 29 constituents of the Dow Jones Industrial Average (DJIA) index obtained from TAQ for the period from January 2001 to December 2017. For the full list of companies we refer to Appendix 3.A. This interval includes the crisis period from the last quarter of 2007 to the second quarter of 2009 (2007:Q4–2009:Q2) which is of particular interest due to increasing dependence among the returns and volatile behavior of the markets. Open-to-close returns are used to compute the realized covariance matrices and overnight returns are not taken into account. Before estimating the models, the dataset was cleaned according to Barndorff-Nielsen et al. (2009): The quotes with normal trading conditions, positive price and volume with the timestamp within trading hours of NYSE are used. After the cleaning procedure, the realized covariance matrices are estimated using the CholCov approach by Boudt et al. (2017) described in Section 3.2, which allows reducing the influence of microstructural noise.

3.5.2 Forecasting results

One-step-ahead forecasting

In this part of the paper, the HAR model (3.2) and the MIDAS model (3.5) are compared in terms of their ability to generate accurate one-step-ahead prediction. Table 3.A.2 of Appendix 3.B reports average values of estimated coefficients separately for diagonal and off-diagonal elements together with average goodness of fit measures. Dynamics of the estimated coefficients over time is presented in Figure 3.A.1 of Appendix 3.B for the HAR model and Figure 3.A.2 of Appendix 3.B for the MIDAS model. The left-hand side plots represent the behavior of the average estimated coefficients for the realized variances, and the right-hand side plots show the average estimated coefficients related to off-diagonal elements of the realized covariance matrix. An average is taken over all assets for volatilities and all pairs of assets in the case of covariances. For the HAR model, there is a strong increase in the contribution of daily, weekly, and monthly components to the prediction while the intercept tends to zero during the crisis. Moreover, for the HAR model, the coefficients assigned to off-diagonal elements are much more volatile than the coefficients of the realized variances. This is an expected result given the unstable behavior of the realized covariances. In contrast to the HAR model, for the MIDAS model, it is observed that the weights of distant lags of variances are strongly reduced during the crisis. Moreover, in

Table 3.1 Empirical bias correction and forecasting accuracy

Table 3.1 shows improvement of one-day-ahead forecasting accuracy for the realized covariance matrix obtained by empirical bias correction, 2003–2017.

Period	Loss	HAR		MIDAS	
		Original	Corrected	Original	Corrected
Crisis	MAE $\times 10^4$	1.2255	1.2103	1.1495	1.1245
	MSE $\times 10^6$	0.1584	0.1479	0.1452	0.1370
	Asym $\times 10^6$	4.0392	4.0787	3.8661	3.8644
	Qike	11.706	10.057	11.569	10.057
Non-crisis	MAE $\times 10^4$	0.2697	0.2721	0.2676	0.2544
	MSE $\times 10^6$	0.0061	0.0058	0.0059	0.0054
	Asym $\times 10^6$	0.0572	0.0566	0.0567	0.0558
	QLike	16.540	14.378	17.092	14.076

contrast to the HAR model, the MIDAS model does not necessarily put the largest weights on the most recent lags of covariances. This conclusion can be drawn by looking at the average estimated coefficients and the corresponding weighting profiles that are shown in Figure 3.1.

After both models are estimated, their performance is compared by computing the four most common loss functions described in Section 3.4.3: MAE, MSE, Asym, and QLike. For both competing models, the empirical bias correction procedure described in Section 3.4.2 is performed. Table 3.1 illustrates the mean losses (3.9)-(3.12) for the considered models before and after the empirical bias correction during the time period from 2003 to 2017. In the majority of cases, both in the crisis and in the non-crisis periods, it is observed that the empirical bias correction reduces the average forecasting error. Therefore, solely bias-corrected forecasts will be considered in the rest part of the paper.

Further on, the behavior of both competing models is studied in two dimensions: Time and industrial sector. First, a relative gain in accuracy over time is described. Second, the elementwise improvements for each element of the covariance matrix are considered. In order to try to investigate the stability of the results over time, a 200-day rolling window analysis is performed. Upper panels of Figure 3.2 illustrate the average loss ratios between the MIDAS and the HAR models for all considered loss functions. The loss ratios are below 1 for the majority of considered rolling windows, indicating that the MIDAS model helps to improve the forecasting accuracy. It is observed that the reduction in forecasting error compared to the HAR model frequently exceeds 10% based on MAE, MSE, and Asym

loss functions, whereas during the crisis, the error is reduced by 30-40%. For ease of interpretation, the crisis period is shaded in grey.

We further compare the performance of the HAR model to that of the MIDAS model and justify the statistical significance of the obtained improvement by applying the corrected Diebold-Mariano test by Harvey et al. (1997) and the fluctuation test by Giacomini and Rossi (2010). The fluctuation test by Giacomini and Rossi (2010) is implemented to address possible problems of multiple-testing and serial correlations among the forecasting errors. For both tests, a zero hypothesis suggests the equal prediction accuracy of the HAR and the MIDAS models, while the alternative hypothesis assumes that the MIDAS model is preferable.

The p-values presented in the lower panels of Figure 3.2 justify the statistical significance of improvement obtained by switching to the MIDAS model. The advantage of the MIDAS approach during the crisis is clearly confirmed by MAE and MSE loss functions. However, it is supported by QLike for short irregular time intervals. The MIDAS model beats the HAR model in terms of the Asym loss which is another kind of loss function that captures asymmetry. On the other hand, QLike loss function speaks in favor of the MIDAS model more frequently during the noncrisis period. A superior predicting ability of the MIDAS model according to three out of four loss functions is supported by the fluctuation test. In Figure 3.2, periods where the MIDAS model is favored according to the fluctuation test are marked in black at the bottom of plots.

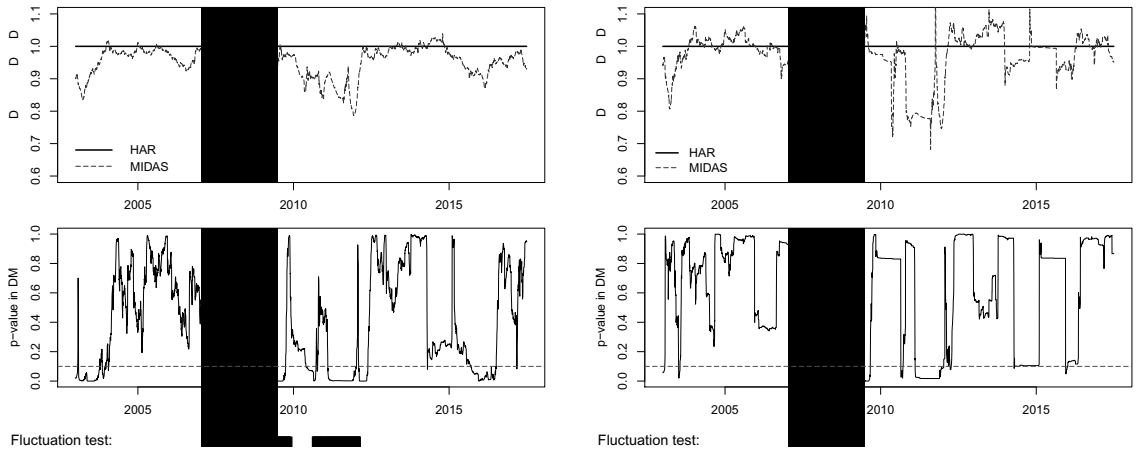
A possible explanation for sometimes controversial conclusions provided by different loss functions is attributable to their construction and is explained in Section 3.4.3. While some of the considered loss functions have been shown to perform better in practice, their superior performance is not theoretically proven. Therefore, an economic analysis is applied for the final justification in Section 3.5.3.

Further on, the gain in forecasting accuracy for the whole considered time period is studied separately for each element of the covariance matrix. This analysis is aimed at showing that the overall improvement provided by the MIDAS approach is homogeneous and is not driven by a few elements of a covariance matrix with certain characteristics.

An elementwise analysis employs MSE for each element of the matrix and a univariate version of QLike for diagonal elements. We do not present detailed results for the two other loss functions since MSE and QLike were shown to be the most robust and, therefore, preferable measures of forecasting accuracy for univariate comparisons (Patton (2011b)). We rely on the MCS procedure in order to clarify which models are included into the set of superior model in forecasting every element of a covariance matrix.

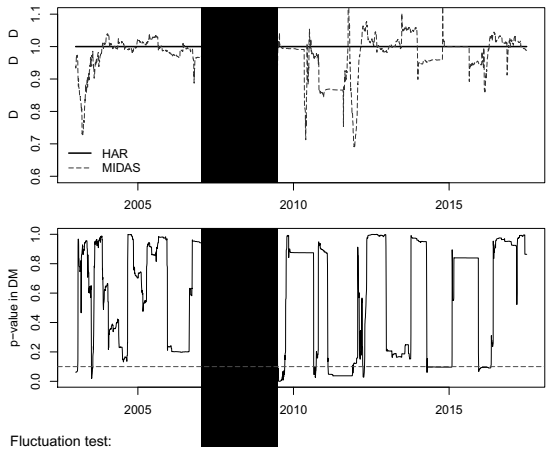
Figure 3.2 Four losses to compare the covariance matrix forecasts

Forecasts of the realized covariance matrix compared by four loss functions: MAE, MSE, Asym, and QLike, 2003–2017. *Upper plot in each pair*: The ratio of losses from the MIDAS model versus the HAR model. *Lower plot in each pair*: p-value in DM test for equal accuracy of two models in a moving window of 200 days. Alternative hypothesis: MIDAS model dominates. The crisis period (2007:Q4–2009:Q2) is shaded. Periods of significant improvement according to the fluctuation test by Giacomini and Rossi (2010) are marked at the bottom.

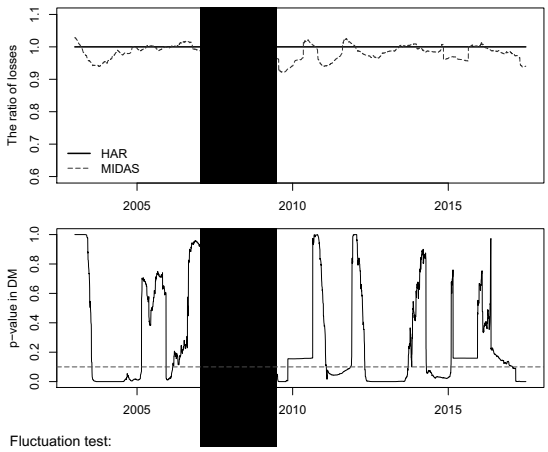


(a) MAE

(b) MSE



(c) Asym



(d) QLike

Figure 3.3 illustrates results of MCS analysis by Hansen et al. (2011) for each element of the covariance matrix based on MSE loss. To facilitate interpretation, we provide the tickers of the companies and the corresponding industries. Elements, for which the MIDAS model is the only model in the superior set, are shaded in white. Grey areas indicate that the HAR and the MIDAS models are both in the set of superior models. Black areas indicate a superior performance of the HAR model. Most elements are shaded in white which indicates that the MIDAS model is the only model in the superior set. 16 out of 17 unique

grey cells are related to financial companies which were mostly exposed to the market turbulence during the recent financial crisis. Therefore, the corresponding covariances are probably harder to predict. For more detailed study on the influence of financial sector on other sectors we refer to Audrino and Tetereva (2019).

In addition to the MSE-based results for all elements of the covariance matrix, Figure 3.3 presents MCS results based on the QLike loss function for diagonal elements of the matrix. The results are presented solely for diagonal elements due to the fact that one-dimensional version of QLike loss is mainly discussed in the literature in the context of volatilities and is not common for individual off-diagonal elements. For 18 out of 29 elements, both models perform equally well. In 9 out of 29 cases, the MIDAS model outperforms the HAR and, in 2 cases out of 29, the HAR model is the only model in the set of superior models.

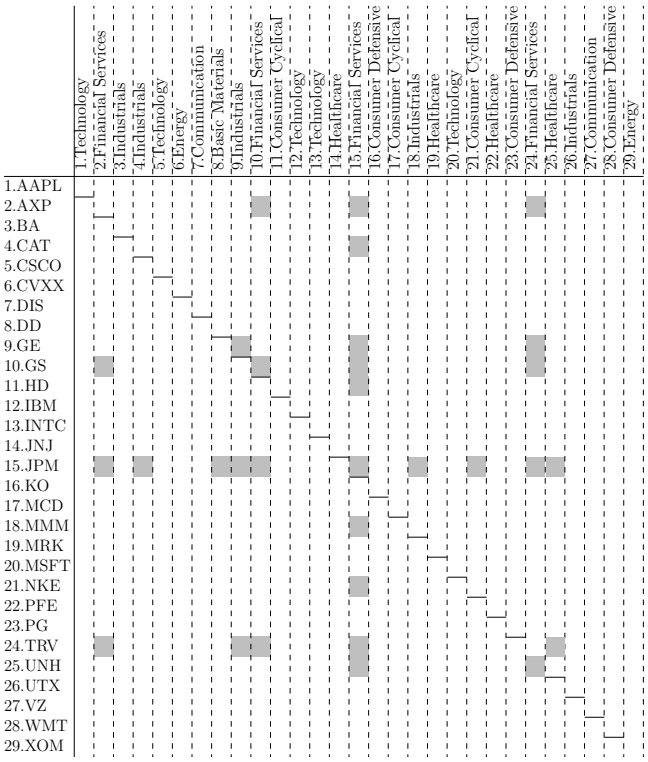
Before turning to multiperiod forecasting, we attempt to explain the improvement obtained by the MIDAS model. As it is well-documented in the literature, the MIDAS approach performs better in forecasting the realized volatilities. This can be explained by taking a closer look at the estimated coefficients presented in Figure 3.A.2 of Appendix 3.B and comparing them to the weighting profiles in Figure 3.1. For volatilities, in contrast to the HAR model, the MIDAS model is able to capture empirical evidence showing that, during the crisis, only recent lags are relevant for making predictions. Such a behavior of coefficients was explicitly justified by Audrino and Knaus (2016) who applied Lasso procedure to select the relevant lags of AR model for the realized volatilities. It was shown that, during the crisis, only the most recent lags are selected by the Lasso procedure. The strength of the MIDAS model is smoothness of its weighting scheme that allows to reduce the weights of distant lags and assign decaying weights to the most recent ones. The HAR model catches up with the MIDAS approach in forecasting volatilities because it is able to replicate this weighing scheme quite well, when β^D is large, β^M is smaller and β^M is close to zero. However, during the crisis, estimated weights related to all lags are increasing which deteriorates forecasting abilities of the HAR model.

Smoothness of the MIDAS weights is even more important and plays a key role in forecasting the covariances. Figure 3.1b exhibits common weighting profiles of the MIDAS regression in modelling covariances. It is harder for the HAR model to suggest a good approximation given its structure, although it tries to construct a similar pattern. Additional advantage of the MIDAS model is smaller variation of its coefficients and more stable forecasts. By construction, the MIDAS model assigns all positive or all negative weights to past realizations of covariances; the sign is governed by β coefficient. This might stabilize the estimated coefficients and result in a higher forecast accuracy. Rather often, coefficients

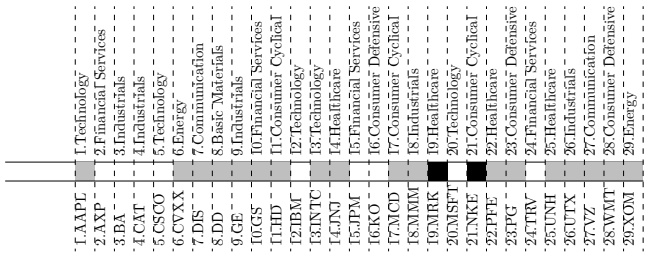
Figure 3.3 Forecasting accuracy for elements of the covariance matrix

Improvement of one-day-ahead forecasting accuracy in terms of MSE and QLike for each element of the realized covariance matrix. Pairs (elements) for which MIDAS is the only model in the set of superior models according to MCS are marked in white; elements for which MIDAS and HAR are both in the set of superior models are marked in grey; elements for which HAR is the only model in the set of superior models are marked in black.

(a) MSE



(b) QLike



of the HAR model take opposite signs which might be attributed to an effort to compensate the lack of flexibility caused by a piecewise constant weighting scheme and results in the lack of stability. Moreover, such behavior of coefficients is harder to interpret.

Multiperiod forecasting

As we have mentioned, multiperiod forecasting of covariances is highly demanded in applications. We compare the performance of the HAR and MIDAS models in the forecasting of the realized covariance matrices by using direct and iterated approaches described in Section 3.4.1. In total, four models are considered: HAR direct (I), HAR iterated (II), MIDAS direct (III), and MIDAS iterated (IV). In order to evaluate the performance of the proposed method, we construct the MCS approach by Hansen et al. (2011). For multiperiod forecasts, we switch from the corrected Diebold-Mariano test approach by Harvey et al. (1997) and fluctuation test by Giacomini and Rossi (2010) to MCS approach to avoid pairwise comparisons for all possible combinations of the models and being able to compare all of them in one step. Table 3.2 compares the accuracy of predictions for the covariance matrices obtained using the considered models. When constructing multiperiod forecasts, we follow guidelines laid out in Section 3.4.1. For completeness, Table 3.2 reports MCS results for one-step-ahead forecasts. Based on the MCS results, it can be concluded that, in the majority of cases, the iterated MIDAS model and the direct HAR remain in the superior set. It is important to note that this is purely an empirical result, and superior performance of one or another approach is not justified theoretically. Therefore, both methods should be considered in practice. It is worth mentioning that MCS-based analysis confirms our prior result for one-step ahead forecasting: the MIDAS regression is the only model in the superior set for almost all cases. For longer horizons, it is more difficult to distinguish among the forecasting abilities of the models. It is in line with the theoretical consideration that the long horizon forecasts provided by all possible models converge to the unconditional mean of the process. Therefore, all models deliver more similar forecasts with the increasing time horizon. Even when several models are included in the set of the best models, the mean error of the iterated MIDAS model (IV) is smaller for the majority of cases. From the above-mentioned analysis, it is evident that the iterated MIDAS model is performing at least as good as the direct HAR model and is outperforming it based on the majority of considered loss functions for shorter forecasting horizons. This conclusion is true for the whole covariance matrix.

It is important to note that superior performance of the MIDAS model is observed not only for the realized covariance matrices but also for the realized correlation matrices. More precise predictions of the correlation matrices are obtained when they are modeled via

covariances, not from past realizations of a correlation matrix: Forecasting the realized covariance matrix and transforming it to the realized correlation matrix results in higher accuracy. This conclusion is consistent with the approach used in González et al. (2012) who suggest to model correlations via covariances and variances rather than directly from correlations. We omit this analysis in order to stay focused on the covariance matrix. Correlation-related results can be obtained from the authors upon request.

3.5.3 Portfolio management application

Portfolio management application: efficient frontier

We now assess the economic value of the realized covariance matrix forecasting when it is conducted using MIDAS. An investor allocating her funds in 29 DJIA components is considered. We are interested in finding an efficient frontier, that is, a subset of the investment opportunity set with the smallest portfolio risk $\sigma_{\text{portfolio}}$ for the required level of portfolio returns μ_p. This task can be formulated as a minimization problem with constrains at time t for z-days investment period:

$$\begin{aligned} \min_{\boldsymbol{w}} \sigma^2_{\text{portfolio}}(\boldsymbol{w}) &= \boldsymbol{w}^\top H_{(t+1):(t+z)} \boldsymbol{w} \\ \text{s.t. } \mu_{\text{portfolio}} &= \boldsymbol{w}^\top \boldsymbol{\mu}^e_{\text{portfolio},(t+1):(t+z)},\ \boldsymbol{w}^\top \boldsymbol{\iota} = 1, \end{aligned} \tag{3.13}$$

where $\boldsymbol{w}$ denotes the weights of individual stocks in the portfolio that add up to one, $\boldsymbol{\mu}^e_{\text{portfolio},(t+1):(t+z)}$ is a vector of 29 expected returns, and $\boldsymbol{\iota}$ is a vector with entries of 1. The expected returns are computed as mean returns for a z-day period during the previous 100 days. $H_{(t+1):(t+z)}$ is replaced by a predicted realized covariance matrix for a z-day-ahead investment period. Prediction is obtained by either direct HAR model or iterated MIDAS model described in Section 3.2.

In our framework, assets are allocated daily with an investment horizon of z days. Solution to system of equations (3.13) is a vector of weights $\boldsymbol{w}_H$ or $\boldsymbol{w}_M$, where $\boldsymbol{w}_H$ are the weights suggested by the HAR model and $\boldsymbol{w}_M$ are weights obtained using MIDAS predicted covariance matrix. Later, these weights are used to determine the actual portfolio returns as $r_{\text{portfolio},H} = \boldsymbol{w}_H^\top \boldsymbol{\mu}_{(t+1):(t+z)}$ and $r_{\text{portfolio},M} = \boldsymbol{w}_M^\top \boldsymbol{\mu}_{(t+1):(t+z)}$, where $\boldsymbol{\mu}_{(t+1):(t+z)}$ is a vector of 29 actual returns. For every level of required returns we report the mean actual portfolio returns and corresponding risk (standard deviation of actual portfolio returns). Superior performance of a MIDAS model is pronounced for the one-day-ahead prediction and longer investment periods, see Figure 3.4.

Table 3.2 Forecasting the realized covariance matrices in 2003–2017.

Period	Performance Measure	HAR		MIDAS		MCS testing	
		Direct (I)	Iterat. (II)	Direct (III)	Iterat. (IV)	Superior set	p-value
Panel 1: 1-day-ahead forecasting, by construction, "Direct" is equivalent to "Iterated"							
Crisis	MAE $\times 10^4$	1.2103		1.1245		(MIDAS)	0.025
	MSE $\times 10^6$	0.1479		0.1370		(MIDAS)	0.085
	Asym $\times 10^6$	4.0787		3.8644		(MIDAS)	0.125
	QLike	10.057		10.057		(MIDAS),(HAR)	0.993
Non-crisis	MAE $\times 10^4$	0.2528		0.0251		(MIDAS)	0.001
	MSE $\times 10^6$	0.0052		0.0049		(MIDAS)	0.087
	Asym $\times 10^6$	0.0585		0.0579		(MIDAS)	0.111
	QLike	14.288		14.018		(MIDAS)	0.001
Panel 2: 5-day-ahead forecasting							
Crisis	MAE $\times 10^4$	8.1385	8.4148	10.645	7.5096	(IV)	0.089
	MSE $\times 10^6$	3.7131	4.0049	6.2154	3.6099	(IV),(I)	0.579
	Asym $\times 10^6$	205.01	215.80	266.42	202.94	(IV),(I),(II)	0.254
	QLike	5.2973	6.3219	15.002	6.2301	(I)	0.018
Non-crisis	MAE $\times 10^4$	1.1105	1.1355	1.2421	1.0665	(IV)	0.166
	MSE $\times 10^6$	0.0902	0.0917	0.1041	0.0856	(IV)	0.143
	Asym $\times 10^6$	0.8298	0.8298	0.8769	0.7924	(IV)	0.104
	QLike	5.7497	7.2427	7.5974	7.0166	(I)	0.001
Panel 3: 10-day-ahead forecasting							
Crisis	MAE $\times 10^4$	17.330	18.293	20.360	16.099	(IV)	0.244
	MSE $\times 10^6$	13.582	15.401	0.2048	13.387	(IV),(I)	0.811
	Asym $\times 10^6$	1201.9	1323.3	1469.8	1200.6	(IV),(I),(III),(II)	0.476
	QLike	5.8009	7.8549	13.482	7.5296	(I),(II),(IV)	0.018
Non-crisis	MAE $\times 10^4$	2.1274	2.2557	2.3377	2.1326	(I),(IV)	0.946
	MSE $\times 10^6$	0.2612	0.2804	0.2986	0.2546	(IV),(I)	0.617
	Asym $\times 10^6$	3.4378	3.6185	3.6493	3.2589	(IV),(I)	0.276
	QLike	4.5361	6.5404	5.8727	6.3243	(I)	0.000
Panel 4: 22-day-ahead forecasting							
Crisis	MAE $\times 10^4$	42.458	46.343	45.629	40.268	(IV),(I),(III)	0.413
	MSE $\times 10^6$	67.957	81.477	85.856	71.830	(I),(IV),(III),(II)	0.415
	Asym $\times 10^6$	10753	12365	12540	11299	(I),(IV),(III),(II)	0.466
	QLike	6.7804	11.210	10.598	11.172	(I),(III),(IV)	0.365
Non-crisis	MAE $\times 10^4$	4.6269	5.2631	4.9779	4.9863	(I)	0.140
	MSE $\times 10^6$	1.0132	1.3260	1.0965	1.0195	(I),(IV)	0.957
	Asym $\times 10^6$	26.098	36.999	24.629	21.219	(IV),(I),(III)	0.580
	QLike	3.6294	6.4359	4.6357	6.3273	(I)	0.012

Note: p-value corresponds to the last test in the sequence of MCS tests. Following Laurent et al. (2012) and Chevallier et al. (2019, p. 273), the significance level in all tests is set to $\alpha = 0.25$ for 10000 bootstrap samples.

Figure 3.4 Efficient frontiers for the crisis and non-crisis periods

Figure 3.4 exhibits efficient frontiers for the investment horizons of 1, 5, 10 and 22 days. The curves are plotted for the crisis period (2007:Q4–2009:Q2) in the left column and the non-crisis period (2003–2007:Q3, 2009:Q3–2017) in the right column.

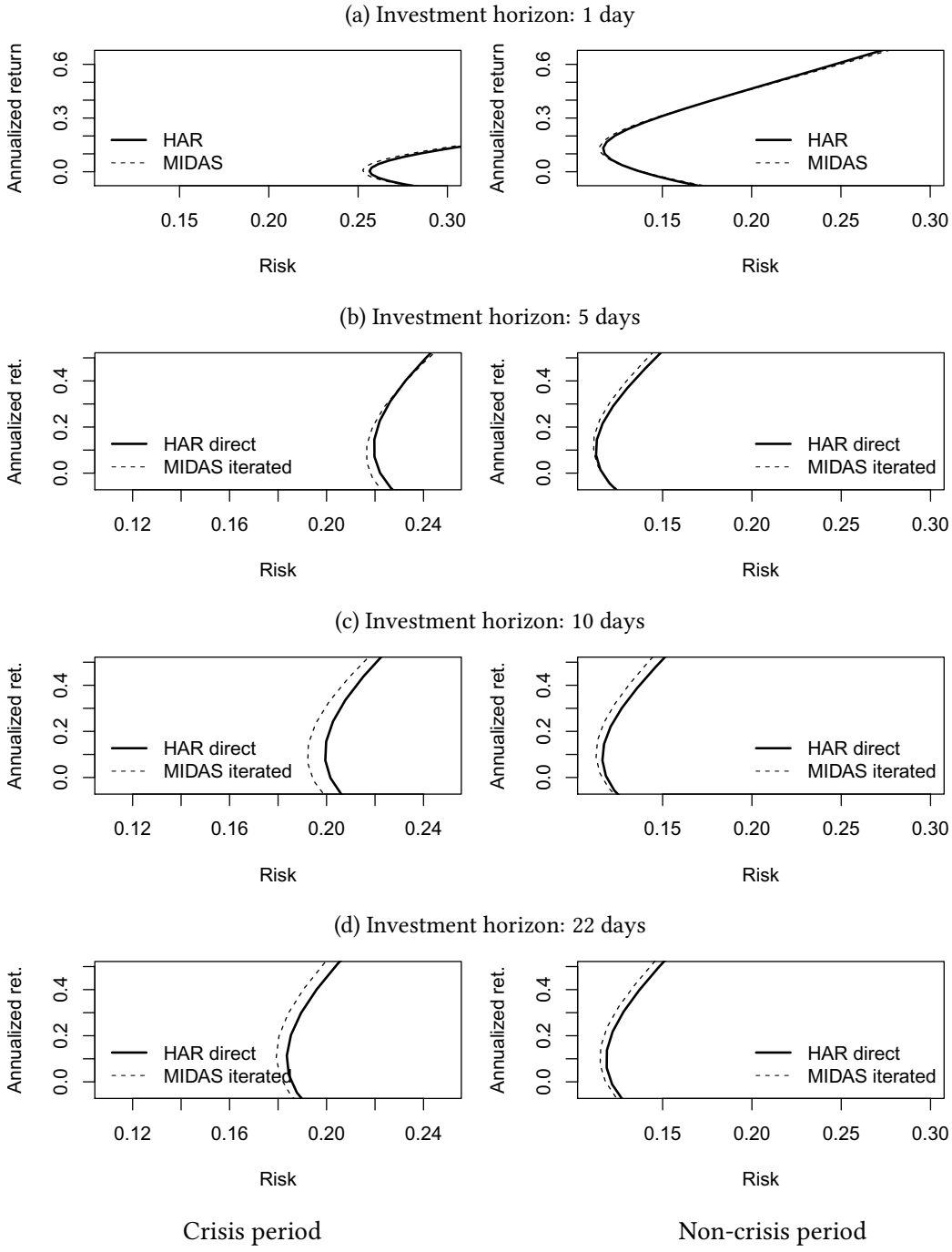


Table 3.3 Performance of GMV portfolio investment

	Crisis period			Non-crisis period		
	Annualized return	Standard deviation	Sharpe ratio	Annualized return	Standard deviation	Sharpe ratio
Panel 1: Investment period of 1 day						
HAR	−0.04853	0.2703	−0.179	0.104233	0.1181	0.882
MIDAS	−0.03652	0.2679	−0.136	0.11116	0.1154	0.963
change	+1.201 pp			+0.692 pp		
Panel 2: Investment period of 5 days						
HAR direct	−0.0562	0.2207	−0.255	0.0976	0.1159	0.842
MIDAS iterated	−0.0330	0.2286	−0.144	0.1027	0.1096	0.937
change	+2.319 pp			+0.509 pp		
Panel 3: Investment period of 10 days						
HAR direct	−0.0601	0.2025	−0.297	0.0959	0.1213	0.791
MIDAS iterated	−0.0282	0.2116	−0.133	0.0984	0.1121	0.878
change	+3.193 pp			+0.248 pp		
Panel 4: Investment period of 22 days						
HAR direct	-0.0541	0.1887	−0.287	0.0926	0.1317	0.703
MIDAS iterated	0.0119	0.1940	0.061	0.1010	0.1171	0.862
change	+6.607 pp			+0.834 pp		

Note: The difference in annualized returns between the MIDAS model and the HAR model is reported in row "change" in percentage points (pp).

Portfolio management application: global minimum variance

An efficient frontier is affected by the use of expected returns with large estimation errors. For this reason, we consider another common application: Investing in the global minimum variance (GMV) portfolio that is based barely on a covariance matrix forecast. The investor minimizes the portfolio variance as follows:

$$\begin{aligned} \min_{\boldsymbol{w}} \sigma^2_{\text{portfolio}}(\boldsymbol{w}) &= \boldsymbol{w}^\top H_{(t+1):(t+z)} \boldsymbol{w} \\ \text{s.t. } \boldsymbol{w}^\top \boldsymbol{\iota} &= 1. \end{aligned}$$

The optimal solution has a closed form

$$\boldsymbol{w}^* = \frac{H^{-1}_{(t+1):(t+z)}\boldsymbol{\iota}}{\boldsymbol{\iota}^\top H^{-1}_{(t+1):(t+z)}\boldsymbol{\iota}}.$$

Table 3.3 reports the annualized mean return during the investment period and the corresponding risk. Clearly, the MIDAS model provides an investor with a more preferable GVM portfolio performance. This result is true for all considered investment horizons and is especially pronounced during the crisis period. For example, using the MIDAS model instead of the HAR model during the crisis allows an investor to switch from negative return to positive when investment horizon is equal to 22 days.

The presented above results are justified by measuring the economic value of more accurate predictions of the realized covariance matrix in the context of the investment decision making by a risk averse investor. Using the utility-based framework of Fleming et al. (2001) and Fleming et al. (2003) we measure the economic value of the performance gains associated with using the MIDAS approach for the realized covariance matrix forecasting. We consider an investor with quadratic utility and risk aversion γ_{risk} who each day invests into the GMV portfolio suggested by the direct HAR and iterated MIDAS models with respective portfolio returns $r_{\text{portfolio},H}$ and $r_{\text{portfolio},M}$. The daily utility generated by this portfolio is

$$U(r_{\text{portfolio}}, \gamma_{\text{risk}}) = (1 + r_{\text{portfolio}}) - \frac{\gamma_{\text{risk}}}{2(1+\gamma_{\text{risk}})}(1 + r_{\text{portfolio}})^2.$$

Investor's utility earned during a periods of T days is equal to the sum of daily utilities. Δ is introduced as the maximum returns the investor is ready to sacrifice for switching from GVM portfolio suggested by the direct HAR model to GVM portfolio suggested by the iterated MIDAS approach. In other words, the economic gain from using the MIDAS model comparing to the HAR model in terms of returns is calculated by solving

$$\sum_{t=1}^{T} U(r_{\text{portfolio},M} - \Delta, \gamma_{\text{risk}}) = \sum_{t=1}^{T} U(r_{\text{portfolio},H}, \gamma_{\text{risk}}).$$

for Δ. Table 3.4 contains the values of Δ computed with standard in the literature coefficients of risk aversion $\gamma_{\text{risk}} = 1$ and $\gamma_{\text{risk}} = 10$ (Fleming et al. (2003), Bollerslev et al. (2018)). The obtained values of the performance gains are consistent with the results presented in Table 3.3.

Table 3.4 Economic value of switching to the MIDAS model, annualized Δ.

Investment horizon	Crisis period		Non-crisis period	
	$\gamma_{\text{risk}} = 1$	$\gamma_{\text{risk}} = 10$	$\gamma_{\text{risk}} = 1$	$\gamma_{\text{risk}} = 10$
1 day	1.326%	1.907%	0.659%	0.948%
5 days	2.274%	0.681%	0.536%	1.189%
10 days	3.195%	1.496%	0.335%	1.351%
22 days	6.843%	5.922%	0.955%	2.757%

3.6 Conclusion

In this study, a novel approach has been introduced for the prediction of the realized covariance matrix. We propose the use of a MIDAS regression with revised optimization to predict the individual elements of the realized covariance matrix. In order to obtain a valid prediction, the matrix logarithm transformation is made in analogy with Bauer and Vorkink (2011), and the empirical bias correction procedure is applied. The suggested MIDAS regression is superior to the HAR model due to its flexibility and smoothness. The parsimonious structure of the MIDAS model keeps the danger of overfitting low. It is important to note, that the suggested estimation procedure makes it possible to estimate the MIDAS model for high-dimensional portfolios with moderate computational costs.

The empirical analysis is performed using 29 constituents of the DJIA index. The considered time period ranges from January 2001 to December 2017. Our results provide a comprehensive comparison of the performance of the proposed MIDAS approach with the benchmark HAR model in one-step-ahead and multistep forecasting. The models are compared using the corrected Diebold-Mariano test by Harvey et al. (1997), the fluctuation test by Giacomini and Rossi (2010), and the MCS procedure by Hansen et al. (2011) based on MSE, MAE, Asym, and QLike loss functions. The results provide evidence for the fact that the proposed model and the estimation procedure significantly improve the prediction of the realized covariance matrix for one-day and multiperiod horizons. In one-step-ahead forecasting, the considered statistical methods unanimously support the predominance of MIDAS. The MIDAS model provides a sound improvement of 30-40% in forecasting accuracy when most needed, during crisis. This result holds in an element-wise analysis of realized covariances, although the performance for diagonal elements is less prominent. In multiperiod forecasting, direct and iterated forecasting methods were applied with the

HAR and MIDAS models. Based on the MCS results, the iterated MIDAS model is found to perform at least as well as the direct HAR model and to outperform it for shorter horizons.

The superior performance of the proposed MIDAS model is economically justified by applying it to the portfolio management problem. The empirical application of the suggested method to the efficient frontier and the GMV problems shows the validity of the results for both, crisis and non-crisis periods. The results are especially good for the crisis period, when the gain of switching to the MIDAS model is always positive and has a maximum value of almost 7 percentage points of annualized return.

Appendices

Appendix 3.A List of 29 companies

Table 3.A.1 29 Dow Jones stocks which are used in the empirical study

	Ticker	Company name
1	MMM	3M
2	AXP	American Express
3	AAPL	Apple
4	BA	Boeing
5	CAT	Caterpillar
6	CVX	Chevron
7	CSCO	Cisco
8	KO	Coca-Cola
9	DIS	Disney
10	DD (DWDP)	DowDuPont Inc
11	XOM	Exxon Mobil
12	GE	General Electric
13	GS	Goldman Sachs
14	HD	Home Depot
15	IBM	IBM
16	INTC	Intel
17	JNJ	Johnson & Johnson
18	JPM	JPMorgan Chase
19	MCD	McDonald's
20	MRK	Merck
21	MSFT	Microsoft
22	NKE	Nike
23	PFE	Pfizer
24	PG	Procter & Gamble
25	TRV	Travelers Companies Inc
26	UTX	United Technologies
27	UNH	UnitedHealth
28	VZ	Verizon
29	WMT	Walmart

Appendix 3.B Estimated models and coefficients

Figure 3.A.1 Estimated coefficients in the HAR model

Figure 3.A.1 shows mean values of estimated coefficients in the rolling windows of 200 days for the HAR model.

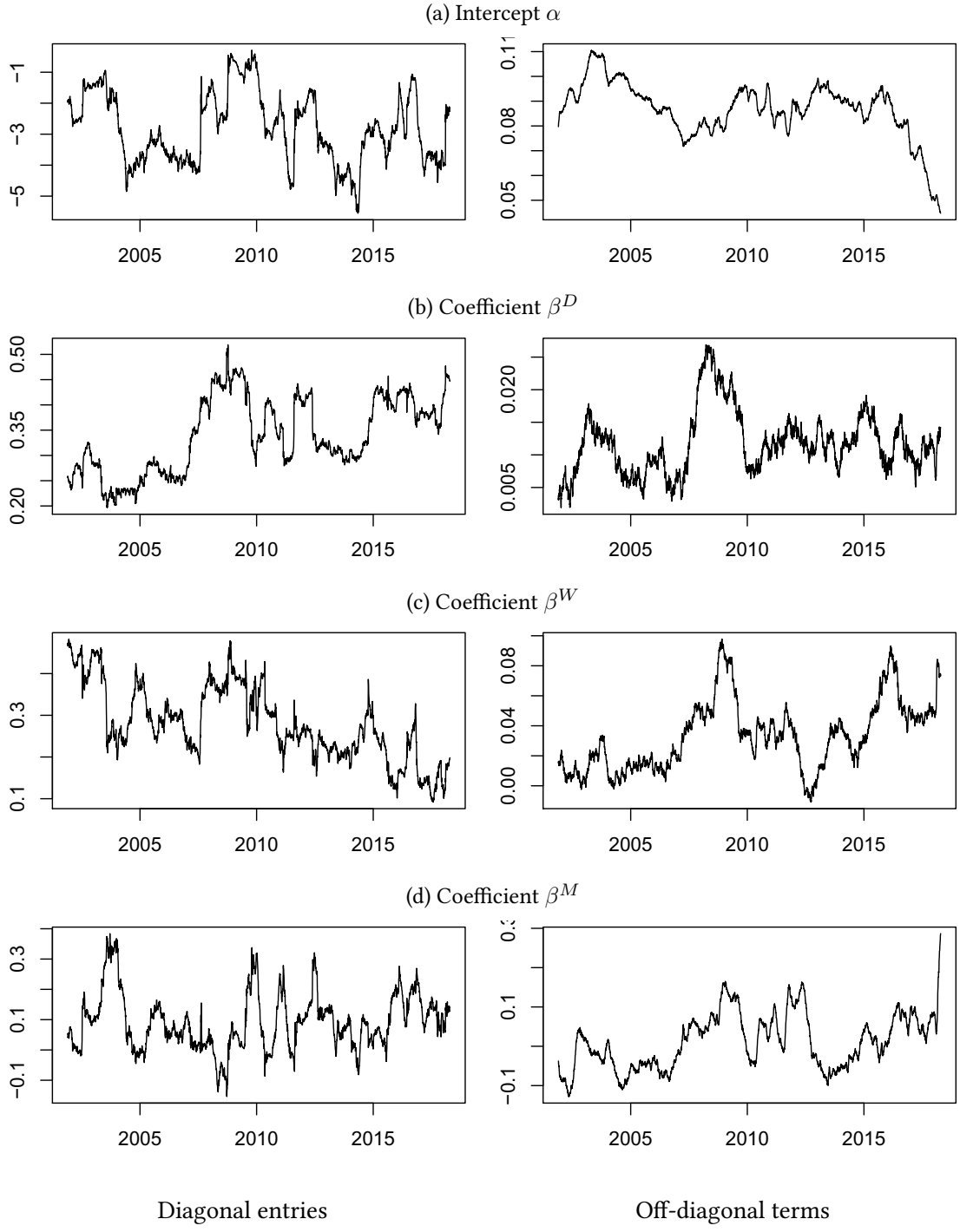
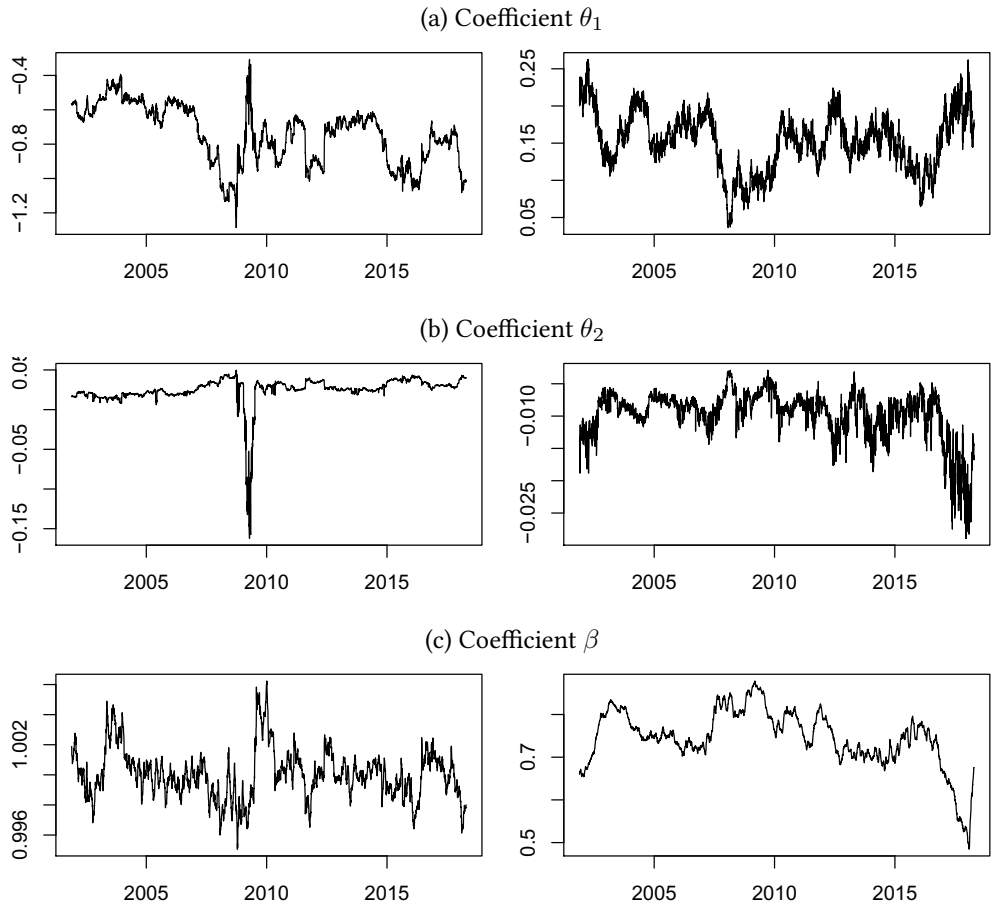


Table 3.A.2 Descriptive statistics for estimated coefficients

Estimated coefficients averaged across all moving windows. Parameters for the MIDAS and HAR models are specified in equations (3.5) and (3.2), respectively.

	The MIDAS model								The HAR model							
	Diagonal terms				Off-diagonal terms				Diagonal terms				Off-diagonal terms			
	Mean	S.d.	Q_{10}	Q_{90}	Mean	S.d.	Q_{10}	Q_{90}	Mean	S.d.	Q_{10}	Q_{90}	Mean	S.d.	Q_{10}	Q_{90}
θ_1	-0.74	0.34	-1.09	-0.44	0.22	0.87	-0.32	1.09								
θ_2	0.02	0.06	0.01	0.04	-0.02	0.20	-0.05	0.01								
β	0.99	0.05	0.98	1.02	0.85	0.15	0.69	0.97								
c									-2.80	1.74	-4.92	-0.91	0.09	0.05	0.03	0.15
β^d									0.34	0.12	0.19	0.49	0.01	0.08	-0.09	0.12
β^w									0.28	0.16	0.06	0.49	0.03	0.20	-0.23	0.29
β^m									0.09	0.18	-0.12	0.31	0.01	0.40	-0.50	0.49
SER	0.49	0.06	0.42	0.58	0.14	0.02	0.12	0.16	0.48	0.06	0.41	0.57	0.14	0.02	0.12	0.16
R^2	0.31	0.18	0.10	0.56	-0.02	0.04	-0.05	0.02	0.34	0.16	0.15	0.57	0.02	0.03	0.00	0.04

Figure 3.A.2 Averaged coefficients in the MIDAS model: Rolling window of 200 days



Chapter 4

Estimating MIDAS regressions via MIDAS-NLS with revised optimization

Alexander Kostrov [1]

[1]University of St. Gallen, Faculty of Mathematics and Statistics, alexander.kostrov@unisg.ch

Abstract

This paper examines a computational problem inherent to existing estimators for MIDAS regressions: In some cases, attempts to estimate the parameters take considerable time and lead to inaccurate solutions. We propose a few innovations, revise an optimization procedure for what is called a MIDAS-NLS estimator to reach a major enhancement in its performance. A simulation study with a focus on Beta lag polynomials confirms that the suggested approach is more accurate and mitigates the numerical pitfalls of standard methods. MIDAS-NLS with revised optimization is particularly beneficial in big data situations and in applications involving multiple estimations of a MIDAS regression. In order to facilitate the use of our approach, we have released the "MIDAS Analytic" extension to the MATLAB MIDAS Toolbox.

4.1 Introduction

In recent years, econometric models that deal with time series sampled at different frequencies have gained considerable attention. This is partially attributed to the emergence of high-frequency data that introduced new challenges in methodological and empirical work. In particular, MIDAS (mixed data sampling) regressions pioneered by Ghysels et al. (2004), in which a lower frequency variable of interest is conditioned on the relevant information observed at higher frequencies, have flourished over the last two decades. Due to its parsimony and flexibility, the MIDAS approach exhibits particularly sound results in Macroeconomics and Financial Econometrics. Its success originates from the smoothness of relations between the economic series: The idea of MIDAS is to assign smooth weights to past realizations of an explanatory variable using a weighting function governed by a few parameters only (Ghysels et al. (2007) and Ghysels et al. (2009)).

In order to run a MIDAS regression, an optimization problem should be solved, in which the objective function is nonlinear in parameters. Typically, the coefficients are estimated via nonlinear least squares (NLS) with good asymptotic properties derived in Andreou

et al. (2010). As a state of the art approach, it is used by the existing statistical software including the MATLAB MIDAS Toolbox[2]. However, Ghysels and Qian (2019) report that the known numerical pitfalls of the MIDAS-NLS estimator render it ill-behaved in some applications: The authors reveal a heavy computational burden and inaccurate estimation of the parameters. The literature on solving this issue is rather limited. Ghysels and Qian (2019) propose a profiling estimator that makes a step forward. Fixing a grid of possible values for the polynomial parameters eliminates nonlinearity in the objective function and paves the way for OLS estimation. Recently, a profiling estimator has been applied for nowcasting tail risk to economic activity using MIDAS (Carriero et al., 2020). The profiling approach inherits certain shortcomings attributed to its grid search origin: There is a trade-off between runtime and accuracy, and under certain conditions, the estimation is not feasible computationally. So far, the computational burden is an obstacle to take advantage of MIDAS in practice, e.g., in the context of volatility forecasting using a rolling window regression analysis in Bollerslev et al. (2018).

In this paper, we mitigate the numerical drawbacks of the original NLS estimation. In MIDAS-NLS with revised optimization, we vectorize the optimization problem that needs to be solved at the time of estimation and the corresponding computations. Our optimization technique relies on analytic derivatives. To promote estimating MIDAS regressions via MIDAS-NLS with revised optimization we release the "MIDAS Analytic extension"[3] to the MIDAS MATLAB Toolbox which currently supports several MIDAS models. The focus of this paper is on a univariate MIDAS regression with Beta lags.

We demonstrate performance of MIDAS-NLS with revised optimization using simulated and real data. In a simulation study, three tools are compared in terms of computational time and accuracy of the solution: The original implementation of MIDAS-NLS in the MATLAB MIDAS Toolbox, OLS with polynomial parameter profiling, and MIDAS-NLS with revised optimization, a novel approach we propose. Our focus is specifically on the Beta lag polynomial which is commonly used in the MIDAS literature. The computational results reveal that the suggested method is simultaneously many times faster and more accurate as compared to the existing methods. Gauging the sources of improvement, we discover that the use of analytic derivatives in optimization makes the major contribution. Another strong side of our approach is that it allows relaxing restrictions on parameters in the MIDAS weighting scheme, which are not always supported by empirical evidence. It is worth mentioning that MIDAS-NLS with revised optimization works well in a big data

[2]The MIDAS MATLAB Toolbox is available on the MATLAB Central File Exchange (No. 45150).
[3]The "MIDAS Analytic" extension is available on the MATLAB Central File Exchange (No. 81243).

setting, where even a small error in the estimated Beta polynomial parameters becomes critical. We show that our approach retains its advantages when applied to realized volatility forecasting.

The paper proceeds as follows. Section 4.2 introduces the MIDAS model of interest and existing estimators. Subsequently, we explain MIDAS-NLS with revised optimization, a method we propose. Section 4.3 contains a simulation study with our comments on results and key findings. In Section 4.4, we apply a novel approach to realized volatility forecasting. The final section concludes this paper.

4.2 Reconsidering the MIDAS-NLS estimation

4.2.1 MIDAS regressions

In this paper, the focus is on the factor-augmented MIDAS model with the Beta polynomials in which a lower frequency outcome variable is predicted using the relevant information observed at higher frequency. Following the standard notations, a MIDAS regression with a single MIDAS-weighted regressor and factors is expressed as

$$y_{t+1}^{L} = \alpha + \sum_{j=1}^{k} \rho_j v_{j,t}^{L} + \beta w(L^{1/m}, \theta_1, \theta_2, \theta_3) z_t^{H} + \epsilon_{t+1}^{L}, \tag{4.1}$$

where the explanatory variable z^H is sampled m times per a low-frequency period, $w(L^{1/m}, \theta_1, \theta_2, \theta_3)$ is the Beta weight function, and v_j^L, $j \in \{1, \dots, k\}$ are factors. For the sake of generality, we switch to the matrix representation of (4.1) and omit time subscripts:

$$\boldsymbol{y} = V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2, \theta_3) + \boldsymbol{\epsilon}, \tag{4.2}$$

where weights $\boldsymbol{w}$ are assigned using the three-parameter Beta polynomials:

$$\begin{aligned} &\boldsymbol{w}(\theta_1, \theta_2, \theta_3) = \boldsymbol{x}^{\theta_1 - 1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2 - 1} \cdot \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} + \theta_3, \\ &\boldsymbol{x} = [1, 2, 3 \dots, p]^\top / (p+1), \text{and} \\ &\Gamma(\xi) = \int_0^\infty \exp^{-t} t^{\xi - 1}\, dt \text{ – the Gamma function.} \end{aligned} \tag{4.3}$$

The matrix V contains a column-vector of ones and factors. Columns of the matrix Z are p lagged vectors z^H, where p is the number of (high-frequency) MIDAS lags. This regression

Table 4.1 Beta lag polynomials

Table 4.1 explains Beta lag polynomials with restrictions on parameters (fixing thetas to some values) which are commonly used in the MIDAS literature.

Parameter	Unrestricted case	Restricted cases		
		#1	#2	#3
$\theta_1 :=$	θ_1	1	θ_1	1
$\theta_2 :=$	θ_2	θ_2	θ_2	θ_2
$\theta_3 :=$	θ_3	θ_3	0	0

model (4.2) has the expanded representation (4.4):

$$\begin{bmatrix} y_1^L \\ y_2^L \\ \vdots \\ y_T^L \end{bmatrix} = \begin{bmatrix} 1 & v_{1,0}^L & \cdots & v_{k,0}^L \\ 1 & v_{1,1}^L & \cdots & v_{k,1}^L \\ \vdots & \vdots & \ddots & \vdots \\ 1 & v_{1,T-1}^L & \cdots & v_{k,T-1}^L \end{bmatrix} \begin{bmatrix} \alpha \\ \rho_1 \\ \vdots \\ \rho_k \end{bmatrix} + \beta \begin{bmatrix} z_0^H & z_{0-\frac{1}{m}}^H & \cdots & z_{0-\frac{p-1}{m}}^H \\ z_1^H & z_{1-\frac{1}{m}}^H & \cdots & z_{1-\frac{p-1}{m}}^H \\ \vdots & \vdots & \ddots & \vdots \\ z_{T-1}^H & z_{T-1\frac{1}{m}}^H & \cdots & z_{T-1-\frac{p-1}{m}}^H \end{bmatrix} \begin{bmatrix} w_1(\theta_1,\theta_2,\theta_3) \\ w_2(\theta_1,\theta_2,\theta_3) \\ \vdots \\ w_p(\theta_1,\theta_2,\theta_3) \end{bmatrix} + \begin{bmatrix} \varepsilon_1^L \\ \varepsilon_2^L \\ \vdots \\ \varepsilon_T^L \end{bmatrix}, \quad (4.4)$$

where T is the number of (low-frequency) observations. Note that autoregressive components might be included in V. Mixing of frequencies can be ignored in (4.2), without loss of generality for the model estimation. While parameters θ_1 and θ_2 are driving the shape of the weighting scheme, parameter θ_3 allows for non-zero last lag. The assumption of zero-last-lag, which is implied by fixing θ_3 to zero, is too restrictive in regression specifications with a small MIDAS temporal aggregation period p (Ghysels, 2013, pp. 5–6). It is worth noting that model (4.2) is non-linear in parameters. The use of the Beta polynomial makes necessary estimation of the regression coefficients by minimizing the following objective function:

$$\underset{(\theta_1,\theta_2,\theta_3,\beta,\boldsymbol{\rho})}{\operatorname{argmin}} \; (\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1,\theta_2,\theta_3))^\top (\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1,\theta_2,\theta_3)). \quad (4.5)$$

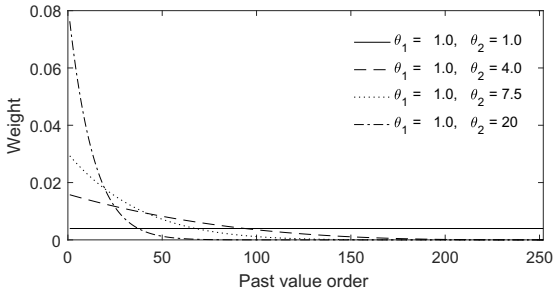
Whereas the optimization problem (4.5) does not have an explicit solution, numerical methods should be applied. Best practices in the MIDAS literature on estimating regressions rely on two pillars: Imposing constraints on the weighting scheme in order to reduce the number of parameters and the use of parameter profiling to estimate the model with the restricted Beta polynomial.

Table 4.1 defines three restricted ways to parameterize the weighting scheme (4.3) so that *the number of parameters to estimate is reduced.* In the restricted cases 1 and 2, two

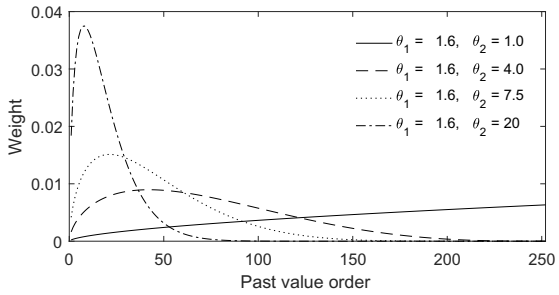
Figure 4.1 Beta MIDAS weights with zero-last-lag ($\theta_3 = 0$)

Variability in θ_1 makes inverted U-shaped and steeply decaying profiles feasible.

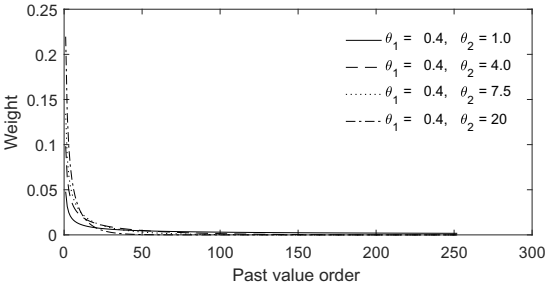
(a) $\theta_1 = 1$



(b) $\theta_1 = 1.6$



(c) $\theta_1 = 0.4$



parameters have to be estimated. In order to promote further simplification it is possible to turn to a single parameter problem, as in the restricted case 3. It is notable that fixing θ_1 to a constant equal to one is common in the MIDAS literature (Becker et al. (2010), (Ghysels and Marcellino, 2018, p. 461)). In combination with θ_2 greater than one, this choice yields decreasing weights. However, these constraints are not purely data-driven, may contradict empirical evidence and result in poor goodness of fit. To some extent, unjustified restrictions may restrain the main advantage of MIDAS, its flexibility. Figure 4.1 shows that the variability in θ_1 allows for a wider range of weighting profiles. *Estimation of the restricted model* is typically performed using software packages[4] such as the MIDAS Toolbox in MATLAB (Ghysels, 2013) and the midasr package in R (Ghysels, Kvedaras and Zemlys, 2016) which are designed to solve the optimization problem using NLS. Ghysels and Qian (2019) report some numerical issues in using MIDAS-NLS estimator and propose the use of OLS with polynomial parameter profiling as a remedy. Profiling approach combines searching across a grid of polynomial parameters to remove nonlinearity in (4.2)

[4]We refer the reader to http://eghysels.web.unc.edu/ for the most recent codes for running MIDAS regressions.

and subsequent estimation via OLS. This algorithm has a list of advantages over standard MIDAS-NLS: Explosive behavior of Beta parameters is impossible, the estimation goes faster and delivers more accurate solutions. A potential user should be aware of some peculiarities of the profiling approach. Clearly, there is a trade-off between time and accuracy as in any grid search algorithm. Once the upper and lower bounds on the Beta parameters are defined, the set of candidate solutions should be provided. The use of a coarse grid entails faster estimation with lower accuracy. Similarly, a small grid step requires higher computational efforts but results in precise solutions. Generally speaking, the smallest grid step that keeps the estimation computationally feasible is most appealing. While this issue is a minor one for a single parameter Beta polynomial (restricted case 3 in Table 4.1), the cost of accuracy measured in computational time soars when we switch to the two- and three- dimensional profiling.

Consequently, in some applications this two-pillar principle combined with standard MIDAS-NLS or profiling estimator delivers unsatisfactory solutions and remains computationally expensive. In their recent study, Bollerslev et al. (2018) apply a MIDAS regression to predict volatility. The authors explain that the rolling window MIDAS estimation is not feasible in that situation. The same pattern of high computational complexity is present in multistep forecasting and in cases where many MIDAS estimations should be completed.

We revise optimization behind MIDAS-NLS and compare its performance to NLS estimation in the MIDAS MATLAB Toolbox and to the profiling approach. This comparison is performed using simulated data for restricted and unrestricted parametrization of the Beta polynomial.

4.2.2 Revised approach to MIDAS-NLS

In our attempt to revise the MIDAS-NLS approach, we introduce a few modifications. First, we vectorize the optimization problem intrinsic to MIDAS estimation and the corresponding computations. Second, we rely on analytic derivatives in solving it with numerical methods. This idea is associated with Audrino, Kostrov and Ortega (2019) and Kostrov and Tetereva (2020) where the analytical gradient is used for estimating a MIDAS regression with the exponential Almon lag form. Simultaneous use of the analytical gradient and the analytical Hessian matrix in optimization allows for a straightforward enhancement. Unlike our approach, the original implementation of nonlinear optimization in statistical packages is based on the numerical gradient, which is then used to approximate the Hessian at each iteration using the BFGS formula. Finally, we employ the interior-point optimization algorithm (Byrd et al., 1999) which benefits from the user-specified analytical gradient and

the analytical Hessian that we derive. It should be borne in mind that our approach allows to mitigate unnecessary fixing of Beta parameters: We let the data decide.

We have made available the "MIDAS Analytic extension" which allows to estimate MIDAS regressions via MIDAS-NLS with revised optimization. Its current version supports a classical univariate MIDAS regression, a multivariate MIDAS regression, the MIDAS logit model (Audrino, Kostrov and Ortega, 2019), and the MIDAS VAR model (Ghysels, 2016). Documentation to the "MIDAS Analytic extension"explains how to use the main functions and contains the computation of the analytic derivatives.

4.3 Simulation study

We now move to a correctly specified simulation study where we explore the ability of the three methods under consideration to find a true pattern in data. The design is developed to address different cases of the Beta polynomials. Our experiment considers the effects of sample size, the number of MIDAS lags, and grid step on the recovery of true coefficients and their impact on the goodness of fit for different combinations of true parameters. We pay some extra attention to the results in a big data setting.

4.3.1 Data generating process

Data is generated to address a reasonably diverse set of parameters which represent combinations typically encountered in applications. Broadly speaking, we proceed in accordance with the following steps. First, we generate nine i.i.d. normally distributed factors in the matrix V, a random normal $\boldsymbol{z}$ – a MIDAS-weighted variable which is used to construct the matrix Z, and a Gaussian noise $\boldsymbol{\varepsilon}$. The choice of thetas $\theta_1^{\text{dgp}}, \theta_2^{\text{dgp}}$, and θ_3^{dgp} in the dgp is fixed, while elements of a parameter vector $(\boldsymbol{\rho}^{\text{dgp}}, \beta^{\text{dgp}})$ are picked randomly from values between 0 and 10. The output variable is assigned by (4.6), for the selected sample size T and MIDAS temporal aggregation period p.

$$\boldsymbol{y}^{\text{dgp}} := V\boldsymbol{\rho}^{\text{dgp}} + \beta^{\text{dgp}} Z\boldsymbol{w}(\theta_1, \theta_2, \theta_3) + \boldsymbol{\varepsilon}. \tag{4.6}$$

Second, parameters are estimated using the MATLAB MIDAS Toolbox, OLS with polynomial parameter profiling, and MIDAS-NLS with revised optimization. After that, we compute and store the performance measures (4.7)–(4.9). Finally, steps 1 and 2 are repeated 500 times. When computations are ready, we report averaged time on estimation and accuracy indicators as well as respective percentiles Q_α, $\alpha \in \{25\%, 75\%, 99\%\}$ in order to check

the stability of results and the presence of outliers. We analyze scenarios by varying $\theta_1^{\text{dgp}} = \{0.4082, \underline{1}, 1.5358\}$, $\theta_2^{\text{dgp}} = \{1.3752, 6.7339, 16.3983\}$, and $\theta_3^{\text{dgp}} = \{\underline{0}, 0.7623\}$. The profiling approach needs upper and lower boundaries to be imposed on the grid parameters: $\theta_1 \in [0.1, 1.9], \theta_2 \in [1.2, 20]$, and $\theta_3 \in [0, 3]$. Very wide ranges treat the profiling approach unfairly in the sense of overstating the computational effort, thus we specify conservative restrictions on Beta polynomial parameters. In the simulation study, we analyze the performance of candidate methods using 500 generated data samples in the restricted and unrestricted cases of the Beta polynomial.

4.3.2 Performance measures

The evaluation of competing methods is based on computation time and accuracy measures. Following Ghysels and Qian (2019), bias squared is used to characterize the deviation of estimated parameters from the true values in the data generating process (dgp). So, bias squared in the weighting scheme is

$$Bias_{ws}^2 = \sum_{i=1}^{p} \left(\hat{\beta} \omega_i(\widehat{\theta_1}, \widehat{\theta_2}) - \beta^{\text{dgp}} \omega_i(\theta_1^{\text{dgp}}, \theta_2^{\text{dgp}}) \right)^2, \tag{4.7}$$

and bias squared in the factor coefficients (including an intercept) is

$$Bias_f^2 = \sum_{i=0}^{k} \left(\widehat{\rho}_i - \rho_i^{\text{dgp}} \right)^2. \tag{4.8}$$

Another measure is the sum of squared residuals (RSS) which coincides with the optimal value of the target function (4.5) behind estimation:

$$RSS = \sum_{t=1}^{T} \left(\widehat{y}_t - y_t^{\text{dgp}} \right)^2. \tag{4.9}$$

Bias in the Beta polynomial parameters, in thetas, is employed for the supplementary analysis:

$$Bias_\theta = \sum_i |\widehat{\theta}_i - \theta_i^{\text{dgp}}|.$$

4.3.3 Findings

Findings are discussed for the single-parameter (Table 4.2) and the two-parameter (Tables 4.3 and 4.4) versions of the Beta polynomial. Three-parameter case, this is the Beta specification without restrictions, is illustrated by Table 4.A.2 of Appendix 4.C.

Unlike a single-parameter specification, the use of two or three thetas in the Beta polynomial hits the computational constraint (Figure 4.2) which prohibits the use of a smaller grid step. The simulation study shows that MIDAS approach with revised optimization exhibits high accuracy, robustness to DGP settings, and stability of results in three restricted cases. Importantly, the use of MIDAS-NLS with revised optimization allows for nearly instant MIDAS estimation. The profiling approach has limited practical implications in a big data setup, which becomes apparent when the Beta polynomial is governed by more than one parameters. This drawback is overcome with the suggested approach.

Simulation results in Tables 4.2–4.4 share a few common patterns. So, when the true θ_2^{dgp} is closer to one (refer to Panel 1 in each table), the optimization problem is more sophisticated and tends to show an explosive behavior. Performance for larger values of θ_2^{dgp} is rather similar which stands behind our decision to report most results for $\theta_2^{\text{dgp}} = \{1.3752, 6.7339\}$ and omit the case of $\theta_2^{\text{dgp}} = 16.3983$. Unsurprisingly, the estimation for a larger sample size is time-demanding.

Several findings emerge from simulation results for a MIDAS regression with the Beta polynomial which is defined using one parameter (Table 4.2). It is true that profiling approach is a strong competitor. However, OLS with polynomial parameter profiling does not reach the accuracy of MIDAS-NLS with revised optimization. One might blame a coarse grid of Beta parameters with step $\eta = 0.065$ that we apply. This guess is partially true: 4.B illustrates that a very soft grid is feasible for a single-parameter case and enhances profiling performance. Results are approaching the accuracy of MIDAS-NLS with revised optimization. Another interesting observation is that profiling performance is adversely affected by using many lags of a MIDAS-weighted variable, refer to Figure 4.A.1(c) in Appendix 4.B. The reason is that small errors in Beta polynomial parameters embedded in a grid with fixed step lead to growing bias in assigned MIDAS weight. Accordingly, its contribution to the RSS growth becomes considerable. Moreover, this issue is scaled up by the use of large data samples.

The computational efficiency of the MIDAS MATLAB Toolbox and MIDAS-NLS with revised optimization can be compared from an alternative standview. Table 4.A.1 of 4.A reports the number of iterations taken before the solution is found. The suggested approach performs better: Fewer iterations are required. It is notable that in the big data setting

Table 4.2 MIDAS-NLS with revised optimization: Optimization by θ_2

Comparison of MIDAS-NLS with revised optimization ("Revised") to the existing methods: "Toolb." stands for the MATLAB MIDAS Toolbox, "Profil." – profiling. DGP is parametrized using $(\theta_1, \theta_2, \theta_3) = (1, \theta_2^{\text{dgp}}, 0)$.

T		t, time in seconds			$Bias^2_{ws}$			$Bias^2_f$			RSS		
p		Toolb.	Profil.	Revised	Toolb.	Profil.	Revised	Toolb.	Profil.	Revised	Toolb.	Profil.	Revised
Panel 1: Simulation results when $\theta_2^{\text{dgp}} = 1.3752$													
200	mean	0.197	0.015	0.017	9.759	0.041	0.011	0.641	0.060	0.058	1970	192.8	186.7
8	$Q_{25\%}$	0.172	0.013	0.013	1.721	0.015	0.003	0.128	0.041	0.040	473.3	177.4	171.7
	$Q_{75\%}$	0.214	0.016	0.016	15.78	0.061	0.016	0.862	0.072	0.070	2975	207.6	200.5
	$Q_{99\%}$	0.267	0.022	0.050	35.58	0.133	0.052	3.575	0.155	0.147	7319	248.2	235.7
200	mean	0.251	0.015	0.015	5.122	0.146	0.013	0.378	0.068	0.059	1118	212.7	186.8
30	$Q_{25\%}$	0.221	0.014	0.013	0.917	0.039	0.003	0.107	0.045	0.041	346.7	187.7	172.1
	$Q_{75\%}$	0.281	0.015	0.015	8.291	0.235	0.018	0.512	0.084	0.071	1694	231.2	199.6
	$Q_{99\%}$	0.332	0.019	0.026	14.81	0.396	0.075	1.526	0.178	0.151	3202	335.9	235.5
2000	mean	11.57	0.058	0.264	1.191	1.170	0.001	0.014	0.032	0.006	4359	4211	1987
250	$Q_{25\%}$	10.23	0.055	0.241	0.233	0.232	0.000	0.007	0.008	0.004	2461	2402	1944
	$Q_{75\%}$	13.05	0.061	0.282	1.990	1.955	0.002	0.018	0.032	0.007	5981	5640	2028
	$Q_{99\%}$	14.66	0.069	0.377	3.469	3.361	0.008	0.048	0.264	0.013	8965	10492	2130
Panel 2: Simulation results when $\theta_2^{\text{dgp}} = 6.7339$													
200	mean	0.201	0.016	0.013	0.028	0.013	0.012	0.059	0.058	0.058	189.7	186.7	186.5
8	$Q_{25\%}$	0.168	0.015	0.011	0.012	0.003	0.003	0.041	0.040	0.040	174.9	171.9	171.9
	$Q_{75\%}$	0.233	0.017	0.014	0.040	0.018	0.017	0.071	0.070	0.069	204.0	200.2	200.2
	$Q_{99\%}$	0.281	0.020	0.023	0.083	0.063	0.057	0.151	0.150	0.150	240.0	235.2	235.2
200	mean	0.338	0.017	0.015	0.019	0.011	0.011	0.059	0.058	0.058	188.3	187.7	186.9
30	$Q_{25\%}$	0.287	0.016	0.013	0.009	0.003	0.003	0.042	0.040	0.040	173.3	173.3	172.2
	$Q_{75\%}$	0.395	0.018	0.016	0.024	0.013	0.015	0.071	0.071	0.069	201.4	201.4	200.0
	$Q_{99\%}$	0.446	0.021	0.023	0.067	0.069	0.055	0.146	0.144	0.143	237.7	236.0	235.3
2000	mean	13.75	0.063	0.255	0.002	0.02	0.001	0.006	0.006	0.006	1989	2018	1987
250	$Q_{25\%}$	12.95	0.059	0.223	0.001	0.004	0.000	0.004	0.004	0.004	1947	1970	1945
	$Q_{75\%}$	15.27	0.068	0.277	0.003	0.027	0.002	0.007	0.008	0.007	2031	2063	2029
	$Q_{99\%}$	16.33	0.077	0.406	0.008	0.048	0.006	0.013	0.014	0.013	2130	2176	2131
Panel 3: Simulation results when $\theta_2^{\text{dgp}} = 16.3983$													
200	mean	0.180	0.016	0.015	0.011	0.011	0.011	0.058	0.058	0.058	186.7	186.7	186.6
8	$Q_{25\%}$	0.158	0.014	0.013	0.003	0.003	0.003	0.040	0.040	0.040	171.9	172.0	171.9
	$Q_{75\%}$	0.199	0.017	0.015	0.015	0.015	0.015	0.070	0.070	0.070	200.3	200.2	200.2
	$Q_{99\%}$	0.237	0.021	0.026	0.055	0.056	0.054	0.153	0.151	0.151	234.4	234.7	234.7
200	mean	0.345	0.017	0.015	0.012	0.012	0.011	0.058	0.058	0.058	187.0	187.0	186.7
30	$Q_{25\%}$	0.296	0.015	0.013	0.004	0.003	0.003	0.041	0.041	0.040	172.3	172.5	172.1
	$Q_{75\%}$	0.403	0.018	0.016	0.017	0.018	0.015	0.070	0.070	0.070	199.7	199.8	199.5
	$Q_{99\%}$	0.443	0.022	0.025	0.058	0.055	0.057	0.143	0.142	0.143	234.2	235.4	234.8
2000	mean	14.22	0.064	0.274	0.001	0.010	0.001	0.006	0.006	0.006	1988	2006	1987
250	$Q_{25\%}$	13.30	0.059	0.245	0.001	0.003	0.000	0.004	0.004	0.004	1946	1963	1944
	$Q_{75\%}$	15.71	0.068	0.293	0.002	0.016	0.001	0.007	0.007	0.007	2029	2045	2029
	$Q_{99\%}$	16.89	0.079	0.422	0.006	0.028	0.006	0.013	0.013	0.013	2131	2162	2130

Note: Results are based on 500 simulated samples. In the MATLAB MIDAS Toolbox, this is "beta" case with 1 parameter. In profiling, grid step η equals 0.065, $\theta_2 \in [1.2, 20]$.

Table 4.3 MIDAS-NLS with revised optimization: Optimization by θ_2 and θ_3

Comparison of MIDAS-NLS with revised optimization ("Revised") to the existing methods: "Toolb." stands for the MATLAB MIDAS Toolbox, "Profil." – profiling. DGP is parametrized using $(\theta_1, \theta_2, \theta_3) = (1, \theta_2^{\text{dgp}}, \theta_3^{\text{dgp}})$.

T		t, time in seconds			$Bias^2_{ws}$			$Bias^2_f$			RSS		
p		Toolb.	Profil.	Revised	Toolb.	Profil.	Revised	Toolb.	Profil.	Revised	Toolb.	Profil.	Revised
Panel 1: Simulation results when $\theta_2^{\text{dgp}} = 1.3752$ and $\theta_3^{\text{dgp}} = 0.7623$													
200	mean	1.167	5.377	0.037	0.030	0.018	0.016	0.059	0.059	0.059	186.0	186.1	185.8
8	$Q_{25\%}$	1.043	5.314	0.031	0.016	0.007	0.006	0.041	0.040	0.040	171.3	171.2	170.9
	$Q_{75\%}$	1.290	5.413	0.042	0.039	0.024	0.020	0.072	0.070	0.070	200.3	200.5	199.9
	$Q_{99\%}$	1.477	5.815	0.067	0.089	0.077	0.068	0.148	0.150	0.150	236.8	235.0	234.7
200	mean	1.651	5.402	0.034	20.26	0.025	0.019	1.101	0.059	0.059	3603	187.4	185.8
30	$Q_{25\%}$	1.503	5.365	0.030	0.044	0.010	0.008	0.046	0.040	0.040	184.0	173.0	171.4
	$Q_{75\%}$	1.832	5.438	0.038	0.218	0.029	0.025	0.097	0.072	0.072	231.7	201.6	199.4
	$Q_{99\%}$	2.352	5.603	0.051	206.6	0.133	0.093	14.26	0.149	0.149	37836	238.0	235.1
2000	mean	52.70	8.205	0.608	535.8	0.092	0.002	4.491	0.006	0.006	11×10^5	2167	1960
250	$Q_{25\%}$	45.46	8.147	0.477	0.075	0.019	0.001	0.007	0.004	0.004	2160	2031	1944
	$Q_{75\%}$	56.26	8.253	0.722	978.2	0.152	0.002	6.277	0.008	0.007	20×10^5	2280	2027
	$Q_{99\%}$	95.79	8.481	0.923	1949	0.266	0.009	28.48	0.015	0.013	49×10^5	2621	2129
Panel 2: Simulation results when $\theta_2^{\text{dgp}} = 6.7339$ and $\theta_3^{\text{dgp}} = 0.7623$													
200	mean	1.236	5.428	0.016	0.027	0.040	0.017	0.059	0.060	0.059	185.3	190.2	185.6
8	$Q_{25\%}$	1.164	5.366	0.013	0.013	0.019	0.007	0.041	0.041	0.041	170.1	175.6	170.3
	$Q_{75\%}$	1.333	5.458	0.017	0.034	0.055	0.024	0.071	0.072	0.070	198.8	205.2	199.4
	$Q_{99\%}$	1.504	5.841	0.031	0.077	0.153	0.059	0.153	0.156	0.150	235.8	242.4	235.1
200	mean	1.626	5.471	0.018	2.333	0.104	0.018	0.245	0.067	0.059	627.1	201.9	185.9
30	$Q_{25\%}$	1.367	5.424	0.015	0.014	0.033	0.006	0.041	0.045	0.041	170.8	182.6	171.1
	$Q_{75\%}$	1.851	5.509	0.020	0.038	0.151	0.024	0.073	0.082	0.071	198.3	216.7	198.6
	$Q_{99\%}$	2.327	5.730	0.028	0.112	0.407	0.079	0.157	0.193	0.148	237.8	278.2	235.1
2000	mean	46.89	8.324	0.290	1657	0.736	0.002	26.34	0.028	0.006	35×10^5	3344	1986
250	$Q_{25\%}$	42.86	8.251	0.249	0.002	0.150	0.001	0.004	0.007	0.004	1947	2246	1943
	$Q_{75\%}$	49.36	8.358	0.315	0.004	1.217	0.002	0.008	0.029	0.007	2040	4178	2027
	$Q_{99\%}$	75.41	8.826	0.502	56120	2.106	0.008	719.3	0.191	0.013	12×10^7	7206	2126
Panel 3: Simulation results when $\theta_2^{\text{dgp}} = 16.3983$ and $\theta_3^{\text{dgp}} = 0.7623$													
200	mean	1.224	5.503	0.019	0.023	0.029	0.017	0.060	0.059	0.059	185.1	187.9	185.7
8	$Q_{25\%}$	1.150	5.456	0.016	0.010	0.013	0.006	0.041	0.041	0.041	169.7	173.7	170.8
	$Q_{75\%}$	1.312	5.548	0.021	0.028	0.035	0.022	0.070	0.071	0.070	198.2	201.3	199.3
	$Q_{99\%}$	1.513	5.656	0.031	0.074	0.152	0.061	0.153	0.151	0.151	234.3	238.3	234.6
200	mean	1.555	5.485	0.017	131.0	0.193	0.018	6.768	0.081	0.059	21714	215.0	185.8
30	$Q_{25\%}$	1.254	5.427	0.015	0.011	0.050	0.006	0.041	0.048	0.041	171.2	186.4	171.4
	$Q_{75\%}$	1.837	5.539	0.019	0.034	0.301	0.026	0.074	0.098	0.071	199.1	232.9	197.6
	$Q_{99\%}$	2.274	5.721	0.030	6917	0.552	0.080	326.3	0.271	0.146	11×10^5	355.5	234.7
2000	mean	48.97	8.324	0.303	1773	1.444	0.002	21.54	0.109	0.006	36×10^5	4519	1986
250	$Q_{25\%}$	43.85	8.234	0.266	0.001	0.287	0.001	0.004	0.010	0.004	1943	2463	1943
	$Q_{75\%}$	49.76	8.384	0.330	0.004	2.366	0.002	0.007	0.111	0.007	2029	5953	2028
	$Q_{99\%}$	109.5	8.684	0.466	87308	4.195	0.007	1021	0.993	0.013	19×10^7	13145	2127

Note: Results are based on 500 simulated samples. In the MATLAB MIDAS Toolbox, this is "betann" case with 2 parameters. In profiling, grid step η equals 0.065, $\theta_2 \in [1.2, 20]$, and $\theta_3 \in [0, 3]$.

Table 4.4 MIDAS-NLS with revised optimization: Optimization by θ_1 and θ_2

Comparison of MIDAS-NLS with revised optimization ("Revised") to the existing methods: "Toolb." stands for the MATLAB MIDAS Toolbox, "Profil." – profiling. DGP is parametrized using $(\theta_1, \theta_2, \theta_3) = (\theta_1^{\text{dgp}}, \theta_2^{\text{dgp}}, 0)$.

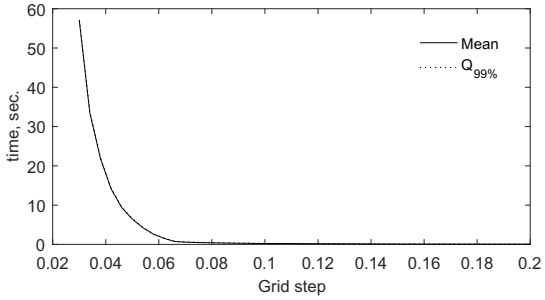
T		t, time in seconds			$Bias^2_{ws}$			$Bias^2_f$			RSS		
p		Toolb.	Profil.	Revised	Toolb.	Profil.	Revised	Toolb.	Profil.	Revised	Toolb.	Profil.	Revised
Panel 1: Simulation results when $\theta_1^{\text{dgp}} = 0.4082$ and $\theta_2^{\text{dgp}} = 1.3752$													
200	mean	0.649	1.200	0.024	2.647	0.023	0.017	0.212	0.059	0.059	659.9	187.5	185.6
8	$Q_{25\%}$	0.620	1.179	0.018	0.461	0.009	0.007	0.081	0.040	0.040	270.0	173.3	170.8
	$Q_{75\%}$	0.683	1.217	0.027	4.255	0.027	0.023	0.291	0.070	0.071	953.7	201.1	199.5
	$Q_{99\%}$	0.783	1.298	0.051	7.657	0.126	0.065	0.773	0.148	0.148	1752	237.8	235.1
200	mean	1.020	1.218	0.020	34.11	0.205	0.018	2.636	0.073	0.059	5845	218.6	185.8
30	$Q_{25\%}$	0.974	1.192	0.016	6.438	0.052	0.006	0.427	0.046	0.041	1138	189.0	171.4
	$Q_{75\%}$	1.076	1.245	0.023	55.49	0.321	0.025	3.795	0.093	0.072	9230	238.0	198.3
	$Q_{99\%}$	1.198	1.305	0.036	99.81	0.676	0.082	13.33	0.193	0.150	21388	334.4	235.2
2000	mean	37.51	2.882	0.363	1966	7.354	0.002	83.52	0.065	0.006	35×10^5	16267	1986
250	$Q_{25\%}$	35.78	2.857	0.327	388.7	1.434	0.001	5.901	0.016	0.004	69×10^4	4991	1943
	$Q_{75\%}$	39.72	2.901	0.398	3188	12.27	0.002	74.42	0.088	0.007	57×10^5	25549	2027
	$Q_{99\%}$	45.45	3.070	0.538	5732	21.25	0.008	1001	0.351	0.013	12×10^6	47324	2130
Panel 2: Simulation results when $\theta_1^{\text{dgp}} = 0.4082$ and $\theta_2^{\text{dgp}} = 6.7339$													
200	mean	0.638	1.177	0.047	0.020	0.016	0.016	0.059	0.058	0.059	186.1	185.7	185.6
8	$Q_{25\%}$	0.612	1.155	0.033	0.009	0.006	0.006	0.040	0.041	0.041	170.6	171.0	170.9
	$Q_{75\%}$	0.671	1.197	0.057	0.026	0.023	0.023	0.070	0.069	0.069	199.4	199.2	199.2
	$Q_{99\%}$	0.758	1.289	0.095	0.063	0.061	0.060	0.156	0.152	0.152	236.0	234.2	234.1
200	mean	1.020	1.235	0.025	2.569	0.038	0.017	0.230	0.060	0.059	639.0	190.5	185.8
30	$Q_{25\%}$	0.964	1.212	0.018	0.493	0.017	0.006	0.080	0.042	0.040	266.5	174.9	171.2
	$Q_{75\%}$	1.077	1.255	0.031	4.136	0.050	0.023	0.323	0.073	0.070	931.6	203.8	199.0
	$Q_{99\%}$	1.247	1.347	0.043	7.449	0.165	0.064	1.014	0.150	0.143	1676	238.4	234.7
2000	mean	39.79	3.014	0.417	1348	5.765	0.002	22.40	0.062	0.006	26×10^5	13326	1986
250	$Q_{25\%}$	37.73	2.983	0.365	263.7	1.121	0.001	2.548	0.014	0.004	52×10^4	4290	1943
	$Q_{75\%}$	41.77	3.040	0.470	2244	9.660	0.002	26.35	0.081	0.007	43×10^5	20889	2028
	$Q_{99\%}$	46.57	3.141	0.623	3915	16.71	0.007	191.1	0.390	0.013	80×10^5	36629	2130
Panel 3: Simulation results when $\theta_1^{\text{dgp}} = 1.5358$ and $\theta_2^{\text{dgp}} = 1.3752$													
200	mean	0.575	1.168	0.019	4.692	0.086	0.017	0.369	0.063	0.059	1026	197.5	185.7
8	$Q_{25\%}$	0.526	1.151	0.016	0.512	0.028	0.007	0.075	0.043	0.040	270.0	180.5	170.9
	$Q_{75\%}$	0.639	1.186	0.021	3.571	0.126	0.022	0.249	0.077	0.070	847.1	213.7	200.0
	$Q_{99\%}$	0.755	1.237	0.030	87.04	0.331	0.074	6.857	0.152	0.150	15368	258.5	234.6
200	mean	0.815	1.230	0.019	19.25	0.296	0.039	1.338	0.075	0.060	3518	240	188.2
30	$Q_{25\%}$	0.751	1.212	0.016	3.764	0.074	0.007	0.245	0.047	0.041	801.6	194.4	171.4
	$Q_{75\%}$	0.917	1.247	0.020	31.60	0.467	0.024	1.827	0.094	0.072	5573	270.3	199.3
	$Q_{99\%}$	1.053	1.310	0.029	55.17	0.821	0.092	7.282	0.202	0.155	10684	479.8	235.9
2000	mean	32.62	2.968	0.358	3.831	2.370	0.012	0.033	0.027	0.006	9597	6644	1996
250	$Q_{25\%}$	30.67	2.938	0.303	0.751	0.471	0.001	0.010	0.010	0.004	3534	2878	1943
	$Q_{75\%}$	35.88	2.989	0.405	6.406	4.020	0.003	0.046	0.035	0.007	14876	9707	2027
	$Q_{99\%}$	39.85	3.112	0.491	11.12	6.905	0.010	0.150	0.140	0.015	24489	19258	2136
Panel 4: Simulation results when $\theta_1^{\text{dgp}} = 1.5358$ and $\theta_2^{\text{dgp}} = 6.7339$													
200	mean	0.654	1.191	0.018	0.052	0.021	0.017	0.060	0.059	0.059	191.8	186.4	185.5
8	$Q_{25\%}$	0.611	1.173	0.014	0.024	0.005	0.006	0.042	0.041	0.040	176.5	171.7	170.8
	$Q_{75\%}$	0.703	1.207	0.019	0.078	0.030	0.024	0.072	0.070	0.069	207.0	200.0	198.9
	$Q_{99\%}$	0.795	1.258	0.042	0.135	0.090	0.061	0.151	0.154	0.152	245.1	236.4	234.4
200	mean	1.025	1.249	0.016	23.77	0.041	0.017	1.529	0.061	0.059	4450	191.0	185.8
30	$Q_{25\%}$	0.982	1.231	0.013	4.127	0.017	0.007	0.290	0.042	0.041	906.0	176.0	171.2
	$Q_{75\%}$	1.090	1.265	0.018	38.86	0.054	0.023	2.258	0.074	0.070	7069	204.2	198.5
	$Q_{99\%}$	1.172	1.316	0.031	68.41	0.1782	0.084	7.257	0.140	0.142	14165	241.6	234.3
2000	mean	34.47	2.995	0.266	47.40	0.260	0.002	0.259	0.008	0.006	96354	2502	1986
250	$Q_{25\%}$	33.20	2.962	0.239	9.283	0.053	0.001	0.049	0.005	0.004	20673	2101	1943
	$Q_{75\%}$	36.61	3.023	0.281	79.25	0.433	0.003	0.411	0.010	0.007	16×10^4	2819	2027
	$Q_{99\%}$	38.81	3.205	0.426	137.5	0.752	0.007	1.269	0.021	0.013	28×10^4	3687	2126

Note: Results are based on 500 simulated samples. In the MATLAB MIDAS Toolbox, this is "beta" case with 2 parameters. In profiling, grid step η equals 0.065, $\theta_1 \in [0.1, 1.9]$, and $\theta_2 \in [1.2, 20]$.

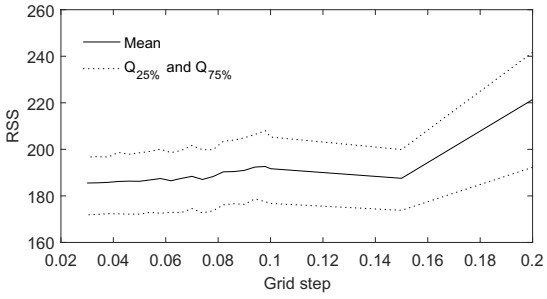
Figure 4.2 Relation between a grid step and the computational complexity

Figure 4.2 describes the Beta polynomial governed by two parameters: Relation between a grid step, the computational complexity, and the target function value. Characteristics of dgp are similar to those in Panel 1 of Table 4.4: $\theta_1^{\text{dgp}} = 0.4082$, $\theta_2^{\text{dgp}} = 1.3752$, sample size $T = 200$, and the number of MIDAS-weighted lags $p = 8$. Results are based on 500 simulated samples.

(a) Grid step and the computational effort.



(b) The impact of grid step on RSS.



(Panel 3 of Table 4.A.1), when the MIDAS MATLAB Toolbox is particularly inaccurate, its algorithm often terminates after less than 5 iterations near the starting point, far away from the optimal solution.

4.3.4 Importance of fast estimation

Table 4.A.2 of Appendix 4.C demonstrates the simulation results for the unrestricted Beta polynomial with three parameters. This case is not supported by the MATLAB MIDAS Toolbox and becomes computationally extensive for profiling when a grid step is not small enough to deliver a satisfactory level of accuracy. So, one estimation with a grid step of 0.065 as in the restricted cases (Tables 4.2–4.4) is computationally not feasible: One run takes about 2 hours when $T = 200$, in a small sample.

Talking about MIDAS-NLS estimator, (Ghysels et al., 2020, pp. 137–138) point out that NLS problem intrinsic to MIDAS estimation might be sensitive to the choice of a starting point. If this pattern emerges in applications, the numerical stability can be improved by running a solver repeatedly starting from different initial points. In prior research, this solution was less attractive due to much slower NLS optimization. Fast estimation is of great value in applications where MIDAS estimation should be completed hundreds of times, e.g., for bootstrapping (Aastveit et al., 2017), rolling window estimation (Bollerslev et al., 2018), and forecasting in high dimensions (Kostrov and Tetereva, 2020).

4.4 Volatility forecasting application

MIDAS regressions were originally introduced in the context of volatility forecasting (Ghysels et al., 2006), which is a central topic in Financial Econometrics. In this section, we put forward the related questions: "Does MIDAS-NLS with revised optimization perform well in realized volatility forecasting?" and "Is it the best choice to use the common single-parameter Beta specification in realized volatility forecasting?"

The MIDAS model for realized volatility (RV) is:

$$V^h_{(t+h):(t+1)} = a + \beta \sum_{q=1}^{p} w(q, \theta_1, \theta_2, \theta_3) V^d_{t-q+1} + e^h_{(t+h):(t+1)}, \tag{4.10}$$

where the logarithmic RV at day t is $V^d_t = \log(RV^d_t)$, $V^h_{(t+h):(t+1)}$ denotes volatility for the h-day period from day $t+1$ to day $t+h$, weights w are defined using the Beta polynomial, and the number of MIDAS lags p is set to 22.

We estimate the coefficients using the MATLAB MIDAS Toolbox, the profiling approach with a very small grid step, and MIDAS-NLS with revised optimization. Profiling must exhibit high accuracy by design and provides a natural benchmark for the MIDAS model (4.10). The HAR model (Corsi, 2009), which can be considered as a particular case of MIDAS with step weighting functions, serves as another reference point:

$$V^h_{(t+h):(t+1)} = c + \beta^d V^d_t + \beta^w V^w_t + \beta^m V^m_t + e^h_{(t+h):(t+1)}, \tag{4.11}$$

where $V^w_t = \frac{1}{5}\sum_{q=1}^{5} V^d_{t-q+1}$, and $V^m_t = \frac{1}{22}\sum_{q=1}^{22} V^d_{t-q+1}$. According to the realized volatility literature, the HAR model makes accurate RV forecasts. We fit the models using the realized volatility data for the S&P 500 index, the Dow Jones 30 index, and the Nasdaq 100 index from the Oxford-Man Realized Library[5]. This dataset contains daily realized volatility series obtained from 5-min intra-day squared returns and spans from January 2000 to December 2017.

We estimate regressions (4.10) and (4.11) in a rolling window of 500 observations. Out-of-sample forecasting performance is evaluated using the robust loss functions, MSE and QLike (Patton, 2011). Our results in Tables 4.5 and 4.6 allow to comment on the questions raised at the beginning of this section. First, the performance of MIDAS-NLS with revised optimization is identical to the use of profiling with a small grid step: It is the best possible result for the MIDAS estimators. There is a major improvement in accuracy as compared to

[5]The Final version of Realized Library (Version 0.2) is used.

Table 4.5 Out-of-sample RV forecasting, MSE loss function

Horizon	1 day ahead, $\times 10^{-8}$				1 week ahead, $\times 10^{-7}$				1 month ahead, $\times 10^{-6}$			
Beta parametrization	1 θ_2 0	1 θ_2 θ_3	θ_1 θ_2 0	θ_1 θ_2 θ_3	1 θ_2 0	1 θ_2 θ_3	θ_1 θ_2 0	θ_1 θ_2 θ_3	1 θ_2 0	1 θ_2 θ_3	θ_1 θ_2 0	θ_1 θ_2 θ_3
Panel 1: Realized volatility of S&P 500 index												
Toolb.	4.287	3.365	3.328	–	6.348	4.898	4.859	–	10.93	9.701	10.28	–
Profil.	3.192	3.142	3.171	–	4.724	4.694	4.638	–	9.293	9.569	9.350	–
Revised	3.192	3.143	3.171	3.144	4.724	4.677	4.637	4.651	9.293	9.577	9.351	9.541
HAR	3.193				4.700				9.586			
Panel 2: Realized volatility of Dow Jones 30 index												
Toolb.	5.346	4.250	4.383	–	7.237	5.805	5.769	–	11.89	10.74	10.87	–
Profil.	4.233	4.151	4.190	–	5.659	5.638	5.561	–	9.949	10.54	10.31	–
Revised	4.233	4.155	4.190	4.154	5.660	5.632	5.561	5.576	9.948	10.54	10.30	10.51
HAR	4.188				5.633				10.63			
Panel 3: Realized volatility of Nasdaq 100 index												
Toolb.	1.752	1.325	1.354	–	2.653	2.076	1.997	–	4.525	4.030	4.092	–
Profil.	1.285	1.268	1.273	–	1.902	1.891	1.877	–	3.924	3.960	3.921	–
Revised	1.285	1.268	1.273	1.268	1.902	1.887	1.877	1.884	3.924	3.957	3.920	3.948
HAR	1.281				1.885				3.962			

Table 4.6 Out-of-sample RV forecasting, QLike loss function

Horizon	1 day ahead				1 week ahead				1 month ahead			
Beta parametrization	1 θ_2 0	1 θ_2 θ_3	θ_1 θ_2 0	θ_1 θ_2 θ_3	1 θ_2 0	1 θ_2 θ_3	θ_1 θ_2 0	θ_1 θ_2 θ_3	1 θ_2 0	1 θ_2 θ_3	θ_1 θ_2 0	θ_1 θ_2 θ_3
Panel 1: Realized volatility of S&P 500 index												
Toolb.	1090	763.0	763.7	–	698.3	518.7	518.7	–	669.3	610.7	600.0	–
Profil.	749.4	745.7	749.4	–	495.3	489.7	489.3	–	582.6	574.1	574.9	–
Revised	749.4	745.9	749.3	747.0	495.4	489.0	489.3	491.4	582.6	574.0	574.8	573.2
HAR	758.0				496.4				587.9			
Panel 2: Realized volatility of Dow Jones 30 index												
Toolb.	1438	960.0	954.4	–	1004	718.0	699.4	–	1065	927.0	938.0	–
Profil.	951.5	946.6	950.3	–	688.1	682.1	682.1	–	909.5	899.1	901.0	–
Revised	951.5	947.1	950.3	949.9	688.1	682.7	681.8	683.1	909.5	899.1	901.0	900.5
HAR	966.3				691.3				924.3			
Panel 3: Realized volatility of Nasdaq 100 index												
Toolb.	1022	619.3	630.0	–	734.6	511.3	515.1	–	732.2	641.8	637.2	–
Profil.	592.9	590.1	586.2	–	488.5	487.1	481.8	–	620.5	615.0	614.9	–
Revised	592.9	590.3	586.4	586.1	488.5	487.1	481.8	484.9	620.5	615.1	614.9	614.3
HAR	589.1				486.5				623.2			

the MIDAS MATLAB Toolbox. Second, the conventional choice of the single-parameter Beta weights is not confirmed in our application. In the majority of cases, switching to the alternative specifications leads to the reduction in errors. Moreover, the parametrization affects ranking of the MIDAS model as compared to the HAR model. Our results hold across the stock indices and the forecasting horizons that we analyze.

4.5 Conclusion

MIDAS regressions are instrumental to harness data with different frequencies. Due to the advantages it offers over the alternative methods, the MIDAS approach gained prominence in academic research and in practical applications: It is more efficient than aggregation, robust to multicollinearity, and prevents the proliferation of parameters.

However, there is a nonlinear optimization problem underlying the MIDAS estimation with the Beta lag polynomial. We revise optimization for the MIDAS-NLS estimator using the analytical gradient and the analytical Hessian matrix which are supplied to the interior-point optimization algorithm. Using simulated data, we discover a major advantage in estimation speed and considerable improvement in accuracy of the estimated coefficient.

Our findings remain valid in the case of the Beta polynomial with three parameters. Previously, its estimation was extremely challenging because of numerical concerns. A simulation study indicates both reduction in the computational effort and improvement in accuracy, perfect numerical stability of results is observed.

MIDAS-NLS with revised optimization may be instrumental in a wide range of applications that involve the estimation of MIDAS regressions, including the volatility forecasting exercise that we carry on. The suggested approach is a sole leader in a big data setting.

Hardware and software

Simulations are performed using a dedicated computer HP EliteDesk 800 G2 SFF; CPU: Intel Core i7-6700 (3.40 GHz); RAM: 32 Gb, DDR4 (2133 MHz). Software: Windows 10 Pro, 64-bit; MATLAB R2016b.

Appendices

Appendix 4.A Measure of the computational efficiency

Table 4.A.1 Computational efficiency of MIDAS-NLS with revised optimization

Table 4.A.1 compares the computational efficiency of MIDAS-NLS with revised optimization ("Revised") to the MATLAB MIDAS Toolbox ("Toolb."). The number of iterations in the optimization procedure intrinsic to MIDAS estimation is an alternative measure of computational efficiency.

Optimization		$(1,\theta_2^{\text{dgp}},0)$: by θ_2			$(1,\theta_2^{\text{dgp}},\theta_3^{\text{dgp}})$: by θ_2,θ_3			$(\theta_1^{\text{dgp}},\theta_2^{\text{dgp}},0)$: by θ_1,θ_2			
Beta parameters in the DGP	θ_1^{dgp}	1	1	1	1	1	1	0.4082	0.4082	1.5358	1.5358
	θ_2^{dgp}	1.3752	6.7339	16.3983	1.3752	6.7339	16.3983	1.3752	6.7339	1.3752	6.7339
	θ_3^{dgp}	0	0	0	0.7623	0.7623	0.7623	0	0	0	0
Panel 1: Simulation results when $T=200$ and $p=8$											
Toolb.	mean	30.6	36.6	32.2	109.2	120.6	122.1	135.2	141.7	104.6	141.9
	$Q_{25\%}$	26	23	24	101	116	116	129	135	92	129
	$Q_{75\%}$	35	49	41	121	135	135	145	151	122	156
	$Q_{99\%}$	47.5	62	51	146	152.5	155.5	164	166.5	141	176.5
Revised	mean	10.4	8.4	11.9	37.3	11.7	14.9	16.8	48.4	14.4	13.9
	$Q_{25\%}$	9	7.5	11	30	10	13	13	31	12	11
	$Q_{75\%}$	11	9	12	43	13	16	20	61	16	14
	$Q_{99\%}$	15	13	19.5	75	18	24	30.5	104.5	21	31.5
Panel 2: Simulation results when $T=200$ and $p=30$											
Toolb.	mean	39.8	66.2	68.5	99.1	98.8	87.4	160.7	160.2	115.2	161.6
	$Q_{25\%}$	28	50.5	53.3	93	61.5	40	152	148	103	153.5
	$Q_{75\%}$	51	84	86.5	128.5	136.5	136	173	172	134	174
	$Q_{99\%}$	67	97	98.5	156	176.5	169.5	193	205	155	188.5
Revised	mean	11.1	9.3	10.1	28.5	11.8	11.9	13.7	19.5	13.5	11.3
	$Q_{25\%}$	10	8	9	25	10	11	12	13	12	10
	$Q_{75\%}$	12	10	11	33	13	12	16	25	15	12
	$Q_{99\%}$	16	13	13	47	18	17	25	30	19.5	20
Panel 3: Simulation results when $T=2000$ and $p=30$											
Toolb.	mean	54.8	75.1	80.3	26.3	13.9	15.6	132.5	137.3	105.3	121.1
	$Q_{25\%}$	44.5	69	73.5	11	4	4	124.5	129	97	116
	$Q_{75\%}$	68	87	93	31	19	18	142	147.5	119	131
	$Q_{99\%}$	83	95.5	100.5	121	62.5	142.5	165	163	134	140.5
Revised	mean	11	9.6	10.9	23.2	11.2	11.9	14.2	15.6	14.4	10.3
	$Q_{25\%}$	10	9	10	17	10	11	13	14	12	9
	$Q_{75\%}$	12	10	12	28	12	12	15	18	16	11
	$Q_{99\%}$	14	12	15	34	17	16	19	20	19	16.5

Note: Results are based on 500 simulated samples.

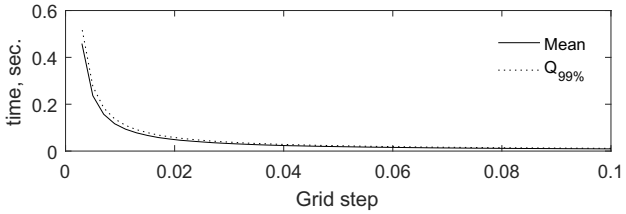
Appendix 4.B Sensitivity analysis for profiling

In this section, we explore the behavior of a profiling estimator for a MIDAS regression with a single-parameter Beta polynomial.

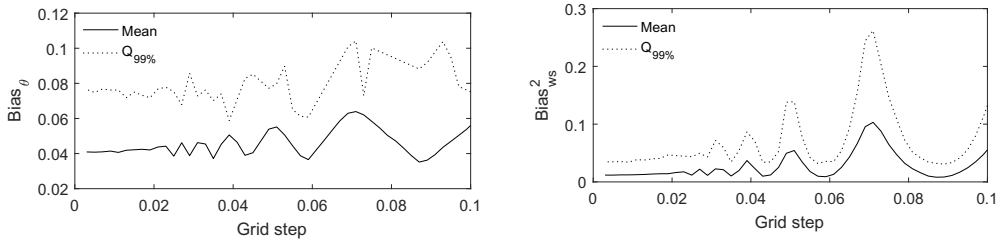
Figure 4.A.1 Sensitivity analysis for the Beta polynomial

Figure 4.A.1 contains results of the sensitivity analysis for the simple-parameter Beta polynomial associated with dgp as in Panel 1 of Table 4.2: $\theta_2^{\text{dgp}} = 1.3752$, sample size $T = 200$, the number of MIDAS-weighted lags $p = 8$, and grid step $\eta = 0.065$. Results are based on 500 simulated samples.

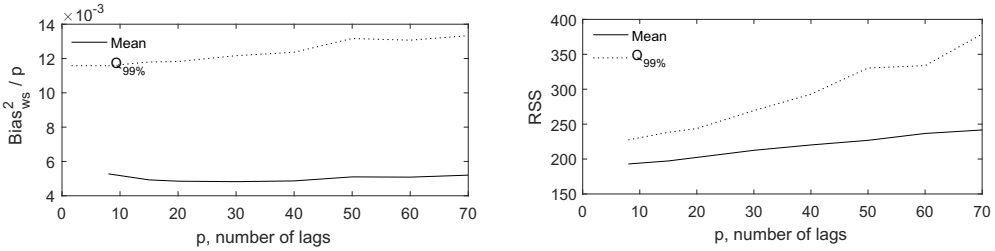
(a) Other things being equal, a smaller grid step requires longer estimation time. However, the computation is not heavy.



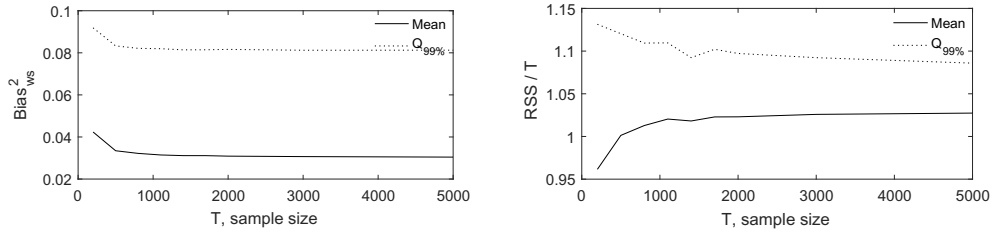
(b) Ceteris paribus, a soft grid leads to smaller bias in the Beta parameters and MIDAS weights.



(c) Ceteris paribus, the use of more MIDAS-weighted lags deteriorates accuracy.



(d) Ceteris paribus, the use of larger samples does not resolve the accuracy issue.



Appendix 4.C Three-parameter Beta case

Table 4.A.2 MIDAS-NLS with revised optimization: Optimization by θ_1, θ_2, and θ_3

Comparison of MIDAS-NLS with revised optimization ("Revised") to profiling ("Profil.") DGP is parametrized using $(\theta_1^{\text{dgp}}, \theta_2^{\text{dgp}}, \theta_3^{\text{dgp}})$. Unrestricted Beta parametrization is not supported by the MATLAB MIDAS Toolbox.

T		t, time in seconds			$Bias^2_{ws}$			$Bias^2_f$			RSS		
p		Toolb.	Profil.	Revised	Toolb.	Profil.	Revised	Toolb.	Profil.	Revised	Toolb.	Profil.	Revised
Panel 1: Simulation results when $\theta_1^{\text{dgp}} = 0.4082$, $\theta_2^{\text{dgp}} = 1.3752$, and $\theta_3^{\text{dgp}} = 0.7623$													
200	mean	–	7.057	0.038	–	0.104	0.020	–	0.063	0.059	–	198.1	185.1
8	$Q_{25\%}$	–	6.996	0.031	–	0.034	0.009	–	0.043	0.041	–	180.4	170.0
	$Q_{75\%}$	–	7.071	0.043	–	0.161	0.026	–	0.078	0.071	–	212.2	198.8
	$Q_{99\%}$	–	7.637	0.065	–	0.359	0.066	–	0.165	0.149	–	261.9	234.6
200	mean	–	7.118	0.033	–	1.201	0.022	–	0.121	0.059	–	185.0	185.7
30	$Q_{25\%}$	–	7.083	0.028	–	0.248	0.010	–	0.062	0.041	–	170.5	170.5
	$Q_{75\%}$	–	7.143	0.037	–	1.996	0.029	–	0.158	0.072	–	197.5	197.2
	$Q_{99\%}$	–	7.347	0.052	–	3.596	0.089	–	0.398	0.148	–	234.5	234.5
2000	mean	–	10.74	0.605	–	151.4	0.002	–	3.039	0.006	–	27×10^4	1985
250	$Q_{25\%}$	–	10.70	0.539	–	28.88	0.001	–	0.306	0.004	–	53819	1942
	$Q_{75\%}$	–	10.77	0.663	–	244.0	0.003	–	3.050	0.007	–	44×10^4	2025
	$Q_{99\%}$	–	10.90	0.870	–	436.1	0.009	–	34.76	0.013	–	87×10^4	2128
Panel 2: Simulation results when $\theta_1^{\text{dgp}} = 0.4082$, $\theta_2^{\text{dgp}} = 1.3752$, and $\theta_3^{\text{dgp}} = 0.0000$													
200	mean	–	6.991	0.034	–	0.104	0.019	–	0.063	0.059	–	201.4	185.2
8	$Q_{25\%}$	–	6.960	0.028	–	0.035	0.008	–	0.042	0.041	–	182.8	169.9
	$Q_{75\%}$	–	7.018	0.038	–	0.152	0.027	–	0.075	0.071	–	217.7	198.9
	$Q_{99\%}$	–	7.109	0.059	–	0.354	0.066	–	0.161	0.150	–	271.2	235.0
200	mean	–	7.095	0.028	–	2.811	0.021	–	0.213	0.059	–	655.0	185.4
30	$Q_{25\%}$	–	7.061	0.024	–	0.555	0.009	–	0.081	0.041	–	263.2	171.0
	$Q_{75\%}$	–	7.123	0.032	–	4.627	0.028	–	0.303	0.072	–	950.8	198.2
	$Q_{99\%}$	–	7.223	0.044	–	8.303	0.098	–	0.834	0.150	–	1874	235.2
2000	mean	–	10.65	0.594	–	162.9	0.002	–	1.117	0.006	–	30×10^4	1985
250	$Q_{25\%}$	–	10.62	0.536	–	32.36	0.001	–	0.173	0.004	–	64886	1943
	$Q_{75\%}$	–	10.68	0.652	–	268.2	0.003	–	1.550	0.007	–	49×10^4	2026
	$Q_{99\%}$	–	10.81	0.858	–	476.7	0.008	–	8.086	0.013	–	95×10^4	2128
Panel 3: Simulation results when $\theta_1^{\text{dgp}} = 1.5358$, $\theta_2^{\text{dgp}} = 6.7339$, and $\theta_3^{\text{dgp}} = 0.7623$													
200	mean	–	7.119	0.023	–	0.186	0.022	–	0.069	0.059	–	211.9	184.6
8	$Q_{25\%}$	–	7.066	0.018	–	0.049	0.011	–	0.046	0.041	–	188.1	169.7
	$Q_{75\%}$	–	7.140	0.025	–	0.292	0.030	–	0.087	0.071	–	231.9	198.0
	$Q_{99\%}$	–	7.556	0.046	–	0.613	0.067	–	0.184	0.150	–	302.0	233.8
200	mean	–	7.188	0.022	–	0.929	0.024	–	0.138	0.059	–	345.4	184.8
30	$Q_{25\%}$	–	7.158	0.018	–	0.182	0.011	–	0.066	0.041	–	223.6	170.3
	$Q_{75\%}$	–	7.220	0.024	–	1.537	0.032	–	0.166	0.071	–	431.5	196.9
	$Q_{99\%}$	–	7.304	0.035	–	2.826	0.094	–	0.562	0.145	–	813.4	234.3
2000	mean	–	10.85	0.326	–	9.752	0.002	–	0.285	0.006	–	20241	1985
250	$Q_{25\%}$	–	10.81	0.279	–	1.857	0.001	–	0.030	0.004	–	5515	1942
	$Q_{75\%}$	–	10.89	0.354	–	15.58	0.003	–	0.284	0.007	–	31810	2025
	$Q_{99\%}$	–	11.08	0.552	–	28.34	0.009	–	2.839	0.013	–	61181	2124
Panel 4: Simulation results when $\theta_1^{\text{dgp}} = 1.5358$, $\theta_2^{\text{dgp}} = 6.7339$, and $\theta_3^{\text{dgp}} = 0.0000$													
200	mean	–	6.955	0.020	–	0.023	0.020	–	0.060	0.589	–	188.7	185.1
8	$Q_{25\%}$	–	6.920	0.015	–	0.009	0.008	–	0.040	0.040	–	173.9	170.2
	$Q_{75\%}$	–	6.989	0.022	–	0.027	0.028	–	0.071	0.070	–	202.6	198.5
	$Q_{99\%}$	–	7.087	0.062	–	0.136	0.063	–	0.157	0.151	–	238.5	234.0
200	mean	–	7.029	0.018	–	0.422	0.021	–	0.826	0.060	–	263.7	185.3
30	$Q_{25\%}$	–	6.983	0.015	–	0.093	0.008	–	0.051	0.041	–	203.9	170.9
	$Q_{75\%}$	–	7.051	0.019	–	0.667	0.028	–	0.104	0.070	–	307.1	197.9
	$Q_{99\%}$	–	7.269	0.030	–	1.189	0.094	–	0.224x	0.140	–	468.0	234.3
2000	mean	–	10.76	0.322	–	5.277	0.003	–	0.038	0.006	–	12489	1986
250	$Q_{25\%}$	–	10.71	0.278	–	1.030	0.001	–	0.011	0.004	–	4126	1943
	$Q_{75\%}$	–	10.79	0.351	–	8.787	0.003	–	0.051	0.007	–	19341	2027
	$Q_{99\%}$	–	11.22	0.553	–	15.28	0.009	–	0.138	0.013	–	35419	2129

Note: Results are based on 500 simulated samples. In profiling, grid step η equals 0.2, $\theta_1 \in [0.1, 1.9], \theta_2 \in [1.2, 20]$, and $\theta_3 \in [0, 3]$.

Chapter 5

"MIDAS Analytic" extension to the MIDAS MATLAB Toolbox (software)

Alexander Kostrov [1]

[1]University of St. Gallen, Faculty of Mathematics and Statistics, alexander.kostrov@unisg.ch

Description: "MIDAS Analytic" extension supports estimating MIDAS regressions via MIDAS-NLS with revised optimization (introduced in Chapter 4). As compared to alternatives, our method exhibits better performance in terms of computation time and accuracy of estimation. "MIDAS Analytic" extension can be used autonomously or alongside the MIDAS MATLAB Toolbox[2] written by Eric Ghysels (Ghysels, 2013).

Download link: MATLAB Central File Exchange (No. 81243)

Details: The "MIDAS Analytic" extension contains functions to perform MIDAS regression analysis with Beta or Exponential Almon lags and to make predictions based on estimated parameters.

- The main function to handle a classical univariate MIDAS regression (including DL-MIDAS, ADL-MIDAS, and FADL-MIDAS) is reg_midas_analytic.
- The main function to apply a multivariate MIDAS regression (with more than one MIDAS-weighted variable) is reg_mvmidas_analytic.
- The main function to run a MIDAS logit regression is logit_midas_analytic.
- The main function to use a MIDAS VAR regression is var_midas_analytic.

This chapter explains the computation of the analytic derivatives which are used in the optimization behind MIDAS-NLS. The "MIDAS Analytic" extension contains Example_model_name.m files, one for each MIDAS model, which demonstrate how to simulate data, estimate regression coefficients, and make forecasts.

[2]The MIDAS MATLAB Toolbox is available on the MATLAB Central File Exchange (No. 45150).

5.1 Classical univariate MIDAS model

Model and input data

In order to describe the input data objects, we consider the ADL-MIDAS model (Ghysels and Marcellino, 2018, Chapter 12). Following the standard notations, a univariate ADL-MIDAS regression is expressed as

$$y_{t+1}^{L} = \mu + \lambda y_t^{L} + \beta w(L^{1/m}, \theta) x_t^{H} + \epsilon_{t+1}^{L}, \tag{5.1}$$

where x is sampled at a higher-frequency, m times per a low-frequency period, and $w(L^{1/m}, \theta)$ is a weight function which depends on the parameter(s) θ. For the sake of generality, we switch to the matrix representation of (5.1):

$$\boldsymbol{y} = V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\boldsymbol{\theta}) + \boldsymbol{\epsilon}. \tag{5.2}$$

In this particular example of ADL-MIDAS, the matrix V contains a column-vector of ones and the past values of y. Respectively, the parameter vector $\boldsymbol{\rho} = (\mu, \lambda)$. The coefficient β is a scalar, Z is a matrix of p lagged vectors x_t in columns, where p is the number of MIDAS lags (also known as MIDAS temporal aggregation period).

Consider an application of (5.1) from the MIDAS MATLAB Toolbox documentation, in which quarterly y is predicted using its past realization and 6 monthly lags of x ($p = 6$), for the period from 1-Jan-1985 to 1-Jan-2009. The corresponding matrix representation is

$$\begin{bmatrix} y_{\text{1-Jan-85}} \\ y_{\text{1-Apr-85}} \\ y_{\text{1-Jul-85}} \\ \vdots \\ y_{\text{1-Jan-09}} \end{bmatrix}_{T\times 1} = \begin{bmatrix} 1 & y_{\text{1-Oct-84}} \\ 1 & y_{\text{1-Jan-85}} \\ 1 & y_{\text{1-Apr-85}} \\ \vdots & \vdots \\ 1 & y_{\text{1-Oct-08}} \end{bmatrix}_{T\times k} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}_{k\times 1}$$
$$+ \beta \begin{bmatrix} x_{\text{1-Oct-85}} & x_{\text{1-Sep-85}} & \cdots & x_{\text{1-May-85}} \\ x_{\text{1-Jan-85}} & x_{\text{1-Dec-84}} & \cdots & x_{\text{1-Aug-84}} \\ x_{\text{1-Apr-85}} & x_{\text{1-Mar-85}} & \cdots & x_{\text{1-Nov-84}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\text{1-Oct-08}} & x_{\text{1-Sep-08}} & \cdots & x_{\text{1-May-08}} \end{bmatrix}_{T\times p} \begin{bmatrix} w_1(\boldsymbol{\theta}) \\ w_2(\boldsymbol{\theta}) \\ \vdots \\ w_p(\boldsymbol{\theta}) \end{bmatrix}_{p\times 1} + \begin{bmatrix} \varepsilon_{\text{1-Jan-85}} \\ \varepsilon_{\text{1-Apr-85}} \\ \varepsilon_{\text{1-Jul-85}} \\ \vdots \\ \varepsilon_{\text{1-Jan-09}} \end{bmatrix}_{T\times 1}, \tag{5.3}$$

where MIDAS weights $\boldsymbol{w}(\boldsymbol{\theta})$ are defined by the Beta or the Exponential Almon scheme. Entries in V and Z must be finite and numeric in order to pass a validity check in the "MIDAS Analytic" extension. Note that MixFreqData function from the MIDAS MATLAB Toolbox might be instrumental to construct Z in the required format from a time series.

reg_midas_analytic

Estimates a univariate MIDAS regression with Beta or Exponential Almon weights.
args_hat = reg_midas_analytic(...);
[args_hat, RSS, comptime] = reg_midas_analytic(...);

(Required) Data objects $\boldsymbol{y}$, V, and Z. Refer to Section 5.1 for details.

(Required) Weighting function: Beta 'beta' or Exponential Almon 'expalm'.

Beta weights:

$$\boldsymbol{w}(\theta_1, \theta_2, \theta_3) = \boldsymbol{x}^{\theta_1 - 1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2 - 1} \cdot \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} + \theta_3. \tag{5.4}$$

Exponential Almon weights:

$$\boldsymbol{w}(\theta_1, \theta_2) = \exp(\theta_1 \boldsymbol{\xi} + \theta_2 \boldsymbol{\xi}^2). \tag{5.5}$$

(Optional) Restrict parametrization of the Beta weights, 'betacase'.
Default value of 'betacase' is 3.

		Restricted cases		
'betacase'	0	1	2	3
$\theta_1 :=$	θ_1	1	θ_1	1
$\theta_2 :=$	θ_2	θ_2	θ_2	θ_2
$\theta_3 :=$	θ_3	θ_3	0	0

(Optional) Specify the starting point 'point' as a column-vector. Mind the order of the parameters: $(\theta_1, \theta_2, \theta_3, \beta, \boldsymbol{\rho})$ in the 'beta' case and $(\theta_1, \theta_2, \beta, \boldsymbol{\rho})$ in the 'expalm' case.
Default value for 'beta': [1; 5; 0; 1; zeros(k,1)].
Default value for 'expalm': [0.25; -0.25; 1; zeros(k,1)].

(Optional) Try multiple random starting points, 'multistart'.
Default value of 'multistart' is 0.

'multistart'	Meaning
0	Do not use multistart
1	Use multistart

(Optional) Distribute starting points to multiple workers. 'multistart' should be previously activated.
Default value of 'run_parallel' is 1.

'run_parallel'	Meaning
0	Do not use parallel computations
1	Run code in parallel. Parallel Computing Toolbox in MATLAB is needed to use this option.

(Optional) Specify the number of random starting points in multistart, 'runs_number'. Default value of 'runs_number' is 16.

Output

- args_hat - a vector of estimated parameters: $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\beta}, \hat{\boldsymbol{\rho}})$ in the 'beta' case and $(\hat{\theta}_1, \hat{\theta}_2, \hat{\beta}, \hat{\boldsymbol{\rho}})$ in the 'expalm' case.
- RSS, optimal value of the target function, the sum of squared residuals.
- comptime, time in seconds spent on solving the minimization problem.

Examples

- args_hat = reg_midas_analytic(y, V, Z, 'expalm', 'multistart', 1, 'run_parallel', 1, 'runs_number', 8);
- [args_hat, RSS] = reg_midas_analytic(y, V, Z, 'beta', 'betacase', 3);

predict_reg_midas

Computes prediction of the output variable using the estimated coefficients.
y_hat = predict_reg_midas(...);

Parameters and Inputs

(Required) A vector of estimated parameters args_hat, an output of reg_midas_analytic function.

(Required) Data objects V and Z. Refer to Section 5.1 for details.

(Required) Weighting function: Beta 'beta' or Exponential Almon 'expalm'.

Output

- y_hat - predicted output variable.

Examples

- y_hat = predict_reg_midas(args_hat, V, Z, 'beta');
- y_hat = predict_reg_midas(args_hat, V, Z, 'expalm');

predict_weights

Computes prediction for the MIDAS weights using the estimated coefficients.
[w, bw] = predict_weights(...);
w = predict_weights(...);

Parameters and Inputs

(Required) A vector of estimated parameters args_hat, an output of reg_midas_analytic function.

(Required) Weighting function: Beta 'beta' or Exponential Almon 'expalm'.

(Required) Specify the number of MIDAS lags p.

Output

- w - vector of MIDAS weights, $\boldsymbol{w}(\boldsymbol{\theta})$ in terms of Equation (5.2).
- bw - vector of MIDAS weights multiplied by the slope coefficient, $\beta\boldsymbol{w}(\boldsymbol{\theta})$ in terms of Equation (5.2).

Examples

- w = predict_weights(args_hat, 'beta', 8);
- [w, bw] = predict_weights(args_hat, 'expalm', 15);

sim_reg_midas

Simulates data for the MIDAS model.
[y, V, Z] = sim_reg_midas(...);

Parameters and Inputs

(Required) Weighting function: Beta 'beta' or Exponential Almon 'expalm'.

(Required) Specify the number of MIDAS lags (MIDAS temporal aggregation period). As many as p lags of a random normal vector are stacked in columns into the output matrix Z.

(Required) Specify the number of column-vectors in V, k: a single column-vector of ones and $(k-1)$ i.i.d. random normal variables.

(Required) Define the number of observations T.

(Optional) Parameters in the data generating process: 'thetas', 'beta', and 'rho'.
Default values for the Beta case:
'thetas'=[1; 16.3983; 0.7623]; 'beta'=rand(1,1)*10; 'rho'=rand(k,1)*10;
Default values for the Exponential Almon case:
'thetas'=[0.5; -0.5]; 'beta'=rand(1,1)*10; 'rho'=rand(k,1)*10;

Output

- Simulated objects $\boldsymbol{y}$, V, and Z. Refer to Section 5.1 for details.

Examples

- p=30; k=10; T=2000;
  thetas_dgp = [1.34; 6.73; 0.03];
  beta_dgp = rand(1,1)*10;
  rho_dgp = rand(k,1)*10;
  [y, V, Z] = sim_reg_midas('beta', p, k, T, 'thetas', thetas_dgp, 'beta', beta_dgp, 'rho', rho_dgp);
- [y, V, Z] = sim_reg_midas('expalm', p, k, T);

Example of usage, technical functions

Technical functions:

- estimate_reg_beta: Computes value and the analytical gradient of the target function which is minimized during the MIDAS model estimation with Beta weights.
- AnalyticalHessian_reg_beta: Computes the analytical Hessian matrix of the target function which is minimized during the MIDAS model estimation with Beta weights.
- estimate_reg_expalm: Computes value, the analytical gradient, and the analytical Hessian matrix of the target function which is minimized during the MIDAS model estimation with Exponential Almon weights.

File Example_reg_MIDAS.m demonstrates the use of functions related to the classical univariate MIDAS regression which are described in Section 5.1.

5.2 Multivariate MIDAS model

Model and input data

The main feature of the multivariate MIDAS model with the Exponential Almon weights is that the regression equation includes more than one MIDAS-weighted variable in the RHS part as compared to Equation (5.2):

$$\boldsymbol{y} = V\boldsymbol{\rho} + \sum_{i=1}^{\text{numZ}} \beta_i Z_i \boldsymbol{w}_i(\theta_{1,i}, \theta_{2,i}) + \boldsymbol{\epsilon}, \tag{5.6}$$

where numZ is the number of MIDAS-weighted variables. Each of them has the individual number of MIDAS lags, which coincides with the number of columns in each matrix Z_i (refer to Section 5.1 for details).

Consider the case of two MIDAS-weighed variables with 5 and 7 lags, respectively. Auxiliary matrix B, vector $\tilde{\boldsymbol{\xi}}$, and vector $\tilde{\boldsymbol{\xi}}^2$ are introduced as

$$B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \hline 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}_{(5+7)\times 2}, \tilde{\boldsymbol{\xi}} = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ \hline 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{bmatrix}_{(5+7)\times 1}, \tilde{\boldsymbol{\xi}}^2 = \begin{bmatrix} \xi_1^2 \\ \xi_2^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 9 \\ 16 \\ 25 \\ \hline 1 \\ 4 \\ 9 \\ 16 \\ 25 \\ 36 \\ 49 \end{bmatrix}_{(5+7)\times 1}. \tag{5.7}$$

Then, parameters $\theta_{1,i}$, $\theta_{2,i}$, and β_i are grouped into column-vectors of parameters:

$$\boldsymbol{\theta_1} = \begin{bmatrix} \theta_{1,1} \\ \theta_{1,2} \end{bmatrix}_{2\times 1}, \boldsymbol{\theta_2} = \begin{bmatrix} \theta_{2,1} \\ \theta_{2,2} \end{bmatrix}_{2\times 1}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}_{2\times 1}. \tag{5.8}$$

Finally, Z_i matrices are stacked column-to-column into a united large matrix Z:

$$[Z_{1\ T\times 5}, Z_{2\ T\times 7}] = Z_{T\times(5+7)} \tag{5.9}$$

Using the introduced objected, Equation (5.6) can be reformulated as

$$\boldsymbol{y} = V\boldsymbol{\rho} + Z(B\boldsymbol{\beta} \odot \boldsymbol{w}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})) + \boldsymbol{\epsilon}, \tag{5.10}$$

where

$$\boldsymbol{w}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \exp(B\boldsymbol{\theta_1} \odot \tilde{\boldsymbol{\xi}} + B\boldsymbol{\theta_2} \odot \tilde{\boldsymbol{\xi}}^2). \tag{5.11}$$

reg_mvmidas_analytic

Estimates a multivariate MIDAS regression with Exponential Almon weights.
args_hat = reg_mvmidas_analytic(...);
[args_hat, RSS, comptime] = reg_mvmidas_analytic(...);

Parameters and Inputs

(Required) Data objects $\boldsymbol{y}$, V, and Z. Refer to Section 5.2 for details. Note that Z is constructed as a "united" matrix.

(Required) Specify the number of MIDAS lags for each MIDAS-weighed variable in the order or stacking into the "united" matrix Z (starting from Z_1): vector 'pZ'.

(Required) Weighting function: Exponential Almon 'expalm' is the only available specification at the moment.

(Optional) Specify the starting point 'point' as a column-vector. Mind the order of the parameters: Vector of thetas 1, vector of thetas 2, vector of betas, and rho: $(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \beta, \boldsymbol{\rho})$.
Default value for 'expalm':
[repelem(0.25,numZ,1); repelem(-0.25,numZ,1); repelem(1,numZ,1); zeros(k,1)];

(Optional) Try multiple random starting points, 'multistart'.
Default value of 'multistart' is 0.

'multistart'	Meaning
0	Do not use multistart
1	Use multistart

(Optional) Distribute starting points to multiple workers. 'multistart' should be previously activated.
Default value of 'run_parallel' is 1.

'run_parallel'	Meaning
0	Do not use parallel computations
1	Run code in parallel. Parallel Computing Toolbox in MATLAB is needed to use this option.

(Optional) Specify the number of random starting points in multistart, 'runs_number'.
Default value of 'runs_number' is 16.

Output

- args_hat - a vector of estimated parameters: $(\hat{\boldsymbol{\theta}}_{\mathbf{1}}, \hat{\boldsymbol{\theta}}_{\mathbf{2}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\rho}})$.
- RSS, optimal value of the target function, the sum of squared residuals.
- comptime, time in seconds spent on solving the minimization problem.

Examples

- args_hat = reg_mvmidas_analytic(y, V, Z, pZ, 'expalm', 'multistart', 1, 'run_parallel', 1, 'runs_number', 16);
- [args_hat, RSS] = reg_mvmidas_analytic(y, V, Z, pZ, 'expalm');

predict_reg_mvmidas

Computes prediction of the output variable using the estimated coefficients.
y_hat = predict_reg_mvmidas(...);

Parameters and Inputs

(Required) A vector of estimated parameters args_hat, an output of reg_mvmidas_analytic function.

(Required) Data objects V and Z. Refer to Section 5.2 for details.

(Required) Specify the number of MIDAS lags for each MIDAS-weighed variable in the order or stacking into the "united" matrix Z (starting from Z_1): vector 'pZ'.

(Required) Exponential Almon 'expalm' is the only available specification at the moment.

Output

- y_hat - predicted output variable.

Examples

- y_hat = predict_reg_midas(args_hat, V, Z, pZ, 'expalm');

predict_weights_mv

Computes prediction for the MIDAS weights using the estimated coefficients.
[w_mat, bw_mat] = predict_weights_mv(...);
w_mat = predict_weights_mv(...);

Parameters and Inputs

(Required) A vector of estimated parameters args_hat, an output of reg_mvmidas_analytic function.

(Required) Exponential Almon 'expalm' is the only available specification at the moment.

(Required) Specify the number of MIDAS lags for each MIDAS-weighed variable in the order or stacking into the "united" matrix Z (starting from Z_1): vector 'pZ'.

Output

- w_mat - a matrix of MIDAS weights, each column i contains weights $\boldsymbol{w_i}(\theta_{1,i}, \theta_{2,i})$ in terms of Equation (5.6).
- bw_mat - a matrix of MIDAS weights multiplied by the slope coefficient, each column i contains $\beta_i \boldsymbol{w_i}(\theta_{1,i}, \theta_{2,i})$ in terms of Equation (5.6).

Examples

- w_mat = predict_weights_mv(args_hat, 'expalm', [5, 7]);
- [w_mat, bw_mat] = predict_weights_mv(args_hat, 'expalm', [5, 7]);

sim_reg_mvmidas

Simulates data for the multivariate MIDAS model.
[y, V, Z] = sim_reg_mvmidas(...);

Parameters and Inputs

(Required) Exponential Almon 'expalm' is the only available specification at the moment.

(Required) Specify the number of MIDAS lags for each MIDAS-weighed variable in the order or stacking into the "united" matrix Z (starting from Z_1): vector 'pZ'.

(Required) Specify the number of column-vectors in V, k: a single column-vector of ones and $(k-1)$ i.i.d. random normal variables.

(Required) Define the number of observations T.

(Optional) Parameters in the data generating process: 'thetas1', 'thetas2', 'betas', and 'rho'. Default values for the Exponential Almon case:
'thetas1'=repelem(0.5,numZ,1);
'thetas2'=repelem(-0.5,numZ,1);

'betas'=rand(numZ,1)*10;
'rho'=rand(k,1)*10;

Output

- Simulated objects $\boldsymbol{y}$, V, and Z. Refer to Section 5.2 for details.

Examples

- pZ = [25,17]; T = 4000; k = 1;
  thetas1_dgp = [0.123; 0.321];
  thetas2_dgp = [-0.278; -0.437];
  betas_dgp = rand(2,1)*10;
  rho_dgp = rand(k,1)*10;
- [y, V, Z] = sim_reg_mvmidas('expalm', pZ, k, T, 'thetas1', thetas1_dgp, ...
  'thetas2', thetas2_dgp, 'betas', betas_dgp, 'rho', rho_dgp);

Example of usage, technical functions

Technical function:

- estimate_reg_expalm_mn: Computes value and the analytical gradient of the target function which is minimized in order to estimate coefficients of the multivariate MIDAS model with the Exponential Almon weights.

File Example_reg_mvMIDAS.m demonstrates the use of functions related to the multivariate MIDAS model which are described in Section 5.2.

5.3 MIDAS logit model

Model and input data

Input data comply with requirements in Section 5.1. Note that the vector $\boldsymbol{y}$ has binary entries in this case. We refer the reader to Audrino, Kostrov and Ortega (2019) for details about the model.

logit_midas_analytic

Estimates a univariate MIDAS logit regression.
args_hat = logit_midas_analytic(...);
[args_hat, LogL, comptime] = logit_midas_analytic(...);

Parameters and Inputs

(Required) Data objects $\boldsymbol{y}$, V, and Z. Refer to Section 5.1 for details.

(Required) Weighting function: Exponential Almon 'expalm' is the only available specification at the moment.

$$\boldsymbol{w}(\theta_1, \theta_2) = \exp(\theta_1 \boldsymbol{\xi} + \theta_2 \boldsymbol{\xi}^2). \tag{5.12}$$

(Optional) Specify the starting point 'point' as a column-vector. Mind the order of the parameters: $(\theta_1, \theta_2, \beta, \boldsymbol{\rho})$.
Default value for 'expalm': [0.25; -0.25; 1; zeros(k,1)].

(Optional) Try multiple random starting points, 'multistart'.
Default value of 'multistart' is 0.

'multistart'	Meaning
0	Do not use multistart
1	Use multistart

(Optional) Distribute starting points to multiple workers. 'multistart' should be previously activated.
Default value of 'run_parallel' is 1.

'run_parallel'	Meaning
0	Do not use parallel computations
1	Run code in parallel. Parallel Computing Toolbox in MATLAB is needed to use this option.

(Optional) Specify the number of random starting points in multistart, 'runs_number'. Default value of 'runs_number' is 16.

Output

- args_hat - a vector of estimated parameters: $(\hat{\theta}_1, \hat{\theta}_2, \hat{\beta}, \hat{\boldsymbol{\rho}})$.
- LogL - optimal value of the target function, minimized (minus) log-likelihood.
- comptime - time in seconds spent on solving the minimization problem.

Examples

- args_hat = logit_midas_analytic(y, V, Z, 'expalm', 'multistart', 1, 'run_parallel', 1, 'runs_number', 16);
- [args_hat, LogL] = logit_midas_analytic(y, V, Z, 'expalm');

predict_logit_midas

Computes prediction of the output variable using the estimated coefficients.
y_hat = predict_logit_midas(...);

Parameters and Inputs

(Required) A vector of estimated parameters args_hat, an output of logit_midas_analytic function.

(Required) Data objects V and Z. Refer to Section 5.1 for details.

(Required) Weighting function. Exponential Almon 'expalm' is the only available specification at the moment.

predict_weights

Computes prediction for the MIDAS weights using the estimated coefficients.
[w, bw] = predict_weights(...);
w = predict_weights(...);
The function is described in Section 5.1.

Output

- y_hat - a vector of predicted probabilities.

Examples

- y_hat = predict_logit_midas(args_hat, V, Z, 'expalm');

sim_logit_midas

Simulates data for the MIDAS logit model.
[y, V, Z] = sim_logit_midas(...);

Parameters and Inputs

(Required) Weighting function: Exponential Almon 'expalm' is the only available specification at the moment.

(Required) Specify the number of MIDAS lags. As many as p lags of a random normal vector are stacked in columns into the output matrix Z.

(Required) Specify the number of column-vectors in V, k: a single column-vector of ones and $(k-1)$ i.i.d. random normal variables.

(Required) Define the number of observations T.

(Optional) Parameters in the data generating process: 'thetas', 'beta', and 'rho'. Default values for the Exponential Almon case:
'thetas'=[0.5; -0.5]; 'beta'=rand(1,1)*10; 'rho'=rand(k,1)*10;

Output

- Simulated objects $\boldsymbol{y}$, V, and Z. Refer to Section 5.1 for details.

Examples

- p=22; k=1; T=2000;
  thetas_dgp = [0.22; -0.37];
  beta_dgp = rand(1,1)*10;
  rho_dgp = rand(k,1)*10;
  [y, V, Z] =sim_logit_midas('expalm', p, k, T, 'thetas', thetas_dgp, 'beta', beta_dgp, 'rho', rho_dgp);
- [y, V, Z] = sim_logit_midas('expalm', p, k, T);

Example of usage, technical functions

Technical function:

- estimate_logit_expalm: Computes value and the analytical gradient of the target function which is minimized in order to estimate coefficients of the MIDAS logit model with the Exponential Almon weights.

File Example_logit_MIDAS.m demonstrates the use of functions related to the multivariate MIDAS model which are described in Section 5.3.

5.4 MIDAS VAR model (mixed frequency VAR)

Ghysels (2016) introduced a mixed frequency VAR model which involves the use of MIDAS. Equation (5.13), that is Equation (2.9) in Ghysels (2016), specifies the MIDAS VAR model with p lags for a single low-frequency time series which is observed every m fixed periods of time:

$$\begin{aligned}\begin{bmatrix} x_H(\tau,1) \\ \vdots \\ x_H(\tau,1) \\ x_L(\tau_L) \end{bmatrix} = &\begin{bmatrix} 0 & \dots & \rho & a \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \rho^m & a\sum_{j=0}^{m-1}\rho^j \\ \omega(\gamma)_m & \dots & \omega(\gamma)_1 & \alpha_1 \end{bmatrix} \times \begin{bmatrix} x_H(\tau-1,1) \\ \vdots \\ x_H(\tau-1,1) \\ x_L(\tau_L-1) \end{bmatrix} \\ &+\sum_{j=2}^{p} \begin{bmatrix} 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \\ \omega(\gamma)_{jm} & \dots & \omega(\gamma)_{(j-1)m+1} & \alpha_j \end{bmatrix} \times \begin{bmatrix} x_H(\tau-1,1) \\ \vdots \\ x_H(\tau-1,1) \\ x_L(\tau_L-1) \end{bmatrix} + \epsilon(\tau_L)\end{aligned} \tag{5.13}$$

Model and input data

MIDAS VAR is a multi-equation model. In the following example, we consider a particular case of $m = 4$ when four quarterly time series $q_i, i \in \{1 \dots 4\}$ are stacked with y, which is sampled annually, and the number of low-frequency lags p in the VAR model is 3.

$$Y = CX + U, \tag{5.14}$$

where

$$Y = \begin{bmatrix} x_H(\tau,1) \\ \vdots \\ x_H(\tau,1) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} q_{1,1} & \dots & q_{1,T} \\ q_{2,1} & \dots & q_{2,T} \\ q_{3,1} & \dots & q_{3,T} \\ q_{4,1} & \dots & q_{4,T} \\ y_1 & \dots & y_T \end{bmatrix},$$

$$U = \begin{bmatrix} u_{1,1} & \dots & u_{1,T} \\ u_{2,1} & \dots & u_{2,T} \\ u_{3,1} & \dots & u_{3,T} \\ u_{4,1} & \dots & u_{4,T} \\ u_{5,1} & \dots & u_{5,T} \end{bmatrix},$$

$$CX = \underbrace{\begin{bmatrix} 0 & 0 & 0 & \rho & a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho^2 & a(1+\rho) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho^3 & a(1+\rho+\rho^2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho^4 & a(1+\rho+\rho^2+\rho^3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \omega_4 & \omega_3 & \omega_2 & \omega_1 & \alpha_1 & \omega_8 & \omega_7 & \omega_6 & \omega_5 & \alpha_2 & \omega_{12} & \omega_{11} & \omega_{10} & \omega_9 & \alpha_3 \end{bmatrix}}_{\text{Lag 1}} \underbrace{\qquad\qquad}_{\text{Lag 2}} \underbrace{\qquad\qquad}_{\text{Lag 3}} \times \begin{bmatrix} q_{1,0} & \cdots & q_{1,T-1} \\ q_{2,0} & \cdots & q_{2,T-1} \\ q_{3,0} & \cdots & q_{4,T-1} \\ q_{4,0} & \cdots & q_{1,T-1} \\ y_0 & \cdots & y_{T-1} \\ q_{1,-1} & \cdots & q_{1,T-2} \\ q_{2,-1} & \cdots & q_{2,T-2} \\ q_{3,-1} & \cdots & q_{4,T-2} \\ q_{4,-1} & \cdots & q_{1,T-2} \\ y_{-1} & \cdots & y_{T-2} \\ q_{1,-2} & \cdots & q_{1,T-3} \\ q_{2,-2} & \cdots & q_{2,T-3} \\ q_{3,-2} & \cdots & q_{4,T-3} \\ q_{4,-2} & \cdots & q_{1,T-3} \\ y_{-2} & \cdots & y_{T-3} \end{bmatrix}.$$

Equation (5.14) can be reorganized (Lütkepohl, 2005, pp. 70–71) as

$$\begin{aligned} \text{vec}(Y) &= \text{vec}(CX) + \text{vec}(U), \\ &= (X^\top \otimes I)\text{vec}(C) + \text{vec}(U), \\ \boldsymbol{y} &= (X^\top \otimes I)\boldsymbol{c} + \boldsymbol{u}, \end{aligned} \tag{5.15}$$

where vec is the column stacking operator and $\otimes$ denotes the Kronecker product, $\boldsymbol{y} = \text{vec}(Y)$, $\boldsymbol{c} = \text{vec}(C)$, and $\boldsymbol{u} = \text{vec}(U)$. The OLS estimation of $\boldsymbol{c}$ means to choose the value that minimizes

$$\boldsymbol{u}^\top \boldsymbol{u} = \left(\boldsymbol{y} - (X^\top \otimes I)\boldsymbol{c}\right)^\top \left(\boldsymbol{y} - (X^\top \otimes I)\boldsymbol{c}\right). \tag{5.16}$$

Further vectorization of the target function (5.15) requires the coefficient matrix C to be represented as follows:

$$C = \rho\left(R_1 + aR_{a1}\right) + \rho^2\left(R_2 + aR_{a2}\right) + \rho^3\left(R_3 + aR_{a3}\right) + \rho^4 R_4 + aR_{a0} + \boldsymbol{i}_\omega \boldsymbol{\omega}^\top R_\omega + \boldsymbol{i}_\omega \boldsymbol{\alpha}^\top R_\alpha, \tag{5.17}$$

where a, ρ, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3]^\top$, and $\boldsymbol{\omega} = [\omega_1, \omega_2, \ldots, \omega_{12}]^\top$ are parameters from coefficient matrix C. In order to compute the first five terms in C, the below mentioned objects are demanded.

$$R_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; R_{a1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$R_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; R_{a2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$R_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; R_{a3} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$R_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$R_{a0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The term $\boldsymbol{i}_\omega \boldsymbol{\alpha}^\top R_\alpha$ requires the use of $\boldsymbol{i}_\omega = [0,0,0,0,1]^\top$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3]^\top$, and

$$R_\alpha = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Finally, the computation of $\boldsymbol{i}_\omega \boldsymbol{\omega}^\top R_\omega$ relies on supplementary objects $\boldsymbol{i}_\omega = [0,0,0,0,1]^\top$, $\boldsymbol{\omega} = [\omega_1, \omega_2, \ldots, \omega_{12}]^\top$, and

$$R_\omega = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{5.18}$$

var_midas_analytic

Estimates a MIDAS VAR regression with Beta weights.
args_hat = var_midas_analytic(...);
[args_hat, RSS, comptime] = var_midas_analytic(...);

Parameters and Inputs

(Required) Data objects Y and X. Refer to Section 5.4 for details.

(Required) Weighting function: Beta 'beta' is the only available specification at the moment.

$$\boldsymbol{w}(\theta_1, \theta_2, \theta_3) = \boldsymbol{x}^{\theta_1 - 1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2 - 1} \cdot \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} + \theta_3.$$

(Optional) Restrict parametrization of the Beta weights, 'betacase'.
Default value of 'betacase' is 3.

(Optional) Specify the starting point 'point' as a column-vector. Mind the order of the parameters: $(\theta_1, \theta_2, \theta_3, \beta, \rho, a, \boldsymbol{\alpha})$, vector $\boldsymbol{\alpha}$ has p entries.
Default value for 'beta': [1; 5; 0; repelem(0.5, 3+p,1)].

(Optional) Try multiple random starting points, 'multistart'.
Default value of 'multistart' is 0.

'multistart'	Meaning
0	Do not use multistart
1	Use multistart

(Optional) Distribute starting points to multiple workers. 'multistart' should be previously activated.
Default value of 'run_parallel' is 1.

'run_parallel'	Meaning
0	Do not use parallel computations
1	Run code in parallel. Parallel Computing Toolbox in MATLAB is needed to use this option.

(Optional) Specify the number of random starting points in multistart, 'runs_number'.
Default value of 'runs_number' is 16.

Output

- args_hat - a vector of estimated parameters: $(\hat{\theta_1}, \hat{\theta_2}, \hat{\theta_3}, \hat{\beta}, \hat{\rho}, \hat{a}, \hat{\boldsymbol{\alpha}})$.
- RSS, optimal value of the target function, the sum of squared residuals.
- comptime, time in seconds spent on solving the minimization problem.

Examples

- args_hat = var_midas_analytic(Y, X, 'beta', 'multistart', 1, 'run_parallel', 1, 'runs_number', 8);
- [args_hat, RSS] = var_midas_analytic(Y, X, 'beta', 'betacase', 1);

predict_weights

Computes prediction for the MIDAS weights using the estimated coefficients.
[w, bw] = predict_weights(...);
w = predict_weights(...);
The function is described in Section 5.1.

Examples

- [w, bw] = predict_weights(args_hat,'beta', p*m);

sim_var_midas

Simulates data for the MIDAS VAR model with a single low-frequency variable.
[Y, X] = sim_var_midas(...);

Parameters and Inputs

(Required) Weighting function: Beta 'beta' or Exponential Almon 'expalm'.

(Required) Specify p, the number of low-frequency lags in the VAR model.

(Required) Specify m, the number of high-frequency periods per a low-frequency period.

(Required) Define the number of (low-frequency) observations T.

(Optional) Parameters in the data generating process: 'thetas', 'beta', 'rho', 'a', 'alph'.
Default values for the Beta case:
'thetas'=[1.2; 4; 0.03]; 'beta'=0.053; 'rho'=0.45; 'a'=0.05;
temp_alph = rand(p, 1); 'alph'= temp_alph /sum(temp_alph)*(0.3+0.5*rand);

Output

- Simulated matrices X and Y. Refer to Section 5.4 for details.

Examples

- thetas_dgp = [1.34; 6.73; 0.03];
  beta_dgp = 0.048;
  rho_dgp = 0.32;
  a_dgp = 0.061;
  [Y, X] = sim_var_midas('beta', p, m, T, 'thetas', thetas_dgp, 'beta', beta_dgp, 'rho', rho_dgp, 'a', a_dgp);

Example of usage, technical functions

Technical functions:

- estimate_var_beta: Computes value and the analytical gradient of the target function which is minimized during the MIDAS VAR model estimation with Beta weights.
- AnalyticalHessian_var_beta: Computes the analytical Hessian matrix of the target function which is minimized during the MIDAS VAR model estimation with Beta weights.

File Example_var_MIDAS.m demonstrates the use of functions related to the classical univariate MIDAS regression which are described in Section 5.4.

5.5 Notations

Mathematical objects

- Scalar value: "a" – a regular symbol.
- Vector: "$\boldsymbol{a}$" – a bold symbol.
- Matrix: "A" – a capital symbol.

Mathematical operations and operators

- "$\odot$" stands for element-wise multiplication.
- "$\otimes$" denotes the Kronecker product.
- "$\langle \cdot, \cdot \rangle$" stands for scalar (dot) multiplication.

- "∇f" is the gradient of a function f.
- "$A^\top$" is used for the transpose of an object A.
- diag – the operator that transforms a given vector into a diagonal matrix with its entries in the main diagonal: $\text{diag}(\boldsymbol{a}_k) = A_{k\times k}$.
- T^* – the adjoint of the tangent map T:

$$\langle T^*(\alpha), v\rangle = \langle \alpha, T(v)\rangle$$

Common auxiliary elements

- p – a temporal aggregation period in the MIDAS regression.
- k – the number of factors.
- Z – a matrix of p consequent lagged versions of a MIDAS variable.
- V – a matrix of k variables with the first column of 1s for the intercept.
- T – the number of observations (rows in V, Z, and $\boldsymbol{y}$.).
- $\boldsymbol{\iota}$ – a column vector of 1s with p entries.
- $\boldsymbol{\xi}$ – a column vector, $[1, 2, 3, \ldots, p]^\top$.
- $\boldsymbol{x}$ – a column vector, $[1, 2, 3, \ldots, p]^\top/(p+1)$.
- $\Gamma(\xi) = \int_0^\infty \exp^{-t} t^{\xi-1}\, dt$, the Gamma function.
- $\psi(\xi)$ and $\psi_1(\xi)$ are the digamma function and trigamma functions, respectively: $\psi_{(n-1)}(\xi) = \dfrac{\partial^n}{\partial \xi^n} \ln\Gamma(\xi)$, the n^{th} logarithmic derivative of the Gamma function.
- $\tilde{\boldsymbol{w}}(\theta_1, \theta_2) = \boldsymbol{x}^{\theta_1-1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2-1} \cdot \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}$.

Appendices

Appendix 5.A Analytic derivatives for the classical univariate MIDAS model

This section explains the derivation of the analytical gradient and of the analytical Hessian matrix of the objective function in the optimization problem related to estimating the coefficients of a classical MIDAS regression (5.2) via non-linear least squares. The MIDAS-NLS estimator requires minimizing the target function (5.19):

$$\text{RSS} = f(\theta_1, \theta_2, (\theta_3), \beta, \boldsymbol{\rho}) = (\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1, \theta_2, (\theta_3)))^\top \cdot (\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1, \theta_2, (\theta_3))) \,. \tag{5.19}$$

We explain the unrestricted case of the Beta weights when optimization is performed with respect to the parameter vector $(\theta_1, \theta_2, \theta_3, \beta, \boldsymbol{\rho})$. It is straightforward to adjust derivations for the three restricted cases. Analytic derivatives for the Beta case were originally computer for Chapter 4. Derivation for the Exponential Almon case is based on results in Chapter 1.

The analytical gradient (Beta polynomial)

The above-mentioned objective function (5.19) is presented as a combination of operators:

$$f(\theta_1, \theta_2, \theta_3, \beta, \boldsymbol{\rho}) = \mathcal{G}\left(\mathcal{C}(\boldsymbol{\rho}, \beta, \mathcal{W}(\theta_1, \theta_2, \theta_3))\right).$$

$$\begin{aligned}
1)\mathcal{W} &: \mathbb{R} \times \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}^p \\
&\quad (\theta_1, \theta_2, \theta_3) \longmapsto \boldsymbol{w}, \\
2)\,\mathcal{C} &: \mathbb{R}^k \times \mathbb{R} \times \mathbb{R}^p \longrightarrow \mathbb{R}^T \\
&\quad (\boldsymbol{\rho}, \beta, \boldsymbol{w}) \longmapsto \boldsymbol{c}, \\
3)\,\mathcal{G} &: \mathbb{R}^T \longrightarrow \mathbb{R} \\
&\quad \boldsymbol{c} \longmapsto \text{RSS}.
\end{aligned}$$

Consider the directional derivative of the target function f in the direction of λ. We aim to find an explicit expression for the gradient.

$$\begin{aligned}
\langle \nabla f, \boldsymbol{\lambda} \rangle &= \left\langle \begin{pmatrix} \nabla_{\theta_1} f \\ \nabla_{\theta_2} f \\ \nabla_{\theta_3} f \\ \nabla_{\beta} f \\ \nabla_{\boldsymbol{\rho}} f \end{pmatrix}, \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ (\lambda_i)_{i \in \{5,\dots,4+k\}} \end{pmatrix} \right\rangle \\
&= \left\langle \begin{pmatrix} \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \end{pmatrix}, \begin{pmatrix} (T_{\boldsymbol{w}} \mathcal{C}) \cdot (T_{\theta_1} \mathcal{W}) \cdot \lambda_1 \\ (T_{\boldsymbol{w}} \mathcal{C}) \cdot (T_{\theta_2} \mathcal{W}) \cdot \lambda_2 \\ (T_{\boldsymbol{w}} \mathcal{C}) \cdot (T_{\theta_3} \mathcal{W}) \cdot \lambda_3 \\ (T_{\beta} \mathcal{C}) \cdot \lambda_4 \\ (T_{\boldsymbol{\rho}} \mathcal{C}) \cdot (\lambda_i)_{i \in \{5,\dots,4+k\}} \end{pmatrix} \right\rangle \\
&= \left\langle \begin{pmatrix} \left(T^*_{\theta_1} \mathcal{W}\right) \cdot (T^*_{\boldsymbol{w}} \mathcal{C}) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\theta_2} \mathcal{W}\right) \cdot (T^*_{\boldsymbol{w}} \mathcal{C}) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\theta_3} \mathcal{W}\right) \cdot (T^*_{\boldsymbol{w}} \mathcal{C}) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\beta} \mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\boldsymbol{\rho}} \mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \end{pmatrix}, \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ (\lambda_i)_{i \in \{5,\dots,4+k\}} \end{pmatrix} \right\rangle
\end{aligned}$$

The components of the gradient are separately computed.

Component $\nabla_{\boldsymbol{c}} \mathcal{G}$

$$\langle \nabla_{\boldsymbol{c}} \mathcal{G}, \boldsymbol{\delta c} \rangle = \frac{d}{dt}\bigg|_{t=0} \mathcal{G}(\boldsymbol{c} + t\boldsymbol{\delta c}) = \frac{d}{dt}\bigg|_{t=0} (\boldsymbol{c} + t\boldsymbol{\delta c})^\top (\boldsymbol{c} + t\boldsymbol{\delta c}) = 2\boldsymbol{c}^\top \cdot \boldsymbol{\delta c} = \langle \boldsymbol{\delta c}, 2\boldsymbol{c} \rangle$$

$$\boxed{\nabla_{\boldsymbol{c}} \mathcal{G} = 2\boldsymbol{c}}$$

Calculating components $T^*_{\boldsymbol{\rho}} \mathcal{C}$, $T^*_{\beta} \mathcal{C}$, and $T^*_{\boldsymbol{w}} \mathcal{C}$

$$T_{\boldsymbol{\rho}} \mathcal{C}(\boldsymbol{\delta \rho}) = \frac{d}{dt}\bigg|_{t=0} \mathcal{C}(\boldsymbol{\rho} + t\boldsymbol{\delta \rho}, \beta, \boldsymbol{w}) = \frac{d}{dt}\bigg|_{t=0} \boldsymbol{y} - V(\boldsymbol{\rho} + t\boldsymbol{\delta \rho}) - \beta Z \boldsymbol{w} = -V \boldsymbol{\delta \rho}$$

$$T_{\beta} \mathcal{C}(\delta\beta) = \frac{d}{dt}\bigg|_{t=0} \mathcal{C}(\boldsymbol{\rho}, \beta + t\delta\beta, \boldsymbol{w}) = \frac{d}{dt}\bigg|_{t=0} \boldsymbol{y} - V\boldsymbol{\rho} - Z\boldsymbol{w}(\beta + t\delta\beta) = -Z\boldsymbol{w}\delta\beta$$

$$T_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta w}) = \frac{d}{dt}\Big|_{t=0} \mathcal{C}(\boldsymbol{\rho}, \beta, \boldsymbol{w} + t\boldsymbol{\delta w}) = \frac{d}{dt}\Big|_{t=0} \boldsymbol{y} - V\boldsymbol{\rho} - \beta Z(\boldsymbol{w} + t\boldsymbol{\delta w}) = -\beta Z\boldsymbol{\delta w}$$

$$\langle T^*_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta c}), \boldsymbol{\delta\rho}\rangle = \langle \boldsymbol{\delta c}, T_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta\rho})\rangle = \langle \boldsymbol{\delta c}, -V\boldsymbol{\delta\rho}\rangle = \langle -V^\top \boldsymbol{\delta c}, \boldsymbol{\delta\rho}\rangle$$
$$\boxed{T^*_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta c}) = -V^\top}\boldsymbol{\delta c}$$

$$\langle T^*_{\beta}\mathcal{C}(\boldsymbol{\delta c}), \delta\beta\rangle = \langle \boldsymbol{\delta c}, T_{\beta}\mathcal{C}(\delta\beta)\rangle = \langle \boldsymbol{\delta c}, -Z\boldsymbol{w}\delta\beta\rangle = \langle -\boldsymbol{w}^\top Z^\top \boldsymbol{\delta c}, \delta\beta\rangle$$
$$\boxed{T^*_{\beta}\mathcal{C}(\boldsymbol{\delta c}) = -\boldsymbol{w}^\top Z^\top}\boldsymbol{\delta c}$$

$$\langle T^*_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta c}), \boldsymbol{\delta w}\rangle = \langle \boldsymbol{\delta c}, T_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta w})\rangle = \langle \boldsymbol{\delta c}, -\beta Z\boldsymbol{\delta w}\rangle = \langle -\beta Z^\top \boldsymbol{\delta c}, \boldsymbol{\delta w}\rangle$$
$$\boxed{T^*_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta c}) = -\beta Z^\top}\boldsymbol{\delta c}$$

Calculating components $T^*_{\theta_1}\mathcal{W}, T^*_{\theta_2}\mathcal{W}, T^*_{\theta_3}\mathcal{W}$

$$\boldsymbol{w}(\theta_1, \theta_2, \theta_3) = \boldsymbol{x}^{\theta_1 - 1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2 - 1} \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} + \theta_3.$$

When $i \in \{1, 2\}$ the following statements hold:

$$\begin{aligned}
\frac{d\boldsymbol{w}(\theta_1, \theta_2, \theta_3)}{d\theta_i} &= \frac{d\tilde{\boldsymbol{w}}(\theta_1, \theta_2)}{d\theta_i} \\
\frac{d\ln\tilde{\boldsymbol{w}}(\theta_1, \theta_2)}{d\theta_i} &= \frac{1}{\tilde{\boldsymbol{w}}(\theta_1, \theta_2)} \odot \frac{d\tilde{\boldsymbol{w}}(\theta_1, \theta_2)}{d\theta_i}, \\
\frac{d\tilde{\boldsymbol{w}}(\theta_1, \theta_2)}{d\theta_i} &= \tilde{\boldsymbol{w}}(\theta_1, \theta_2) \odot \frac{d\ln\tilde{\boldsymbol{w}}(\theta_1, \theta_2)}{d\theta_i},
\end{aligned}$$

$$\ln\tilde{\boldsymbol{w}}(\theta_1, \theta_2) = (\theta_1 - 1)\ln(\boldsymbol{x}) + (\theta_2 - 1)\ln(\boldsymbol{\iota} - \boldsymbol{x}) + \ln\Gamma(\theta_1 + \theta_2) - \ln\Gamma(\theta_1) - \ln\Gamma(\theta_2).$$

$$\begin{aligned}
T_{\theta_1}\ln\tilde{\boldsymbol{w}}(\delta\theta_1) &= \frac{d}{dt}\Big|_{t=0} \ln\tilde{\boldsymbol{w}}(\theta_1 + t\delta\theta_1, \theta_2) \\
&= \frac{d}{dt}\Big|_{t=0} (\theta_1 + t\delta\theta_1 - 1)\ln(\boldsymbol{x}) + (\theta_2 - 1)\ln(\boldsymbol{\iota} - \boldsymbol{x}) \\
&\qquad + \ln\Gamma(\theta_1 + t\delta\theta_1 + \theta_2) - \ln\Gamma(\theta_1 + t\delta\theta_1) - \ln\Gamma(\theta_2) \\
&\qquad = \ln(\boldsymbol{x})\delta\theta_1 + \psi(\theta_1 + \theta_2)\delta\theta_1 - \psi(\theta_1)\delta\theta_1 \\
&\qquad = (\ln(\boldsymbol{x}) + \psi(\theta_1 + \theta_2) - \psi(\theta_1))\,\delta\theta_1 \\
T_{\theta_1}\mathcal{W}(\delta\theta_1) &= (\tilde{\boldsymbol{w}}(\theta_1, \theta_2) \odot (\ln(\boldsymbol{x}) + \psi(\theta_1 + \theta_2) - \psi(\theta_1)))\delta\theta_1
\end{aligned}$$

$$
\begin{aligned}
T_{\theta_2} \ln \tilde{\boldsymbol{w}}(\delta\theta_2) &= \left.\frac{d}{dt}\right|_{t=0} \ln \tilde{\boldsymbol{w}}(\theta_1, \theta_2 + t\delta\theta_2) \\
&= \left.\frac{d}{dt}\right|_{t=0} \theta_1 \ln(\boldsymbol{x}) + (\theta_2 + t\delta\theta_2 - 1)\ln(\boldsymbol{\iota} - \boldsymbol{x}) \\
&\qquad + \ln\Gamma(\theta_1 + \theta_2 + t\delta\theta_2) - \ln\Gamma(\theta_1) - \ln\Gamma(\theta_2 + t\delta\theta_2) \\
&= \ln(\boldsymbol{\iota} - \boldsymbol{x})\delta\theta_2 + \psi(\theta_1 + \theta_2)\delta\theta_2 - \psi(\theta_2)\delta\theta_2 \\
&= (\ln(\boldsymbol{\iota} - \boldsymbol{x}) + \psi(\theta_1 + \theta_2) - \psi(\theta_2))\,\delta\theta_2 \\
T_{\theta_2}\mathcal{W}(\delta\theta_2) &= (\tilde{\boldsymbol{w}}(\theta_1, \theta_2) \odot (\ln(\boldsymbol{\iota} - \boldsymbol{x}) + \psi(\theta_1 + \theta_2) - \psi(\theta_2)))\delta\theta_2
\end{aligned}
$$

$$
T_{\theta_3}\boldsymbol{w}(\delta\theta_3) = \left.\frac{d}{dt}\right|_{t=0} \boldsymbol{w}(\theta_1, \theta_2, \theta_3 + t\delta\theta_3) = \left.\frac{d}{dt}\right|_{t=0} \tilde{\boldsymbol{w}}(\theta_1, \theta_2) + \boldsymbol{\iota}(\theta_3 + t\delta\theta_3) = \boldsymbol{\iota}\delta\theta_3
$$

$$
\begin{aligned}
\langle T^*_{\theta_1}(\boldsymbol{\delta w}), \delta\theta_1\rangle &= \langle \boldsymbol{\delta w}, T_{\theta_1}\boldsymbol{w}(\delta\theta_1)\rangle = \langle \boldsymbol{\delta w}, (\tilde{\boldsymbol{w}}(\theta_1, \theta_2) \odot (\ln(\boldsymbol{x}) + \psi(\theta_1 + \theta_2) - \psi(\theta_1)))\,\delta\theta_1\rangle \\
&= \langle (\tilde{\boldsymbol{w}}(\theta_1, \theta_2) \odot (\ln(\boldsymbol{x}) + \psi(\theta_1 + \theta_2) - \psi(\theta_1)))^\top \boldsymbol{\delta w}, \delta\theta_1\rangle \\
\boxed{T^*_{\theta_1}\mathcal{W}(\boldsymbol{\delta w}) = (\tilde{\boldsymbol{w}}(\theta_1, \theta_2) \odot (\ln(\boldsymbol{x}) + \psi(\theta_1 + \theta_2) - \psi(\theta_1)))^\top}\boldsymbol{\delta w}.
\end{aligned}
$$

$$
\begin{aligned}
\langle T^*_{\theta_2}(\boldsymbol{\delta w}), \delta\theta_2\rangle &= \langle \boldsymbol{\delta w}, T_{\theta_2}\mathcal{W}(\delta\theta_2)\rangle = \langle \boldsymbol{\delta w}, (\tilde{\boldsymbol{w}}(\theta_1, \theta_2) \odot (\ln(\boldsymbol{\iota} - \boldsymbol{x}) + \psi(\theta_1 + \theta_2) - \psi(\theta_2)))\delta\theta_2\rangle \\
&= \langle (\tilde{\boldsymbol{w}}(\theta_1, \theta_2) \odot (\ln(\boldsymbol{\iota} - \boldsymbol{x}) + \psi(\theta_1 + \theta_2) - \psi(\theta_2)))^\top \boldsymbol{\delta w}, \delta\theta_2\rangle \\
\boxed{T^*_{\theta_2}\mathcal{W}(\boldsymbol{\delta w}) = (\tilde{\boldsymbol{w}}(\theta_1, \theta_2) \odot (\ln(\boldsymbol{\iota} - \boldsymbol{x}) + \psi(\theta_1 + \theta_2) - \psi(\theta_2)))^\top}\boldsymbol{\delta w}
\end{aligned}
$$

$$
\begin{aligned}
\langle T^*_{\theta_3}(\boldsymbol{\delta w}), \delta\theta_3\rangle &= \langle \boldsymbol{\delta w}, T_{\theta_3}\mathcal{W}(\delta\theta_3)\rangle = \langle \boldsymbol{\delta w}, \boldsymbol{\iota}\delta\theta_3\rangle \\
&= \langle \boldsymbol{\iota}^\top \boldsymbol{\delta w}, \delta\theta_3\rangle \\
\boxed{T^*_{\theta_3}\mathcal{W}(\boldsymbol{\delta w}) = \boldsymbol{\iota}^\top}\boldsymbol{\delta w}
\end{aligned}
$$

The analytical Hessian (Beta polynomial)

In the previous section, we have derived the analytical gradient of the objective function.

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial \theta_1} \\ \frac{\partial f}{\partial \theta_2} \\ \frac{\partial f}{\partial \theta_3} \\ \frac{\partial f}{\partial \beta} \\ \frac{\partial f}{\partial \boldsymbol{\rho}} \end{pmatrix} = \begin{pmatrix} [\tilde{\boldsymbol{w}}(\theta_1,\theta_2) \odot (\ln(\boldsymbol{x}) + \psi(\theta_1+\theta_2) - \psi(\theta_1))]^\top \cdot [-\beta Z^\top] \cdot [2\,(\boldsymbol{y} - V\boldsymbol{\delta}\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1,\theta_2,\theta_3))] \\ [\tilde{\boldsymbol{w}}(\theta_1,\theta_2) \odot (\ln(\boldsymbol{\iota} - \boldsymbol{x}) + \psi(\theta_1+\theta_2) - \psi(\theta_2))]^\top \cdot [-\beta Z^\top] \cdot [2\,(\boldsymbol{y} - V\boldsymbol{\delta}\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1,\theta_2,\theta_3))] \\ [\boldsymbol{\iota}]^\top \cdot [-\beta Z^\top] \cdot [2\,(\boldsymbol{y} - V\boldsymbol{\delta}\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1,\theta_2,\theta_3))] \\ [-\boldsymbol{w}^\top(\theta_1,\theta_2,\theta_3) Z^\top] \cdot [2\,(\boldsymbol{y} - V\boldsymbol{\delta}\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1,\theta_2,\theta_3))] \\ [-V^\top] \cdot [2\,(\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1,\theta_2,\theta_3))] \end{pmatrix}$$

Whereas the Hessian matrix is symmetric, we derive the upper triangular entries including elements on the main diagonal.

$$H_f = \nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 f}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 f}{\partial \theta_1 \partial \theta_3} & \frac{\partial^2 f}{\partial \theta_1 \partial \beta} & \left(\frac{\partial^2 f}{\partial \theta_1 \partial \boldsymbol{\rho}}\right)^\top \\ & \frac{\partial^2 f}{\partial \theta_2 \partial \theta_2} & \frac{\partial^2 f}{\partial \theta_2 \partial \theta_3} & \frac{\partial^2 f}{\partial \theta_2 \partial \beta} & \left(\frac{\partial^2 f}{\partial \theta_2 \partial \boldsymbol{\rho}}\right)^\top \\ & & \frac{\partial^2 f}{\partial \theta_3 \partial \theta_3} & \frac{\partial^2 f}{\partial \theta_3 \partial \beta} & \left(\frac{\partial^2 f}{\partial \theta_3 \partial \boldsymbol{\rho}}\right)^\top \\ & & & \frac{\partial^2 f}{\partial \beta \partial \beta} & \left(\frac{\partial^2 f}{\partial \beta \partial \boldsymbol{\rho}}\right)^\top \\ & & & & \frac{\partial^2 f}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}} \end{pmatrix}.$$

We extensively use the following rules of matrix differentiation:

$$\begin{aligned} \mathrm{d}(P \cdot Q) &= \mathrm{d}P \cdot Q + P \cdot \mathrm{d}Q \\ \mathrm{d}(P \odot Q) &= \mathrm{d}P \odot Q + P \odot \mathrm{d}Q \end{aligned}$$

The 1^{st} row of the Hessian matrix

$$\frac{\partial f}{\partial \theta_1} = \overbrace{[\underbrace{\tilde{\boldsymbol{w}}(\theta_1,\theta_2)}_{C_1} \odot \underbrace{(\ln(\boldsymbol{x}) + \psi(\theta_1+\theta_2) - \psi(\theta_1))}_{D_1}]^\top}^{A_1} \cdot \overbrace{\underbrace{[-\beta Z^\top]}_{F_1} \cdot \underbrace{[2\,(\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1,\theta_2,\theta_3))]}_{E_1}}^{B_1}$$

$$\begin{aligned}\mathrm{d}(A_1^\top \cdot B_1) &= \mathrm{d}A_1^\top \cdot B_1 + A_1^\top \cdot \mathrm{d}B_1 = \mathrm{d}(C_1 \odot D_1)^\top \cdot B_1 + A_1^\top \cdot \mathrm{d}B_1 \\ &= (\mathrm{d}C_1 \odot D_1 + C_1 \odot \mathrm{d}D_1)^\top \cdot B_1 + A_1^\top \cdot \mathrm{d}B_1\end{aligned}$$

Element 1.1: $\dfrac{\partial^2 f}{\partial\theta_1\partial\theta_1} = \left(\dfrac{\partial C_1}{\partial\theta_1} \odot D_1 + C_1\dfrac{\partial D_1}{\partial\theta_1}\right)^\top \cdot B_1 + A_1^\top \cdot \dfrac{\partial B_1}{\partial\theta_1}$, where

$$\begin{aligned}\frac{\partial C_1}{\partial\theta_1} &= C_1 \odot D_1, \\ \frac{\partial D_1}{\partial\theta_1} &= \psi_1(\theta_1+\theta_2) - \psi_1(\theta_1), \\ \frac{\partial B_1}{\partial\theta_1} &= F_1 \cdot [-2\beta Z\frac{\partial \boldsymbol{w}(\theta_1,\theta_2,\theta_3)}{\partial\theta_1}] = F_1 \cdot [-2\beta Z\frac{\partial C_1}{\partial\theta_1}].\end{aligned}$$

Element 1.2: $\dfrac{\partial^2 f}{\partial\theta_1\partial\theta_2} = \left(\dfrac{\partial C_1}{\partial\theta_2} \odot D_1 + C_1\dfrac{\partial D_1}{\partial\theta_2}\right)^\top \cdot B_1 + A_1^\top \cdot \dfrac{\partial B_1}{\partial\theta_2}$, where

$$\begin{aligned}\frac{\partial C_1}{\partial\theta_2} &= C_1 \odot (\ln(\boldsymbol{\iota} - \boldsymbol{x}) + \psi(\theta_1+\theta_2) - \psi(\theta_2)), \\ \frac{\partial D_1}{\partial\theta_2} &= \psi_1(\theta_1+\theta_2), \\ \frac{\partial B_1}{\partial\theta_2} &= F_1 \cdot [-2\beta Z\frac{\partial \boldsymbol{w}(\theta_1,\theta_2,\theta_3)}{\partial\theta_2}] = F_1 \cdot [-2\beta Z\frac{\partial C_1}{\partial\theta_2}].\end{aligned}$$

Element 1.3: $\dfrac{\partial^2 f}{\partial\theta_1\partial\theta_3} = \dfrac{\partial(A_1^\top \cdot F_1 \cdot E_1)}{\partial\theta_3} = A_1^\top \cdot F_1 \cdot \dfrac{\partial E_1}{\partial\theta_3} = A_1^\top \cdot F_1 \cdot [-2\beta Z\boldsymbol{\iota}]$

$$A_1^\top \cdot \mathrm{d}(F_1 \cdot E_1) = A_1^\top \cdot (\mathrm{d}F_1 \cdot E_1 + F_1 \cdot \mathrm{d}E_1)$$

Element 1.4: $\dfrac{\partial^2 f}{\partial\theta_1\partial\beta} = A_1^\top \cdot \left(\dfrac{\partial F_1}{\partial\beta} \cdot E_1 + F_1\dfrac{\partial E_1}{\partial\beta}\right)^\top$, where

$$\begin{aligned}\frac{\partial F_1}{\partial\beta} &= -Z^\top, \\ \frac{\partial E}{\partial\beta} &= -2Z\boldsymbol{w}(\theta_1,\theta_2,\theta_3).\end{aligned}$$

Element 1.5: $\dfrac{\partial^2 f}{\partial\theta_1\partial\boldsymbol{\rho}} = \dfrac{\partial(A_1^\top \cdot F_1 \cdot E_1)}{\partial\boldsymbol{\rho}} = A_1^\top \cdot F_1 \cdot \dfrac{\partial E_1}{\partial\boldsymbol{\rho}} = A_1^\top \cdot F_1 \cdot [-2V]$

The 2^{nd} row of the Hessian matrix

$$\frac{\partial f}{\partial \theta_2} = \overbrace{[\underbrace{\tilde{\boldsymbol{w}}(\theta_1,\theta_2)}_{C_2} \odot \underbrace{(\ln(\boldsymbol{\iota}-\boldsymbol{x})+\psi(\theta_1+\theta_2)-\psi(\theta_2))}_{D_2}]^\top}^{A_2} \cdot \overbrace{\underbrace{[-\beta Z^\top]}_{F_2} \cdot \underbrace{[2\,(\boldsymbol{y}-V\boldsymbol{\rho}-\beta Z\boldsymbol{w}(\theta_1,\theta_2,\theta_3))]}_{E_2}}^{B_2}$$

$$\begin{aligned}\mathrm{d}(A_2^\top \cdot B_2) &= \mathrm{d}A_2^\top \cdot B_2 + A_2^\top \cdot \mathrm{d}B_2 = \mathrm{d}(C_2 \odot D_2)^\top \cdot B_2 + A_2^\top \cdot \mathrm{d}B_2 \\ &= (\mathrm{d}C_2 \odot D_2 + C_2 \odot \mathrm{d}D_2)^\top \cdot B_2 + A_2^\top \cdot \mathrm{d}B_2\end{aligned}$$

Element 2.2: $\dfrac{\partial^2 f}{\partial\theta_2\partial\theta_2} = \left(\dfrac{\partial C_2}{\partial\theta_2} \odot D_2 + C_2\dfrac{\partial D_2}{\partial\theta_2}\right)^\top \cdot B_2 + A_2^\top \cdot \dfrac{\partial B_2}{\partial\theta_2}$, where

$$\begin{aligned}\frac{\partial C_2}{\partial\theta_2} &= C_2 \odot D_2, \\ \frac{\partial D_2}{\partial\theta_2} &= \psi_1(\theta_1+\theta_2) - \psi_1(\theta_2), \\ \frac{\partial B_2}{\partial\theta_2} &= \frac{\partial B_1}{\partial\theta_2}.\end{aligned}$$

Element 2.3: $\dfrac{\partial^2 f}{\partial\theta_2\partial\theta_3} = \dfrac{\partial(A_2^\top \cdot F_2 \cdot E_2)}{\partial\theta_3} = A_2^\top \cdot F_2 \cdot \dfrac{\partial E_2}{\partial\theta_3} = A_2^\top \cdot F_2 \cdot \dfrac{\partial E_1}{\partial\theta_3}$

$$A_2^\top \cdot \mathrm{d}(F_2 \cdot E_2) = A_2^\top \cdot (\mathrm{d}F_2 \cdot E_2 + F_2 \cdot \mathrm{d}E_2)$$

Element 2.4: $\dfrac{\partial^2 f}{\partial\theta_2\partial\beta} = A_2^\top \cdot \left(\dfrac{\partial F_2}{\partial\beta} \cdot E_2 + F_2\dfrac{\partial E_2}{\partial\beta}\right)^\top = A_2^\top \cdot \left(\dfrac{\partial F_1}{\partial\beta} \cdot E_2 + F_2\dfrac{\partial E_1}{\partial\beta}\right)^\top$

Element 2.5: $\dfrac{\partial^2 f}{\partial\theta_1\partial\boldsymbol{\rho}} = \dfrac{\partial(A_2^\top \cdot F_2 \cdot E_2)}{\partial\boldsymbol{\rho}} = A_2^\top \cdot F_2 \cdot \dfrac{\partial E_2}{\partial\boldsymbol{\rho}} = A_2^\top \cdot F_2 \cdot \dfrac{\partial E_1}{\partial\boldsymbol{\rho}}$

The 3^{rd} row of the Hessian matrix

$$\frac{\partial f}{\partial \theta_3} = \overbrace{[\boldsymbol{\iota}^\top] \cdot [-\beta Z^\top]}^{A_3} \cdot \overbrace{[2\,(\boldsymbol{y}-V\boldsymbol{\rho}-\beta Z\boldsymbol{w}(\theta_1,\theta_2,\theta_3))]}^{B_3}$$

Element 3.3: $\dfrac{\partial^2 f}{\partial\theta_3\partial\theta_3} = \dfrac{\partial(A_3 \cdot B_3)}{\partial\theta_3} = A_3 \cdot \dfrac{\partial B_3}{\partial\theta_3} = A_3 \cdot \dfrac{\partial E_1}{\partial\theta_3}$

$$\mathrm{d}(A_3 \cdot B_3) = \mathrm{d}A_3 \cdot B_3 + A_3 \cdot \mathrm{d}B_3$$

Element 3.4: $\dfrac{\partial^2 f}{\partial \theta_3 \partial \beta} = \dfrac{\partial A_3}{\partial \beta} \cdot B_3 + A_3 \cdot \dfrac{\partial B_3}{\partial \beta} = \left(-\boldsymbol{\iota}^\top \cdot Z^\top\right) B_3 + A_3 \cdot \dfrac{\partial E_1}{\partial \beta}$

Element 3.5: $\dfrac{\partial^2 f}{\partial \theta_3 \partial \boldsymbol{\rho}} = A_3 \cdot \dfrac{\partial B_3}{\partial \boldsymbol{\rho}} = A_3 \cdot \dfrac{\partial E_1}{\partial \boldsymbol{\rho}}$

The 4^{th} row of the Hessian matrix

$$\frac{\partial f}{\partial \beta} = \overbrace{[-\boldsymbol{w}^\top(\theta_1, \theta_2, \theta_3) Z^\top]}^{A_4} \cdot \overbrace{[2\left(\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1, \theta_2, \theta_3)\right)]}^{B_4}$$

Element 4.4: $\dfrac{\partial^2 f}{\partial \beta \partial \beta} = A_4 \cdot \dfrac{\partial B_4}{\partial \beta} = A_4 \cdot \dfrac{\partial E_1}{\partial \beta}$

Element 4.5: $\dfrac{\partial^2 f}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}} = A_4 \cdot \dfrac{\partial B_4}{\partial \boldsymbol{\rho}} = A_4 \cdot \dfrac{\partial E_1}{\partial \boldsymbol{\rho}}$

The 5^{th} row of the Hessian matrix

$$\frac{\partial f}{\partial \boldsymbol{\rho}} = \overbrace{[-V^\top]}^{A_5} \cdot \overbrace{[2\left(\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z \boldsymbol{w}(\theta_1, \theta_2, \theta_3)\right)]}^{B_5}$$

Element 5.5: $\dfrac{\partial^2 f}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}} = A_5 \cdot \dfrac{\partial B_5}{\partial \boldsymbol{\rho}} = A_5 \cdot \dfrac{\partial E_1}{\partial \boldsymbol{\rho}}$

The analytical gradient (Exponential Almon case)

The above-mentioned objective function (5.19) is presented as a combination of operators:

$$f(\theta_1, \theta_2, \beta, \boldsymbol{\rho}) = \mathcal{G}\left(\mathcal{C}(\boldsymbol{\rho}, \beta, \mathcal{W}(\theta_1, \theta_2))\right).$$

$$
\begin{aligned}
1)\mathcal{W} :&\mathbb{R}\times\mathbb{R}\longrightarrow\mathbb{R}^p\\
&(\theta_1,\theta_2)\longmapsto \boldsymbol{w},\\
2)\,\mathcal{C} :&\mathbb{R}^k\times\mathbb{R}\times\mathbb{R}^p\longrightarrow\mathbb{R}^T\\
&(\boldsymbol{\rho},\beta,\boldsymbol{w})\longmapsto \boldsymbol{c},\\
3)\,\mathcal{G} :&\mathbb{R}^T\longrightarrow\mathbb{R}\\
&\boldsymbol{c}\longmapsto \text{RSS}.
\end{aligned}
$$

Consider the directional derivative of the target function f in the direction of λ. We aim to find an explicit expression for the gradient.

$$
\begin{aligned}
\langle\nabla f,\boldsymbol{\lambda}\rangle &= \left\langle \begin{pmatrix} \nabla_{\theta_1} f\\ \nabla_{\theta_2} f\\ \nabla_{\beta} f\\ \nabla_{\boldsymbol{\rho}} f \end{pmatrix}, \begin{pmatrix} \lambda_1\\ \lambda_2\\ \lambda_3\\ (\lambda_i)_{i\in\{4,\dots,3+k\}} \end{pmatrix} \right\rangle\\
&= \left\langle \begin{pmatrix} \nabla_{\boldsymbol{c}}\mathcal{G}\\ \nabla_{\boldsymbol{c}}\mathcal{G}\\ \nabla_{\boldsymbol{c}}\mathcal{G}\\ \nabla_{\boldsymbol{c}}\mathcal{G} \end{pmatrix}, \begin{pmatrix} (T_{\boldsymbol{w}}\mathcal{C})\cdot(T_{\theta_1}\mathcal{W})\cdot\lambda_1\\ (T_{\boldsymbol{w}}\mathcal{C})\cdot(T_{\theta_2}\mathcal{W})\cdot\lambda_2\\ (T_{\beta}\mathcal{C})\cdot\lambda_3\\ (T_{\boldsymbol{\rho}}\mathcal{C})\cdot(\lambda_i)_{i\in\{4,\dots,3+k\}} \end{pmatrix} \right\rangle\\
&= \left\langle \begin{pmatrix} \left(T^*_{\theta_1}\mathcal{W}\right)\cdot(T^*_{\boldsymbol{w}}\mathcal{C})\cdot\nabla_{\boldsymbol{c}}\mathcal{G}\\ \left(T^*_{\theta_2}\mathcal{W}\right)\cdot(T^*_{\boldsymbol{w}}\mathcal{C})\cdot\nabla_{\boldsymbol{c}}\mathcal{G}\\ \left(T^*_{\beta}\mathcal{C}\right)\cdot\nabla_{\boldsymbol{c}}\mathcal{G}\\ \left(T^*_{\boldsymbol{\rho}}\mathcal{C}\right)\cdot\nabla_{\boldsymbol{c}}\mathcal{G} \end{pmatrix}, \begin{pmatrix} \lambda_1\\ \lambda_2\\ \lambda_3\\ (\lambda_i)_{i\in\{4,\dots,3+k\}} \end{pmatrix} \right\rangle
\end{aligned}
$$

The components of the gradient are separately computed.

Component $\nabla_{\boldsymbol{c}}\mathcal{G}$

$$
\langle\nabla_{\boldsymbol{c}}\mathcal{G},\boldsymbol{\delta c}\rangle = \left.\frac{d}{dt}\right|_{t=0}\mathcal{G}(\boldsymbol{c}+t\boldsymbol{\delta c}) = \left.\frac{d}{dt}\right|_{t=0}(\boldsymbol{c}+t\boldsymbol{\delta c})^\top(\boldsymbol{c}+t\boldsymbol{\delta c}) = 2\boldsymbol{c}^\top\cdot\boldsymbol{\delta c} = \langle\boldsymbol{\delta c},2\boldsymbol{c}\rangle
$$

$$
\boxed{\nabla_{\boldsymbol{c}}\mathcal{G} = 2\boldsymbol{c}}
$$

Calculating components $T^*_{\boldsymbol{\rho}}\mathcal{C}$, $T^*_{\beta}\mathcal{C}$, and $T^*_{\boldsymbol{w}}\mathcal{C}$

$$T_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta\rho}) = \frac{d}{dt}\Big|_{t=0} \mathcal{C}(\boldsymbol{\rho}+t\boldsymbol{\delta\rho},\beta,\boldsymbol{w}) = \frac{d}{dt}\Big|_{t=0} \boldsymbol{y} - V(\boldsymbol{\rho}+t\boldsymbol{\delta\rho}) - \beta Z\boldsymbol{w} = -V\boldsymbol{\delta\rho}$$

$$T_{\beta}\mathcal{C}(\delta\beta) = \frac{d}{dt}\Big|_{t=0} \mathcal{C}(\boldsymbol{\rho},\beta+t\delta\beta,\boldsymbol{w}) = \frac{d}{dt}\Big|_{t=0} \boldsymbol{y} - V\boldsymbol{\rho} - Z\boldsymbol{w}(\beta+t\delta\beta) = -Z\boldsymbol{w}\delta\beta$$

$$T_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta w}) = \frac{d}{dt}\Big|_{t=0} \mathcal{C}(\boldsymbol{\rho},\beta,\boldsymbol{w}+t\boldsymbol{\delta w}) = \frac{d}{dt}\Big|_{t=0} \boldsymbol{y} - V\boldsymbol{\rho} - \beta Z(\boldsymbol{w}+t\boldsymbol{\delta w}) = -\beta Z\boldsymbol{\delta w}$$

$$\langle T^*_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta c}),\boldsymbol{\delta\rho}\rangle = \langle \boldsymbol{\delta c}, T_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta\rho})\rangle = \langle \boldsymbol{\delta c}, -V\boldsymbol{\delta\rho}\rangle = \langle -V^\top\boldsymbol{\delta c},\boldsymbol{\delta\rho}\rangle$$
$$\boxed{T^*_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta c}) = -V^\top}\boldsymbol{\delta c}$$

$$\langle T^*_{\beta}\mathcal{C}(\boldsymbol{\delta c}),\delta\beta\rangle = \langle \boldsymbol{\delta c}, T_{\beta}\mathcal{C}(\delta\beta)\rangle = \langle \boldsymbol{\delta c}, -Z\boldsymbol{w}\delta\beta\rangle = \langle -\boldsymbol{w}^\top Z^\top\boldsymbol{\delta c},\delta\beta\rangle$$
$$\boxed{T^*_{\beta}\mathcal{C}(\boldsymbol{\delta c}) = -\boldsymbol{w}^\top Z^\top}\boldsymbol{\delta c}$$

$$\langle T^*_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta c}),\boldsymbol{\delta w}\rangle = \langle \boldsymbol{\delta c}, T_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta w})\rangle = \langle \boldsymbol{\delta c}, -\beta Z\boldsymbol{\delta w}\rangle = \langle -\beta Z^\top\boldsymbol{\delta c},\boldsymbol{\delta w}\rangle$$
$$\boxed{T^*_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta c}) = -\beta Z^\top}\boldsymbol{\delta c}$$

Calculating components $T^*_{\theta_1}\mathcal{W}$ and $T^*_{\theta_2}\mathcal{W}$

$$\boldsymbol{w}(\theta_1,\theta_2) = \exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)$$

$$\begin{aligned} T_{\theta_1}\mathcal{W}(\delta\theta_1) &= \frac{d}{dt}\Big|_{t=0} \boldsymbol{w}(\theta_1+t\delta\theta_1,\theta_2) = \frac{d}{dt}\Big|_{t=0} \exp(\boldsymbol{\xi}(\theta_1+t\delta\theta_1)+\boldsymbol{\xi}^2\theta_2) \\ &= \Big(\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}\Big)\,\delta\theta_1 \end{aligned}$$

$$\begin{aligned} T_{\theta_2}\mathcal{W}(\delta\theta_2) &= \frac{d}{dt}\Big|_{t=0} \boldsymbol{w}(\theta_1,\theta_2+t\delta\theta_2) = \frac{d}{dt}\Big|_{t=0} \exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^2(\theta_2+t\delta\theta_2)) \\ &= \Big(\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2\Big)\,\delta\theta_2 \end{aligned}$$

$$\begin{aligned}\langle T^*_{\theta_1}(\boldsymbol{\delta w}), \delta\theta_1\rangle &= \langle \boldsymbol{\delta w}, T_{\theta_1}\boldsymbol{w}(\delta\theta_1)\rangle = \langle \boldsymbol{\delta w}, \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}\Big)\,\delta\theta_1\rangle \\ &= \langle \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}\Big)^\top \boldsymbol{\delta w}, \delta\theta_1\rangle \\ \boxed{T^*_{\theta_1}\mathcal{W}(\boldsymbol{\delta w}) &= \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}\Big)^\top}\boldsymbol{\delta w}.\end{aligned}$$

$$\begin{aligned}\langle T^*_{\theta_2}(\boldsymbol{\delta w}), \delta\theta_2\rangle &= \langle \boldsymbol{\delta w}, T_{\theta_2}\boldsymbol{w}(\delta\theta_2)\rangle = \langle \boldsymbol{\delta w}, \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2\Big)\,\delta\theta_2\rangle \\ &= \langle \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2\Big)^\top \boldsymbol{\delta w}, \delta\theta_2\rangle \\ \boxed{T^*_{\theta_2}\mathcal{W}(\boldsymbol{\delta w}) &= \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2\Big)^\top}\boldsymbol{\delta w}.\end{aligned}$$

The analytical Hessian (Exponential Almon case)

In the previous section, we have derived the analytical gradient of the objective function.

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial\theta_1} \\ \frac{\partial f}{\partial\theta_2} \\ \frac{\partial f}{\partial\beta} \\ \frac{\partial f}{\partial\boldsymbol{\rho}} \end{pmatrix} = \begin{pmatrix} [\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}]^\top \cdot [-\beta Z^\top] \cdot [2\,(\boldsymbol{y} - V\boldsymbol{\delta\rho} - \beta Z\boldsymbol{w}(\theta_1,\theta_2))] \\ [\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2]^\top \cdot [-\beta Z^\top] \cdot [2\,(\boldsymbol{y} - V\boldsymbol{\delta\rho} - \beta Z\boldsymbol{w}(\theta_1,\theta_2))] \\ [-\boldsymbol{w}^\top(\theta_1,\theta_2)Z^\top] \cdot [2\,(\boldsymbol{y} - V\boldsymbol{\delta\rho} - \beta Z\boldsymbol{w}(\theta_1,\theta_2))] \\ [-V^\top] \cdot [2\,(\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z\boldsymbol{w}(\theta_1,\theta_2))] \end{pmatrix}$$

Whereas the Hessian matrix is symmetric, we derive the upper triangular entries including elements on the main diagonal.

$$H_f = \nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial\theta_1\partial\theta_1} & \frac{\partial^2 f}{\partial\theta_1\partial\theta_2} & \frac{\partial^2 f}{\partial\theta_1\partial\beta} & \left(\frac{\partial^2 f}{\partial\theta_1\partial\boldsymbol{\rho}}\right)^\top \\ & \frac{\partial^2 f}{\partial\theta_2\partial\theta_2} & \frac{\partial^2 f}{\partial\theta_2\partial\beta} & \left(\frac{\partial^2 f}{\partial\theta_2\partial\boldsymbol{\rho}}\right)^\top \\ & & \frac{\partial^2 f}{\partial\beta\partial\beta} & \left(\frac{\partial^2 f}{\partial\beta\partial\boldsymbol{\rho}}\right)^\top \\ & & & \frac{\partial^2 f}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}} \end{pmatrix}.$$

The 1^{st} row of the Hessian matrix

$$\frac{\partial f}{\partial\theta_1} = \overbrace{[\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}]^\top \cdot [-\beta Z^\top]}^{A_1} \cdot \overbrace{[2\,(\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z\boldsymbol{w}(\theta_1,\theta_2))]}^{B_1}$$

$$\mathrm{d}(A_1 \cdot B_1) = \mathrm{d}A_1 \cdot B_1 + A_1 \cdot \mathrm{d}B_1$$

Element 1.1: $\dfrac{\partial^2 f}{\partial\theta_1\partial\theta_1} = \dfrac{\partial A_1}{\partial\theta_1}\cdot B_1 + A_1 \cdot \dfrac{\partial B_1}{\partial\theta_1}$, where

$$\frac{\partial A_1}{\partial\theta_1} = [\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2]^\top \cdot [-\beta Z^\top],$$
$$\frac{\partial B_1}{\partial\theta_1} = [-2\beta Z\left(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}\right)].$$

Element 1.2: $\dfrac{\partial^2 f}{\partial\theta_1\partial\theta_2} = \dfrac{\partial A_1}{\partial\theta_2}\cdot B_1 + A_1 \cdot \dfrac{\partial B_1}{\partial\theta_2}$, where

$$\frac{\partial A_1}{\partial\theta_2} = [\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^3]^\top \cdot [-\beta Z^\top],$$
$$\frac{\partial B_1}{\partial\theta_2} = [-2\beta Z\left(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2\right)].$$

Element 1.3: $\dfrac{\partial^2 f}{\partial\theta_1\partial\beta} = A_1 \cdot \dfrac{\partial B_1}{\partial\beta} + B_1 \cdot \dfrac{\partial A_1}{\partial\beta}$, where

$$\frac{\partial A_1}{\partial\beta} = [\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}]^\top \cdot [-Z^\top],$$
$$\frac{\partial B_1}{\partial\beta} = -2Z\boldsymbol{w}(\theta_1,\theta_2).$$

Element 1.4: $\dfrac{\partial^2 f}{\partial\theta_1\partial\boldsymbol{\rho}} = A_1\dfrac{\partial B_1}{\partial\boldsymbol{\rho}} = A_1 \cdot [-2V]$

The 2^{nd} row of the Hessian matrix

$$\frac{\partial f}{\partial\theta_2} = \overbrace{[\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2]^\top \cdot [-\beta Z^\top]}^{A_2}\cdot\overbrace{[2\left(\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z\boldsymbol{w}(\theta_1,\theta_2)\right)]}^{B_2}$$

$$\mathrm{d}(A_2 \cdot B_2) = \mathrm{d}A_2 \cdot B_2 + A_2 \cdot \mathrm{d}B_2$$

Element 2.2: $\dfrac{\partial^2 f}{\partial\theta_2\partial\theta_2} = \dfrac{\partial A_2}{\partial\theta_2}\cdot B_2 + A_2 \cdot \dfrac{\partial B_2}{\partial\theta_2}$, where

$$\frac{\partial A_2}{\partial\theta_2} = [\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^4]^\top \cdot [-\beta Z^\top],$$
$$\frac{\partial B_2}{\partial\theta_2} = [-2\beta Z\left(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2\right)].$$

Element 2.3: $\dfrac{\partial^2 f}{\partial\theta_2\partial\beta} = A_2 \cdot \dfrac{\partial B_2}{\partial\beta} + B_2 \cdot \dfrac{\partial A_2}{\partial\beta}$, where

$$\frac{\partial A_1}{\partial\beta} = [\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}^2]^\top \cdot [-Z^\top],$$
$$\frac{\partial B_1}{\partial\beta} = -2Z\boldsymbol{w}(\theta_1, \theta_2).$$

Element 2.4: $\dfrac{\partial^2 f}{\partial\theta_2\partial\boldsymbol{\rho}} = A_2\dfrac{\partial B_2}{\partial\boldsymbol{\rho}} = A_2 \cdot [-2V]$

The 3^{rd} row of the Hessian matrix

$$\frac{\partial f}{\partial\beta} = \overbrace{[-\boldsymbol{w}^\top(\theta_1,\theta_2)Z^\top]}^{A_3} \cdot \overbrace{[2\,(\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z\boldsymbol{w}(\theta_1,\theta_2))]}^{B_3}$$

Element 3.3: $\dfrac{\partial^2 f}{\partial\beta\partial\beta} = A_3 \cdot \dfrac{\partial B_3}{\partial\beta} = A_3 \cdot \dfrac{\partial B_1}{\partial\beta}$

Element 3.4: $\dfrac{\partial^2 f}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}} = A_3 \cdot \dfrac{\partial B_3}{\partial\boldsymbol{\rho}} = A_3 \cdot \dfrac{\partial B_1}{\partial\boldsymbol{\rho}}$

The 4^{th} row of the Hessian matrix

$$\frac{\partial f}{\partial\boldsymbol{\rho}} = \overbrace{[-V^\top]}^{A_4} \cdot \overbrace{[2\,(\boldsymbol{y} - V\boldsymbol{\rho} - \beta Z\boldsymbol{w}(\theta_1,\theta_2,\theta_3))]}^{B_4}$$

Element 4.4: $\dfrac{\partial^2 f}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}} = A_4 \cdot \dfrac{\partial B_4}{\partial\boldsymbol{\rho}} = A_4 \cdot \dfrac{\partial B_1}{\partial\boldsymbol{\rho}}$

Appendix 5.B Analytic derivatives for the multivariate MIDAS model

This section explains the derivation of the analytical gradient and of the analytical Hessian matrix of the objective function in the optimization problem related to estimating the coefficients of a multivariate MIDAS regression (5.2) via non-linear least squares. The MIDAS-NLS estimator requires minimizing the target function, RSS (5.20):

$$f(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \beta, \boldsymbol{\rho}) = (\boldsymbol{y} - V\boldsymbol{\rho} - Z(B\beta \odot \boldsymbol{w}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})))^\top \cdot (\boldsymbol{y} - V\boldsymbol{\rho} - Z(B\beta \odot \boldsymbol{w}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}))). \quad (5.20)$$

Analytic derivatives for the Exponential Almon case are based on the results in Chapters 1 and 3.

The analytical gradient (Exponential Almon case)

The objective function (5.20) can be represented as a combination of operators as follows:

$$\begin{aligned}
1)\mathcal{W} :& \mathbb{R}^{\text{numZ}} \times \mathbb{R}^{\text{numZ}} \longrightarrow \mathbb{R}^{\Sigma \boldsymbol{p_Z}} \\
& (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \longmapsto \boldsymbol{w}, \\
2)\ \mathcal{C} :& \mathbb{R}^k \times \mathbb{R}^{\text{numZ}} \times \mathbb{R}^{\Sigma \boldsymbol{p_Z}} \longrightarrow \mathbb{R}^T \\
& (\boldsymbol{\rho}, \boldsymbol{\beta}, \boldsymbol{w}) \longmapsto \boldsymbol{c}, \\
3)\ \mathcal{G} :& \mathbb{R}^T \longrightarrow \mathbb{R} \\
& \boldsymbol{c} \longmapsto \text{RSS}.
\end{aligned}$$

Note that the vector $\boldsymbol{p_Z}$ specifies the number of MIDAS lags for each MIDAS-weighed variable and numZ is the number of MIDAS-weighed variables (refer to Section 5.2 for details). Consider the directional derivative of the target function RSS in the direction of λ. We reorganize terms to find an explicit expression for the gradient.

$$\langle \nabla \text{RSS}, \boldsymbol{\lambda} \rangle = \left\langle \begin{pmatrix} \nabla_{\boldsymbol{\theta_1}} \text{RSS} \\ \nabla_{\boldsymbol{\theta_2}} \text{RSS} \\ \nabla_{\beta} \text{RSS} \\ \nabla_{\boldsymbol{\rho}} \text{RSS} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\lambda_1} \\ \boldsymbol{\lambda_2} \\ \boldsymbol{\lambda_3} \\ \boldsymbol{\lambda_4} \end{pmatrix} \right\rangle$$

$$= \left\langle \begin{pmatrix} \nabla_{\boldsymbol{c}}\mathcal{G} \\ \nabla_{\boldsymbol{c}}\mathcal{G} \\ \nabla_{\boldsymbol{c}}\mathcal{G} \\ \nabla_{\boldsymbol{c}}\mathcal{G} \end{pmatrix}, \begin{pmatrix} (T_{\boldsymbol{w}}\mathcal{C}) \cdot (T_{\boldsymbol{\theta_1}}\mathcal{W}) \cdot \boldsymbol{\lambda_1} \\ (T_{\boldsymbol{w}}\mathcal{C}) \cdot (T_{\boldsymbol{\theta_2}}\mathcal{W}) \cdot \boldsymbol{\lambda_2} \\ (T_{\boldsymbol{\beta}}\mathcal{C}) \cdot \boldsymbol{\lambda_3} \\ (T_{\boldsymbol{\rho}}\mathcal{C}) \cdot \boldsymbol{\lambda_4} \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} \left(T^*_{\boldsymbol{\theta_1}}\mathcal{W}\right) \cdot (T^*_{\boldsymbol{w}}\mathcal{C}) \cdot \nabla_{\boldsymbol{c}}\mathcal{G} \\ \left(T^*_{\boldsymbol{\theta_2}}\mathcal{W}\right) \cdot (T^*_{\boldsymbol{w}}\mathcal{C}) \cdot \nabla_{\boldsymbol{c}}\mathcal{G} \\ \left(T^*_{\boldsymbol{\beta}}\mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}}\mathcal{G} \\ \left(T^*_{\boldsymbol{\rho}}\mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}}\mathcal{G} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\lambda_1} \\ \boldsymbol{\lambda_2} \\ \boldsymbol{\lambda_3} \\ \boldsymbol{\lambda_4} \end{pmatrix} \right\rangle.$$

Finally, the components of the gradient are separately computed.

Calculating Component $\nabla_{\boldsymbol{c}}\mathcal{G}$

$$\langle \nabla_{\boldsymbol{c}}\mathcal{G}, \boldsymbol{\delta c}\rangle = \frac{d}{dt}\Big|_{t=0} \mathcal{G}(\boldsymbol{c}+t\boldsymbol{\delta c}) = \frac{d}{dt}\Big|_{t=0} (\boldsymbol{c}+t\boldsymbol{\delta c})^\top(\boldsymbol{c}+t\boldsymbol{\delta c}) = 2\boldsymbol{c}^\top \cdot \boldsymbol{\delta c} = \langle \boldsymbol{\delta c}, 2\boldsymbol{c}\rangle$$
$$\boxed{\nabla_{\boldsymbol{c}}\mathcal{G} = 2\boldsymbol{c}}$$

Calculating Components $T^*_{\boldsymbol{\rho}}\mathcal{C}$, $T^*_{\boldsymbol{\beta}}\mathcal{C}$ and $T^*_{\boldsymbol{w}}\mathcal{C}$

$$T_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta\rho}) \;=\; \frac{d}{dt}\Big|_{t=0} \mathcal{C}(\boldsymbol{\rho}+t\boldsymbol{\delta\rho}, \boldsymbol{\beta}, \boldsymbol{w}) \;=\; \frac{d}{dt}\Big|_{t=0} \boldsymbol{y} - V(\boldsymbol{\rho}+t\boldsymbol{\delta\rho}) - Z(B\boldsymbol{\beta} \odot \boldsymbol{w}) \;=\; -V\boldsymbol{\delta\rho}.$$

$$T_{\boldsymbol{\beta}}\mathcal{C}(\boldsymbol{\delta\beta}) = \frac{d}{dt}\Big|_{t=0} \mathcal{C}(\boldsymbol{\rho}, \boldsymbol{\beta}+t\boldsymbol{\delta\beta}, \boldsymbol{w}) = \frac{d}{dt}\Big|_{t=0} \boldsymbol{y} - V\boldsymbol{\rho} - Z(B(\boldsymbol{\beta}+t\boldsymbol{\delta\beta}) \odot \boldsymbol{w}) = -Z(B\boldsymbol{\delta\beta} \odot \boldsymbol{w})$$
$$= -Z\,\mathrm{diag}(\boldsymbol{w})B\boldsymbol{\delta\beta}.$$

$$T_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta w}) = \frac{d}{dt}\Big|_{t=0} \mathcal{C}(\boldsymbol{\rho}, \boldsymbol{\beta}, \boldsymbol{w}+t\boldsymbol{\delta w}) = \frac{d}{dt}\Big|_{t=0} \boldsymbol{y} - V\boldsymbol{\rho} - Z(B\boldsymbol{\beta} \odot (\boldsymbol{w}+t\boldsymbol{\delta w})) = -Z(B\boldsymbol{\beta} \odot \boldsymbol{\delta w})$$
$$= -Z\,\mathrm{diag}(B\boldsymbol{\beta})\boldsymbol{\delta w}.$$

$$\langle T^*_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta c}), \boldsymbol{\delta\rho}\rangle = \langle \boldsymbol{\delta c}, T_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta\rho})\rangle = \langle \boldsymbol{\delta c}, -V\boldsymbol{\delta\rho}\rangle = \langle -V^\top\boldsymbol{\delta c}, \boldsymbol{\delta\rho}\rangle$$
$$\boxed{T^*_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta c}), \boldsymbol{\delta\rho} = -V^\top}\boldsymbol{\delta c}$$

$$\langle T^*_{\boldsymbol{\beta}}\mathcal{C}(\boldsymbol{\delta c}), \boldsymbol{\delta\beta}\rangle = \langle \boldsymbol{\delta c}, T_{\boldsymbol{\beta}}\mathcal{C}(\boldsymbol{\delta\beta})\rangle = \langle \boldsymbol{\delta c}, -Z\,\mathrm{diag}(\boldsymbol{w})B\boldsymbol{\delta\beta}\rangle = \langle -B^\top\,\mathrm{diag}(\boldsymbol{w})Z^\top\boldsymbol{\delta c}, \boldsymbol{\delta\beta}\rangle$$
$$\boxed{T^*_{\boldsymbol{\beta}}\mathcal{C}(\boldsymbol{\delta c}) = -B^\top\,\mathrm{diag}(\boldsymbol{w})Z^\top}\boldsymbol{\delta c}$$

$$\langle T_{\boldsymbol{w}}^{*}\mathcal{C}(\boldsymbol{\delta c}),\boldsymbol{\delta w}\rangle = \langle \boldsymbol{\delta c}, T_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta w})\rangle = \langle \boldsymbol{\delta c}, -Z\,\mathrm{diag}(B\boldsymbol{\beta})\boldsymbol{\delta w}\rangle = \langle -\mathrm{diag}(B\boldsymbol{\beta})Z^{\top}\boldsymbol{\delta c}, \boldsymbol{\delta w}\rangle$$
$$\boxed{T_{\boldsymbol{w}}^{*}\mathcal{C}(\boldsymbol{\delta z}) = -\mathrm{diag}(B\boldsymbol{\beta})Z^{\top}}\boldsymbol{\delta c}$$

Calculating Components $T_{\boldsymbol{\theta_1}}^{*}\mathcal{W}, T_{\boldsymbol{\theta_2}}^{*}\mathcal{W}$

$$\boldsymbol{w}(\boldsymbol{\theta_1},\boldsymbol{\theta_2}) = \exp(B\boldsymbol{\theta_1}\odot\tilde{\boldsymbol{\xi}} + B\boldsymbol{\theta_2}\odot\tilde{\boldsymbol{\xi}}^{2}).$$

$$T_{\boldsymbol{\theta_1}}\mathcal{W}(\boldsymbol{\delta\theta_1}) = \frac{d}{dt}\Big|_{t=0}\mathcal{W}(\boldsymbol{\theta_1}+t\boldsymbol{\delta\theta_1},\boldsymbol{\theta_2}) = \frac{d}{dt}\Big|_{t=0}\exp(B(\boldsymbol{\theta_1}+t\boldsymbol{\delta\theta_1})\odot\tilde{\boldsymbol{\xi}}+B\boldsymbol{\theta_2}\odot\tilde{\boldsymbol{\xi}}^{2}) = \left(\boldsymbol{w}\odot\tilde{\boldsymbol{\xi}}\right)B\boldsymbol{\delta\theta_1}$$

$$T_{\boldsymbol{\theta_2}}\mathcal{W}(\boldsymbol{\delta\theta_2}) = \frac{d}{dt}\Big|_{t=0}\mathcal{W}(\boldsymbol{\theta_1},\boldsymbol{\theta_2}+t\boldsymbol{\delta\theta_2}) = \frac{d}{dt}\Big|_{t=0}\exp(B\boldsymbol{\theta_1}\odot\tilde{\boldsymbol{\xi}}+B(\boldsymbol{\theta_2}+t\boldsymbol{\delta\theta_2})\odot\tilde{\boldsymbol{\xi}}^{2}) = \left(\boldsymbol{w}\odot\tilde{\boldsymbol{\xi}}^{\boldsymbol{2}}\right)B\boldsymbol{\delta\theta_2}$$

$$\langle T_{\boldsymbol{\theta_1}}^{*}\mathcal{W}(\boldsymbol{\delta w}),\boldsymbol{\delta\theta_1}\rangle = \langle \boldsymbol{\delta w}, T_{\boldsymbol{\theta_1}}\mathcal{W}(\boldsymbol{\delta\theta_1})\rangle = \langle \boldsymbol{\delta w}, \left(\boldsymbol{w}\odot\tilde{\boldsymbol{\xi}}\right)B\boldsymbol{\delta\theta_1}\rangle =$$
$$\langle B^{\top}\mathrm{diag}(\boldsymbol{w}\odot\tilde{\boldsymbol{\xi}})\boldsymbol{\delta\theta_1},\boldsymbol{\delta\theta_1}\rangle$$
$$\boxed{T_{\boldsymbol{\theta_1}}^{*}\mathcal{W}(\boldsymbol{\delta w}) = B^{\top}\mathrm{diag}(\boldsymbol{w}\odot\tilde{\boldsymbol{\xi}})}\boldsymbol{\delta w}.$$

$$\langle T_{\boldsymbol{\theta_2}}^{*}\mathcal{W}(\boldsymbol{\delta w}),\boldsymbol{\delta\theta_2}\rangle = \langle \boldsymbol{\delta w}, T_{\boldsymbol{\theta_2}}\mathcal{W}(\boldsymbol{\delta\theta_2})\rangle = \langle \boldsymbol{\delta w}, \left(\boldsymbol{w}\odot\tilde{\boldsymbol{\xi}}^{2}\right)B\boldsymbol{\delta\theta_2}\rangle =$$
$$\langle B^{\top}\mathrm{diag}(\boldsymbol{w}\odot\tilde{\boldsymbol{\xi}}^{2})\boldsymbol{\delta\theta_2},\boldsymbol{\delta\theta_2}\rangle$$
$$\boxed{T_{\boldsymbol{\theta_2}}^{*}\mathcal{W}(\boldsymbol{\delta w}) = B^{\top}\mathrm{diag}(\boldsymbol{w}\odot\tilde{\boldsymbol{\xi}}^{2})}\boldsymbol{\delta w}.$$

Appendix 5.C Analytic derivatives for the MIDAS logit model

This section explains the derivation of the analytical gradient and of the analytical Hessian matrix of the objective function in the optimization problem related to estimating the coefficients of a MIDAS logit regression via non-linear least squares. The MIDAS-NLS estimator requires minimizing the target function, (minus) log-likelihood:

$$\mathrm{LL}(\boldsymbol{\rho},\beta,\theta_1,\theta_2) = -\left(\boldsymbol{y}^{T}(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2)) - \boldsymbol{\iota}^{T}\log(\boldsymbol{\iota}+\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2)))\right)$$

Analytic derivatives for the Exponential Almon case should be considered as simplified and extended results in Audrino, Kostrov and Ortega (2019).

The analytical gradient (Exponential Almon case)

The above-mentioned objective function is presented as a combination of operators:

$$f(\theta_1, \theta_2, \beta, \boldsymbol{\rho}) = \mathcal{G}\left(\mathcal{C}(\boldsymbol{\rho}, \beta, \mathcal{W}(\theta_1, \theta_2))\right).$$

$$\begin{aligned}
1)\mathcal{W} &: \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}^p \\
&\quad (\theta_1, \theta_2) \longmapsto \boldsymbol{w}, \\
2)\,\mathcal{C} &: \mathbb{R}^k \times \mathbb{R} \times \mathbb{R}^p \longrightarrow \mathbb{R}^T \\
&\quad (\boldsymbol{\rho}, \beta, \boldsymbol{w}) \longmapsto \boldsymbol{c}, \\
3)\,\mathcal{G} &: \mathbb{R}^T \longrightarrow \mathbb{R} \\
&\quad \boldsymbol{c} \longmapsto \text{LL}.
\end{aligned}$$

Consider the directional derivative of the target function f in the direction of λ. We aim to find an explicit expression for the gradient.

$$\begin{aligned}
\langle \nabla f, \boldsymbol{\lambda} \rangle &= \left\langle \begin{pmatrix} \nabla_{\theta_1} f \\ \nabla_{\theta_2} f \\ \nabla_{\beta} f \\ \nabla_{\boldsymbol{\rho}} f \end{pmatrix}, \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ (\lambda_i)_{i \in \{4,\dots,3+k\}} \end{pmatrix} \right\rangle \\
&= \left\langle \begin{pmatrix} \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \end{pmatrix}, \begin{pmatrix} (T_{\boldsymbol{w}} \mathcal{C}) \cdot (T_{\theta_1} \mathcal{W}) \cdot \lambda_1 \\ (T_{\boldsymbol{w}} \mathcal{C}) \cdot (T_{\theta_2} \mathcal{W}) \cdot \lambda_2 \\ (T_{\beta} \mathcal{C}) \cdot \lambda_3 \\ (T_{\boldsymbol{\rho}} \mathcal{C}) \cdot (\lambda_i)_{i \in \{4,\dots,3+k\}} \end{pmatrix} \right\rangle \\
&= \left\langle \begin{pmatrix} \left(T^*_{\theta_1} \mathcal{W}\right) \cdot (T^*_{\boldsymbol{w}} \mathcal{C}) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\theta_2} \mathcal{W}\right) \cdot (T^*_{\boldsymbol{w}} \mathcal{C}) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\beta} \mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\boldsymbol{\rho}} \mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \end{pmatrix}, \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ (\lambda_i)_{i \in \{4,\dots,3+k\}} \end{pmatrix} \right\rangle
\end{aligned}$$

The components of the gradient are separately computed.

Component $\nabla_{\boldsymbol{c}}\mathcal{G}$

$$\begin{aligned}\langle\nabla_{\boldsymbol{c}}\mathcal{G},\boldsymbol{\delta c}\rangle &= \frac{d}{dt}\Big|_{t=0}\mathcal{G}(\boldsymbol{c}+t\boldsymbol{\delta c}) = \frac{d}{dt}\Big|_{t=0} -\boldsymbol{y}^\top(\boldsymbol{c}+t\boldsymbol{\delta c})+\boldsymbol{\iota}^\top\log(\boldsymbol{\iota}+\exp(\boldsymbol{c}+t\boldsymbol{\delta c}))\\ &= -\boldsymbol{y}^\top\boldsymbol{\delta c}+\left(\frac{\exp(\boldsymbol{c})}{\boldsymbol{\iota}+\exp(\boldsymbol{c})}\right)^\top\boldsymbol{\delta c} = \left(-\boldsymbol{y}^\top+\left(\frac{\exp(\boldsymbol{c})}{\boldsymbol{\iota}+\exp(\boldsymbol{c})}\right)^\top\right)\boldsymbol{\delta c}\\ &= \langle\boldsymbol{\delta c},-\boldsymbol{y}+\frac{\exp(\boldsymbol{c})}{\boldsymbol{\iota}+\exp(\boldsymbol{c})}\rangle\end{aligned}$$

$$\boxed{\nabla_{\boldsymbol{c}}\mathcal{G} = -\boldsymbol{y}+\frac{\exp(\boldsymbol{c})}{\boldsymbol{\iota}+\exp(\boldsymbol{c})}}$$

Calculating components $T^*_{\boldsymbol{\rho}}\mathcal{C}$, $T^*_{\beta}\mathcal{C}$, and $T^*_{\boldsymbol{w}}\mathcal{C}$

$$T_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta\rho}) = \frac{d}{dt}\Big|_{t=0}\mathcal{C}(\boldsymbol{\rho}+t\boldsymbol{\delta\rho},\beta,\boldsymbol{w}) = \frac{d}{dt}\Big|_{t=0}V(\boldsymbol{\rho}+t\boldsymbol{\delta\rho})+\beta Z\boldsymbol{w} = V\boldsymbol{\delta\rho}$$

$$T_{\beta}\mathcal{C}(\delta\beta) = \frac{d}{dt}\Big|_{t=0}\mathcal{C}(\boldsymbol{\rho},\beta+t\delta\beta,\boldsymbol{w}) = \frac{d}{dt}\Big|_{t=0}V\boldsymbol{\rho}+Z\boldsymbol{w}(\beta+t\delta\beta) = Z\boldsymbol{w}\delta\beta$$

$$T_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta w}) = \frac{d}{dt}\Big|_{t=0}\mathcal{C}(\boldsymbol{\rho},\beta,\boldsymbol{w}+t\boldsymbol{\delta w}) = \frac{d}{dt}\Big|_{t=0}V\boldsymbol{\rho}+\beta Z(\boldsymbol{w}+t\boldsymbol{\delta w}) = \beta Z\boldsymbol{\delta w}$$

$$\langle T^*_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta c}),\boldsymbol{\delta\rho}\rangle = \langle\boldsymbol{\delta c},T_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta\rho})\rangle = \langle\boldsymbol{\delta c},V\boldsymbol{\delta\rho}\rangle = \langle V^\top\boldsymbol{\delta c},\boldsymbol{\delta\rho}\rangle$$

$$\boxed{T^*_{\boldsymbol{\rho}}\mathcal{C}(\boldsymbol{\delta c}) = V^\top}\boldsymbol{\delta c}$$

$$\langle T^*_{\beta}\mathcal{C}(\boldsymbol{\delta c}),\delta\beta\rangle = \langle\boldsymbol{\delta c},T_{\beta}\mathcal{C}(\delta\beta)\rangle = \langle\boldsymbol{\delta c},Z\boldsymbol{w}\delta\beta\rangle = \langle\boldsymbol{w}^\top Z^\top\boldsymbol{\delta c},\delta\beta\rangle$$

$$\boxed{T^*_{\beta}\mathcal{C}(\boldsymbol{\delta c}) = \boldsymbol{w}^\top Z^\top}\boldsymbol{\delta c}$$

$$\langle T^*_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta c}),\boldsymbol{\delta w}\rangle = \langle\boldsymbol{\delta c},T_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta w})\rangle = \langle\boldsymbol{\delta c},\beta Z\boldsymbol{\delta w}\rangle = \langle\beta Z^\top\boldsymbol{\delta c},\boldsymbol{\delta w}\rangle$$

$$\boxed{T^*_{\boldsymbol{w}}\mathcal{C}(\boldsymbol{\delta c}) = \beta Z^\top}\boldsymbol{\delta c}$$

Calculating components $T^*_{\theta_1}\mathcal{W}$ and $T^*_{\theta_2}\mathcal{W}$

$$\boldsymbol{w}(\theta_1,\theta_2) = \exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^2\theta_2)$$

$$
\begin{aligned}
T_{\theta_1}\mathcal{W}(\delta\theta_1) &= \frac{d}{dt}\Big|_{t=0} \boldsymbol{w}(\theta_1 + t\delta\theta_1, \theta_2) = \frac{d}{dt}\Big|_{t=0} \exp(\boldsymbol{\xi}(\theta_1 + t\delta\theta_1) + \boldsymbol{\xi}^2\theta_2) \\
&= \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}\Big)\, \delta\theta_1
\end{aligned}
$$

$$
\begin{aligned}
T_{\theta_2}\mathcal{W}(\delta\theta_2) &= \frac{d}{dt}\Big|_{t=0} \boldsymbol{w}(\theta_1, \theta_2 + t\delta\theta_2) = \frac{d}{dt}\Big|_{t=0} \exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2(\theta_2 + t\delta\theta_2)) \\
&= \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}^2\Big)\, \delta\theta_2
\end{aligned}
$$

$$
\begin{aligned}
\langle T^*_{\theta_1}(\boldsymbol{\delta w}), \delta\theta_1\rangle &= \langle \boldsymbol{\delta w}, T_{\theta_1}\boldsymbol{w}(\delta\theta_1)\rangle = \langle \boldsymbol{\delta w}, \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}\Big)\, \delta\theta_1\rangle \\
&= \langle \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}\Big)^\top \boldsymbol{\delta w}, \delta\theta_1\rangle \\
\boxed{T^*_{\theta_1}\mathcal{W}(\boldsymbol{\delta w}) &= \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}\Big)^\top} \boldsymbol{\delta w}.
\end{aligned}
$$

$$
\begin{aligned}
\langle T^*_{\theta_2}(\boldsymbol{\delta w}), \delta\theta_2\rangle &= \langle \boldsymbol{\delta w}, T_{\theta_2}\boldsymbol{w}(\delta\theta_2)\rangle = \langle \boldsymbol{\delta w}, \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}^2\Big)\, \delta\theta_2\rangle \\
&= \langle \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}^2\Big)^\top \boldsymbol{\delta w}, \delta\theta_2\rangle \\
\boxed{T^*_{\theta_2}\mathcal{W}(\boldsymbol{\delta w}) &= \Big(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}^2\Big)^\top} \boldsymbol{\delta w}.
\end{aligned}
$$

The analytical Hessian (Exponential Almon case)

We have previously derived the analytical gradient of the objective function.

$$
\nabla f = \begin{pmatrix} \frac{\partial f}{\partial \theta_1} \\ \frac{\partial f}{\partial \theta_2} \\ \frac{\partial f}{\partial \beta} \\ \frac{\partial f}{\partial \boldsymbol{\rho}} \end{pmatrix} = \begin{pmatrix} [\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}]^\top \cdot [\beta Z^\top] \cdot [-\boldsymbol{y} + \frac{\exp(V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2))}{\boldsymbol{\iota} + \exp(V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2))}] \\ [\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2) \odot \boldsymbol{\xi}^2]^\top \cdot [\beta Z^\top] \cdot [-\boldsymbol{y} + \frac{\exp(V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2))}{\boldsymbol{\iota} + \exp(V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2))}] \\ [\boldsymbol{w}^\top(\theta_1, \theta_2) Z^\top] \cdot [-\boldsymbol{y} + \frac{\exp(V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2))}{\boldsymbol{\iota} + \exp(V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2))}] \\ [V^\top] \cdot [-\boldsymbol{y} + \frac{\exp(V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2))}{\boldsymbol{\iota} + \exp(V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2))}] \end{pmatrix}
$$

Whereas the Hessian matrix is symmetric, we derive the upper triangular entries including elements on the main diagonal.

$$H_f = \nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial\theta_1\partial\theta_1} & \frac{\partial^2 f}{\partial\theta_1\partial\theta_2} & \frac{\partial^2 f}{\partial\theta_1\partial\beta} & \left(\frac{\partial^2 f}{\partial\theta_1\partial\boldsymbol{\rho}}\right)^\top \\ & \frac{\partial^2 f}{\partial\theta_2\partial\theta_2} & \frac{\partial^2 f}{\partial\theta_2\partial\beta} & \left(\frac{\partial^2 f}{\partial\theta_2\partial\boldsymbol{\rho}}\right)^\top \\ & & \frac{\partial^2 f}{\partial\beta\partial\beta} & \left(\frac{\partial^2 f}{\partial\beta\partial\boldsymbol{\rho}}\right)^\top \\ & & & \frac{\partial^2 f}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}} \end{pmatrix}.$$

The 1^{st} row of the Hessian matrix

$$\frac{\partial f}{\partial\theta_1} = \overbrace{[\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}]^\top\cdot[\beta Z^\top]}^{A_1}\cdot\overbrace{[-\boldsymbol{y} + \frac{\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2))}{\boldsymbol{\iota}+\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2))}]}^{B_1}$$

$$\mathrm{d}(A_1\cdot B_1) = \mathrm{d}A_1\cdot B_1 + A_1\cdot\mathrm{d}B_1$$

Element 1.1: $\frac{\partial^2 f}{\partial\theta_1\partial\theta_1} = \frac{\partial A_1}{\partial\theta_1}\cdot B_1 + A_1\cdot\frac{\partial B_1}{\partial\theta_1}$, where

$$\frac{\partial A_1}{\partial\theta_1} = [\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2]^\top\cdot[\beta Z^\top],$$
$$\frac{\partial B_1}{\partial\theta_1} = [\frac{\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2))}{(\boldsymbol{\iota}+\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2)))^2}\cdot\beta Z\cdot\frac{\partial\boldsymbol{w}(\theta_1,\theta_2)}{\partial\theta_1}],$$ refer to $T_{\theta_1}\mathcal{W}(\delta\theta_1)$.

Element 1.2: $\frac{\partial^2 f}{\partial\theta_1\partial\theta_2} = \frac{\partial A_1}{\partial\theta_2}\cdot B_1 + A_1\cdot\frac{\partial B_1}{\partial\theta_2}$, where

$$\frac{\partial A_1}{\partial\theta_2} = [\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^3]^\top\cdot[\beta Z^\top],$$
$$\frac{\partial B_1}{\partial\theta_2} = [\frac{\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2))}{(\boldsymbol{\iota}+\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2)))^2}\cdot\beta Z\cdot\frac{\partial\boldsymbol{w}(\theta_1,\theta_2)}{\partial\theta_2}],$$ refer to $T_{\theta_2}\mathcal{W}(\delta\theta_2)$.

Element 1.3: $\frac{\partial^2 f}{\partial\theta_1\partial\beta} = A_1\cdot\frac{\partial B_1}{\partial\beta} + B_1\cdot\frac{\partial A_1}{\partial\beta}$, where

$$\frac{\partial A_1}{\partial\beta} = [\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}]^\top\cdot[Z^\top],$$
$$\frac{\partial B_1}{\partial\beta} = [\frac{\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2))}{(\boldsymbol{\iota}+\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2)))^2}\cdot Z\boldsymbol{w}].$$

Element 1.4: $\dfrac{\partial^2 f}{\partial\theta_1\partial\boldsymbol{\rho}} = A_1\dfrac{\partial B_1}{\partial\boldsymbol{\rho}} = A_1\cdot\operatorname{diag}\left(\dfrac{\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2))}{(\boldsymbol{\iota}+\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2)))^2}\right)\cdot V$

The 2^{nd} row of the Hessian matrix

$$\frac{\partial f}{\partial\theta_2} = \overbrace{[\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2]^\top\cdot[\beta Z^\top]}^{A_2}\cdot\overbrace{[-\boldsymbol{y}+\frac{\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2))}{\boldsymbol{\iota}+\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2))}]}^{B_2}$$

$$\mathrm{d}(A_2\cdot B_2) = \mathrm{d}A_2\cdot B_2 + A_2\cdot\mathrm{d}B_2$$

Element 2.2: $\dfrac{\partial^2 f}{\partial\theta_2\partial\theta_2} = \dfrac{\partial A_2}{\partial\theta_2}\cdot B_2 + A_2\cdot\dfrac{\partial B_2}{\partial\theta_2}$, where

$$\frac{\partial A_2}{\partial\theta_2} = [\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^4]^\top\cdot[\beta Z^\top],$$
$$\frac{\partial B_2}{\partial\theta_2} = \frac{\partial B_1}{\partial\theta_2}.$$

Element 2.3: $\dfrac{\partial^2 f}{\partial\theta_2\partial\beta} = A_2\cdot\dfrac{\partial B_2}{\partial\beta} + B_2\cdot\dfrac{\partial A_2}{\partial\beta}$, where

$$\frac{\partial A_1}{\partial\beta} = [\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^2\theta_2)\odot\boldsymbol{\xi}^2]^\top\cdot[Z^\top],$$
$$\frac{\partial B_2}{\partial\beta} = \frac{\partial B_1}{\partial\beta}.$$

Element 2.4: $\dfrac{\partial^2 f}{\partial\theta_2\partial\boldsymbol{\rho}} = A_2\dfrac{\partial B_2}{\partial\boldsymbol{\rho}} = A_2\cdot\dfrac{\partial B_1}{\partial\boldsymbol{\rho}}$.

The 3^{rd} row of the Hessian matrix

$$\frac{\partial f}{\partial\beta} = \overbrace{[\boldsymbol{w}^\top(\theta_1,\theta_2)Z^\top]}^{A_3}\cdot\overbrace{[-\boldsymbol{y}+\frac{\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2))}{\boldsymbol{\iota}+\exp(V\boldsymbol{\rho}+\beta Z\boldsymbol{w}(\theta_1,\theta_2))}]}^{B_3}$$

Element 3.3: $\dfrac{\partial^2 f}{\partial\beta\partial\beta} = A_3\cdot\dfrac{\partial B_3}{\partial\beta} = A_3\cdot\dfrac{\partial B_1}{\partial\beta}$

Element 3.4: $\dfrac{\partial^2 f}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}} = A_3\cdot\dfrac{\partial B_3}{\partial\boldsymbol{\rho}} = A_3\cdot\dfrac{\partial B_1}{\partial\boldsymbol{\rho}}$

The 4^{th} row of the Hessian matrix

$$\frac{\partial f}{\partial \boldsymbol{\rho}} = \overbrace{[V^\top]}^{A_4} \cdot \overbrace{[-\boldsymbol{y} + \frac{\exp(V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2))}{\boldsymbol{\iota} + \exp(V\boldsymbol{\rho} + \beta Z \boldsymbol{w}(\theta_1, \theta_2))}]}^{B_4}$$

Element 4.4: $\frac{\partial^2 f}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}} = A_4 \cdot \frac{\partial B_4}{\partial \boldsymbol{\rho}} = A_4 \cdot \frac{\partial B_1}{\partial \boldsymbol{\rho}}$

Appendix 5.D Analytic derivatives for the MIDAS VAR model

This section explains the derivation of the analytical gradient and of the analytical Hessian matrix of the objective function in the optimization problem related to estimating the coefficients of a MIDAS VAR regression (5.13) via non-linear least squares. The MIDAS-NLS estimator requires minimizing the target function (5.16) which can be reformulated as:

$$\text{RSS} = \boldsymbol{u}^\top \boldsymbol{u} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c}).$$

We explain the unrestricted case of the Beta weights, where optimization is performed with respect to the parameter vector $(\theta_1, \theta_2, \theta_3, \beta, \rho, a, \boldsymbol{\alpha})$. It is straightforward to adjust derivations for the three restricted cases.

The analytical gradient (Beta polynomial)

$$\text{RSS} = f(\beta, \theta_1, \theta_2, \theta_3, \rho, a, \boldsymbol{\alpha}) = \mathcal{G}\left(\mathcal{C}(\mathcal{W}(\beta, \theta_1, \theta_2, \theta_3), \rho, a, \boldsymbol{\alpha})\right):$$

$$\begin{aligned}
&1)\mathcal{W}: \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}^{m \cdot p} \\
&\qquad (\beta, \theta_1, \theta_2, \theta_3) \longmapsto \boldsymbol{w}, \\
&2)\ \mathcal{C}: \mathbb{R}^{m \cdot p} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^p \longrightarrow \mathbb{R}^{(m+1)^2 \cdot p} \\
&\qquad (\boldsymbol{w}, \rho, a, \boldsymbol{\alpha}) \longmapsto \boldsymbol{c}, \\
&3)\mathcal{G}: \mathbb{R}^{(m+1)^2 \cdot p} \longrightarrow \mathbb{R} \\
&\qquad \boldsymbol{c} \longmapsto \text{RSS}.
\end{aligned}$$

Consider the directional derivative of the target function f in the direction of λ. We aim to find an explicit expression for the gradient.

$$\langle \nabla f, \boldsymbol{\lambda} \rangle = \left\langle \begin{pmatrix} \nabla_{\theta_1} f \\ \nabla_{\theta_2} f \\ \nabla_{\theta_3} f \\ \nabla_{\beta} f \\ \nabla_{\rho} f \\ \nabla_{a} f \\ \nabla_{\boldsymbol{\alpha}} f \end{pmatrix}, \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ (\lambda_i)_{i \in \{7,\dots,6+p\}} \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \\ \nabla_{\boldsymbol{c}} \mathcal{G} \end{pmatrix}, \begin{pmatrix} (T_{\theta_1} \mathcal{C}) \cdot \lambda_1 \\ (T_{\theta_2} \mathcal{C}) \cdot \lambda_2 \\ (T_{\theta_3} \mathcal{C}) \cdot \lambda_3 \\ (T_{\beta} \mathcal{C}) \cdot \lambda_4 \\ (T_{\rho} \mathcal{C}) \cdot \lambda_4 \\ (T_{a} \mathcal{C}) \cdot \lambda_4 \\ (\lambda_i)_{i \in \{7,\dots,6+p\}} \end{pmatrix} \right\rangle$$

$$= \left\langle \begin{pmatrix} \left(T^*_{\theta_1} \mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\theta_2} \mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\theta_3} \mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\beta} \mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ \left(T^*_{\rho} \mathcal{C}\right) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ (T^*_{a} \mathcal{C}) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \\ (T^*_{\boldsymbol{\alpha}} \mathcal{C}) \cdot \nabla_{\boldsymbol{c}} \mathcal{G} \end{pmatrix}, \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ (\lambda_i)_{i \in \{7,\dots,6+p\}} \end{pmatrix} \right\rangle.$$

The components of the gradient are separately computed.

Component $\nabla_{\boldsymbol{c}} \mathcal{G}$

$$\langle \nabla_{\boldsymbol{c}} \mathcal{G}, \boldsymbol{\delta c} \rangle = \frac{d}{dt}\Big|_{t=0} \mathcal{G}(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{c} + t\boldsymbol{\delta c})) = \frac{d}{dt}\Big|_{t=0} (\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{c} + t\boldsymbol{\delta c}))^\top (\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{c} + t\boldsymbol{\delta c})) =$$
$$2(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c})^\top (-\boldsymbol{X}) \cdot \boldsymbol{\delta c} = \langle \boldsymbol{\delta c}, -2\boldsymbol{X}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c}) \rangle$$
$$\boxed{\nabla_{\boldsymbol{c}} \mathcal{G} = -2\boldsymbol{X}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c})}$$

Component $T^*_{\theta_1}\mathcal{C}$

$$\frac{\partial C}{\partial \theta_1} = \frac{\partial\left(\boldsymbol{i}_\omega \boldsymbol{\omega}^\top R_\omega\right)}{\partial \theta_1} = \boldsymbol{i}_\omega \frac{\partial\left(\boldsymbol{\omega}^\top\right)}{\partial \theta_1} R_\omega = \left[\frac{\partial \boldsymbol{\omega}}{\partial \theta_1} \text{relies on } T_{\theta_1}\mathcal{W}(\delta\theta_1) \text{ from Section 5.A}\right]$$
$$= \boldsymbol{i}_\omega \left(\beta \cdot \boldsymbol{x}^{\theta_1 - 1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2 - 1} \cdot \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} \odot (\ln(\boldsymbol{x}) + \psi(\theta_1+\theta_2) - \psi(\theta_1))\right)^\top R_\omega.$$

$$\boxed{T^*_{\theta_1}\mathcal{C}(\delta\theta_1) = \left(\operatorname{vec}\left(\frac{\partial C}{\partial \theta_1}\right)\right)^\top \delta\theta_1}$$

Component $T^*_{\theta_2}\mathcal{C}$

$$\frac{\partial C}{\partial \theta_2} = \frac{\partial\left(\boldsymbol{i}_\omega \boldsymbol{\omega}^\top R_\omega\right)}{\partial \theta_2} = \boldsymbol{i}_\omega \frac{\partial\left(\boldsymbol{\omega}^\top\right)}{\partial \theta_2} R_\omega = \left[\frac{\partial \boldsymbol{\omega}}{\partial \theta_2} \text{relies on } T_{\theta_2}\mathcal{W}(\delta\theta_2) \text{ from Section 5.A}\right]$$
$$= \boldsymbol{i}_\omega \left(\beta \cdot \boldsymbol{x}^{\theta_1 - 1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2 - 1} \cdot \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} \odot (\ln(\boldsymbol{\iota} - \boldsymbol{x}) + \psi(\theta_1+\theta_2) - \psi(\theta_2))\right)^\top R_\omega.$$

$$\boxed{T^*_{\theta_2}\mathcal{C}(\delta\theta_2) = \left(\operatorname{vec}\left(\frac{\partial C}{\partial \theta_2}\right)\right)^\top \delta\theta_2}$$

Component $T^*_{\theta_3}\mathcal{C}$

$$\frac{\partial C}{\partial \theta_3} = \frac{\partial\left(\boldsymbol{i}_\omega \boldsymbol{\omega}^\top R_\omega\right)}{\partial \theta_3} = \boldsymbol{i}_\omega \frac{\partial\left(\boldsymbol{\omega}^\top\right)}{\partial \theta_3} R_\omega = \boldsymbol{i}_\omega \beta \boldsymbol{\iota}^\top R_\omega.$$

$$\boxed{T^*_{\theta_3}\mathcal{C}(\delta\theta_3) = \left(\operatorname{vec}\left(\frac{\partial C}{\partial \theta_3}\right)\right)^\top \delta\theta_3}$$

Component $T^*_{\beta}\mathcal{C}$

$$\frac{\partial C}{\partial \beta} = \frac{\partial\left(\boldsymbol{i}_\omega \boldsymbol{\omega}^\top R_\omega\right)}{\partial \beta} = \boldsymbol{i}_\omega \frac{\partial\left(\boldsymbol{\omega}^\top\right)}{\partial \beta} R_\omega = \boldsymbol{i}_\omega \left(\boldsymbol{x}^{\theta_1 - 1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2 - 1} \cdot \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} + \theta_3\right)^\top R_\omega.$$

$$\boxed{T^*_{\beta}\mathcal{C}(\delta\beta)) = \left(\operatorname{vec}\left(\frac{\partial C}{\partial \beta}\right)\right)^\top \delta\beta}$$

Component $T^*_\rho \mathcal{C}$

$$\begin{aligned}\frac{\partial C}{\partial \rho} &= \frac{\partial \left(\rho \left(R_1 + aR_{a1}\right) + \rho^2 \left(R_2 + aR_{a2}\right) + \rho^3 \left(R_3 + aR_{a3}\right) + \rho^4 R_4\right)}{\partial \beta} \\ &= \left(R_1 + aR_{a1}\right) + 2\rho \left(R_2 + aR_{a2}\right) + 3\rho^2 \left(R_3 + aR_{a3}\right) + 4\rho^3 R_4.\end{aligned}$$

$$\boxed{T^*_\rho \mathcal{C}(\delta\rho)) = \left(\operatorname{vec}\left(\frac{\partial C}{\partial \rho}\right)\right)^\top \delta\rho}$$

Component $T^*_a \mathcal{C}$

$$\begin{aligned}\frac{\partial C}{\partial a} &= \frac{\partial \left(\rho \left(R_1 + aR_{a1}\right) + \rho^2 \left(R_2 + aR_{a2}\right) + \rho^3 \left(R_3 + aR_{a3}\right) + \rho^4 R_4 + aR_{a0}\right)}{\partial a} \\ &= \rho R_{a1} + \rho^2 R_{a2} + \rho^3 R_{a3} + R_{a0}.\end{aligned}$$

$$\boxed{T^*_a \mathcal{C}(\delta a)) = \left(\operatorname{vec}\left(\frac{\partial C}{\partial a}\right)\right)^\top \delta a}$$

Component $T^*_{\boldsymbol{\alpha}} \mathcal{C}$

The following rule (Lütkepohl, 2005, p. 662) is instrumental to find the derivative of interest:

$$\operatorname{vec}(ABC) = (C^\top \otimes A) \operatorname{vec}(B), \text{where}$$

$A, B,$ and C are matrices with appropriate dimensions.

$$\operatorname{vec}\left(\frac{\partial C}{\partial \boldsymbol{\alpha}}\right) = \frac{\partial \operatorname{vec}\left(\boldsymbol{i}_\omega \boldsymbol{\alpha}^\top R_\alpha\right)}{\partial \boldsymbol{\alpha}} = \frac{\partial (R_\alpha^\top \otimes \boldsymbol{i}_\omega) \operatorname{vec}\left(\boldsymbol{\alpha}^\top\right)}{\partial \boldsymbol{\alpha}} = R_\alpha^\top \otimes \boldsymbol{i}_\omega$$

$$\boxed{T^*_{\boldsymbol{\alpha}} \mathcal{C}(\delta\boldsymbol{\alpha})) = \left(\operatorname{vec}\left(\frac{\partial C}{\partial \boldsymbol{\alpha}}\right)\right)^\top \delta\boldsymbol{\alpha}.}$$

The analytical Hessian (Beta polynomial)

In the previous section, we have derived the analytical gradient of the objective function.

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial \theta_1} \\ \frac{\partial f}{\partial \theta_2} \\ \frac{\partial f}{\partial \theta_3} \\ \frac{\partial f}{\partial \beta} \\ \frac{\partial f}{\partial \rho} \\ \frac{\partial f}{\partial a} \\ \frac{\partial f}{\partial \boldsymbol{\alpha}} \end{pmatrix} = \begin{pmatrix} [\text{vec}\left(\boldsymbol{i}_\omega \left(\beta \cdot \boldsymbol{x}^{\theta_1 - 1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2 - 1} \cdot \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} \odot (\ln(\boldsymbol{x}) + \psi(\theta_1+\theta_2) - \psi(\theta_1))\right)^\top R_\omega\right)]^\top \cdot [-2\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c})] \\ [\text{vec}\left(\boldsymbol{i}_\omega \left(\beta \cdot \boldsymbol{x}^{\theta_1 - 1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2 - 1} \cdot \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} \odot (\ln(\boldsymbol{\iota} - \boldsymbol{x}) + \psi(\theta_1+\theta_2) - \psi(\theta_2))\right)^\top R_\omega\right)]^\top \cdot [-2\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c})] \\ [\text{vec}\left(\boldsymbol{i}_\omega \beta \boldsymbol{\iota}^\top R_\omega\right)]^\top \cdot [-2\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c})] \\ [\text{vec}\left(\boldsymbol{i}_\omega \left(\boldsymbol{x}^{\theta_1 - 1} \odot (\boldsymbol{\iota} - \boldsymbol{x})^{\theta_2 - 1} \cdot \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} + \theta_3\right)^\top R_\omega\right)]^\top \cdot [-2\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c})] \\ [\text{vec}\left((R_1 + aR_{a1}) + 2\rho(R_2 + aR_{a2}) + 3\rho^2(R_3 + aR_{a3}) + 4\rho^3 R_4\right)]^\top \cdot [-2\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c})] \\ [\text{vec}\left(\rho R_{a1} + \rho^2 R_{a2} + \rho^3 R_{a3} + R_{a0}\right)]^\top \cdot [-2\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c})] \\ [\text{vec}\left(R_\alpha^\top \otimes \boldsymbol{i}_\omega\right)]^\top \cdot [-2\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{c})] \end{pmatrix}$$

Whereas the Hessian matrix is symmetric, we derive the upper triangular entries including elements on the main diagonal.

$$H_f = \nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 f}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 f}{\partial \theta_1 \partial \theta_3} & \frac{\partial^2 f}{\partial \theta_1 \partial \beta} & \frac{\partial^2 f}{\partial \theta_1 \partial \rho} & \frac{\partial^2 f}{\partial \theta_1 \partial a} & \left(\frac{\partial^2 f}{\partial \theta_1 \partial \boldsymbol{\alpha}}\right)^\top \\ & \frac{\partial^2 f}{\partial \theta_2 \partial \theta_2} & \frac{\partial^2 f}{\partial \theta_2 \partial \theta_3} & \frac{\partial^2 f}{\partial \theta_2 \partial \beta} & \frac{\partial^2 f}{\partial \theta_2 \partial \rho} & \frac{\partial^2 f}{\partial \theta_2 \partial a} & \left(\frac{\partial^2 f}{\partial \theta_2 \partial \boldsymbol{\alpha}}\right)^\top \\ & & \frac{\partial^2 f}{\partial \theta_3 \partial \theta_3} & \frac{\partial^2 f}{\partial \theta_3 \partial \beta} & \frac{\partial^2 f}{\partial \theta_3 \partial \rho} & \frac{\partial^2 f}{\partial \theta_3 \partial a} & \left(\frac{\partial^2 f}{\partial \theta_3 \partial \boldsymbol{\alpha}}\right)^\top \\ & & & \frac{\partial^2 f}{\partial \beta \partial \beta} & \frac{\partial^2 f}{\partial \beta \partial \rho} & \frac{\partial^2 f}{\partial \beta \partial a} & \left(\frac{\partial^2 f}{\partial \beta \partial \boldsymbol{\alpha}}\right)^\top \\ & & & & \frac{\partial^2 f}{\partial \rho \partial \rho} & \frac{\partial^2 f}{\partial \rho \partial a} & \left(\frac{\partial^2 f}{\partial \rho \partial \boldsymbol{\alpha}}\right)^\top \\ & & & & & \frac{\partial^2 f}{\partial a \partial a} & \left(\frac{\partial^2 f}{\partial a \partial \boldsymbol{\alpha}}\right)^\top \\ & & & & & & \frac{\partial^2 f}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}} \end{pmatrix}.$$

We do not explain how to compute every element of the Hessian matrix to keep this document brief. The derivation is based on the same principle as in Section 5.A. Interested readers may follow the derivation directly in the code: Refer to AnalyticalHessian_var_beta.m file.

Bibliography

Aastveit, K. A., Foroni, C. and Ravazzolo, F. (2017). Density forecasts with MIDAS models, *Journal of Applied Econometrics* **32**(4): 783–801.

Acharya, V. V. and Yorulmazer, T. (2007). Too many to fail–An analysis of time-inconsistency in bank closure policies, *Journal of Financial Intermediation* **16**(1): 1–31.

Agarwal, S., Lucca, D., Seru, A. and Trebbi, F. (2014). Inconsistent regulators: Evidence from banking, *The Quarterly Journal of Economics* **129**(2): 889–938.

Agarwal, V. and Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models, *Journal of Banking and Finance* **32**: 1541–1551.

Aït-Sahalia, Y., Fan, J. and Xiu, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data, *Journal of the American Statistical Association* **105**(492): 1504–1517.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures, *Econometrica* **33**(1): 178–196.

Amendola, A. and Storti, G. (2015). Model uncertainty and forecast combination in high-dimensional multivariate volatility prediction, *Journal of Forecasting* **34**(2): 83–91.

Anatolyev, S. and Kobotaev, N. (2018). Modeling and forecasting realized covariance matrices with accounting for leverage, *Econometric Reviews* **37**(2): 114–139.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001). The distribution of realized stock return volatility, *Journal of Financial Economics* **61**(1): 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility, *Econometrica* **71**(2): 579–625.

Andreou, E., Ghysels, E. and Kourtellos, A. (2010). Regression models with mixed sampling frequencies, *Journal of Econometrics* **158**(2): 246–261.

Aubuchon, C. P. and Wheelock, D. C. (2010). The geographic distribution and characteristics of U.S. bank failures, 2007-2010: Do bank failures still reflect local economic conditions?, *Federal Reserve Bank of St. Louis Review* **92**(5): 395–415.

Audrino, F. (2014). Forecasting correlations during the late-2000s financial crisis: The short-run component, the long-run component, and structural breaks, *Computational Statistics and Data Analysis* **76**: 43–60.

Audrino, F., Huang, C. and Okhrin, O. (2019). Flexible HAR model for realized volatility, *Studies in Nonlinear Dynamics & Econometrics* **23**(3).

Audrino, F. and Knaus, S. D. (2016). Lassoing the HAR model: A model selection perspective on realized volatility dynamics, *Econometric Reviews* **35**(8-10): 1485–1521.

Audrino, F., Kostrov, A. and Ortega, J.-P. (2019). Predicting U.S. bank failures with MIDAS logit models, *Journal of Financial and Quantitative Analysis* **54**(6): 2575–2603.

Audrino, F. and Tetereva, A. (2019). Sentiment spillover effects for US and European companies, *Journal of Banking & Finance* **106**: 542–567.

Balla, E., Prescott, E. S. and Walter, J. R. (2015). Did the financial reforms of the early 1990s fail? A comparison of bank failures and FDIC losses in the 1986-92 and 2007-13 periods. FRB Richmond Working Paper, WP15-05.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2009). Realized kernels in practice: Trades and quotes, *The Econometrics Journal* **12**(3): C1–C32.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics* **2**(1): 1–37.

Bauer, G. H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility, *Journal of Econometrics* **160**(1): 93–101.

Becker, R., Clements, A. E., Doolan, M. and Hurn, A. (2015). Selecting volatility forecasting models for portfolio allocation purposes, *International Journal of Forecasting* **31**(3): 849–861.

Becker, R., Clements, A. and O'Neill, R. (2010). A Cholesky-MIDAS model for predicting stock portfolio volatility. NCER Working Paper.

Bennett, R. L. and Unal, H. (2014). The effects of resolution methods and industry stress on the loss on assets from bank failures, *Journal of Financial Stability* **15**: 18–31.

Berger, A. N. and Bouwman, C. H. S. (2013). How does capital affect bank performance during financial crises?, *Journal of Financial Economics* **109**(1): 146–176.

Berger, A. N. and DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks, *Journal of Banking and Finance* **21**(6): 849–870.

Betz, F., Oprică, S., Peltonen, T. A. and Sarlin, P. (2014). Predicting distress in European banks, *Journal of Banking and Finance* **45**: 225–241.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**(3): 307–327.

Bollerslev, T., Hood, B., Huss, J. and Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility, *The Review of Financial Studies* **31**(7): 2729–2773.

Bollerslev, T., Patton, A. J. and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting, *Journal of Econometrics* **192**(1): 1–18.

Bologna, P. (2015). Structural funding and bank failures, *Journal of Financial Services Research* **47**(1): 81–113.

Boudt, K., Laurent, S., Lunde, A., Quaedvlieg, R. and Sauri, O. (2017). Positive semidefinite integrated covariance estimation, factorizations and asynchronicity, *Journal of Econometrics* **196**(2): 347–367.

Breitung, J. and Roling, C. (2015). Forecasting inflation rates using daily data: A nonparametric MIDAS approach, *Journal of Forecasting* **34**(7): 588–603.

Brown, C. O. and Dinç, I. S. (2011). Too many to fail? Evidence of regulatory forbearance when the banking sector is weak, *The Review of Financial Studies* **24**(4): 1378–1405.

Byrd, R. H., Hribar, M. E. and Nocedal, J. (1999). An interior point algorithm for large-scale nonlinear programming, *SIAM Journal on Optimization* **9**(4): 877–900.

Carriero, A., Clark, T. E. and Marcellino, M. (2020). Nowcasting tail risks to economic activity with many indicators. SSRN Working paper.

Chen, X. and Ghysels, E. (2010). News – good or bad – and its impact on volatility predictions over multiple horizons, *The Review of Financial Studies* **24**(1): 46–81.

Chen, X. and Ghysels, E. (2011). News – good or bad – and its impact on volatility predictions over multiple horizons, *Review of Financial Studies* **24**(1): 46–81.

Chernykh, L. and Cole, R. A. (2015). How should we measure bank capital adequacy for triggering prompt corrective action? A (simple) proposal, *Journal of Financial Stability* **20**: 131–143.

Chevallier, J., Goutte, S., Guerreiro, D., Saglio, S. and Sanhaji, B. (2019). *Financial Mathematics, Volatility and Covariance Modelling: Volume 2*, Routledge.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility, *Journal of Applied Econometrics* **26**(6): 922–947.

Chodorow-Reich, G. (2014). The employment effects of credit market disruptions: Firm-level evidence from the 2008–9 financial crisis, *The Quarterly Journal of Economics* **129**(1): 1–59.

Clements, M. P., Galvão, A. B. and Kim, J. H. (2008). Quantile forecasts of daily exchange rate returns from forecasts of realized volatility, *Journal of Empirical Finance* **15**(4): 729–750.

Colacito, R., Engle, R. F. and Ghysels, E. (2011). A component model for dynamic correlations, *Journal of Econometrics* **164**(1): 45–59.

Cole, R. A., Taylor, J. and Wu, Q. (2020). Predicting bank failures using a simple dynamic hazard model. Available at SSRN: https://ssrn.com/abstract=1460526.

Cole, R. A. and White, L. J. (2012). Déjà vu all over again: The causes of U.S. commercial bank failures this time around, *Journal of Financial Services Research* **42**: 5–29.

Cole, R. A. and White, L. J. (2017). When time is not on our side: The costs of regulatory forbearance in the closure of insolvent banks, *Journal of Banking and Finance* **80**(C): 235–249.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* **7**(2): 174–196.

Corsi, F., Peluso, S. and Audrino, F. (2015). Missing in asynchronicity: A Kalman-EM approach for multivariate realized covariance estimation, *Journal of Applied Econometrics* **30**(3): 377–397.

Demirgüç-Kunt, A. and Detragiache, E. (2000). Monitoring banking sector fragility: A multivariate logit approach, *The World Bank Economic Review* **14**(2): 287–307.

Demyanyk, Y. and Hasan, I. (2010). Financial crises and bank failures: A review of prediction methods, *Omega* **38**: 315–324.

DeYoung, R. and Torna, G. (2013). Nontraditional banking activities and bank failures during the financial crisis, *Journal of Financial Intermediation* **22**(3): 397–421.

Engle, R. (2009). *Anticipating correlations: A new paradigm for risk management*, Princeton University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**(4): 987–1007.

Engle, R. F., Ghysels, E. and Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals, *Review of Economics and Statistics* **95**(3): 776–797.

Fengler, M. and Okhrin, O. (2016). Managing risk with a realized copula parameter, *Computational Statistics and Data Analysis* **100**: 131–152.

Ferrara, L., Marsilli, C. and Ortega, J.-P. (2014). Forecasting growth during the Great Recession: Is financial volatility the missing ingredient?, *Economic Modelling* **36**: 44–50.

Flanagan, T. and Purnanandam, A. (2020). Why do banks hide losses? Available at SSRN: https://ssrn.com/abstract=3329953.

Fleming, J., Kirby, C. and Ostdiek, B. (2001). The economic value of volatility timing, *Journal of Finance* **LVI**(1): 329–352.

Fleming, J., Kirby, C. and Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility, *Journal of Financial Economics* **67**: 473–509.

Foroni, C. and Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates, *International Journal of Forecasting* **30**(3): 554–568.

Foroni, C., Marcellino, M. and Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials, *Journal of the Royal Statistical Society* **178**(1): 57–82.

Francq, C. and Zakoian, J. M. (2011). *GARCH models: structure, statistical inference and financial applications*, John Wiley & Sons.

Freitag, L. (2014). Default probabilities, CDS premiums and downgrades: A probit-MIDAS analysis. GSBE Research Memoranda: https://cris.maastrichtuniversity.nl/portal/files/523024/content.

Freitag, L. (2016). *Credit Rating Agencies and the European Sovereign Debt Crisis*, PhD thesis, Maastricht University.

García, V., Sánchez, J. S. and Mollineda, R. A. (2013). On the effectiveness of preprocessing methods when dealing with different levels of class imbalance, *Knowledge-Based Systems* **25**(1): 13–21.

Ghysels, E. (2013). Matlab toolbox for mixed sampling frequency data analysis using MIDAS regression models. Available on MATLAB Central at http://www.mathworks.com/matlabcentral/fileexchange/45150-midas-regression.

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data, *Journal of Econometrics* **193**(2): 294–314.

Ghysels, E., Kvedaras, V. and Zemlys-Balevičius, V. (2020). Mixed data sampling (MIDAS) regression models, *Handbook of Statistics*, Vol. 42, Elsevier, chapter 4, pp. 117–153.

Ghysels, E., Kvedaras, V. and Zemlys, V. (2016). Mixed frequency data sampling regression models: the R package midasr, *Journal of Statistical Software* pp. 1–35.

Ghysels, E. and Marcellino, M. (2018). Forecasting volatility, *Applied Economic Forecasting Using Time Series Methods*, Oxford University Press, chapter 14, pp. 531–557.

Ghysels, E., Plazzi, A. and Valkanov, R. (2016). Why invest in emerging markets? The role of conditional return asymmetry, *Journal of Finance* **71**(5): 2145–2192.

Ghysels, E., Plazzi, A., Valkanov, R., Rubia, A. and Dossani, A. (2019). Direct versus iterated multiperiod volatility forecasts, *Annual Review of Financial Economics* **11**: 173–195.

Ghysels, E. and Qian, H. (2019). Estimating MIDAS regressions via OLS with polynomial parameter profiling, *Econometrics and Statistics* **9**: 1–16.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models. CIRANO Working Paper: http://www.cirano.qc.ca/files/publications/2004s-20.pdf.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2005). There is a risk-return trade-off after all, *Journal of Financial Economics* **76**(3): 509–548.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies, *Journal of Econometrics* **131**(1-2): 59–96.

Ghysels, E., Sinko, A. and Valkanov, R. (2007). MIDAS regressions: Further results and new directions, *Econometric Reviews* **26**(1): 53–90.

Ghysels, E., Valkanov, R. and Rubia, A. (2009). Multi-period forecasts of volatility: Direct, iterated, and mixed-data approaches. Available at SSRN: https://papers.ssrn.com/abstract=1344742.

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments, *Journal of Applied Econometrics* **25**(4): 595–620.

Gogas, P., Papadimitriou, T. and Agrapetidou, A. (2018). Forecasting bank failures and stress testing: A machine learning approach, *International Journal of Forecasting* **34**(3): 440–455.

Golosnoy, V., Gribisch, B. and Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility, *Journal of Econometrics* **167**(1): 211–223.

González, M., Nave, J. and Rubio, G. (2012). The cross section of expected returns with MIDAS betas, *Journal of Financial and Quantitative Analysis* **47**(1): 115–135.

Gorgi, P., Koopman, S. J. and Li, M. (2019). Forecasting economic time series using score-driven dynamic models with mixed-data sampling, *International Journal of Forecasting* **35**(4): 1735–1747.

Gouriéroux, C., Jasiak, J. and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility, *Journal of Econometrics* **150**(2): 167–181.

Guérin, P. and Marcellino, M. (2013). Markov-switching MIDAS models, *Journal of Business and Economic Statistics* **31**(1): 45–56.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1, 1), *Journal of Applied Econometrics* **20**(7): 873–889.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The model confidence set, *Econometrica* **79**(2): 453–497.

Harvey, D., Leybourne, S. and Newbold, P. (1997). Testing the equality of prediction mean squared errors, *International Journal of forecasting* **13**(2): 281–291.

Hautsch, N., Kyj, L. M. and Oomen, R. C. (2012). A blocking and regularization approach to high-dimensional realized covariance estimation, *Journal of Applied Econometrics* **27**(4): 625–645.

Hu, B. G. and Dong, W. M. (2014). A study on cost behaviors of binary classification measures in class-imbalanced problems. Available at arXiv: https://arxiv.org/pdf/1403.7100.pdf.

Iturriaga, F. J. L. and Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks, *Expert Systems with Applications* **42**(6): 2857–2869.

James, C. (1991). The losses realized in bank failures, *The Journal of Finance* **46**(4): 1223–1242.

Japkowicz, N. and Shah, M. (2011). *Evaluating learning algorithms: a classification perspective*, Cambridge University Press.

Japkowicz, N. et al. (2000). Learning from imbalanced data sets: A comparison of various strategies, *AAAI workshop on learning from imbalanced data sets*, Vol. 68, pp. 10–15.

Jin, J. Y., Kanagaretnam, K. and Lobo, G. J. (2011). Ability of accounting and audit quality variables to predict bank failure during the financial crisis, *Journal of Banking and Finance* **35**(11): 2811–2819.

Jin, J. Y., Kanagaretnam, K., Lobo, G. J. and Mathieu, R. (2013). Impact of FDICIA internal controls on bank risk taking, *Journal of Banking and Finance* **37**(2): 614–624.

Jordan, D. J., Rice, D., Sanchez, J., Walker, C. and Wort, D. H. (2010). Predicting bank failures: Evidence from 2007 to 2010. Available at SSRN: https://ssrn.com/abstract=1899581.

Kang, A., Lowery, R. and Wardlaw, M. (2015). The costs of closing failed banks: A structural estimation of regulatory incentives, *The Review of Financial Studies* **28**(4): 1060–1102.

Karminsky, A. M. and Kostrov, A. (2017). The back side of banking in russia: Forecasting bank failures with negative capital, *International Journal of Computational Economics and Econometrics* **7**(1-2): 170–209.

Kerstein, J. and Kozberg, A. (2013). Using accounting proxies of proprietary FDIC ratings to predict bank failures and enforcement actions during the recent financial crisis, *Journal of Accounting, Auditing, and Finance* **28**(2): 128–151.

Kolari, J., Glennon, D., Shin, H. and Caputo, M. (2002). Predicting large U.S. commercial bank failures, *Journal of Economics and Business* **54**(4): 361–387.

Kostrov, A. (2020). Estimating MIDAS regressions via MIDAS-NLS with revised optimization. Working paper.

Kostrov, A. and Mamonov, M. (2019). The formation of hidden negative capital in banking: A product mismatch hypothesis. BOFIT Discussion Paper.

Kostrov, A. and Tetereva, A. (2020). Forecasting realized covariances: a MIDAS approach. Working paper.

Labonte, M. (2020). Who regulates whom?: An overview of the us financial regulatory framework, *Congressional Research Service* **R44918**(7).

Laurent, S., Rombouts, J. V. and Violante, F. (2012). On the forecasting accuracy of multivariate GARCH models, *Journal of Applied Econometrics* **27**(6): 934–955.

Laurent, S., Rombouts, J. V. and Violante, F. (2013). On loss functions and ranking forecasting performances of multivariate volatility models, *Journal of Econometrics* **173**(1): 1–10.

Li, D. C., Liu, C. W. and Hu, S. C. (2010). A learning method for the class imbalance problem with medical data sets, *Computers in Biology and Medicine* **40**(5): 509–518.

Lo Duca, M. and Peltonen, T. A. (2013). Assessing systemic risks and predicting systemic events, *Journal of Banking and Finance* **37**(7): 2183–2195.

Longadge, R., Dongre, S. S. and Malik, L. (2013). Class imbalance problem in data mining: Review, *International Journal of Computer Science and Network* **2**(1): 83–87.

Lu, W. and Whidbee, D. A. (2013). Bank structure and failure during the financial crisis, *Journal of Financial Economic Policy* **5**(3): 281–299.

Lunde, A., Shephard, N. and Sheppard, K. (2016). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice, *Journal of Business & Economic Statistics* **34**(4): 504–518.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*, Springer Science & Business Media.

Malof, M. M., Mazurowski, M. A. and Tourassi, G. D. (2012). The effect of class imbalance on case selection for case-based classifiers: An empirical study in the context of medical decision support, *Neural Networks* **25**: 141–145.

Marcellino, M. and Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP, *Oxford Bulletin of Economics and Statistics* **72**(4): 518–550.

Mayes, D. G. and Stremmel, H. (2013). The effectiveness of capital adequacy measures in predicting bank distress. Available at SSRN: https://ssrn.com/abstract=2191861.

Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A. and Tourassi, G. D. (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance, *Neural Networks* **21**(2-3): 427–436.

McCracken, M. W. (2007). Asymptotics for out of sample tests of granger causality, *Journal of econometrics* **140**(2): 719–752.

McKinsey & Company (2009). Hidden in plain sight: The hunt for banking capital, *McKinsey on Corporate and Investment Banking* **9**(Autumn 2009): 26–35.

Menon, A. K., Narasimhan, H., Agarwal, S. and Chawla, S. (2013). On the statistical consistency of algorithms for binary classification under class imbalance, *Proceedings of Machine Learning Research*, pp. 28 (2013), 603–611.

Miller, A. B., Wall, C., Baines, C. J., Sun, P., To, T. and Narod, S. A. (2014). Twenty five year follow-up for breast cancer incidence and mortality of the canadian national breast screening study: randomised screening trial, *BMJ* **348:g366**.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Pictet, O. V., Olsen, R. B. and Ward, J. R. (1993). Fractals and intrinsic time: A challenge to econometricians. Working paper.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics* **160**(1): 246–256.

Patton, A. J. (2011b). Data-based ranking of realised volatility estimators, *Journal of Econometrics* **161**(2): 284–303.

Sarlin, P. (2013). On policymakers' loss functions and the evaluation of early warning systems, *Economics Letters* **119**(1): 1–7.

Schaeck, K. (2008). Bank liability structure, FDIC loss, and time to failure: A quantile regression approach, *Journal of Financial Services Research* **33**(3): 163–179.

Sun, Y., Kamel, M. S., Wong, A. K. and Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data, *Pattern Recognition* **40**(12): 3358–3378.

Symitsi, E., Symeonidis, L., Kourtis, A. and Markellos, R. (2018). Covariance forecasting in equity markets, *Journal of Banking and Finance* **96**: 153–168.

The Economist (2019). Fixing Europe's zombie banks. Article available at https://www.economist.com/leaders/2019/04/06/fixing-europes-zombie-banks.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1): 267–288.

Vortelinos, D. I. (2016). Realized correlation analysis of contagion, *The Quarterly Review of Economics and Finance* **60**: 138–148.

Wheelock, D. and Wilson, P. (2000). Why do banks disappear? The determinants of U.S. bank failures and acquisitions, *Review of Economics and Statistics* **8**(1): 127–138.

Curriculum vitae



Alexander Kostrov
Bodanstrasse 6
9000 St. Gallen
Switzerland
alexander.kostrov@unisg.ch

Education

2020 - 2021	Research visit, *University of North Carolina at Chapel Hill*, USA
2016 - 2021	PhD in Economics and Finance, *University of St. Gallen*, Switzerland
2010 - 2012	MSc in Financial Economics, *International College of Economics and Finance*, Russia
2005 - 2010	BSc in Economics, *Higher School of Economics*, Russia

Teaching experience

2017–2019	Teaching Assistant: *University of St. Gallen*, Switzerland Data Science Fundamentals, Statistics, Mathematics A and B, Macroeconomics I
2013–2016	Teaching Assistant: *International College of Economics and Finance*, Russia Financial Econometrics, Introduction to Microeconomics, Introduction to Macroeconomics

Conference Talks

Jun 2020	SoFiE Summer School, Chicago, USA
Apr 2020	Econometrics Student Workshop, *University of North Carolina*, USA
Oct 2019	CInSt Banking Workshop, Moscow, Russia
Aug 2019	EEA-ESEM, Manchester, UK (presented by co-author)
Jun 2019	IAAE, Nicosia, Cyprus
Jun 2019	SoFiE (pre-conference), Shanghai, China
Apr 2019	RES, Coventry, UK
May 2019	Econometrics colloquium, *University of Konstanz*, Germany
Dec 2018	CFE-CMStatistics, Pisa, Italy
Sep 2018	Research Seminar, *Higher School of Economics*, Moscow, Russia
Jun 2018	Economic Risk Seminar, *Humboldt University of Berlin*, Germany
Jan 2018	Econometrics colloquium, *University of Konstanz*, Germany
Dec 2017	CFE-CMStatistics, London, UK

Languages

Mothertongue:	Russian
Fluent:	English
Basic Knowledge:	German

Computer Skills

Advanced user :	R, MATLAB, LaTeX, Stata, Eviews, VBA
Intermediate knowledge:	Python, mySQL