Modeling non-stationary and stationary mixed-frequency time series

Citation for published version (APA):

Götz, T. B. (2014). *Modeling non-stationary and stationary mixed-frequency time series*. [Doctoral Thesis, Maastricht University]. Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20140910tg>

Document status and date:

Published: 01/01/2014

DOI:

[10.26481/dis.20140910tg](https://doi.org/10.26481/dis.20140910tg)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

MODELING NON-STATIONARY AND STATIONARY MIXED-FREQUENCY TIME SERIES

Thomas Götz

© Thomas Götz, Maastricht 2014

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing from the author.

This book was typeset by the author using L^AT_EX.

Published by Universitaire Pers Maastricht

ISBN: 978-94-6159-351-1

Printed in The Netherlands by Datawyse Maastricht

Modeling Non-Stationary and Stationary Mixed-Frequency Time Series

DISSERTATION

to obtain the degree of Doctor at
Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. L.L.G. Soete,
in accordance with the decision of the Board of Deans,
to be defended in public
on Wednesday September 10th, 2014 at 10.00 hrs

by

Thomas Burkhard Götz

Supervisor:

Prof. dr. J.R.Y.J. Urbain

Co-Supervisor:

Dr. A.W. Hecq

Assessment Committee:

Prof. dr. F.C. Palm (Chair)

Prof. dr. E. Ghysels (University of North Carolina)

Prof. dr. M. Marcellino (Bocconi University)

Dr. S.J.M. Smeeke

This research was financially supported by the Graduate School of Business and Economics (GSBE).

To my mother, who has always encouraged and supported me.
Thank you.

Acknowledgements

"Forecasting is like trying to drive a car blindfolded and following directions given by a person who is looking out of the back window" [- A. Nonymous]

If I attempted to perform that daring action w.r.t. my future path 5.5 years ago, and the person directing me had spotted one of the various deadlines through the back window, who knows where I would be right now? At that time, while on my study abroad in Australia, I wanted to do my M.Sc. in the UK. I stayed in Maastricht. A year later, I would have bet a fair amount of money on me leaving university in the coming summer.¹ I stayed in Maastricht. Yet another year later, already within my Ph.D., the ominous back-window-guide surely had his doubts about the road to come and whether my driving skills were sufficiently developed.

Engaging in the far less adventurous activity of simply looking at past data, I can happily conclude that things do not always work out as planned. Because if they did, I would have not driven this challenging, exciting, sometimes frustrating, but overall enjoyable road the past 4 years. Many people have contributed to this journey, and I would like to take the opportunity and thank them.

First of all, I thank my wonderful supervisors Alain Hecq and Jean-Pierre Urbain. Not only did you inspire me during my time as a student, which actually caused me to start my Ph.D., you also regularly slapped the aforementioned back-window-guide on his head and taught me that you constantly learn while driving. Needless to say, you implicitly revived my French skills during our/your discussions in Jean-Pierre's office: *Dès le début, quel que soit la question ou le problème, vous*

¹While explicitly pointing out that the information to come appears in smaller letters at the bottom of this page, beneath a bar separating it from the main text, it may be noted that the author's plan at that time was to become a trader or investment banker.

étiez toujours disponible pour me conseiller et m'assister. Ce fut pour moi un plaisir et un privilège de vous avoir tous deux en tant que superviseurs aussi bien sur le plan professionnel que privé. Merci beaucoup pour tout. To Alain's delight (I hope), I finish this paragraph by confessing that you managed to shift my mindset null hypothesis to " H_0 : You are right, I am wrong", whereby α is small.

I would like to thank the reading committee, Franz Palm, Eric Ghysels, Massimiliano Marcellino and Stephan Smeekes, for their time and careful reading of the manuscript. I particularly thank Eric Ghysels for hosting me at the University of North Carolina in Chapel Hill earlier this year, which was a wonderful experience in many aspects. Also, I thank Karin van den Boorn, Haydeé Hallmanns, Yolanda Paulissen and Esther Kockelkoren, for all their help over the years.

The working atmosphere at the Department of Quantitative Economics in particular, and the SBE in general, is a very pleasant and inspiring one. At this stage, I would like to thank all colleagues, especially those on the 3rd and 4th (plus a few on the 2nd and ground) floor of the A-wing, who have made my past 4 years so enjoyable. Special thanks to Artem Duplinskiy, Hande Karabiyik, Martijn van Brink, Lenard Lieb, Greg Csapó, Stephan Smeekes,² Dennis Türk, Andreea Bicu, Benedikt Vogt, Martin Solberger, Lei Wan and many others, too numerous to mention, for tolerating me in your offices, sharing coffees, beers and wines, howling at Dinner tables, working on posters in hotel rooms, using the student discount at Pathé, staring at TV screens in Irish pubs, lifting me through hospital wings, fighting Greek gods, and simply for being awesome people.

Finally, I want to thank my family and friends. I thank my parents for their continuous support, encouragement and never-ending endurance of asking me what I do in my thesis. I hope that, even though my answers were probably never fully illuminating,³ you believe me in saying that what you hold in your hands would not have been possible without you. I thank my girlfriend Ina Fassbender for her superpower of putting me into relaxation mode, especially in stressful times, and for her unconditional support.

Without all of you, this thesis would not have been what it is now. Thank you.

Thomas Götz

Maastricht, May 2014

²Additional credit for helping me in providing a Dutch summary that is actually written in proper Dutch.

³For which I take all the blame.

Contents

Acknowledgements	vii
Contents	ix
1 Introduction	1
2 Forecasting Mixed-Frequency Time Series with ECM-MIDAS Models	9
2.1 Introduction	10
2.2 The MF-ADL-model and ECM Representations	12
2.2.1 Dynamic MF Cointegration at $i = 0$	15
2.2.2 Dynamic MF Cointegration at $0 < i < m - 1$	16
2.2.3 Extending along q_l	17
2.2.4 Remark on the Error Correction Term for Cointegrating Analyses	17
2.3 Monte Carlo Simulations	19
2.3.1 Data Generating Process	19
2.3.2 Forecasting: Approaches and Robustness Tests	20
2.3.3 Results	24
2.3.4 High-Frequency DGP	29
2.4 Application	33
2.5 Conclusion	36

3	Combining Distributions of Real-Time Forecasts: An Application to U.S. Growth	39
3.1	Introduction	40
3.2	Extending the ROF Approach	42
3.2.1	Notation	42
3.2.2	Repeated Observation Forecasting	43
3.2.3	Constructing Densities of Forecasts	45
3.2.4	Application I	46
3.3	Sensitivity of Model 'Horse Races' to Vintages	49
3.3.1	Scoring Rules	49
3.3.2	Combining Densities of Forecasts	50
3.3.3	Remark on Data Uncertainty	51
3.3.4	Application II	52
3.4	Incorporating Revision Process Information into Real-Time Studies	57
3.5	Conclusion	62
3.A	Appendix: Models	65
4	Testing for Common Cycles in Non-Stationary VARs with Varied Frequency Data	69
4.1	Introduction	70
4.2	Mixed-Frequency VECMs	72
4.3	Alternative VAR/VECM Representations	76
4.4	Common Cycles in Mixed-Frequency VARs	80
4.4.1	Definitions	81
4.4.2	Testing Procedure	82
4.5	Applications	86
4.5.1	Common Feature Analysis	86
4.5.2	Forecasting	92
4.6	Monte Carlo Simulations	95
4.7	Conclusion	100
5	Nowcasting Causality in Mixed-Frequency Vector Autoregressive Models	103
5.1	Mixed-Frequency Structural VARs	103
5.2	Nowcasting Causality in MF-VARs	105
5.3	Granger and Nowcasting Causality in Mixed-Frequency Conditional Models and VARs	108

5.3.1	Case 1	109
5.3.2	Case 2	111
5.3.3	Case 3	112
5.4	Monte Carlo Simulations	114
6	Testing for Granger Causality in Large Mixed-Frequency VARs	117
6.1	Introduction	118
6.2	Causality in MF-VARs	119
6.2.1	Notation	119
6.2.2	MF-VARs	120
6.2.3	Granger Causality in MF-VARs	122
6.2.4	Remark on Nowcasting Causality	123
6.3	Parameter Reduction	123
6.3.1	Reduced Rank Restrictions	124
6.3.2	Bayesian MF-VARs	128
6.3.3	Common Low-Frequency VARs	136
6.3.4	Unrestricted VARs	138
6.4	Monte Carlo Simulations	138
6.4.1	Size	141
6.4.2	Power	143
6.5	Application	145
6.6	Conclusion	149
6.A	Appendix: Construction of the Selection Matrix S	151
7	Conclusion	153
	Bibliography	157
	Nederlandse samenvatting	167
	Valorization	171
	Curriculum Vitae	177

Chapter 1

Introduction

Time series data are available at mixed frequencies. This simple fact did not only serve as a foundation for an entire branch of econometric research, but also for this thesis. In view of the huge amount of distinct data that is available to researchers nowadays, one may alter the initial sentence to 'time series data are available at very mixed frequencies'. Indeed, CO2 emissions,¹ health outlooks or national income accounts appear on an annual basis, national account variables, e.g., the gross domestic product, are often available at a quarterly frequency, many economic indicators, such as the industrial production index or the unemployment rate, get supplied each month, money stock measures provide an example of weekly data, whereas many financial data get published daily, intra-daily or even on a tick-by-tick basis.

When attempting to perform an econometric analysis of variables that are sampled at mixed frequencies, one faces the dilemma of deciding how to make use of the high-frequency observations at one's disposal. On the one extreme, one may choose to simply use all of these observations directly. While this approach may be feasible for small frequency discrepancies among the variables under consideration, it is less attractive if the variables become available in time intervals of greater difference. Not only do the high-frequency observations become more 'noisy' as a variable's frequency increases, researchers quickly face the issue of parameter proliferation, i.e., the amount of parameters becomes too large given usually available sample sizes.

¹At least for data from the World Bank.

On the other extreme, one can temporally aggregate the high-frequency variables in order to achieve a common low frequency for all time series under consideration. The study of temporal aggregation in the econometric literature dates back almost half a century, with contributions by Amemiya and Wu (1972), Tiao (1972), Nijman and Palm (1990) and Marcellino (1999) among many others. Silvestrini and Veredas (2008) provide an excellent survey on this topic. Due to its simplicity, temporal aggregation has also received a lot of attention in the mixed-frequency literature and is still often considered as a benchmark approach. The basis for the method's simplicity is, however, also its weakness: The temporal aggregation scheme is *a priori* fixed.² If this aggregation scheme differs from the unknown aggregation scheme that is truly underlying the data, researchers may not only discard potentially useful information, they may also end up with biased estimates.

As a means to address this dilemma, Ghysels et al. (2004) introduced MI(xed) DA(ta) S(ampling) models. Their main virtue is their parsimonious specification, which allows researchers to make use of potentially all high-frequency observations with a relatively small number of parameters to estimate. While initial work on MIDAS models has focused on volatility (see Ghysels et al., 2006 or Forsberg and Ghysels, 2007 among others), much work has been devoted to improving quarterly (or monthly) macroeconomic forecasting.³ Additionally, MIDAS has been employed in the context of other regression applications⁴ or within dynamic correlation models (Colacito et al., 2011). In this thesis, however, only the former two applications are considered. As far as different extensions and variants of MIDAS regression models are concerned, i.e., different weight polynomial specifications, autoregressive augmentations, non-linear and multivariate MIDAS regression models among others, Ghysels et al. (2007) provide an excellent overview.

It is worth mentioning that the traditional approach to deal with mixed-frequency data (apart from temporally aggregating the high-frequency variables) has been to interpolate the low-frequency variable to achieve a common high frequency for the variables under consideration. This method rests on the assumption that all data are truly sampled at the high frequency, whereby the high-frequency observations of the variables sampled at lower frequencies are simply missing.

²Often a simple average of the high-frequency observations or a specific high-frequency observation is chosen as its temporally aggregated counterpart.

³Clements and Galvão (2008), Andreou et al. (2010) and Forni et al. (2012) to only name a few.

⁴For example, Ghysels et al. (2007) investigate commercial real estate valuation

The model is then cast in state-space form and the Kalman filter is used to obtain parameter estimates. Handling mixed-frequency data along these lines has been considered in Zadrozny (1988), Mariano and Murasawa (2003) and Giannone et al. (2008) among others.⁵ Note that there is a close connection between MIDAS regressions and the Kalman filter in the sense that the former are either exact representations or very accurate approximations of the latter (Bai et al., 2013). In this thesis, though, only MIDAS regressions models (and their system variants) are dealt with.

As far as the econometric analysis of MIDAS regression models is concerned, using a mixed-frequency data generating process, Andreou et al. (2010) derive the asymptotic properties of the nonlinear least squares (NLS) estimator and compare them to the estimator based on a temporally aggregated high-frequency variable. Thereby, the authors show that the latter may indeed result in inefficient, biased and inconsistent estimates, whereby the former is asymptotically more efficient. Further work on the econometric analysis of MIDAS models includes Ghysels et al. (2004), the aforementioned paper by Bai et al. (2013), Kvedaras and Rackauskas (2010), Rodriguez and Puggioni (2010), Wohlrabe (2009) and Foroni (2012), whereby the latter includes a survey of econometric methods for mixed-frequency data.

In all of the aforementioned papers, though, the variables under consideration are either assumed or transformed to satisfy covariance-stationarity, i.e., the first and second moments of the parameters have time-invariant (sample) distributions. In view of the aforementioned frequent application of MIDAS models to macroeconomic forecasting, where many of the associated variables are non-stationary, this may be problematic. Merely transforming all eventually non-stationary variables by taking first (or higher-order) differences may have a fatal effect on the model's forecasting performance due to potentially neglecting a common (long-run) stochastic trend in the variables (Clements and Hendry, 1998). Consequently, an extension of MIDAS regression models to the presence of non-stationarities in the data is called for.⁶

⁵Kuzin et al. (2009) compare this approach to MIDAS models in terms of their abilities to now- and forecast GDP growth in the euro area, finding that the two are complements rather than substitutes due to the former providing better results for long horizons and the latter for short horizons.

⁶Note that apart from the literature on state-space models (see, e.g., Seong et al., 2013), few papers have considered non-stationary mixed-frequency variables. Pons and Sansó (2005) discuss the effects of the aforementioned fixed temporal aggregation schemes on inference of cointegrating vectors. For $I(1)$ and $I(2)$ variables, it is shown that the aggregation scheme, which is best in terms of estimation precision, is not necessarily best in terms of hypothesis tests on the

Chapter 2 of this thesis, jointly written with Alain Hecq and Jean-Pierre Urbain, presents a first step in this direction by proposing a mixed-frequency error correction model (ECM) for possibly cointegrated time series sampled at varying frequencies. More specifically, the forecasting performance of these ECM-MIDAS models is compared to models that omit the presence of a long-run relationship and to models obtained after temporally aggregating the high-frequency variable.⁷ Indeed, such an analysis can be important for fore- or nowcasting business cycle indicators, because the gain of including the cointegrating relationship is apparent only for short horizons (Clements and Hendry, 1998). As far as the composition of the long-run term, i.e., *the dynamic mixed-frequency cointegrating relationship*, is concerned, we allow not only the end-of-period observation of the high-frequency variable to enter, but permit the 'neighboring' high-frequency observations to do so instead. While the property of cointegration is obviously not affected by this choice, it has an impact on the short-run dynamics terms of the ECM-MIDAS model at the model representation level.

It is found that ignoring a long-run term in the presence of cointegration lowers the forecasting performance of all approaches under consideration. Furthermore, the choice of a 'neighboring' high-frequency observation to enter the disequilibrium error has no effect on the forecast accuracy provided the structure of the short-run dynamics terms is adapted properly. If the latter requirement is not fulfilled, though, the forecasting performance of the respective approach may deteriorate. More specifically, while the inclusion of too few short-run variables leads to a worsened forecasting performances for all approaches (dynamic misspecification), over-specification of the short-run dynamics does not affect our mixed-frequency ECM models due to the data-driven weight determination underlying the MIDAS specification.

As discussed above, one of the most popular applications of mixed-frequency models is the improvement of macroeconomic forecasts. Aside from the feature of varying frequencies, macroeconomic indicators share another characteristic: Their figures get revised over time. Take quarterly GDP growth as an example. Even

cointegrating vector. Furthermore, the asymptotic distributions of the respective coefficients in the various cointegrated systems are derived. Finally, the effects of different aggregation schemes on estimation are assessed via Monte Carlo studies.

⁷Independent from work presented in this thesis, Miller (2012) derives the asymptotic properties of the NLS estimator in more general mixed-frequency regressions with non-stationary, and possibly cointegrated, variables. So-called CoMIDAS regressions, which are essentially equivalent to our ECM-MIDAS regressions, then arise as special cases of these specifications. In another paper, Miller (2011) focuses on the efficient estimation of the cointegrating vector of a single-equation regression model involving a low-frequency regressand and high-frequency regressors.

though the first figure corresponding to, say, the first quarter gets published in April, a new revised value appears in May, another one in June and so forth. Instead of making use of all of these revised figures, though, researchers often only consider the latest-available data points. Croushore and Stark (2001) argue that, in doing so, a data set is used that is different from the one researchers could have used in real time, and that this could have a drastic impact on the development of economic specifications.

In Chapter 3 of this thesis, also co-written by Alain Hecq and Jean-Pierre Urbain, the analyses of real-time data sets and mixed-frequency variables are combined. In particular, by means of an application to U.S. growth, we extend the repeated observations forecasting approach of Stark and Croushore (2002) to an autoregressive distributed lag setting in which the regressors are of possibly higher frequencies than the dependent variable. Having access to regressors clears the road towards considering a wider range of models than the autoregressive ones in Stark and Croushore (2002). This leads to the question whether selection among a set of competing model specifications is sensitive to the data release chosen.

To investigate this issue, we compute data release-specific forecasts for a given calendar date and approximate the forecasts' distributions by continuous densities, which are subsequently combined along the set of models under consideration. The composition and evolution of the resulting model-specific weights over time provide insights about the aforementioned research question, and allow the incorporation of revision process information into real-time studies. We find that, indeed, model selection is sensitive to the data release chosen, in the sense that the model ranking based on forecast accuracy measures obtained using latest-available data differs from the ranking implied by the aforementioned weights. In terms of incorporating revision process information into real-time studies, it turns out that forecast accuracy measures decrease with the forecast horizon and that forecasts made late in a quarter are more accurate than the ones made earlier.

Note that Chapters 2 and 3 deal with mixed-frequency variables in a single-equation dynamic regression model. If one were to pursue a system approach, such as a vector autoregressive (VAR) model, one would either temporally aggregate the high-frequency variable and work in a common low-frequency VAR or one would cast the model in state-space form, treating the high-frequency observations of the low-frequency variable as missing (Qian, 2013), and obtain a common high-frequency VAR. There has been no middle ground until the work of Ghysels (2012), which introduces an 'observation-driven' (Cox et al., 1981) mixed-frequency system

by stacking the high- and low-frequency variables into a vector and formulating a VAR model in terms of this vector. The term 'observation-driven' refers to the feature that the mixed-frequency VAR does not involve any latent variables, but is formulated entirely in terms of observable data. Consequently, the model's impulse response functions are driven by observable shocks, which stands in contrast to 'parameter-driven' models that contain latent processes and, hence, latent shocks. Note that the latter is an undesirable feature as, e.g., policy shocks, are, of course, observable (Foroni and Marcellino, 2013).

Despite investigating a wide range of econometric issues (parsimony, shocks and estimation among others), the analysis in Ghysels (2012) is restricted to covariance-stationary variables. Similar to the single-equation case, merely transforming the variables to stationarity (by high- or low-frequency differencing) may not only neglect the presence of a cointegrating relationship, it may also lead to inefficiencies when cointegration between the high- and low-frequency variables is absent. Hence, an extension of mixed-frequency VAR models to the non-stationary, possibly cointegrated, case is of great importance. Chapter 4, also jointly written with Alain Hecq and Jean-Pierre Urbain, fills this gap in the literature by deriving a vector error correction model (VECM) representation for non-stationary, possibly cointegrated, mixed-frequency variables.⁸ Furthermore, these representations allow us to extend the common cyclical feature analysis of Engle and Kozicki (1993) or Vahid and Engle (1993) to a mixed-frequency setup.

We show that in the presence of non-stationary mixed-frequency variables, a researcher needs to account for two kinds of long-run relationships in a VAR, (i) a 'trivial' or prespecified set, which stems from the fact that the high-frequency variables are cointegrated with each other, and (ii) the long-run relationship between the low- and high-frequency variables. Incorporating the associated set of restrictions on the dynamics changes the set of variables that has to be used in the subsequent common feature analysis. It is found that our mixed-frequency VECM models perform quite well in practice and that the associated common feature test behaves very well (at least for the correct lag length). Also, common

⁸Seong et al. (2013) provide an example for a 'parameter-driven' VECM, i.e., one that contains latent processes, obtained by casting the model in state-space form. Independent from our work, Ghysels and Miller (2013) investigate cointegration tests for aggregated and mixed-frequency time series. In particular, starting from a common high-frequency VAR generating the data, the familiar trace test is defined and its limiting distribution derived. Subsequently, the authors analyze the impact of temporally aggregating either both variables (common low-frequency case) or only one variable (mixed-frequency case) on cointegration testing. They find that working in mixed frequencies is advantageous due to uncertainty surrounding the additional aggregation of the high-frequency variables.

cycles that are hidden in a common low-frequency model may be detected using their mixed-frequency variants.

Being equipped with a mixed-frequency VAR or VECM sets the stage for an investigation of different sorts of issues, one of which is the notion of Granger causality. Due to the aforementioned VAR vector being obtained by stacking high- and low-frequency variables (corresponding to the same low-frequency time period), Granger causality is defined in terms of the low frequency. In Chapter 5 of this thesis, which is joint work with Alain Hecq, we analyze whether contemporaneous high-frequency observations help to explain, i.e., nowcast, the current value of the low-frequency variable and vice versa. In essence, we extend the notion of instantaneous causality (Lütkepohl, 2005, p.42) to the mixed-frequency setup and refer to it as *nowcasting causality*. We investigate the relationship between the latter and Granger causality in a mixed-frequency VAR and illustrate its impact on the parameters in a single-equation dynamic mixed-frequency regression model.

As far as Granger causality within mixed-frequency VAR models is concerned, Ghysels et al. (2013) develop a set of Granger causality tests that take advantage of the mixed-frequency nature of the variables involved. More precisely, they show that causal relationships (defined in the high frequency) are easier to recover using mixed frequencies compared to the approach based on temporal aggregation (low frequency). Furthermore, power improvements, asymptotically and in finite samples, are documented compared to conventional tests. The authors, however, make the implicit assumption that the frequency discrepancy between the variables involved is rather small, e.g., as in a quarter/month example. Chapter 6, co-written by Alain Hecq, relaxes this assumption and discusses Granger causality testing in a mixed-frequency VAR, where the number of high-frequency observations per low-frequency period is large, e.g., as in a month/working day example.

The proliferation of parameters, resulting from the large number of parameters that need to be estimated, is addressed by two approaches, reduced rank restrictions and a Bayesian mixed-frequency VAR. The former rests on the assumption of a reduced rank within the submatrix of parameters corresponding to the lags of the high-frequency variables. The latter is implemented by properly extending the approach of Banbura et al. (2010) to our mixed-frequency setting. Indeed, due to stacking the high-frequency observations in the mixed-frequency VAR, their approach cannot be directly applied in the mixed-frequency case. Instead, a more careful investigation is required to properly incorporate prior beliefs into auxiliary dummy variables, i.e., into a prior distribution. Subsequently, these techniques are

compared to the unrestricted mixed-frequency VAR and a common low-frequency VAR (based on temporal aggregation) in terms of their Granger (non-)causality testing behavior.

We confirm that the low-frequency VAR may lead to very poor results, due to Granger causality not being invariant to temporal aggregation (see Marcellino, 1999 or Breitung and Swanson, 2002), and that the unrestricted suffers from parameter proliferation for small sample sizes. While the performance of reduced rank restrictions depends crucially on how the observable high-frequency factors are computed (with overall best results for heterogeneous autoregressive type factors), the Bayesian mixed-frequency VAR approach yields size distortions when testing for causality in the direction from the low- to the high-frequency variables. As an empirical illustration, the link between uncertainty in financial markets and economic fluctuations is analyzed, whereby, in general, bi-directional Granger causality is detected.

Chapter 7 provides a short conclusion of the thesis. Although notation is as uniform as possible across chapters, slight differences occur. As a consequence, relevant notation is defined within each chapter. Whenever applicable, additional information is contained in appendices at the end of the chapters.

Chapter 2

Forecasting Mixed-Frequency Time Series with ECM-MIDAS Models

This chapter proposes a mixed-frequency error correction model for possibly cointegrated non-stationary time series sampled at different frequencies.

We highlight the impact, in terms of model specification, of the choice of the particular high-frequency explanatory variable to be included in the cointegrating relationship, which we call a dynamic mixed-frequency cointegrating relationship. The forecasting performances of aggregated models and several mixed-frequency regressions are compared in a set of Monte Carlo experiments. In particular, we look at both the unrestricted mixed-frequency model and at a more parsimonious MIDAS regression.

Whilst the existing literature has only investigated the potential improvements of the MIDAS framework for stationary time series, our study emphasizes the need to include the relevant cointegrating vectors in the non-stationary case. Furthermore, it is illustrated that the choice of the dynamic mixed-frequency cointegrating relationship does not matter as long as the short-run dynamics are adapted accordingly. Finally, the unrestricted model is shown to suffer from parameter pro-

lification for samples of relatively small size whereas MIDAS forecasts are robust to over-parameterization. We illustrate our results for the U.S. inflation rate.¹

2.1 Introduction

In economics we are often concerned with the problem of dealing with variables that are available, and thus sampled, at different frequencies. One example is the task of forecasting or nowcasting (see Giannone et al., 2008) the quarterly gross domestic product using monthly variables such as the industrial production index or daily indicators such as interest rates or stock prices. The classical way to deal with such a situation is the temporal aggregation of the high-frequency variables, i.e., to sample at the common low-frequency. However, this approach might lead to a loss of information due to omitting the high-frequency observations (Andreou et al., 2010). Hence, forecasting performances might improve by making use of these extra information. In many cases however, including all potential unrestricted high-frequency explanatory variables is unfeasible with regard to the number of observations for the low-frequency dependent variable. This has motivated the introduction of (Mi)xed (Da)ta (S)ampling (henceforth MIDAS) which aims at transforming a high-frequency variable so that all information in the high-frequency can be preserved (Ghysels et al., 2004). One of the virtues of the MIDAS approach is its parsimony, meaning that we can reduce the number of parameters to estimate while still employing the information present in the high-frequency variables (Ghysels and Valkanov, 2006).

With the exception of Miller (2012), MIDAS is usually applied to stationary time series or to transformations of non-stationary variables like first differences. For example, Clements and Galvão (2008, 2009) consider leading and coincident indicators as well as the output growth. Galvão (2012) looks at financial variables such as the term spread and the output growth. Andreou et al. (2013) forecast U.S. economic activity employing a large cross-section of transformed daily series. Merely working with first differenced $I(1)$ variables might neglect a possible long-run relationship between the variables. In this paper, we consider mixed-frequency $I(1)$ time series, a low-frequency variable y and high-frequency variables x , that

¹This chapter is based on the paper Götz et al. (2014) published in *Journal of Forecasting*. The authors thank Isaac Miller, two anonymous referees and attendants of the 32nd Annual Symposium on Forecasting in Prague 2011, the 65th European Meeting of the Econometric Society in Oslo 2011 and the 20th SNDE Annual Symposium in Istanbul 2012 for useful and valuable comments on earlier drafts of the paper.

are possibly cointegrated. We compare the forecasting performance of models for Δy when only the first difference terms of y and of x are included, omitting the presence of a long-run relationship, with a mixed-frequency error correction model (ECM hereafter). We also compare our mixed-frequency model with a model resulting from the temporal aggregation of high-frequency variables. In the usual case, i.e., the common-frequency framework, Clements and Hendry (1998) among others have compared models in terms of their ability to predict the levels, the first differences and the long-run relationships. Concerning the first difference, the gain of including the long-run relationship is apparent only for short horizons. In terms of forecasting and nowcasting business cycle indicators such as inflation or the unemployment rate, however, this can be important.

When representing the relationship between a particular high-frequency regressor and a low-frequency regressand in an error correction format we do not only consider the end-of-period observation to enter the long-run term (as done in Miller, 2012), but allow for the 'neighboring' high-frequency observations to do so instead. Despite the fact that it does not matter asymptotically whether a cointegrating relationship is modeled between y and the end-of-period observation of x or, say, mid-period observation of x , since cointegration is a long-run property (Marcellino, 1999), the choice of which particular high-frequency observation enters the disequilibrium error has an impact on the short-run dynamics terms of the ECM at the model representation level. Hence, although irrelevant for cointegration, for empirical analyses characterized by relatively small samples it is of interest to investigate whether an inappropriate specification of the short-run dynamics has an effect on the forecast accuracy.

Seong et al. (2013) also model cointegrated multivariate time series of mixed frequencies. However, they regard the high-frequency observations of y as missing data and cast the model in state-space form. Instead, this paper aims at providing tools for working in a mixed-frequency framework without relying on a state-space form. It turns out that most macroeconomic variables of interest (gross domestic product, inflation, etc) are available at a lower frequency than explanatory indicators, and therefore a regression approach makes sense. With the exception of Miller (2012), the evaluation of a mixed-frequency error correction model for forecasting purpose without adopting a state-space model has not been considered in the current literature and is one of the contributions of this paper.

The rest of the paper is organized as follows. In Section 2.2 we introduce a mixed-frequency ADL model and derive the short-run dynamics implied by alter-

native choices of the variables to be included in a long-run component. A MIDAS framework is proposed to capture the high-frequency structure of the short-run dynamics. The model is compared with the unrestricted approach and classical methods of temporal aggregation using Monte Carlo simulations in Section 2.3. Section 2.4 illustrates the theoretical and the Monte Carlo results via an empirical illustration. A final section concludes.

2.2 The MF-ADL-model and ECM Representations

For expository simplicity we consider a stable (see Boswijk, 1994) mixed-frequency autoregressive distributed lag model, denoted MF-ADL(p_l, q_h, q_l), for two non-stationary $I(1)$ time series y_t , which is sampled at low frequency, and $x_t^{(m)}$, which is available at high frequency. Note that lag orders indexed with l correspond to the variables' low-frequency lag orders whereas the one indexed with h refers to the high-frequency. The index t represents the low frequency and runs from 1 to T . The number of high-frequency observations per t -period equals m such that a high-frequency observation is represented by $x_{t-i/m}^{(m)}$, where i denotes the fraction of high-frequency units per low-frequency time-period we look into the past from the end-of-period observations. In a year/month example we have $m = 12$ and, for $i = 6$, for example $x_{t-1/2}^{(12)}$ denotes the value of x for June in year t . Note further that $x_{t+1/12}^{(12)} \equiv x_{t+1-11/12}^{(12)}$ represents January's x -value of year $t+1$. This notation is standard in the literature on mixed-frequency data (see Clements and Galvão, 2008, 2009, Ghysels et al., 2004 and Miller, 2012 among others).

The MF-ADL(p_l, q_h, q_l) model can then be represented as follows:

$$\begin{aligned}
y_t = & c + \alpha_1 y_{t-1} + \dots + \alpha_{p_l} y_{t-p_l} + \\
& \beta_{00} x_t^{(m)} + \beta_{10} x_{t-1/m}^{(m)} + \dots + \beta_{q_h 0} x_{t-q_h/m}^{(m)} + \\
& \beta_{01} x_{t-1}^{(m)} + \beta_{11} x_{t-1-1/m}^{(m)} + \dots + \beta_{q_h 1} x_{t-1-q_h/m}^{(m)} + \\
& \vdots \\
& \beta_{0q_l} x_{t-q_l}^{(m)} + \beta_{1q_l} x_{t-q_l-1/m}^{(m)} + \dots + \beta_{q_h q_l} x_{t-q_l-q_h/m}^{(m)} + \epsilon_t^{(m)},
\end{aligned} \tag{2.1}$$

or in terms of lag polynomials by

$$A(L)y_t^{(m)} = c + B_0(L^{1/m})x_t^{(m)} + B_1(L^{1/m})x_{t-1}^{(m)} + \dots + B_{q_l}(L^{1/m})x_{t-q_l}^{(m)} + \epsilon_t^{(m)},$$

(2.2)

where ϵ_t is assumed to be a martingale difference sequence with respect to the past information and for simplicity, $x_t^{(m)}$ is assumed to be strongly exogenous for the long run parameters (see Urbain, 1992). This assumption is made to simplify the presentation, extensions to the weakly exogenous case are direct. In the second representation, $A(L) = 1 - \alpha_1 L - \dots - \alpha_{p_l} L^{p_l}$ and $B_i(L^{1/m}) = \beta_{0i} + \beta_{1i} L^{1/m} + \dots + \beta_{q_h i} L^{q_h/m}$, $i = 0, \dots, q_l$, denote polynomials in L and $L^{1/m}$, respectively. The former represents the usual lag operator, i.e., over the low-frequency index t , whereas the latter stands for the lag operator in the high frequency such that $L^{j/m} x_{t-i/m}^{(m)} = x_{t-i/m-j/m}^{(m)}$. Obviously, $L^{1/m} x_{t-(m-1)/m}^{(m)} = x_{t-1}^{(m)}$. The same logic holds for difference operators: Δ represents the usual difference operator in the low frequency, whereas $\Delta^{(j/m)}$ indicates the difference operator in the high frequency and $\Delta^{(j/m)} x_{t-i/m}^{(m)} = (1 - L^{j/m}) x_{t-i/m}^{(m)} = x_{t-i/m}^{(m)} - x_{t-i/m-j/m}^{(m)}$. Note that we abstract from deterministic trends and higher order deterministic components, the inclusion of which would straightforwardly apply in this mixed-frequency framework.

Remark 2.1. In this paper we first consider the MF-ADL as our data generating process (DGP). This DGP has the crucial feature that both processes y and x are truly sampled at different frequencies. It is different from a high-frequency ADL in which both variables are sampled at a common high-frequency and in which the ultimately observed low-frequency variable is obtained as an aggregate version of its high-frequency observations in the respective t -period (see Silvestrini and Veredas, 2008 for aggregation schemes). The high-frequency observations of y are then treated as missing data such that the aggregated versions of y , usually released at the end of each low-frequency period, are the only observed values of y (Seong et al., 2013).

Remark 2.2. Contrary to assuming an aggregation scheme that links the missing, or unobserved, high-frequency observations of y with an observed one, we assume that the dependent variable is truly sampled at the low frequency. Although for many (macro)economic variables, usually characterized as stock or flow, a particular aggregation scheme lends itself straightforwardly (Marcellino, 1999), formulating a model in terms of latent processes demands the user to extract hidden states in order to predict future outcomes (Galvão, 2012). Such models are parameter-driven, whereas we start by relying on an observation-driven model (using the terminology of Cox et al., 1981), i.e., a model formulated exclusively in

terms of observable data. Note that such an MF-ADL is often used as DGP in the MIDAS literature (see, e.g., Andreou et al., 2010 or Galvão, 2012). In order to investigate whether our results are dependent on this assumption, we will consider a high-frequency DGP in Section 2.3.4 of the paper.

For simplicity, we first set $q_l = 0$ in (2.1) and start with an $\text{MF}(p_l, q_h, 0)$, where only the high-frequency lags of x corresponding to period t are present so that (2.1) becomes

$$A(L)y_t = c + B(L^{1/m})x_t^{(m)} + \epsilon_t. \quad (2.3)$$

For convenience, we (temporarily) denote p_l by p and q_h by q , with $q \leq m - 1$. As an example of the latter inequality, we can consider $m = 20$ working days per month but only take the more recent 15 working days ($q = 14$ then). We assume $x_t^{(m)}$ is $I(1)$ and that the series are cointegrated. We can rewrite this mixed-frequency ADL in an error correction format of the following form:

$$A^*(L)\Delta y_t = c - A(1)(y_{t-1} - \frac{B(1)}{A(1)}x_{t-1-i/m}^{(m)}) + B^*(L^{1/m})\Delta^{(1/m)}x_t^{(m)} + \epsilon_t, \quad (2.4)$$

where, by the Beveridge-Nelson decomposition (see, e.g., Davidson, 2000), $A(L) = A(1)L + A^*(L)(1 - L)$ with $A^*(L) = A(1) + A^{**}(L)$, $A^{**}(L) = \sum_{j=0}^{p-1} \alpha_j^{**} L^j$ and $\alpha_j^{**} = -\sum_{k=j+1}^p \alpha_k$, $j = 0, \dots, p-1$, such that $A^*(L)$ is a (low-frequency) lag polynomial of order $p-1$. A similar decomposition is applied to the polynomial $B(L^{1/m})$. The resulting structure of the short-run dynamics, however, i.e., $B^*(L^{1/m})$, depends on which $x_{t-1-i/m}^{(m)}$ -observation enters the long-run term. In a common-frequency case, the error correction term appears naturally following the BN decomposition at lag $t-1$ or $t-p$. In the mixed-frequency scenario, however, there is no reason to solely consider the end-of-period observation of the high-frequency regressors for entering the error correction term (as done in Miller, 2012). Alternative choices are possible without affecting the timing of the cointegration relationship. In particular, we may consider all 'neighboring' observations, i.e., high-frequency observations within the same t -period, leading to a larger set of regressor observations that may enter the cointegrating relationship. In analogy to the concept of dynamic cointegration (see Gourieroux and Monfort, 1990, page 419), we will refer to these as *dynamic mixed-frequency (MF) cointegrating relationships*. More formally, we may consider the following m different cases of a

dynamic MF cointegrating relationship:

$$y_{t-1} - \frac{B(1)}{A(1)} x_{t-1-i/m}^{(m)} \text{ for } i = 0, \dots, m-1. \quad (2.5)$$

Remark 2.3. It is important to point out that we allow precisely one high-frequency observation to enter the long-run term. In Section 2.2.4 we consider an alternative error correction format where the disequilibrium error contains all high-frequency observations of x corresponding to period $t-1$. Note also that we do not consider high-frequency observations of x in period t , i.e., $x_{t-i/m}^{(m)}, i = 1, \dots, m-1$, to enter the long-run relationship. The high-frequency observation entering the dynamic MF cointegrating relationship in (2.5) can be viewed as a particular aggregation (see also Miller, 2011) of all high-frequency observations of x in period $t-1$ since $x_{t-1-i/m}^{(m)} = (0 \dots 1 \dots 0)(x_{t-1}^{(m)} \dots x_{t-1-i/m}^{(m)} \dots x_{t-1-(m-1)/m}^{(m)})'$. Likewise, $x_{t-i/m}^{(m)}, i = 1, \dots, m-1$, is an aggregation of all high-frequency observations of x in period t . In order to keep the cointegration timing unchanged, we specify the error correction term in period $t-1$ using the corresponding low-frequency observation of y and an aggregated version of the x -observations in the same period.

Remark 2.4. The left-hand side of (2.4) involves low-frequency differences (Δ), whereas the short-run dynamics terms on the right-hand side are high-frequency differences ($\Delta^{(1/m)}$). Due to cointegration between both series, the error correction term as well as the aforementioned low- and high-frequency differences are $I(0)$ such that, for each i , the error correction model (2.4) is balanced.

As we illustrate below the choice of i , i.e., the case of dynamic MF cointegrating relationship, while leaving the long run properties unchanged, will impact the short-run dynamics, i.e., $B^*(L^{1/m})$.

2.2.1 Dynamic MF Cointegration at $i = 0$

If we consider the dynamic MF cointegrating relationship for $i = 0$, i.e., $y_{t-1} - \frac{B(1)}{A(1)} x_{t-1}^{(m)}$, the ECM representation can be derived by rewriting $B(L^{1/m})$ as:

$$\begin{aligned}
 B(L^{1/m}) &= \beta_0 + \beta_1 L^{1/m} + \dots + \beta_q L^{q/m} \\
 &= B(1)L^{m/m} + \\
 &\quad \underbrace{\left[\beta_0 + (\beta_0 + \beta_1)L^{1/m} + \dots + B(1)L^{q/m} + \dots + B(1)L^{(m-1)/m} \right]}_{B^*(L^{1/m})} \times \\
 &\quad (1 - L^{1/m}) \\
 &= B(1)L + B^*(L^{1/m})(1 - L^{1/m}),
 \end{aligned}$$

where $B(1) = \sum_{j=1}^q \beta_j$ and $B^*(L^{1/m})$ is a lag polynomial of order $m - 1$. This implies that (2.3) can be written as

$$\begin{aligned}
 A^*(L)\Delta y_t &= c - A(1)(y_{t-1} - \frac{B(1)}{A(1)}Lx_t^{(m)}) + B^*(L^{1/m})(1 - L^{1/m})x_t^{(m)} + \epsilon_t \\
 &= c - A(1)(y_{t-1} - \frac{B(1)}{A(1)}x_{t-1}^{(m)}) + B^*(L^{1/m})\Delta^{(1/m)}x_t^{(m)} + \epsilon_t.
 \end{aligned}$$

Note that the error correction term consists of y and one x -variable only. The short-run, however, is represented by high-frequency differences of all the $x_{t-i/m}^{(m)}$ -observations between t and $t - 1$ leading to m high-frequency terms entering the short-run dynamics in total.

2.2.2 Dynamic MF Cointegration at $0 < i < m - 1$

If $y_{t-1} - \frac{B(1)}{A(1)}x_{t-1-i/m}^{(m)}$, with $0 < i < m$, then the ECM representation becomes:

$$\begin{aligned}
 A^*(L)\Delta y_t &= c - A(1)(y_{t-1} - \frac{B(1)}{A(1)}L^{(m+i)/m}x_t^{(m)}) + \\
 &\quad B^*(L^{1/m})(1 - L^{1/m})x_t^{(m)} + \epsilon_t \\
 &= c - A(1)(y_{t-1} - \frac{B(1)}{A(1)}x_{t-1-i/m}^{(m)}) + B^*(L^{1/m})\Delta^{(1/m)}x_t^{(m)} + \epsilon_t,
 \end{aligned}$$

where $L^{(m+i)/m}x_t^{(m)} = L^{i/m}L^{m/m}x_t^{(m)} = L^{i/m}x_{t-1}^{(m)} = x_{t-1-i/m}^{(m)}$ and

$$\begin{aligned}
 B^*(L^{1/m}) &= \beta_0 + (\beta_0 + \beta_1)L^{1/m} + \dots + B(1)L^{q/m} + \dots + B(1)L^{m/m} + \\
 &\quad \dots + B(1)L^{(m+i-1)/m}.
 \end{aligned}$$

Here, we end up with $m + i$ short-run dynamics terms compared to m in Section 2.2.1.

2.2.3 Extending along q_l

We may allow for more low-frequency lags of x ($q_l > 0$) in the MF-ADL-model, and consider the MF-ADL(p, q_h, q_l) model:

$$A(L)y_t = c + B_0(L^{1/m})x_t^{(m)} + B_1(L^{1/m})x_{t-1}^{(m)} + \dots + B_{q_l}(L^{1/m})x_{t-q_l}^{(m)} + \epsilon_t.$$

Recalling that each of the q_l lag polynomials is of the same order q_h , straightforward application of the techniques employed before leads to the following ECM representation:

$$A^*(L)\Delta y_t = c - A(1) \left[y_{t-1} - \frac{\sum_{j=0}^{q_l} B_j(1)}{A(1)} L^{j/m} x_{t-1}^{(m)} \right] + B^*(L^{1/m})\Delta^{(1/m)} x_t^{(m)} + \epsilon_t,$$

where $B^*(L^{1/m})$ represents an 'aggregate' lag polynomial of order $m q_l + q_h$, whose precise structure depends on which dynamic MF cointegration case we consider.

2.2.4 Remark on the Error Correction Term for Cointegrating Analyses

As pointed out before, the short-run dynamics terms in the error correction models presented before are high-frequency differences of x . Also, each dynamic MF cointegrating relationship in (2.5) consists of one particular $x_{t-1-i/m}^{(m)}$, $i = 0, \dots, m-1$, and y_{t-1} . If we want to allow for several $x_{t-1-i/m}^{(m)}$ to appear in the long-run relationship, we should consider a different error correction representation:

$$\begin{aligned} A(L)y_t &= c + B(L^{1/m})x_t^{(m)} + \epsilon_t \\ \Leftrightarrow A^*(L)\Delta y_t &= c - A(1)y_{t-1} + \beta_0 x_{t-1}^{(m)} + \beta_1 x_{t-1-1/m}^{(m)} + \dots + \beta_q x_{t-1-q/m}^{(m)} + \\ &\quad \beta_0 \Delta x_t^{(m)} + \beta_1 \Delta x_{t-1/m}^{(m)} + \dots + \beta_q \Delta x_{t-q/m}^{(m)} + \epsilon_t \\ &= c - A(1)(y_{t-1} - \frac{1}{A(1)} B(L^{1/m})x_{t-1}^{(m)}) + B(L^{1/m})\Delta x_t^{(m)} + \epsilon_t. \end{aligned} \tag{2.6}$$

Note, however, that the long-run relationship of (2.6) is observationally equivalent to the ones in the previous subsection, e.g., the one obtained in Section 2.2.1:

$$\begin{aligned} &y_{t-1} - \frac{1}{A(1)} B(L^{1/m})x_{t-1}^{(m)} \\ &= y_{t-1} - \frac{1}{A(1)} (\beta_0 x_{t-1}^{(m)} + \beta_1 x_{t-1-1/m}^{(m)} + \dots + \beta_q x_{t-1-q/m}^{(m)}) \end{aligned}$$

$$\begin{aligned}
 &= y_{t-1} - \frac{1}{A(1)}(\beta_0 x_{t-1}^{(m)} + \beta_1 \underbrace{(x_{t-1}^{(m)} - \Delta^{(1/m)} x_{t-1}^{(m)})}_{x_{t-1-1/m}^{(m)}} + \dots \\
 &\quad + \beta_q \underbrace{(x_{t-1}^{(m)} - \Delta^{(1/m)} x_{t-1}^{(m)} - \dots - \Delta^{(1/m)} x_{t-1-(q-1)/m}^{(m)})}_{x_{t-1-q/m}^{(m)}}) \\
 &= \left[y_{t-1} - \frac{B(1)}{A(1)} x_{t-1}^{(m)} \right] + \sum_{i=0}^{q-1} \frac{\sum_{j=i+1}^q \beta_j}{A(1)} \Delta^{(1/m)} x_{t-1-i/m}^{(m)},
 \end{aligned} \tag{2.7}$$

where the first term of the last expression is the same error correction term as in the ECM in Section 2.2.1 and the second expression enters the short-run dynamics structure.

Remark 2.5. Since (2.6) and (2.4) are observationally equivalent, it is a matter of convenience which ECM we choose to work with. In (2.6), the short-run dynamics terms are low-frequency differences of the high-frequency observations of x instead of high-frequency differences as in (2.4). Assuming that x is the monthly industrial production index (IPI) and y is quarterly GDP (so $m = 3$), the estimation of the above model would mean we condition the quarterly change in GDP on *quarterly* changes in IPI, i.e., on the change from October to January ($\Delta x_{t-2/3}^{(3)}$), November to February ($\Delta x_{t-1/3}^{(3)}$) and December to March ($\Delta x_t^{(3)}$), for example. In the case where cointegration is either absent or not taken into account, as, e.g., in Clements and Galvão (2008), we would compute Δy_t and $\Delta^{(1/m)} x_{t-i/m}^{(m)}$, $i = 0, \dots, m-1$, i.e., the high-frequency differences of x . The latter correspond to the short-run dynamics terms in (2.4).

Remark 2.6. In (2.6), the disequilibrium error term $y_{t-1} - \frac{1}{A(1)} B(L^{1/m}) x_{t-1}^{(m)}$ contains all high-frequency observations of x in period $t-1$. Within a Engle-Granger two-step procedure, this would be estimated by regressing y_{t-1} on all $x_{t-1-i/m}^{(m)}$, $i = 0, \dots, m-1$, and subsequently test whether the residuals are $I(0)$ processes (Engle and Granger, 1987). This cointegration regression can be estimated using MIDAS restrictions (to be discussed below). Testing the null of no-cointegration using a MIDAS-type cointegrating regression, is, however, difficult since the model has to be estimated under the null of a spurious regression which leads to computational issues (see Götz, 2010 for details). Note that if we rewrite the disequilibrium error as in (2.7), we end up with only one x -observation in the long-run term and a mixture of low- and high-frequency differences in the short-run dynamics, which may be harder to interpret and justify economically.

Consequently, we focus on the error correction representations in (2.4), which are characterized by only one particular high-frequency observation of x in the error correction term and high-frequency differences of x as short-run dynamics terms.

2.3 Monte Carlo Simulations

2.3.1 Data Generating Process

The DGP is a stable 'observable' MF-ADL(p, q_h, q_l) that is often employed in the MIDAS-literature (see, e.g., Andreou et al., 2010). We consider a MF-ADL($1, m - 1, 0$), where $m = 12$ (think of, e.g., weeks in a quarter or months in a year). This implies that $q_h = 11$, i.e., the corresponding MF-ADL in (2.3) includes all high-frequency observations of x in period t as regressors. The DGP is re-parameterized in ECM form where a specific case of dynamic MF cointegration is imposed:

$$\begin{aligned}\Delta y_t &= c + \delta(y_{t-1} - x_{t-1}^{(12)}) + \alpha \Delta y_{t-1} + \sum_{i=0}^{11} \beta_i \Delta^{(1/12)} x_{t-i/12}^{(12)} + v_t, \\ \Delta^{(1/12)} x_{t-i/12}^{(12)} &= \epsilon_{t-i/12}^{(12)} \text{ for all } i = 0, \dots, 11 \text{ and } t = 1, \dots, T,\end{aligned}$$

where $x_0^{(12)} = u_1$, $\epsilon_{t-i/12}^{(12)} \sim \text{i.i.d. } N(0, 1)$, $u_1 \sim N(0, 1)$ and where $v_t \sim \text{i.i.d. } N(0, 1)$ and $y_t = y_0 + \sum_{i=1}^t \Delta y_i$ for $t > 1$. We fix $y_0 \sim N(0, 1)$ and $y_{-1} = 0$ such that $\Delta y_0 = y_0$. This ensures that we obtain $I(1)$ variables y_t and $x_{t-i/12}^{(12)}$, $i = 0, \dots, 11$ where the former is a low-frequency and the latter are high-frequency variables. Furthermore, y and x are cointegrated. For $-2 < \delta < 0$, this set-up assumes the existence of a dynamic MF cointegrating relationship for $i = 0$, i.e., $y_{t-1} - \gamma x_{t-1}^{(12)}$. Furthermore, we assume that the dynamic MF cointegrating relationship is known exactly, i.e., γ is known. Without loss of generality we set $\gamma = 1$. This is, of course, a strong assumption which should be relaxed in further research. For now, however, we want to focus on the forecasting performance of our mixed-frequency ECMs with and without a long-run term and employing different methods to estimate the short-run dynamics. Alternatively, we could use any consistent estimator of the long-run relationship like, e.g., fully modified OLS.

Throughout the simulations, we set $c = 0.1$, $\alpha = 0.5$ and consider four different values for the error correction coefficient $\delta \in \{-0.75, -0.25, -0.1, 0\}$, the latter value corresponding to the no-cointegration case. The coefficients of the high-frequency differences in $x_{t-i/12}^{(12)}$ are the cumulative ones of the respective $x_{t-i/12}^{(12)}$ -

observations in levels. To be more specific, Section 2.2.1 shows that the coefficient on $\Delta^{(1/m)}x_{t-i/12}^{(12)}$ is the sum of the coefficients on the first i observations of $x_{t-i/12}^{(12)}$ in the underlying MF-ADL model (2.3). We assume that more recent observations of $x_t^{(12)}$ should have a larger impact on y_t implying that the β -coefficients above should be increasing, but with a decreasing 'step size', i.e., if the coefficients were represented by a continuous function, it would be increasing and concave. The sample sizes for the low frequency variable we consider are $T \in (50, 100, 250, 500)$. $T+50$ observations are generated and the first 50 disregarded in the computations. All computations are done using GAUSS10 with 10,000 replications.

2.3.2 Forecasting: Approaches and Robustness Tests

We compare the forecasting performances of eight different approaches that differ in their treatment of the absence/presence of cointegration, the short-run dynamics and the possible temporal aggregation retained. These are:

- (1a) Unrestricted short-run with dynamic MF cointegration,
- (1b) Unrestricted short-run; dynamic MF cointegrating relationship excluded,
- (2a) Restricted short-run (MIDAS) with dynamic MF cointegration,
- (2b) Restricted short-run (MIDAS); dynamic MF cointegrating relationship excluded,
- (3a) Point-in-Time sampling with cointegration,
- (3b) Point-in-Time sampling; cointegrating relationship excluded,
- (4a) Average sampling with cointegration,
- (4b) Average sampling; cointegrating relationship excluded.

'Unrestricted short-run' simply indicates that the high-frequency variables are not temporally aggregated nor are their coefficients estimated via a MIDAS regression; they are estimated via ordinary least squares. These two cases can be seen as cases where no restrictions are imposed on the coefficients of the $\Delta^{(1/12)}x^{(12)}$ -terms. Comparing (1a) and (1b) is essentially what Clements and Hendry (1998) or Engle and Yoo (1987) investigate in the standard framework.

As mentioned in the introduction, MIDAS aims at preserving information present in the high-frequency variables and estimating parameters in a parsimonious way. If we consider the different ECM representations given in Section 2.2,

we see that estimating such models without any restrictions might be unappealing due to parameter proliferation (Andreou et al., 2013). As an example, if the dependent variable is measured at a quarterly frequency and the regressor at a daily frequency, we might have over 50 parameters to estimate. Generally, in a MIDAS model we hyper-parameterize the polynomial lag structure yielding

$$A^*(L)\Delta y_t = c + \delta(y_{t-1} - \frac{B(1)}{A(1)}x_{t-1-i/m}^{(m)}) + \beta \sum_{i=0}^{q^*} w_{i+1}(\theta) \Delta^{(1/m)} x_{t-i/m}^{(m)} + u_t, \quad (2.8)$$

where $w(\theta)$ are weights that sum up to one in order to identify the scale coefficient β and where q^* is the amount of high-frequency lags of x one wants to include in the regression. Note that q^* may, of course, be larger than $m - 1$. Coherent with the DGP in the preceding subsection, $i = 0$ in $y_{t-1} - \frac{B(1)}{A(1)}x_{t-1-i/m}^{(m)}$ such that $q^* = m - 1 = 11$ would be sufficient. Generally speaking, however, one may include as many lags as one desires. The weights are based on an underlying weight function for which different possible specifications are proposed in the literature (see, e.g., Ghysels et al., 2007). In this paper, we employ the exponential Almon Lag polynomial having its roots in the work of Almon (1965) on polynomial interpolation in distributed lag equations. The weights are given by

$$w_k(\theta) = w_k(\theta_1, \theta_2) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=1}^{k^{max}} \exp(\theta_1 k + \theta_2 k^2)}.$$

As discussed in Ghysels et al. (2007), this low-dimensional lag polynomial specification is very flexible as it allows for the weight determination to be completely data-driven, and it allows for various possible shapes of the weight function. Often, decaying weights are detected due to more recent observations being more important. Such a declining structure can be imposed by restricting $\theta_2 \leq 0$. According to Ghysels and Valkanov (2006), forecasts based on the exponential Almon lag polynomial dominate the ones based on other lag polynomial specifications.

Remark 2.7. In the equation above a single weight function is specified for the entire set of short-run variables (in line with Andreou et al., 2013). Two comments have to be made here: First, the purely data-driven determination of the weight function allows us to include more short-run dynamics terms than theoretically necessary; the redundant ones will be assigned a low (optimally zero) weight. This flexibility of MIDAS will prove useful at a later stage. Second, it is also possible to assign one weight function to each low-frequency period of the short-run dynamics

variables. Ghysels et al. (2007) present a generalization of MIDAS models nesting such an approach. To allow for multiple weight functions, we have to amend (2.8) to

$$A^*(L)\Delta y_t = c + \delta(y_{t-1} - \frac{B(1)}{A(1)}x_{t-1-i/m}^{(m)}) + \sum_{j=0}^{q^*} \beta_j \sum_{i=0}^{m-1} w_{i+1}^j(\theta) \Delta^{(1/m)} x_{t-j-i/m}^{(m)} + u_t, \quad (2.9)$$

where β_j and $w^j(\theta)$ are t -period-specific scale coefficients and weight functions, respectively. Note that the number of weights per low-frequency period is now fixed to m , whereas the user can specify the number of low-frequency periods (q^*) he or she wants to include high-frequency lags from in the regression. Such an extension might be interesting to apply in our case since it allows the weights to change more abruptly. Note that a 'joint' weight function will simply act like a smoothing curve of all successive 'single' weight functions.

Approaches (3a)-(4b) are standard methods used to temporally aggregate high-frequency data (see, e.g., Silvestrini and Veredas, 2008 or Marcellino, 1999). Point-in-Time Sampling selects one particular high-frequency observation of a t -period (often the last one, e.g., December or the last trading day) as the corresponding low-frequency observation. As a benchmark, we will apply Point-in-Time Sampling to the last observation, i.e., $x_t^{(PIT)} = x_t^{(12)}$. Average sampling simply selects the average per low-frequency period as the respective low-frequency observation $x_t^{(AV)} = \frac{1}{12} \sum_{i=0}^{11} x_{t-i/12}^{(12)}$.

To summarize the different ECM representations, the four different approaches correspond to the following forecasting models:

$$\begin{aligned} \widehat{\Delta y_t} &= \widehat{c} + \widehat{\delta}(y_{t-1} - x_{t-1-i/12}^{(12)}) + \widehat{\alpha}\Delta y_{t-1} + \sum_{i=0}^{q^*} \widehat{\beta}_i \Delta^{(1/12)} x_{t-i/12}^{(12)} & (\text{Unrestricted}) \\ \widehat{\Delta y_t} &= \widehat{c} + \widehat{\delta}(y_{t-1} - x_{t-1-i/12}^{(12)}) + \widehat{\alpha}\Delta y_{t-1} + \widehat{\beta} \sum_{i=0}^{q^*} w_i(\widehat{\theta}_1, \widehat{\theta}_2) \Delta^{(1/12)} x_{t-i/12}^{(12)} & (\text{MIDAS}) \\ \widehat{\Delta y_t} &= \widehat{c} + \widehat{\delta}(y_{t-1} - x_{t-1}^{(PIT)}) + \widehat{\alpha}\Delta y_{t-1} + \widehat{\beta}\Delta x_t^{(PIT)} & (\text{Point-in-Time}) \\ \widehat{\Delta y_t} &= \widehat{c} + \widehat{\delta}(y_{t-1} - x_{t-1}^{(AV)}) + \widehat{\alpha}\Delta y_{t-1} + \widehat{\beta}\Delta x_t^{(AV)} & (\text{Averaging}). \end{aligned} \quad (2.10)$$

The estimated parameters are then used to compute 36 one-step-ahead forecasts (amounting to 9 or 3 years in a week/quarteror. The Root Mean Squared Error

(RMSE) is computed for every approach in each iteration and the means of the RMSEs are reported. In order to see whether the forecast accuracy of two competing models differs, two test statistics are considered depending on whether the models are nested or non-nested. In the latter case, the classical Diebold-Mariano test (Diebold and Mariano, 1995) is employed whereas the modified Diebold-Mariano test proposed by Harvey et al. (1997) (see also Clark and McCracken, 2001) is conducted for the former case. Note that parameter uncertainty is not taken into account.

The reason for focussing on one-step-ahead forecasts stems from two points. Firstly, as mentioned in the introduction, Clements and Hendry (1998) have shown the gain of including the long-run relationship to be apparent only for short horizons, i.e., from one- until 5-steps ahead. Consequently, we decided to consider only such a short horizon for our simulations. Secondly, we focus on macroeconomic applications for the ECM-MIDAS models such that issues related to short-term forecasting or even nowcasting become important. As such, considering one-step-ahead forecasts is most adequate.

At this stage, we have not been specific about which dynamic MF cointegrating relationship to employ in cases (1a) and (2a). For both approaches the 'optimal' scenario is analyzed meaning that the long-run relationship is specified as in the (usually unknown) DGP, i.e., $y_{t-1} - x_{t-1}^{(12)}$. However, to investigate the impact of choosing an alternative MF cointegration specification, we also compute the RMSEs corresponding to a dynamic MF cointegrating relationship with $0 < i < 11$. For example, if we consider the mid-period observation, then the error correction term is specified as $y_{t-1} - x_{t-1-6/12}^{(12)} = y_{t-1} - x_{t-1-1/2}^{(12)}$. However, in both cases, i.e., using $y_{t-1} - x_{t-1}^{(12)}$ or $y_{t-1} - x_{t-1-1/2}^{(12)}$, the number of terms entering the short-run dynamics is adapted consistently with the specifications in Sections 2.2.1 or 2.2.2, respectively. This implies that we set $q^* = m - 1 = 11$ in (2.10) for the models labeled Unrestricted and MIDAS when the DGP dynamic MF cointegrating relationship, i.e., $y_{t-1} - x_{t-1}^{(12)}$, is considered. For $y_{t-1} - x_{t-1-1/2}^{(12)}$, we set $q^* = m - 1 + i = 17$. Note finally that the ECM-MIDAS model (2.8) is considered, implying that only one weight function is computed instead of several ones as in (2.9).

To sum up, we start by investigating whether the exact composition of the error-correction term matters in finite samples given that we adapt the structure of the short-run dynamics terms accordingly. The aforementioned analysis requires the user, however, to know the true underlying structure of the short-run

dynamics terms, a situation which hardly occurs in practice. In order to assess the sensitivity of the unrestricted and the MIDAS approach to the correct specification of the short-run dynamics structure, the RMSEs for cases (1a)-(2b) are computed employing first less short-run variables than theoretically necessary, and second more than theoretically necessary. In doing so we investigate whether the inclusion of too few or too many variables distorts our forecasts. Here, only the models with the dynamic MF cointegrating relationships corresponding to that in the actual DGP are considered.

2.3.3 Results

Table 2.1 reports the outcomes for MIDAS and the unrestricted approach (Unrestricted). The results for Point-in-Time as well as Average sampling are qualitatively similar and the corresponding tables are therefore skipped for representational ease. The figures represent the (means of the) RMSEs relative to the benchmark model, i.e., the model including the dynamic MF cointegrating relationship as specified in the DGP namely $y_{t-1} - x_{t-1}^{(12)}$. Hence, a value larger than 1 indicates the benchmark to be superior while a value smaller than 1 leads to the opposite conclusion. Furthermore, in order to investigate whether a certain approach significantly outperforms the benchmark case, the percentages in brackets report the frequency with which the Diebold-Mariano tests reject the null of equal forecast accuracy at the 5% significance level. The closer the percentage to 100% (0%), the more (less) the forecasting performance of the model at hand differs from the benchmark case.

Ignoring the issue of which dynamic MF cointegrating relationship is retained, excluding the disequilibrium error in the presence of cointegration ($\delta \neq 0$) has a significant negative impact on the forecasting performance of all approaches (see the corresponding rows labeled 'No long-run'), also of Point-in-Time and Average sampling. Of course, this effect decreases with δ , i.e., the slower the speed of adjustment to the equilibrium. This observation is in line with Clements and Hendry (1998). As mentioned before, in the MIDAS literature, empirical analyzes often only consider first differenced $I(1)$ -variables (see, e.g., Andreou et al., 2013). In the presence of cointegration, the inclusion of the respective relationship might yield considerable forecasting improvements. If the variables are not cointegrated, neglecting the long-run term results in the lowest RMSEs (see the 'No long-run'-row for $\delta = 0$).

Table 2.1: Ratios of RMSEs where the short-run is estimated by MIDAS or left Unrestricted

Long-Run/ T	MIDAS					Unrestricted				
	$\delta = -0.75$					$\delta = -0.75$				
	50	100	250	500		50	100	250	500	
$y_{t-1} - x_{t-1-1/2}^{(12)}$	1.1262 (32.5%)	1.1332 (37.1%)	1.128 (35.6%)	1.1275 (35.4%)		1.1041 (31.7%)	1.0386 (16.6%)	1.0133 (10.5%)	1.0062 (8.6%)	
No long-run	2.7302 (100%)	2.798 (100%)	2.8132 (100%)	2.8205 (100%)		2.7731 (100%)	2.8055 (100%)	2.8157 (100%)	2.8244 (100%)	
$\delta = -0.25$										
$y_{t-1} - x_{t-1-1/2}^{(12)}$	1.0001 (13.9%)	1.0014 (9.1%)	1.0014 (7.9%)	1.0019 (7.7%)		1.1038 (31.5%)	1.0386 (16.4%)	1.0132 (10.5%)	1.0063 (8.8%)	
No long-run	1.3247 (99.4%)	1.3461 (99.6%)	1.3508 (99.7%)	1.3558 (99.7%)		1.3306 (99.1%)	1.3464 (99.5%)	1.3518 (99.6%)	1.3572 (99.7%)	
$\delta = -0.1$										
$y_{t-1} - x_{t-1-1/2}^{(12)}$	1.0063 (14.6%)	1.0039 (9.8%)	1.0034 (8%)	1.003 (8.1%)		1.1039 (32.1%)	1.0384 (16.7%)	1.0132 (10.5%)	1.0063 (8.6%)	
No long-run	1.0746 (74.9%)	1.0939 (78.9%)	1.1009 (80.8%)	1.1039 (81.7%)		1.077 (73.7%)	1.0944 (77.2%)	1.1011 (80.5%)	1.1044 (81.1%)	
$\delta = 0$										
$y_{t-1} - x_{t-1-1/2}^{(12)}$	1.0188 (24.6%)	1.0147 (15.5%)	1.0134 (11.2%)	1.012 (10.6%)		1.1 (35.4%)	1.0373 (17.9%)	1.0133 (10.9%)	1.0062 (8.6%)	
No long-run	0.9054 (15.3%)	0.9653 (11.6%)	0.9903 (7.8%)	0.9956 (7.2%)		0.9106 (15.1%)	0.9663 (11.8%)	0.9908 (8%)	0.9962 (7.5%)	

Note: The figures represent ratios of RMSEs, the numerators of which belong to models with an alternative long-run term ($y_{t-1} - x_{t-1-1/2}^{(12)}$ -cases) or models without a long-run term. The denominators are always the RMSEs corresponding to the model at hand, once MIDAS and once the unrestricted one, including the DGP dynamic MF cointegrating relationship $y_{t-1} - x_{t-1}^{(12)}$, i.e., the benchmark case. The percentages in brackets are the rejection frequencies of Diebold-Mariano tests at the 5% level compared to the aforementioned benchmark. The short-run dynamics are modeled by MIDAS or Unrestricted and for each model under consideration, the number of short-run variables is adapted consistent with the methodology of Section 2.2.

Consider now the impact of including an alternative dynamic MF cointegrating relationship, i.e., $y_{t-1} - x_{t-1-1/2}^{(12)}$ instead of $y_{t-1} - x_{t-1}^{(12)}$ (see the corresponding rows in Table 2.1). It turns out using $y_{t-1} - x_{t-1}^{(12)}$ as in the DGP yields the lowest RMSE for both, MIDAS and Unrestricted. However, the rejection frequencies for the tests of equal forecast accuracy show that the models with $y_{t-1} - x_{t-1-1/2}^{(12)}$ rarely perform significantly worse than the benchmark case. This result is not a surprise since, provided that the short run dynamics is correctly adapted, these models are equally valid. Also, cointegration is a long-run property being invariant to temporal aggregation (Marcellino, 1999) such that it does not matter in terms of forecast accuracy which dynamic MF cointegrating relationship is considered. It is crucial though that the structure of the short-run dynamics is adapted according to Section 2.2, i.e., that a sufficient number of short-run variables is included in the regression. If this is ensured the models are not misspecified, but merely reparameterizations of the underlying ADL model. At the end of the section we investigate the effect of not transforming the short-run terms adequately.

Up until now we have assumed correctly adapted short-run dynamics, i.e., the chosen number of short-run variables is consistent with the methodology of Section 2.2, which requires knowledge of the true underlying DGP. Since the latter is of course unknown in practice, it is of interest to investigate the various approaches' forecasting performances when the short-run dynamics structure is misspecified. Of course, since Point-in-Time and Average sampling are methods circumventing this issue by temporally aggregating the high-frequency variable, we only focus on MIDAS and the unrestricted approach here. We compare our benchmark model, which specifies the dynamic MF cointegrating relationship as $y_{t-1} - x_{t-1}^{(12)}$ and the associated number of 12 short-run terms, to two scenarios: The models with $y_{t-1} - x_{t-1}^{(12)}$, but where (i) only $\frac{1}{2}m = 6$ and (ii) $3m = 36$ short-run variables are included in the regression. Note that as far as MIDAS regressions are concerned, we use (2.8) with $q^* = 5$ for scenario (i) and (2.9) with $q^* = 2$ for scenario (ii). In other words, we compute one weight function for the 6 short-run variables in (i) and three separate weight functions for the $3 \cdot 12 = 36$ short-run terms in (ii). We only display the results for $\delta = -0.75$, since they seem to be insensitive to the choice of δ . Table 2.2 illustrates the outcomes. Again, the figures represent the (means of the) RMSEs of the respective method relative to the ones of the benchmark case whereas the percentages in brackets correspond to rejection frequencies of the corresponding Diebold-Mariano tests.

Table 2.2: The effect of misspecifying the short-run dynamics structure

$\delta = -0.75$		T			
Method	50	100	250	500	
MIDAS	1.0968 (70.2%)	1.1103 (75.6%)	1.1107 (75.6%)	1.1103 (74.4%)	Too few short-run variables
Unrestricted	1.1397 (79.5%)	1.1279 (80.1%)	1.1171 (77.8%)	1.1132 (75.6%)	
MIDAS	1.0557 (9.7%)	1.0291 (6.5%)	1.0132 (5.5%)	1.0072 (5.9%)	Too many short-run variables
Unrestricted	2.193 (86.2%)	1.2463 (32.3%)	1.0745 (8.6%)	1.0336 (5.7%)	

Note: The figures represent ratios of RMSEs, the numerators of which belong to models where the short-run terms have been either left unrestricted or are estimated by a MIDAS regression. However, the number of short-run variables is (i) equal to 6 (too few) or (ii) equal to 36 (too many). As far as MIDAS regressions are concerned, for scenario (i) only one weight function is computed, whereas for scenario (ii) three separate weight functions are estimated. The RMSEs in the denominators are based on the MIDAS model including the DGP dynamic MF cointegrating relationship $y_{t-1} - x_{t-1}^{(12)}$ and the associated number of 12 short-run variables (see Section 2.2). The percentages in brackets are the rejection frequencies of Diebold-Mariano tests at the 5% level compared to the aforementioned benchmark.

As expected, including too few short-run terms leads to a serious dynamic misspecification (serial correlation) of the model and, therefore, worsens the forecasting performances of both, MIDAS and unrestricted approach. On the other hand, including too many variables has significant negative consequences for the unrestricted method when T is small. This reflects the fact that parameter proliferation is exacerbated since 36 short-run variables need to be estimated as opposed to 9 in the case of MIDAS (3 scale coefficients and 6 weight function parameters). This effect decreases as T increases. The forecast accuracy of MIDAS models, however, does not seem to be harmed by including too many parameters, irrespective of the sample size. The reason is the flexible and data-driven weight determination, which ensures that unimportant observations get assigned a small weight. In scenario (ii), the last $2m = 24$ variables entering the short-run dynamics are, in fact, redundant, implying that they should get assigned a zero weight. Figure 2.1 shows the three separate weight functions for $T = 100$ computed for this scenario. For representational ease it is constructed for the first 1,000 of 10,000 simulations only.

Figure 2.1 shows that when the too many short-run dynamics terms are included, MIDAS is flexible enough to assign zero weight to the redundant variables, thereby not forfeiting forecast accuracy. The unrestricted approach, however, suf-

Figure 2.1: Weight functions corresponding to an overspecified short-run dynamics structure

Note: Three consecutive weight functions multiplied by their respective scale coefficients, corresponding to the sets of 12 high-frequency short-run variables for each of the three low-frequency periods considered (36 short-run variables in total), are displayed for each replication. The first 12 are theoretically sufficient according to the methodology of Section 2.2, the latter 24 are theoretically redundant. In each replication considered here, $T = 100$ and $\delta = -0.75$ and the model is estimated by MIDAS regressions where the DGP dynamic MF cointegrating relationship $y_{t-1} - x_{t-1}^{(12)}$ is included.

fers from parameter proliferation for $T = 50$ as presented in Table 2.2. Hence, MIDAS' invariance to over-parameterization of the short run dynamics provides the user with an approach that allows him to include a sufficient amount of short-run variables without knowing the exact dynamic specification of underlying dynamic MF cointegrating relationship. The preserved parsimony of the model adds to the advantages of the MIDAS approach.

Finally, Table 2.3 reports the RMSEs of the unrestricted approach, Point-in-Time and Average sampling, relative to the RMSEs of MIDAS for the benchmark cases, i.e., the model including the dynamic MF cointegrating relationship $y_{t-1} - x_{t-1}^{(12)}$ for $\delta \neq 0$ and the model without an error correction term for $\delta = 0$. Note that for ease of readability rejection frequencies of Diebold-Mariano tests are skipped. As before, only the results for $\delta = -0.75$ are presented since the outcomes seem qualitatively insensitive to the choice of δ .

Table 2.3: Forecasting performances of Unrestricted, Point-in-Time and Averaging Sampling relative to MIDAS

$\delta = -0.75$	T			
MIDAS vs...	50	100	250	500
Unrestricted	1.1118	1.0485	1.0167	1.0076
Point-in-Time	0.9712	0.9914	0.9994	1.0017
Averaging	1.1747	1.2001	1.2074	1.2137

Note: The figures represent ratios of RMSEs, the numerators of which belong to models where the short-run terms have been either left unrestricted, Point-in-Time or Average sampled. The RMSEs in the denominators are based on MIDAS models. If cointegration is present ($\delta \neq 0$), $y_{t-1} - x_{t-1}^{(12)}$ is included and the short-run dynamics structure is adapted consistent with the methodology of Section 2.2. If $\delta = 0$, however, no dynamic MF cointegrating relationship is included.

From Table 2.3, it appears that MIDAS performs better than leaving the short-run unrestricted and using Average sampling. These differences in forecast accuracy are very pronounced for small T illustrating the parameter proliferation problem of the unrestricted short-run method (see Andreou et al., 2013). For $T = 500$, the difference is almost always insignificant. Average sampling yields the worst results for large sample sizes. However, given the DGP, i.e., dynamic MF cointegrating relationship $y_{t-1} - x_{t-1}^{(12)}$ instead of $y_{t-1} - x_{t-1}^{(AV)}$ and non-flat weights on the short-run terms, this is not a surprise. Indeed, when averaging is an incorrect aggregation scheme, it delivers asymptotically biased slope estimates (Andreou et al., 2010). Note that for $T \geq 100$ MIDAS and Point-in-Time Sampling have statistically no different forecast accuracies. For $T = 50$ the latter seems to outperform the former. This is easily understood as the dynamic MF cointegrating relationship specified in the DGP is actually identical to the one used for Point-in-Time Sampling, i.e., $y_{t-1} - x_{t-1}^{(12)} = y_{t-1} - x_{t-1}^{(PIT)}$ since $x_t^{(12)} = x_t^{(PIT)}$. The latter is precisely the observation we assigned the largest impact on y_t to (see Section 2.3.1). Hence, a different Point-in-Time Sampling scheme or respective changes in the DGP (β -coefficients or error correction term) are expected to alter the situation to the disadvantage of Point-in-Time Sampling.

2.3.4 High-Frequency DGP

So far, we have assumed that the DGP is a MF-ADL(p_l, q_h, q_l). As already outlined in Section 2.2, an alternative is to assume a high-frequency ADL, in which both variables are sampled at a common high-frequency and in which the ultimately

observed low-frequency variable is obtained as an aggregate version of its high-frequency observations in the respective t -period. In order to investigate whether our results are sensitive to this form of DGP, we amend the MF-ADL in Section 2.2 to the following HF-ADL(p_h, p_l, q_h, q_l), written only in terms of lag polynomials:

$$\begin{aligned} A_0(L^{1/m})y_t^{(m)} = & c + A_1(L^{1/m})y_{t-1}^{(m)} + \dots + A_{p_l}(L^{1/m})y_{t-p_l}^{(m)} + \\ & B_0(L^{1/m})x_t^{(m)} + B_1(L^{1/m})x_{t-1}^{(m)} + \dots + \\ & B_{q_l}(L^{1/m})x_{t-q_l}^{(m)} + \epsilon_t^{(m)}, \end{aligned} \quad (2.11)$$

where, now, all lag polynomials are high-frequency lag polynomials. A mixed-frequency setup is then achieved by assuming that the observed low-frequency variables $y_{t-j}, j = 0, \dots, p_l$, are aggregated versions of their corresponding high-frequency observations. Silvestrini and Veredas (2008) propose the following aggregation scheme, which nests several popular aggregation choices:

$$y_{t-j} = \sum_{i=0}^{m-1} \alpha_{ij} y_{t-j-i/m}^{(m)},$$

where $\sum_{i=0}^{m-1} \alpha_{ij} = 1 \forall j$. Note that Point-in-Time Sampling on the end-of-period observation, i.e., an HF-ADL($0, p_l, q_h, q_l$), is achieved by setting $\alpha_{0j} = 1$ and $\alpha_{ij} = 0, i \neq 0$, whereas Average sampling is obtained when $\alpha_{ij} = 1/m \forall i$. We consider both of these aggregation schemes for $y_t^{(m)}$ after generating a high-frequency VECM. To do so, we consider a Gaussian VAR(p_h) $z_t^{(m)} = c + \sum_{i=1}^{p_h} \Phi_i z_{t-i/m}^{(m)} + \epsilon_t^{(m)}$ for the vector time series $\{z_t^{(m)} = (y_t^{(m)}, x_t^{(m)})', t = 1, \dots, T\}$. Note that c is a two-dimensional vector of constants. The error term $\epsilon_t^{(m)}$, is an i.i.d. bivariate Gaussian vector with nonsingular covariance matrix Ω . Let us denote $\Phi(L) = I - \sum_{i=1}^{p_h} \Phi_i L^i$. Further, we assume that exactly one unit root is present so that $\text{rank}(\Phi(1)) = 1$ and $\Phi(1)$ can be expressed as $\Phi(1) = -\delta\beta'$ with δ and β being (2×1) vectors of rank 1, where β is the cointegrating vector and δ the error correction term. This can be rewritten in the desired error correction format:

$$\Delta^{(1/m)} z_t^{(m)} = c + \delta\beta' z_{t-1/m}^{(m)} + \sum_{j=1}^{p_h-1} \Phi_j^* \Delta^{(1/m)} z_{t-j/m}^{(m)} + \epsilon_t^{(m)}, \quad (2.12)$$

with $\Phi_0^* = I$ and $\Phi_j^* = -\sum_{k=j+1}^{p_h} \Phi_k, j = 1, \dots, p_h - 1$.

After applying Point-in-Time Sampling on the end-of-period observation or Average sampling to $y_t^{(m)}$ as mentioned before, the remainder of the analysis fol-

lows the previous section. We again assume $x_t^{(m)}$ to be strongly exogenous for the long run parameters so that we impose $\delta = (\delta_1 \ 0)'$. Further, the elements of the bottom rows of the matrices Φ_j^* in Equation (2.12) as well as the corresponding entry of the intercept vector are set to zero to generate $x_t^{(m)}$ as a random walk. The entries of the top row, i.e., the coefficients of the lagged terms of $y_t^{(m)}$ and $x_t^{(m)}$ in the equation for $y_t^{(m)}$ and the corresponding intercept are chosen to behave similarly to those in the previous section, i.e., more recent observations have a stronger impact than more distant ones.

Importantly, with white noise errors, a low-frequency model is consistent with the high-frequency model at hand only if the same aggregation scheme is chosen for both, the dependent and independent variables. Hence, if Point-in-Time or Average sampling is applied to the dependent variable and a different aggregation scheme is chosen for the regressors, an estimation error that enters the error term could be introduced (Miller, 2011). On the contrary, the mixed-frequency DGP considered before assumed that sampling of the regressand does not affect the error term.

Table 2.4 presents the outcomes for a Point-in-Time sampled dependent variable where the short-run variables are modeled by MIDAS. The remaining results are similar to the outcomes of the previous section such that they are not reported for space reasons. They are, however, available on request. In summary, as far as the individual forecasting methods are concerned, the main conclusions do not differ from those reached in the previous analysis. In the presence of cointegration, the inclusion of a dynamic MF cointegrating relationship improves the forecasting performances. Also, the forecast accuracy of models with an alternative long-run term, yet adequately adapted short-run dynamics structure, hardly differs from the benchmark case.

Similarly, the comparison of MIDAS with other approaches does not change drastically. When the regressand is obtained by Point-in-Time Sampling on y_t^h , MIDAS generally outperforms both, Unrestricted and Average sampling. Point-in-Time Sampling performs as well as MIDAS which is not surprising since the dependent variable itself is Point-in-Time sampled. Indeed, Miller (2011) and Chambers (2003) argue that, with white noise errors, the cointegrating vector is most efficiently estimated if the regressand aggregation scheme is matched by the one applied to the regressors. Note that Average sampling performs worst because it is not matching the regressand aggregation scheme. MIDAS is still able to adapt to the Point-in-Time sampled dependent variable due to its flexible data-

Table 2.4: High-Frequency DGP, Short-run estimated by MIDAS

Long-Run/T	MIDAS			
	$\delta = -0.75$			
	50	100	250	500
$y_{t-1} - x_{t-1-1/2}^{(12)}$	1.0026 (11.1%)	1.0014 (8.9%)	1.002 (7.8%)	1.0016 (7.7%)
No long-run	1.3195 (99.6%)	1.3328 (99.8%)	1.3399 (99.8%)	1.3432 (99.8%)
	$\delta = -0.25$			
	50	100	250	500
$y_{t-1} - x_{t-1-1/2}^{(12)}$	0.9657 (16.2%)	0.9659 (14.5%)	0.9667 (14%)	0.9672 (14.4%)
No long-run	1.3875 (99.8%)	1.4031 (99.8%)	1.4122 (99.9%)	1.415 (99.9%)
	$\delta = -0.1$			
	50	100	250	500
$y_{t-1} - x_{t-1-1/2}^{(12)}$	0.9317 (23.1%)	0.9324 (20.8%)	0.9337 (19.8%)	0.9348 (19.5%)
No long-run	1.3736 (99.6%)	1.3902 (99.8%)	1.402 (99.9%)	1.4055 (99.9%)
	$\delta = 0$			
	50	100	250	500
$y_{t-1} - x_{t-1-1/2}^{(12)}$	0.9731 (36.6%)	0.9998 (32.9%)	1.058 (37.6%)	1.1324 (44.4%)
No long-run	0.9208 (27.2%)	0.9742 (16%)	0.9832 (6.8%)	0.9839 (4.6%)

Note: The figures represent ratios of RMSEs, the numerators of which belong to models with an alternative long-run term ($y_{t-1} - x_{t-1-1/2}^{(12)}$ -cases) or models without a long-run term. The denominators are always the RMSEs corresponding to the model including the DGP dynamic MF cointegrating relationship $y_{t-1} - x_{t-1}^{(12)}$, i.e., the benchmark case. The percentages in brackets are the rejection frequencies of Diebold-Mariano tests at the 5% level compared to the aforementioned benchmark. The short-run dynamics are modeled by MIDAS and for each model under consideration the number of short-run variables is adapted consistent with the methodology of Section 2.2. The DGP underlying this table is of a high-frequency-kind, i.e., both x and y are generated as high-frequency variables, where y is aggregated by Point-in-Time Sampling afterwards.

driven weight determination whereas Average sampling imposes an erroneous flat weighting scheme. If $y_t^{(m)}$ is instead temporally aggregated by Average sampling, the results are qualitatively identical, with the obvious exception that the roles of Point-in-Time and Average sampled regressors are interchanged. So, if the regressand aggregation scheme is unknown, applying an erroneous classical aggregation method may degrade the forecasting performance immensely whereas MIDAS is able to mimic the respective aggregation scheme quite well and therefore provides a robust, yet parsimonious solution to the problem.

2.4 Application

As an illustration, we consider the case of forecasting U.S. inflation. All variables have been downloaded from the Federal Reserve Bank of St. Louis at <http://research.stlouisfed.org/fred2/> and cover the period from January 1981 to December 2010. The dependent variable denoted CPI_t is the seasonally adjusted U.S. monthly consumer price index for all urban consumers (all items). Various weekly indicators were considered, but we only report the results obtained for the models with the best fit. Note that interest rates of different maturities, the weighted exchange index over major currencies as well as the stock of money M2 did not improve over the regression presented here. The selected explanatory variables are (i) the U.S. regular all formulations gasoline price (GAS_t) and (ii) the S&P500 index (denoted SP_t). Although gasoline prices have been used before in other studies to model or forecast inflation (see, e.g., de Brouwer and Ericsson, 1998), our interest in analyzing this variable also stems from the fact that in the beginning of 2012 economic news agencies such as CNNMoney² or ActionForex³ have reported increasing gasoline prices to boost inflation making an analysis over a longer period appealing. The role of asset prices has been studied and reviewed in Stock and Watson (2003) who point out that several asset prices help in predicting inflation, motivating our decision to include the S&P500 index rather than single asset prices. We estimate the model on the period ranging from January 2000 to November 2010, amounting to 131 observations. We have focused on that period in order to have a price index series that is well described as $I(1)$ and, hence, to avoid the discussion about the potential $I(2)$ -ness of that series over a longer span.

In contrast to Section 2.3, the cointegrating vector is unknown and hence estimated. The presence of a dynamic MF cointegrating relationship between $\ln(CPI_t)$ and the two regressors is investigated using a static regression where we consider a window of three weeks before the end-of-period X -observations. We thus estimate static cointegrating regressions on $GAS_{t-1-i/4}^{(4)}$ and $SP_{t-1-i/4}^{(4)}$ with $i = 0, 1, 2$ or 3 because $m = 4$ in this month/week setting. This implies that we assume the explanatory variables to enter the regression with the same displacement. This restriction may easily be relaxed but is imposed here to save on computation burden.

²see http://money.cnn.com/2012/02/17/news/economy/cpi_inflation_gas_prices/index.htm

³see <http://www.actionforex.com/analysis/daily-forex-fundamentals/higher-gasoline-prices-boost-headline-inflation-in-february-20120316161615/>

To test whether cointegration is present for all candidate error correction terms, the results of the ADF and Johansen's Trace tests (Johansen, 1991) are reported in Table 2.5 below. Johansen's Trace Test is computed while allowing for a linear deterministic trend in the data and an intercept, but no trend in the cointegrating equation(s). Furthermore, the optimal lag length in a VAR is 2 based on SIC such that 1 lag of differences is chosen for a VECM.

Table 2.5: Testing for the presence of cointegration

Dyn. MF cointegrating relationship	ADF(1)	Johansen's Trace Test	
		for no CE(s)	for at most 1 CE
$y_{t-1} - \gamma_1 GAS_{t-1}^{(4)} - \gamma_2 SP_{t-1}^{(4)}$	-4.61	0.0101	0.4755
$y_{t-1} - \gamma_1 GAS_{t-5/4}^{(4)} - \gamma_2 SP_{t-5/4}^{(4)}$	-4.43	0.0199	0.5114
$y_{t-1} - \gamma_1 GAS_{t-3/2}^{(4)} - \gamma_2 SP_{t-3/2}^{(4)}$	-4.70	0.0047	0.5323
$y_{t-1} - \gamma_1 GAS_{t-7/4}^{(4)} - \gamma_2 SP_{t-7/4}^{(4)}$	-4.84	0.0040	0.6115

Note: The first column illustrates the composition of the EC-term in line with the notation used in the paper. ADF(1)-test statistics and the p-values of Johansen's Trace tests (for no or at most 1 cointegrating equation) are given in columns 2 and 3-4, respectively.

The ADF and Johansen's trace tests reject the null of no-cointegration for all candidate dynamic MF cointegrating relationships. With respect to the regressor variables entering the error correction term we choose the $y_{t-1} - \gamma_1 GAS_{t-1}^{(4)} - \gamma_2 SP_{t-1}^{(4)}$ -case due to its simplicity. Estimating the long-run term by fully modified least squares yields (where lower case letters represent logarithms of the respective variables)

$$\widehat{ecm}_{t-1} = cpi_{t-1} - \underset{(0.201685)}{6.274969} + \underset{(0.029048)}{0.170431} sp_{t-1}^{(4)} - \underset{(0.015342)}{0.275665} gas_{t-1}^{(4)}.$$

Every coefficient is individually significantly different from zero at any sensible significance level (FMOLS standard errors in brackets). Plugging this into an ECM we can compare the forecasts of different modelings for the last 23 months. Note that when forecasting the value of y_{t+1} we employ the realized values of the regressors in period $t + 1$. In other words, we assume that the value of the low-frequency variables appears after the ones of the regressors, which is a reasonable assumption for most macroeconomic applications. Additionally, we assume that we compute the forecast at the end of the t -period, i.e., we actually nowcast y_{t+1} . If a fore- or nowcast should be computed at the beginning or during the t -period, we propose to forecast the values of the regressors using simple ARIMA models.

The RMSEs of the models excluding a dynamic MF cointegrating relationship relative to their EC-term-including counterparts, i.e., MIDAS without long-run term relative to MIDAS with long-run term, Unrestricted without long-run term relative to Unrestricted with long-run term and so on, are 1.1002, 1.1398, 1.629 and 1.15 for Point-in-Time Sampling, Average sampling, the Unrestricted approach and MIDAS, respectively. The Diebold-Mariano test statistics associated with these ratios are 3.186, 3.32, 4.544 and 2.851, respectively. Note that $p_l = 2$ and $q_l = 1$ such that there is one lagged difference of cpi_t as well as 8 short-run terms per regressor in the ECM. Also, several weight functions per low-frequency period were estimated for the MIDAS regressions, i.e., we estimated the regressions in (2.9) with $q^* = 1$, amounting to 4 weight functions (2 per regressor) in total.

As expected, the RMSEs of the models with a dynamic MF cointegrating relationship are lower than those without. Since the differences are statistically significant (at the 5% level), including a long-run term leads to better forecasting performances. Average sampling performs best among the four methods applied here. Relying on (modified) Diebold-Mariano test we find that, when including an EC-term, the forecast accuracies of the four methods are not significantly different from each other. The exception here is Point-in-Time Sampling which yields a significantly larger RMSE than the other approaches.

Foroni et al. (2012) compare, for different frequency combinations, i.e., different m in our notation, the forecast (or nowcast) accuracy of a model where the high-frequency variables are estimated by least squares (unrestrictedly) with a model where a nonlinear MIDAS structure is imposed. They find that for smaller m , e.g., $m = 3$, the unrestricted method dominates traditional MIDAS models whereas for larger m , e.g., $m = 12$ or especially $m = 60$, MIDAS clearly outperforms the unrestricted approach. This could explain why MIDAS does not significantly outperform the unrestricted method here. Note also that the estimation period consists of 109 observations which seems too much for the unrestricted approach to run into parameter proliferation problems. Although MIDAS is not significantly outperforming the other methods, it easily enables us to extend the above analysis. Adding more high-frequency regressors and/or more lags comes at no cost in terms of parsimony whereas the unrestricted method would quickly suffer from the aforementioned problem of parameter proliferation while Average sampling would ignore high-frequency information present in the new variables. Hence, MIDAS remains a robust and parsimonious approach in such real-life situations.

2.5 Conclusion

In this paper, a mixed-frequency error correction model was proposed in order to model non-stationary mixed-frequency variables that are possibly cointegrated. Starting with a mixed-frequency DGP, it was shown that the choice of variables entering the error-correction term has an impact on the short-run dynamics of the mixed-frequency ECM at the model representation level. In particular, we allowed not only the end-of-period observation of a high-frequency regressor to enter the disequilibrium error, but its 'neighboring' high-frequency observations to do so instead. We called these dynamic MF cointegrating relationships and showed in detail how the choice of which relationship is considered affects the short-run dynamics structure of the ECM.

The forecasting performances of four approaches, an unrestricted short-run, a MIDAS regression, Point-in-Time and Average sampling, with and without a dynamic MF cointegrating relationship were assessed in Monte Carlo simulations. It is found that ignoring a cointegrating relationship lowers the forecasting performance of all approaches if the variables are cointegrated, thus extending the results of Clements and Hendry (1998) to the mixed-frequency case. Furthermore, choosing an alternative dynamic MF cointegrating relationship instead of the one set in the DGP has no effect on the forecast accuracy if the number of short-run variables is adequately adapted. If, however, the short-run dynamics structure is misspecified, the forecasting performance of the unrestricted as well as the MIDAS approach may deteriorate. To be more specific, if too few short-run variables are included in the regression, the forecast accuracies of both approaches worsen due to dynamic misspecification. However, if too many short-run dynamics terms are included, the unrestricted approach suffers from parameter proliferation particularly for samples of small and moderate sizes. In contrast, MIDAS produces equally good forecasts as in the correctly specified scenario. The reason is that its flexible data-driven weight determination ensures superfluous short-run variables to get a zero weight.

The results were found to be robust to a high-frequency DGP in which both variables are sampled at a common high-frequency and in which the ultimately observed low-frequency variable is obtained as an aggregate version of its high-frequency observations in the respective t -period. We pointed out that MIDAS is able to mimic the underlying regressand aggregation scheme quite well and therefore lends itself as a robust, yet parsimonious method to deal with a mixed-

frequency data set. A small empirical application, in which U.S. monthly inflation is forecasted employing several weekly variables, was used to illustrate our results.

MIDAS' ability to deal with many regressors without the need for large sample sizes, while still preserving information present in any eventually added high-frequency regressor, make it an attractive tool researchers should consider when forecasting mixed-frequency data.

Chapter 3

Combining Distributions of Real-Time Forecasts: An Application to U.S. Growth

Having compared the previously developed ECM-MIDAS models with a few temporal aggregation based models in terms of their forecasting performances, let us investigate how these approaches fare in a real-time data set, i.e., when variables are subject to revisions. In particular, we extend the repeated observations forecasting (ROF hereafter) analysis of Stark and Croushore (2002) to allow for regressors of possibly higher sampling frequencies than the regressand. For the U.S. GNP quarterly growth rate, we compare the forecasting performances of an AR model with the mixed-frequency models (based on temporal aggregation and MIDAS) of the previous chapter.

Using the additional dimension provided by different data releases, we compute several forecasts for a given calendar date and subsequently approximate the corresponding distribution of forecasts by a continuous density. Scoring rules are then employed to construct combinations of them and analyze the composition and evolution of the implied weights over time. Using this approach, we not only investigate the sensitivity of model selection to the choice of which data release to consider, but also illustrate how to incorporate revision process information into real-time studies. As a consequence of these analyses, we introduce a new weight-

ing scheme that summarizes information contained in the revision process of the variables under consideration.¹

3.1 Introduction

In economics, especially macroeconomics, many variables are subject to revisions. Take, for example, the quarterly growth rate of the U.S. real gross national product (GNP hereafter), for which a revised value is published each month.² Figure 3.1 displays the evolution of U.S. real GNP growth for three dates, 1986Q3, 1991Q4 and 1996Q1.³ Note that even after 20 quarters there are still a lot of revisions taking place in the growth rates. Apart from many large movements being caused by so-called major revisions, or comprehensive benchmarks (Jacobs et al., 2013), also many smaller adjustments are found for all three dates, even after a large amount of data releases have already occurred.

However, instead of making use of the entire history of a revised variable, i.e., instead of using a real-time data set, researchers often solely rely on the latest-available data set, i.e., the values represented by the right end points of the lines in Figure 3.1. In doing so, a data set is used that is different from the one researchers could have used in real time (Croushore and Stark, 2001, Croushore, 2011). This can have a drastic impact on the development of economic specifications, though. Stark and Croushore (2002) investigate how model selection, i.e., the lag order in autoregressive models, changes as different data releases, or vintages, are considered. Also, measures of forecast accuracy may be deceptively lower when latest-available data are dealt with than when real-time data are used. This makes many of the 'horse races' (Diebold, 2012), in which different models compete for the lowest root mean squared error (RMSE), questionable.

The contribution of this paper is threefold. First, in view of the previous argument, we analyze whether model selection outcomes based on latest-available data are robust to the vintage of data that is employed. To this end we extend

¹This chapter is based on the paper Götz et al. (2012). The authors thank Franz Palm, Gianni Amisano, three anonymous referees and participants of the CIRANO workshop on Data Revision in Macroeconomic Forecasting and Policy in Montreal 2011, the 66th European Meeting of the Econometric Society in Malaga 2012, the 2012 Conference on Real-Time Data Analysis, Methods, and Applications in Philadelphia, the 7th Netherlands Econometric Study Group Meeting in Groningen 2012 and the 34th International Symposium on Forecasting in Rotterdam 2014 for useful suggestions and comments on earlier versions of the paper.

²Data releases for the nowadays more commonly modeled gross domestic product are available much later than for GNP, which is why we opted for the latter in this work.

³The data for this chapter were extracted from the databank ALFRED[®] of the Federal Reserve Bank of St. Louis: <http://alfred.stlouis.org>.

Figure 3.1: Revised GNP growth rate along monthly data releases

Note: The GNP growth rates in 1986Q3, 1991Q4 and 1996Q1 are displayed for all applicable data releases until 2010Q4. Data releases progress on the x-axis, whereas values of the GNP growth rates are represented on the y-axis.

the repeated observation forecasting approach proposed by Stark and Croushore (2002) to an autoregressive distributed lag (ADL hereafter) setting. Second, the construction and evaluation of densities of vintage-specific forecasts allow us to measure the sensitivity of these predictions to different data releases. By analyzing this issue, we consider and measure data uncertainty, a feature not often dealt with in the literature, contrary to parameter and model uncertainty. Third, using combinations of the aforementioned densities, we propose a weighting scheme that incorporates information contained in the revision process of the variables under consideration. These weights can then be used for fore- or nowcasting in real time.⁴

The rest of the paper is organized as follows. Section 3.2 contains the notation for a mixed-frequency real-time data set, a description of the repeated observation forecasting approach used in our particular framework and the methodology on how to construct densities of forecasts. Subsequently, the aforementioned tech-

⁴Note that we present an application to the U.S. growth rate involving a very restricted set of models (listed in Appendix 3.A) and variables. Nevertheless, the analysis itself and the corresponding main conclusions can be applied and generalized to any other viable setting, i.e., one in which some (or all) variables are subject to revisions.

niques are illustrated with empirical data. In Section 3.3 we describe how to combine densities of forecasts based on scoring rules, whereby we discuss how to compute the associated weight for each model under consideration. Again, we continue by applying these methods to actual data. In Section 3.4 we show how to use the previous techniques in a real time study. Section 3.5 concludes.

To avoid confusion from the outset, a 'real-time data set' describes a data set that contains variables, which are subject to revisions. The term 'real time' alone, however, refers to using, at each moment in time, only data that has been available at that moment in time. In other words, one may consider only data that is released at or prior to the moment of time under consideration (Clements and Galvão, 2011).

3.2 Extending the ROF Approach

3.2.1 Notation

We observe regularly spaced monthly vintages available for T and $T \times m_j$ ($j = 1, \dots, K$) observations of a low-frequency, here quarterly, variable y and K high-frequency variables $x_1, \dots, x_K$. A vintage is defined to be a moment at which a series is published, for instance 20th of June 2011. Let us denote the first and last available vintages by V_F and V_L , respectively. The index t represents the low-frequency series and runs from 1 to T . For a high-frequency regressor x_j , the number of high-frequency observations per t -period equals m_j . Likewise, the number of data releases within each low-frequency period can be defined as m_v . Due to the low frequency being quarterly and vintages occurring on a monthly basis, we will have $m_v = 3$ here.

Assuming a publication lag of one month for both, the regressors and the regressand,⁵ $y_{t-1}^v, v > t - 1$, denotes the $(t - 1)$ -value of y published in vintage v . The notation becomes more evolved for the high-frequency variables, where the j -index is suppressed now for explanatory convenience. We use $x_{t-i/m}^v, v > t - i/m$, where i selects the corresponding high-frequency observation. To be more precise, if $i = 0$, x_t^v represents the value of x at the end of the low-frequency period t (in

⁵Indeed, for many macroeconomic variables, e.g., GNP, the first figure in quarter t becomes available the first month of quarter $t + 1$. Likewise, we have to wait for the coming month to get a first estimate of a regressor such as the industrial production index in the current month. For regressors that are not revised no publication delay exists in practice. Nevertheless, in order to keep the setup generic, we treat such regressors as being revised over time as well and pretend a publication lag of one month to exist here as well.

vintage v). Consequently, the first high-frequency observation in period t is stored in $x_{t-(m-1)/m}^v$. Trivially, $x_{t-m/m}^v \equiv x_{t-1}^v$. This notation has become standard in the mixed-frequency literature (Clements and Galvão, 2008, 2009 or Götz et al., 2013, 2014) and is here extended to a real-time data set.⁶ Finally, with vintages potentially appearing on a high-frequency basis, v above can be a fraction similar to the subscript of the high-frequency regressors.

Considering a quarter/month-example, $x_{t-1-1/3}^{t-2/3} = x_{t-4/3}^{t-2/3}$ denotes the figure published in the first month of quarter t of the value of x two months earlier, i.e., in the middle month of quarter $t-1$. Table 3.1 illustrates the notation for a quarter/month-example, i.e., for $m=3$.

3.2.2 Repeated Observation Forecasting

A common practice in empirical work is to use the latest-available data set to evaluate forecasts. This means that at period T one collects the historical time series for y_{t-1}^T where $t=2, \dots, T$. Subsequently, a one-step-ahead point forecast for $\hat{y}_T$ may be obtained. Stark and Croushore (2002) proposed an interesting graphical tool, which they call repeated observations forecasting, to evaluate whether the forecast accuracy is sensitive to the vintage chosen. For our variables y and X that are available at different frequencies, the approach works as follows. Let us take a particular low-frequency period t^* , where $\lceil V_F - 1 \rceil \leq t^* \leq \lceil V_L - 1 \rceil$, because no data releases are available prior to V_F , and consider the historical series $y_{t-1}^v, t=2, \dots, t^*$, and X_{t-i/m_j}^v with $t=1, \dots, t^*, i=0, \dots, m_j-1$,⁷ for a set of monthly vintages $v=t^*+1/m_v, \dots, V$. Note that not necessarily $V=V_L$ if one wants to base the ROF analysis on, e.g., a balanced or lower number of vintages. Given a certain model, a one-step-ahead point forecast for $\hat{y}_{t^*}^v$ can be computed for each vintage v such that we end up with a sequence of $V^* \equiv m_v(V-t^*)$, in this case one-step-ahead, forecasts for the same point.

Stark and Croushore (2002) report these forecasts using box-plot-like graphs. In contrast to their study we approximate the corresponding distribution of forecasts by a continuous density. Combining these model-specific densities implies weights for each model, that allow researchers

⁶High-frequency variables are usually labeled as $x_{t-i/m}^{(m)}$ in the mixed-frequency literature. As the superscript is, however, already reserved for the data releases, we deviate from the standard notation slightly in this respect.

⁷Here we assume that the value of X is published before the one of y , which is reasonable for most macroeconomic mixed-frequency data sets.

Table 3.1: Revision triangle for mixed-frequency data

$\frac{Vintage}{Time}$	...	$t-2-1/3$	$t-2$	$t-1-2/3$	$t-1-1/3$	$t-1$	$t-2/3$	$t-1/3$	t
.	.	.	.	.	.	.	.	.	.
$t-2-2/3$		$x_{t-2-1/3}^{t-2-1/3}$	$x_{t-2-2/3}^{t-2}$	$x_{t-2-2/3}^{t-1-2/3}$	$x_{t-2-1/3}^{t-1-1/3}$	$x_{t-2-2/3}^{t-1}$	$x_{t-2/3}^{t-2/3}$	$x_{t-1/3}^{t-1/3}$	$x_{t-2-2/3}^{t-2/3}$
$t-2-1/3$		$x_{t-2-2/3}^{t-2-1/3}$	$x_{t-2-1/3}^{t-2}$	$x_{t-2-1/3}^{t-1-2/3}$	$x_{t-2-1/3}^{t-1-1/3}$	$x_{t-2-1/3}^{t-1}$	$x_{t-2-1/3}^{t-2/3}$	$x_{t-1/3}^{t-1/3}$	$x_{t-2-1/3}^{t-2/3}$
$t-2$			$y_{t-2}^{t-1-2/3}$	$y_{t-2}^{t-1-2/3}; x_{t-2}^{t-1-2/3}$	$y_{t-2}^{t-1-1/3}; x_{t-2}^{t-1-1/3}$	y_{t-2}^{t-1}	$y_{t-2}^{t-2/3}; x_{t-2}^{t-2/3}$	$y_{t-2}^{t-1/3}; x_{t-2}^{t-1/3}$	$y_{t-2}^{t-2/3}; x_{t-2}^{t-2/3}$
$t-1-2/3$				$x_{t-1-2/3}^{t-1-2/3}$	$x_{t-1-2/3}^{t-1-1/3}$	$x_{t-1-2/3}^{t-1}$	$x_{t-1-2/3}^{t-2/3}$	$x_{t-1-2/3}^{t-1/3}$	$x_{t-1-2/3}^{t-2/3}$
$t-1-1/3$						x_{t-1}^{t-1}	$x_{t-1-1/3}^{t-2/3}$	$x_{t-1-1/3}^{t-1/3}$	$x_{t-1-1/3}^{t-2/3}$
$t-1$						x_{t-1}^{t-1}	$y_{t-1}^{t-2/3}; x_{t-1}^{t-2/3}$	$y_{t-1}^{t-1/3}; x_{t-1}^{t-1/3}$	$y_{t-1}^{t-2/3}; x_{t-1}^{t-2/3}$
$t-2/3$								$x_{t-2/3}^{t-2/3}$	$x_{t-2/3}^{t-2/3}$
$t-1/3$								$x_{t-1/3}^{t-1/3}$	$x_{t-1/3}^{t-1/3}$
t									

Note: A quarter/month-example ($m = 3$) of a mixed-frequency real-time data set is depicted. x is the high-frequency (monthly) variable, whereas y is the low-frequency (quarterly) variable. Both variables are updated on a monthly basis, i.e., $m_y = 3$. A publication delay of one month is assumed for both, x and y .

- (i) to assess whether model 'horse races' are vintage-sensitive, and
- (ii) to include information contained in the revision process into real-time studies.

Importantly, for (i) we set $V = V_L$ and thereby directly extend the analysis of Stark and Croushore (2002) by investigating the sensitivity of forecasts to the data release considered. For (ii), however, we fix $V = T^*$, where we denote by T^* the currently last available vintage. In other words, to mimic what a researcher could have done at a specific moment in time, we only consider data that is released no later than T^* , i.e., 'today' (Clements and Galvão, 2011). Note that in the analysis to come we only consider one-step-ahead forecasts, whereby an extension towards h -step-ahead (direct or indirect) forecasts is straightforwardly done.

Stark and Croushore (2002) discuss the issue of which observations to use as realizations, or 'actuals', in order to compare different methods graphically or to compute forecast errors. Depending on the choice of the realization, the outcome of an analysis may change substantially. Stark and Croushore (2002) consider three possibilities for the actual value, namely (a) the respective observation four quarters later, (b) the corresponding figure in the latest-available data set or (c) the first observation after a comprehensive benchmark has occurred. In this paper we propose a different set of observations as 'actuals', namely the realization in the respective vintage itself. To be more precise, the one-step-ahead forecast of y_{t^*} corresponding to vintage v , i.e., $\hat{y}_{t^*}^v$, is compared to $y_{t^*}^v$ and a forecast error is computed as $e_{t^*}^v = \hat{y}_{t^*}^v - y_{t^*}^v$. We refer to this set of 'actuals' as same-vintage realizations or same-vintage 'actuals'.

3.2.3 Constructing Densities of Forecasts

Given a model specification, we obtain a sequence of V^* one-step-ahead forecasts, i.e., $\hat{\mathbf{y}}_{t^*} = (\hat{y}_{t^*}^{t^*+1/m_v}, \dots, \hat{y}_{t^*}^{V})'$, using the ROF approach as described above. Remember that $V = V_L$, i.e., the last data release of our data set, or $V = T^*$, i.e., the last available vintage in a real-time analysis, depending on whether we analyze case (i) or (ii) above (see Sections 3.3.4 and 3.4, respectively). The distribution of these forecasts is then approximated by a continuous density, i.e., $\hat{f}_{t^*}$, via kernel density estimation techniques. To be more precise, assuming the V^* one-step-ahead forecasts to have some underlying unknown distribution f , the

kernel density estimator is

$$\hat{f}_{t^*}(y) = \frac{1}{(V^*)h} \sum_{i=t^*+1/m_v}^V K\left(\frac{y - \hat{y}_{t^*}^i}{h}\right), \quad (3.1)$$

where $K(\cdot)$ is the kernel function and h is the bandwidth parameter. In this paper, the Gaussian kernel, $K(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)$ is considered. The bandwidth is chosen of the order $\hat{\sigma}(V^*)^{-1/5}$, where $\hat{\sigma}$ is an estimator of the standard deviation.⁸

3.2.4 Application I

Let us illustrate the techniques introduced thus far using actual data. To this end, we consider the quarterly real gross national product (ref. GNPC96 and referred to as *GNP* hereafter), seasonally adjusted, as dependent, low-frequency, variable y . The series is observed from 1960Q1 ($t = 1$) until 2010Q3 ($T = 203$).⁹ For the regressors we consider, for the same time span, the monthly seasonally adjusted industrial production index (ref. INDPRO and labeled *IPI* hereafter) and the daily S&P 500 stock index (ref. SP500, *SP* hereafter). Consequently, $m_{IPI} = 3$ and $m_{SP} = 60$.¹⁰ Note that all variables are measured in logarithms. There are monthly data releases from July 1986 ($V_F = 107 - 2/3$) to December 2010 ($V_L = 204$), which implies $m_v = 3$.¹¹ *SP* is not revised at all such that along different vintages new data becomes available, but existing ones do not change. Note that one could easily generate daily vintages as well, we however facilitate the analysis by only considering monthly vintages for both regressors. A final remark concerns the fact that figures of both series are computed with reference to a base year. Due to changes in the base years, the values jump at the respective dates. Note that because we forecast growth rates of output, this is not an issue for our analysis.

⁸Note that the results are robust to the use of other kernels, e.g., the Parzen or Triangle kernel (see, for example, Greene, 2008).

⁹For this exercise we have extracted the series in February 2011

¹⁰In order to have the same number of working days per vintage we take the maximum amount of working days that is available in each month over the time period considered. This leads to 20 daily observations per month and, hence, 60 per quarter.

¹¹The vintages do not match exactly for the three series. The first mismatch comes from the fact that they are not published on the same day. We do not do anything about that. When there are two vintages for the same month (this happens for *IPI*), we take the vintage at the half of the month. The second problem is that there are, at different releases, missing vintages for the two regressors. To facilitate the computation we create missing data releases by assigning missing values their respective values in the previous vintage. The alternative would have been to delete uncommon vintages.

Before actually conducting the ROF analysis outlined in Section 3.2.2, let us give a short comment on the models we use for obtaining the forecasts, from which densities are constructed subsequently. We consider three models that contain high-frequency regressors and one autoregressive model that serves as a benchmark (Stark and Croushore, 2002). With respect to the former, two of the models temporally aggregate the high-frequency regressors prior to estimation, whereas the MI(xed) DA(ta) S(ampling), or MIDAS hereafter, regression model directly employs the high-frequency observations. For those three model specifications, we take into account the possibility of cointegration between the variables by considering them with and without a cointegrating relationship (the former contains an 'a' in the label, the latter a 'b'). This amounts to 7 models, labeled 1a, 1b, 2a, 2b, 3a, 3b and 4 in Appendix 3.A, where detailed descriptions of each specification are provided. Note that we abstract from other approaches such as factor models which present a popular choice when forecasting (or nowcasting) GNP growth rates (Giannone et al., 2008). Although the methodology presented in this paper can just as well be applied to any other model, also with a possibly large set of regressors as in Andreou et al. (2013), one has to obtain and clean a real-time data set for each regressor concerned, i.e., correct for possible mismatches or dealing with multiple vintages for one time period. This task may be very time-consuming or even impossible due to, e.g., non-existent data.

In order to analyze whether model comparisons are sensitive to the data release considered, let us conduct the aforementioned ROF analysis: We consider calendar dates $106 \leq t^* \leq 203$, corresponding to the range 1986Q2-2010Q3, and the historical series $GNP_{t-1}^v, t = 2, \dots, t^*, IPI_{t-i/3}^v$ with $t = 1, \dots, t^*, i = 0, 1, 2$, and $SP_{t-i/20}^v$ with $t = 1, \dots, t^*, i = 0, \dots, 19$ for a set of monthly vintages $v = t^* + 1/3, \dots, V_L$. For each model $i \in \{1a, 1b, 2a, 2b, 3a, 3b, 4\}$, a set of $V^* = 3(V_L - t^*)$ one-step-ahead point forecasts can be computed, i.e., $\hat{\mathbf{y}}_{t^*}^i = (\hat{y}_{t^*}^{t^*+1/3}, \dots, \hat{y}_{t^*}^{V_L})'$, which are used to construct densities of forecasts using (3.1).

Let us consider two individual dates $t^* = 107$ (1986Q3) and 128 (1991Q4) as illustrative examples.¹² Note that $V^* = 291$ and 228, respectively, for our example dates. Averaging (models 2a and 2b) is denoted by AV, Point-in-Time sampling (models 3a and 3b) by PIT and the same-vintage 'actuals' by D_Q. Whenever a long-run relationship is included, COINT is appearing in the label. The densities of forecasts produced by the respective seven models are plotted in

¹²In comparison to Figure 3.1 the outcomes for 1996Q1 are not presented to save on space, but are available upon request.

Figures 3.2 and 3.3. Additionally, histograms of the same-vintage realizations are displayed in the top right corners. At this stage, we ask the reader to disregard the density displayed by a solid black line, which we get back to at a later stage in the paper.

Figure 3.2: Densities of forecasts and same-vintage 'actuals'

Note: This figure graphs the densities of forecasts corresponding to the seven forecasting approaches under consideration (see Appendix 3.A). Values of the GNP growth rate (multiplied by 100) are displayed on the x-axis. The density displayed by a solid black line corresponds to the combination density with weights computed as in (3.4). Corresponding same-vintage 'actuals' are depicted in the histogram in the top right corner.

Just as their corresponding realizations (see Figure 3.1), forecasts of the GNP growth rates differ substantially depending on which vintage we consider. This becomes apparent from the wide range of data the densities cover. Furthermore, the forecasts of each model result in quite a different density, i.e., the GNP growth rate forecasts differ not only with respect to the data vintage (something we call 'data uncertainty'), but also the model (model uncertainty) under consideration. In the case of 1986Q3, for example, PIT.COINT produces a rather wide and flat density whereas the one of PIT is concentrated on a smaller region. Similar observations can be made in Figure 3.3.

Figure 3.3: Densities of forecasts and same-vintage 'actuals'

Note: See Figure 3.3

3.3 Sensitivity of Model 'Horse Races' to Vintages

3.3.1 Scoring Rules

To evaluate the densities of vintage-specific forecasts, we make use of scoring rules, an avenue often followed in this field of literature (see Amisano and Giacomini, 2007 or Diebold and Lopez, 1996). A scoring rule is a loss function whose value depends on the density, $\hat{f}_{t^*}$, and the vector of 'actuals', i.e., $\mathbf{y}_{t^*}^{\bar{V}} = (y_{t^*}^{t^*+1/m_v}, \dots, y_{t^*}^{\bar{V}})'$. Without going into too much detail, $\bar{V}$ is not necessarily equal to V , i.e., we do not have to plug all same-vintage 'actuals' corresponding to $\hat{\mathbf{y}}_{t^*}$ into the scoring function to evaluate our densities. A scoring rule is hence denoted as $S(\hat{f}_{t^*}, \mathbf{y}_{t^*}^{\bar{V}})$ and is constructed in such a way that a density mimicking the shape of the same-vintage 'actuals' better than another one is assigned a higher score.

To clarify, if the density for model A assigns a probability of (practically) zero to the occurrence of a particular same-vintage 'actual', i.e., if $\hat{f}_{t^*}^A(y_{t^*}^v) = 0$ for some v ($t^* + 1/m_v \leq v \leq \bar{V}$), whereas model B 's density assigns a positive probability to the same event, i.e., $\hat{f}_{t^*}^B(y_{t^*}^v) > 0$, then model B is assigned a higher score than model A for that particular same-vintage 'actual'. Indeed, $\hat{f}_{t^*}^A(y_{t^*}^v) = 0$ implies that

the one-step-ahead forecasts of y_{t^*} generated by model A do not assign a positive probability to the event that the actual value of y_{t^*} equals $y_{t^*}^v$. Consequently, model A is not able to forecast that same-vintage 'actual' and receives a lower score than a model that places positive probability on that outcome to occur, i.e., model B . Doing this for each same-vintage realization under consideration yields a vector of scores for each model. We will, as Amisano and Giacomini (2007), employ the logarithmic scoring rule

$$S_l(\hat{f}_{t^*}, \mathbf{y}_{t^*}^{\bar{v}}) = \log[\hat{f}_{t^*}(\mathbf{y}_{t^*}^{\bar{v}})]. \quad (3.2)$$

At this stage, let us argue why we opted for our same-vintage 'actuals' instead of, e.g., their final-vintage counterparts. If we had used one of the options described at the end of Section 3.2.2, the vector of 'actuals' would have consisted of only one element, e.g., $y_{t^*}^T$ with final-vintage 'actuals'. Employing same-vintage 'actuals' instead, the scoring vectors measure how well a model's forecasts resemble the revision process of a variable.

3.3.2 Combining Densities of Forecasts

It is well known in the literature for point (see, e.g., Timmermann, 2006) or density forecasting (Wallis, 2005) that combinations of forecasts may offer diversification gains making it attractive to combine the information present in the different forecasting models. Being equipped with densities produced by different models, a density combination approach is a natural avenue to follow (Wallis, 2005 or Aastveit et al., 2011).

Similar to many of the aforementioned studies we consider a so-called linear opinion pool, i.e., defining the combined density of forecasts as a linear combination of the individual densities:

$$\hat{f}_{t^*}^{Com}(y) = \sum_{i=1}^M \phi_i \hat{f}_{t^*}^i(y), \quad (3.3)$$

where M is the number of models to combine. When focusing on one date individually, we define the recursive log score weights of model $i = 1, \dots, M$ as

$$\phi_i = \frac{\exp(\bar{S}_{t^*}^i)}{\sum_{i=1}^M \exp(\bar{S}_{t^*}^i)}, \quad (3.4)$$

where,

$$\bar{S}_{t^*}^i = \frac{1}{\bar{V}^*} \sum_{v=t^*+1/m_v}^{\bar{V}} \log[\hat{f}_{t^*}^i(y_{t^*}^v)] = \frac{1}{\bar{V}^*} S_t(\hat{f}_{t^*}, \mathbf{y}_{t^*}^{\bar{V}})' \iota_{\bar{V}^*},$$

where $\bar{V}^* = m_v(\bar{V} - t^*)$ and ι_N denotes an $N \times 1$ vector of ones.

When interest lies not only in the weights for one date individually, but the models' ability to mimic the revision process over all or many consecutive dates in time, we propose to compute what we call revision-process-robust (RPR hereafter) weights. Note that these are adapted versions of the real-time optimal pool weights found in Amisano and Geweke (2013). As the name suggests, the latter can be interpreted as the optimal action a researcher can perform in real time. In our context, however, the RPR weights depend on the previously constructed densities of forecasts that may employ information of future data releases (if $V > T^*$), making the expression 'real-time' inappropriate. For each calendar date, the RPR weights summarize the extent to which the various models are able to resemble the revision process up until the date under consideration. In particular, the RPR weights at date t^* maximize the following criterion:

$$\mathbf{w}_{t^*}^{RPR} = \arg \max_{\mathbf{w}_{t^*}} \sum_{j=\lceil V_F - 1 \rceil}^{t^*} \log \left[\sum_{i=1}^M w_{i,t^*} \exp(\bar{S}_j^i) \right], \quad (3.5)$$

where $\mathbf{w}_{t^*}$ is bound to be in the M -dimensional unit simplex. Note that the weights at any date t^* depend on the average scores of all previous dates (remember that $\lceil V_F - 1 \rceil$ is the first calendar date to consider). Consequently, they take into account how well the models resembled the revision process of previous dates. Stated differently, for each calendar date they find the optimal combination of forecasting models based on how they mimicked the revision process thus far.

3.3.3 Remark on Data Uncertainty

Until now we have almost exclusively used the term 'density of forecasts' and not 'forecast density'. The reason is, of course, that our densities are merely a nonparametric way to approximate the distribution of the one-step-ahead forecasts corresponding to each vintage. These are no density forecasts in the sense of Tay and Wallis (2000) in particular or the forecasting literature in general.

Consequently, we are only focusing on point forecasts itself and do not take parameter uncertainty into account. We, however, obtain a measure of data uncertainty, i.e., how much the revised observations differ from an underlying 'true' or final value. Let us introduce data uncertainty by setting $y_t = y_t^v - \mu_t^v$ with $\mu_t^v \sim N(0, \sigma_\mu^2)$ for all t . Note that we are not concerned with the news or noise issue (Mankiw and Shapiro, 1986 or Jacobs and van Norden, 2011), but merely claim that the true underlying value of y_t differs (potentially a lot) from the one observed at vintage v . Using our nonparametric distribution approximations we get an idea about the sensitivity of the point forecasts to different vintages and thereby a measure of data uncertainty.

An alternative approach would be to construct a forecast density for each vintage and subsequently combine these densities to deal with data uncertainty. This allows the researcher to analyze certain regions of the densities, not only their means as we do. We, however, leave such an approach for further research.

3.3.4 Application II

Let us return to our empirical analysis, which we started in Section 3.2.4. After having constructed densities of forecast for each model under consideration, we combine these densities using (3.3) and the weights in (3.4) or (3.5), depending on whether we look at, respectively, one calendar date individually or all of them consecutively. The scores used to evaluate the densities (and compute the corresponding weights) are based on all same-vintage 'actuals' that are at our disposal, i.e., $\bar{V} = V_L$ such that $S_l(\hat{f}_{t^*}, \mathbf{y}_{t^*}^{\bar{V}}) = S_l(\hat{f}_{t^*}, \mathbf{y}_{t^*}^{V_L})$ in (3.2). Figure 3.4 illustrates the entire approach.

Before combining the model-specific densities for all calendar dates, let us focus on our two example dates, 1986Q3 and 1991Q1. Note that the weights in (3.4) are used to combine the densities in these cases, because we deal with individual calendar dates. The densities graphed by solid black lines in Figures 3.2 and 3.3 represent the corresponding combination densities. While each individual model-based density does a relatively poor job of mimicking the shape of the respective realizations distribution, the combination densities replicate their shapes quite well: In the case of 1986Q3 it mimics the tri-modality of the histograms and for 1991Q4 it detects the minimum at 0.005 as well as the maximum at roughly 0.0065 almost perfectly. As far as the weight composition is concerned, for $t^* = 107$ (1986Q3) the weights turn out to be 0.13, 0, 0.74, 0, 0.14, 0 and 0 for PIT.COINT,

Figure 3.4: The ROF approach illustrated

Note: The graph shows that for each date t^* under consideration a forecast is computed for each vintage available. Out of these forecasts (depicted as dark grey bars) densities are constructed using (3.1). These densities are evaluated using scores that are based on all same-vintage 'actuals' available (depicted as light grey bars). Note that the dark and light grey bars cover the same areas for each t^* visualizing that construction and evaluation of the densities of forecasts are based on all vintages at our disposal.

PIT, AV_COINT, AV, MIDAS_COINT, MIDAS and AR, respectively, whereas for $t^* = 128$ (1991Q4) they are 0.19, 0.02, 0.64, 0.13, 0.01, 0.01 and 0, respectively.

Note that the AR-model is assigned a zero weight for both dates. It turns out that, when considering all calendar dates, the autoregressive model is very often outperformed by a majority of the competing approaches. Hence, adding explanatory variables seems to improve the forecasting performance across different vintages. This is mainly due to the weakness of the AR model not to incorporate mid-quarter information. As an example, *GNP* grew from 2000Q3 until 2001Q2 followed by a decline in 2001Q3. The autoregressive model, however, is not able to predict this decline whereas other models, taking into account mid-quarter information, are better able to do so. The relatively poor performance of

the AR-model is, of course, likely to change for multi-step-ahead forecasts because regressors would have to be forecasted as well.

So far, we have merely looked at two individual dates. Let us inspect how the weight distribution over our seven models evolves over all calendar dates by computing the RPR weights in (3.5). Figure 3.5 graphs the evolution of the RPR weights associated with our seven models.

It turns out that all models except AV possess non-zero RPR weights. It is striking that while the AR-model had a weight of up to 33% assigned in the beginning of the analysis, its weight quickly dropped to 5 to 10% where it remained until the end. The model with the largest average weight over the whole period is PIT_COINT (45%), followed by MIDAS_COINT (15%), MIDAS and AV_COINT (both 12%). Interestingly, MIDAS_COINT has a much larger average weight in the first half of the graph than in the second half, whereas the opposite is true for the MIDAS approach without a cointegrating term.

Note that the RPR weights graphed in Figure 3.5 enable us to rank the models under consideration. More importantly, such a ranking would take into account how well the each model's forecasts resembled the revision process of the variable of interest. In order to investigate whether model rankings are sensitive to the vintage chosen, we should compare the ranking implied by the RPR weights in Figure 3.5 with one that is based on final-vintage data only. To this end, we conduct the following forecasting exercise: For each model and using only data from the latest-available vintage, we compute 20 one-step-ahead out-of-sample forecasts of the U.S. growth rate for two different time periods, (1) 1995Q1-1999Q4 and (2) 2005Q4-2010Q3. Subsequently, the models are ranked based on their RMSEs. The results are summarized in Table 3.2.

To obtain the ranking implied by the RPR weights in Figure 3.5, we rely on the average RPR weight of each model during the periods in question, (1) 1995Q1-1999Q4 and (2) 2005Q4-2010Q3. The outcomes are presented in Table 3.3. It turns out that several models appear to result in much better forecasts when only final-vintage data instead of all data releases are used, i.e., their ranking implied by the RPR weights is lower than it was in Table 3.2. Examples are MIDAS_COINT and AR for period (1) and AV for period (2). Conversely, some models get assigned a deceptively low ranking based on latest-available data only, i.e., their average RPR weight suggests a higher ranking than it was in Table 3.2 (see PIT_COINT and AV_COINT for period (1) or MIDAS for period (2)).

Figure 3.5: RPR weights for all calendar dates

Note: The figure displays the RPR weights assigned to the seven models for all the calendar dates in the sample. The different dates are represented on the horizontal axis, whereas the weight itself is plotted on the vertical axis.

Table 3.2: RMSEs and model ranks using final-vintage data

Model	(1) 1995Q1 - 1999Q4		(2) 2005Q4 - 2010Q3	
	RMSE	Rank	RMSE	Rank
PIT_COINT	0.0205	4	0.0219	1
PIT	0.0206	5	0.0230	4
AV_COINT	0.0233	7	0.0220	2
AV	0.0232	6	0.0229	3
MIDAS_COINT	0.0182	1	0.0275	6
MIDAS	0.0185	2	0.0275	6
AR	0.0193	3	0.0377	7

Note: The figures represent the RMSEs corresponding to a forecasting exercise conducted using only the final data release. 20 one-step-ahead out-of-sample forecasts of U.S. GNP growth were computed using seven different models for the time periods (1) 1995Q1-1999Q4 and (2) 2005Q4-2010Q3. Subsequently, the RMSEs and associated ranks were computed.

Table 3.3: Model ranks implied by RPR weights

Model	(1) 1995Q1 - 1999Q4		(2) 2005Q4 - 2010Q3	
	<i>RPR weight</i>	Rank	<i>RPR weight</i>	Rank
PIT_COINT	0.498	1	0.329	1
PIT	0.086	5	0.113	4
AV_COINT	0.093	3	0.2	2
AV	0.001	7	0.001	7
MIDAS_COINT	0.092	4	0.103	5
MIDAS	0.159	2	0.194	3
AR	0.071	6	0.06	6

Note: The figures represent the average RPR weights (see Figure 3.5) for the time periods (1) 1995Q1-1999Q4 and (2) 2005Q4-2010Q3 and the associated ranks of seven different models.

Hence, model 'horse races' are sensitive to the data vintage employed. Tables 3.2 and 3.3 illustrate that the ranking of several models based on RMSEs computed using only final-vintage data differs from the ranking implied by the RPR weights, which are based on all data releases available to us. To be more precise, using only latest-available data, some models seem to possess a better (or worse) forecasting performance than when their forecast accuracy is evaluated also using past data releases. In this sense, our results extend the work of Stark and Croushore (2002) who concluded for autoregressive models that "measures of forecast error, such as

root-mean-squared error or mean absolute error, can be deceptively lower when using latest-available data [...]” (p. 507).¹³

3.4 Incorporating Revision Process Information into Real-Time Studies

The results presented thus far do not provide guidance as to what a researcher should do in real time. Let us assume that we are at month T^* such that, coherent with Section 3.2.2, the most up-to-date data set is released at $v = T^*$. Now, we consider calendar dates t^* , where $(1986Q2 =) 106 \leq t^* \leq \lceil T^* - 1 \rceil$, and again deal with the series $GNP_{t-1}^v, t = 2, \dots, t^*, IPI_{t-i/3}^v$ with $t = 1, \dots, t^*, i = 0, 1, 2$, and $SP_{t-i/20}^v$ with $t = 1, \dots, t^*, i = 0, \dots, 19$ for vintages $v = t^* + 1/3, \dots, T^*$, i.e., $V = T^*$. Similar to before, we compute $V^* = 3(T^* - t^*)$ one-step-ahead forecasts for each model in order to construct densities of forecasts using (3.1). The scores used to evaluate and combine the model-specific densities of forecasts are, however, only based on maximally the first 12 (equivalent to 1 year) same-vintage ‘actuals’. In other words, $\bar{V} = \min\{t^* + 4, T^*\}$ such that the vector of ‘actuals’ is written as

$$y_{t^*}^{\bar{V}} = \begin{cases} (y_{t^*}^{t^*+1/3}, \dots, y_{t^*}^{t^*+4})' & \text{if } t^* + 4 < T^* \\ (y_{t^*}^{t^*+1/3}, \dots, y_{t^*}^{T^*})' & \text{else.} \end{cases}$$

To clarify, we construct our densities of forecasts using as many data releases as possible in order to deal with a reliable kernel density estimation. However, when it comes to evaluating and combining the model-based densities we only check how well the models predict the first 12 releases of the variable of interest, i.e., when data is still revised a lot. Figure 3.6 illustrates the approach just outlined.

In essence, we redo the ROF analysis in the previous subsection restricted to the real-time data set available at T^* and to using at most 12 same-vintage ‘actuals’. This implies that as time progresses we obtain a (possibly different) set of RPR weights, i.e., for each T^* we get a graph in the same fashion as the

¹³When considering more recent dates, the construction of our densities is based on only a few vintages, i.e., V^* is rather small. However, the RPR weights are based on the average scores of all previous time periods, and are thereby smoothing the weights corresponding to individual dates. In other words, the effect of basing the densities corresponding to more recent dates on only a small number of vintages is mitigated by constructing the RPR weights as cumulated weighted averages of the individual weights in (3.4). If one still suspects negative effects of the aforementioned decreased number of data releases, one could either forecast weights using a VAR model or forecast revisions as done in Clements and Galvão (2011).

Figure 3.6: The approach for real-time studies illustrated

Note: The densities are evaluated using scores that are based on maximally the first 12 same-vintage 'actuals' (depicted as light grey bars). Hence, the evaluation and combination of model-specific densities is based on their forecasting performance with respect to the first 12 releases of the variable of interest, i.e., when data is revised a lot. For the rest, see Figure 3.4.

one in Figure 3.5. Suppose a researcher having obtained a new data release at moment T^* , which provides her or him with a first or revised estimate of $y_{\lceil T^*-1 \rceil}$ (depending on whether T^* is the first month of quarter $\lceil T^* \rceil$ or not), aims at using the newly published data to now- or forecast the next 4 quarters of y , i.e., $\hat{y}_{\lceil T^* \rceil}^{T^*}$ (nowcast), $\hat{y}_{\lceil T^*+1 \rceil}^{T^*}$, $\hat{y}_{\lceil T^*+2 \rceil}^{T^*}$ and $\hat{y}_{\lceil T^*+3 \rceil}^{T^*}$. The question arises how she or he can make use of the corresponding RPR weights in order to incorporate revision process information (up to month T^*) into these now- and forecasts.

We propose the following: As before, we estimate our models (1a)-(4)¹⁴ on the period from $t = 2$ to $\lceil T^* - 1 \rceil$. Note that we only employ vintage $v = T^*$ though. Subsequently, for each model we compute 4 one-step-ahead out-of-

¹⁴Remember that models (1a)-(3b) are nested in equation (3.7) with $\alpha(L) = 1 - \alpha L$, $K_l = 1$ and two regressors (see Appendix 3.A).

sample dynamic forecasts to obtain $\hat{y}_{[T^*+h]}^{T^*}$, $h = 0, 1, 2, 3$.¹⁵ For each h , these model-specific forecasts are then combined using the RPR weights corresponding to $[T^* - 1]$. Indeed, as stated at the end of Section 3.3.2, the RPR weights summarize the extent to which our various models were able to resemble the revision process up until the date under consideration. Furthermore, as became evident from Figure 3.5, the RPR weights stabilize along time, which, together with the previous argument, makes it reasonable to use the latest-available RPR weights to combine those forecasts.

As an illustration, let us consider a researcher conducting the aforementioned real-time analysis from July 1987 ($T^* = V_F = 107 - 2/3$) until December 2009 ($T^* = 200$). In July 1987 she or he computes i) $\widehat{GNP}_{107}^{107-2/3}$, the first estimate of which becomes available 3 months later, ii) $\widehat{GNP}_{108}^{107-2/3}$, getting assigned a first estimate 6 months later, iii) $\widehat{GNP}_{109}^{107-2/3}$, in this sense a 9-month-ahead forecast and iv) $\widehat{GNP}_{110}^{107-2/3}$, a 12-month-ahead forecast. In August 1987, $\widehat{GNP}_{107}^{107-1/3}$, $\widehat{GNP}_{108}^{107-1/3}$, $\widehat{GNP}_{109}^{107-1/3}$ and $\widehat{GNP}_{110}^{107-1/3}$ constitute 2-, 5-, 8- and 11-month-ahead forecasts, respectively, whereas the corresponding now- and forecasts made in September 1987 are 1-, 4-, 7- and 10-month-ahead forecasts, respectively. As the example period consists of 23.5 years, we obtain a 1-, ..., 12-month-ahead forecast for each of the 94 quarters under consideration. By comparing the now-/forecasts to their corresponding first releases, i.e., to $GNP_{[T^*+h]}^{[T^*+h]+1/3}$ for $h = 0, 1, 2, 3$, we can compute 12 RMSEs for the 1-, ..., 12-month-ahead forecasts. This allows us to investigate in what sense prediction accuracy of the first release depends on T^* and h .¹⁶ The corresponding 12 RMSEs are given in Table 3.4.

It becomes apparent that, with only one exception (from 6- to 5-months-ahead), the RMSEs are decreasing with the 'forecast horizon'. Note that the largest difference can be observed from 4- to 3-months-ahead, i.e., when turning from fore- to nowcasting. This implies the obvious conclusion that the closer a researcher is to

¹⁵Whenever necessary, we use a simple AR(1), respectively AR(m_j), model to forecast future (w.r.t. T^*) values of X . To be more precise, for Point-in-Time or Average sampled regressors an AR(1) is employed, i.e., $\hat{x}_t = \hat{\mu}^{(AV/PIT)} + \hat{\rho}^{(AV/PIT)}x_{t-1}$, whereas for MIDAS-restricted regressors an AR(m_j) is used, i.e., $\hat{x}_{t-k/m_j} = \hat{\mu}^{(MIDAS)} + \sum_{i=1}^{m_j} \hat{\rho}_i^{(MIDAS)}x_{t-(k+i)/m_j}$ for $k = 0, \dots, m_j - 1$. The parameters $\hat{\rho}^{(AV/PIT)}$ and $\hat{\rho}^{(MIDAS)}$ are computed on the estimation period $t = 2, \dots, [T^* - 1]$ as well.

¹⁶In order to compute RMSEs for each of the forecasts under consideration, we neglected vintages beyond December 2009. In January 2010, for example, the 12-month-ahead forecast ($\widehat{GNP}_{204}^{201-2/3}$) would need to be compared to the figure published in January 2011 ($v = 204\frac{1}{3}$), a time period that is beyond our sample since the last vintage at our disposal is December 2010 ($V_L = 204$).

Table 3.4: RMSEs for 1- up to 12-month-ahead forecasts

...months-ahead	RMSE
12-...	0.00578
11-...	0.00577
10-...	0.00576
9-...	0.00575
8-...	0.00574
7-...	0.00569
6-...	0.00568
5-...	0.00570
4-...	0.00545
3-...	0.00477
2-...	0.00457
1-...	0.00441

Note: The figures represent the RMSEs for the period July 1987 to December 2009 corresponding to forecasts of U.S. GNP growth made in real-time 1- up to 12-months ahead with respect to the first release of the respective figure. See the previous paragraph for details.

the moment of the first release, the better her or his forecast. However, inspecting the forecast sensitivity to a particular vintage within a quarter, i.e., comparing 12-, 11- and 10-month-ahead forecasts, 9-, 8- and 7-month-ahead forecasts, and so on, reveals an interesting observation: Except for the 10-month-ahead case, forecasts made in the last month of a quarter, i.e., 1-, 4- and 7-months-ahead, are statistically significantly more accurate than those made in the first or second month of a quarter.¹⁷ This implies that in January 2010, for example, we should wait for the data release of March to become available in order to compute forecasts of GNP growth in 2010Q1-2010Q3.¹⁸

As mentioned before, for each T^* we obtain a (potentially) different set of RPR weights. With $T^* = 106\frac{1}{3}, 106\frac{2}{3}, \dots, 200$ in our example, we obtain 282 (amount of months in 23.5 years) graphs similar to the one in Figure 3.5. As an illustration of how the RPR weights change throughout the real-time analysis, Figure 3.7 displays the corresponding graphs made at six distant moments in time.

Note how the set of calendar dates, for which RPR weights are computed, extends as we move forward in time and how the final graph, i.e., the one correspond-

¹⁷As computed via respective Diebold and Mariano (1995) tests. Results available upon request.

¹⁸Note that we cannot formally compare, e.g., 3- and 4-month-ahead or 6- and 7-month-ahead forecasts, as the set of first releases, to which we need to compare the forecasts, changes.

Figure 3.7: RPR weights at 6 vintages in real time

Note: The figure displays the RPR weights assigned to the seven models at 6 different moments in time: 1986M9, 1991M12, 1996M3, 2001M6, 2005M9 and 2009M12. Depending on the moment of time at which the graph is produced the calendar dates, to which the RPR apply, differ. For example, in 1991M12 only calendar dates from 1986Q2 to 1991Q3 get weights assigned. Calendar dates are represented on the horizontal axis, whereas the weight itself is plotted on the vertical axis (ranging from 0 to 1).

ing to 2009M12, almost coincides with the one in Figure 3.5.¹⁹ Another interesting observation is that while the PIT model receives a relatively large weight in the top three panels, its cointegration-including counterpart, PIT_COINT, seems to take over in the bottom three panels.

Finally, when computing now- or forecasts in real time as outlined above, we proposed to employ the latest-available RPR weights at each corresponding moment in time. In other words, for the six example vintages in Figure 3.7 we would have used the weights corresponding to the right end points of each panel. For our real-time sample from 1987M7 ($T^* = V_F = 107 - 2/3$) to 2009M12 ($T^* = 200$) Figure 3.8 plots the corresponding latest-available RPR weights. It thus shows the evolution of the models' RPR weights that are ultimately used for now- or forecasting in real time.

Here, too, we can see how PIT_COINT replaces PIT as the dominant Point-in-Time sampling model after roughly half of the real-time periods have been considered. Furthermore, it seems that the latest-available RPR weights stabilize after roughly a third of the real-time sample with PIT_COINT and MIDAS claiming the largest weights (approximately 30% each).

3.5 Conclusion

In this paper we combined the issues of working with real-time data sets and dealing with variables that are sampled at mixed frequencies. As such, the repeated observations forecasting approach of Stark and Croushore (2002), was extended to an ADL setting where the regressors are possibly sampled at higher frequencies than the regressand. Furthermore, we discussed how to use our approach in real-time analyses. The ideas were illustrated by means of an empirical analysis involving quarterly U.S. GNP growth.

As far as the extension of the ROF approach is concerned, the distribution of one-step-ahead forecasts, of each model, was non-parametrically approximated by a density function summarizing the extent to which forecasts of a particular calendar date differ across vintages. In other words, we considered and measured uncertainty surrounding the data. Combinations of these model-specific densities were subsequently constructed using what we called revision-process-robust weights, which summarize the extent to which the various models are able to mimic

¹⁹Even at vintage V_L they would not coincide, though, since $\bar{V}$ in Section 3.3.4 differs from $\bar{V}$ in our real-time analysis.

Figure 3.8: Latest-available RPR weights in real time

Note: The figure displays the latest-available RPR weights of our seven models for each vintage in our real-time sample period, i.e., for each $T^* \in [107 - 2/3, 200]$. At a given T^* (plotted on the horizontal axis), the vertical length of a shaded area corresponds to the respective model's latest-available RPR weight. Of course, the total vertical length equals 1 for each T^* .

the revision process up until the date under consideration. The RPR weights implied a ranking of the different models under consideration, which was compared to usual RMSE-based rankings obtained using final-vintage data only. In this way it was investigated whether model rankings are sensitive to data releases.

Empirically, it was found that forecasts of U.S. GNP growth are sensitive to both the model chosen and the data release considered and that adding exogenous variables improves forecast accuracy. More importantly, model 'horse races' were sensitive to the data release employed, i.e., the model ranking based on RMSEs obtained using final-vintage data differed from the ranking implied by the RPR weights. To be more precise, using only latest-available data, some models seemed to possess a better (or worse) forecasting performance than when their forecast accuracy was evaluated also using past data releases.

As for the incorporation of revision process information into real-time studies, we proposed to conduct the ROF analysis from before (with emphasis on the models' ability to forecast early releases) up until the current moment in time. Assuming constancy of the RPR weights in the future, we used the latest-available RPR weights to combine out-of-sample forecasts of U.S. GNP growth. It turned out that RMSEs decrease with the 'forecast horizon' and that forecasts made in the last month of a quarter seem more accurate than the ones based on data from the first or second month. Finally, the latest-available RPR weights stabilized after roughly a third of our real-time sample with PIT_COINT and MIDAS as dominating models.

The preceding analysis could be extended along several lines. Most importantly, the set of regressors and models could be increased. Including variables of different and higher frequencies as well as a larger set of models might lead to new insights in several aspects, notably the combination approach, its associated weight structure and the use of higher-frequency vintages. Independent from the scale of the models and the variables' sampling frequencies, however, our analysis showed that model rankings are potentially very sensitive to the data release chosen. It also illustrates, though, that researchers can guard themselves against such critique by incorporating revision process information into their (real-time) studies. The computation of our proposed RPR weights is, however, dependent on the existence of a real-time data set, emphasizing the importance of making them available to empirical researchers.

3.A Appendix: Models

Below we list the models which we use for forecasting. As mentioned in the main body of the paper, the possibility of cointegration between the variables is taken into account by considering the models with and without a cointegrating relationship. Although we aim at forecasting GNP growth rates, it is well possible that *GNP* is cointegrated with one or more of the regressors. As shown in Götz et al. (2014), disregarding (including) a long-run relationship in the presence (absence) of cointegration leads to a considerably worsened forecasting performance. Hence, all models, except (4), i.e., the autoregressive one, are in an error-correction format in which a long-run term (z_{t-1}^v below) is either in- or excluded (labeled by 'a' and 'b', respectively).

- (1) **Restricted short-run (ECM-MIDAS):** MIDAS regression models have been introduced by Ghysels et al. (2004) and aim at preserving information present in the high-frequency variables while estimating parameters in a parsimonious way. Estimating standard linear regression models unrestrictedly might be unappealing due to parameter proliferation (Andreou et al., 2013). If y_t^v is a quarterly variable and x_{t-i/m_j}^v is daily, we might have to estimate over 50 parameters. In an (ECM-)MIDAS model we hyper-parameterize the polynomial lag structure yielding

$$\alpha(L)\Delta y_t^v = c + \delta z_{t-1}^v + \beta \sum_{i=0}^{K_h} w_{i+1}(\theta) \Delta^{(1/m_j)} x_{t-i/m_j}^v + u_t, \quad (3.6)$$

where $\alpha(L)$ is a lag polynomial allowing for autoregressive terms, z_{t-1}^v denotes the disequilibrium error in the previous period and $w_{i+1}(\theta)$ are weights that sum up to one in order to identify the scale coefficient β . Note that $\Delta^{(1/m_j)}$ represents the high-frequency difference operator, $\Delta^{(1/m_j)} x_{t-i/m_j}^v = x_{t-i/m_j}^v - x_{t-(i+1)/m_j}^v$, whereas Δ represents its low-frequency counterpart, $\Delta y_t^v = y_t^v - y_{t-1}^v$. The weights are based on an underlying function for which different specifications are proposed in the literature (see Ghysels et al., 2007 for a good overview). We employ the two-dimensional exponential Almon lag polynomial which is shown to be extremely flexible allowing the weight-determination to be completely data-driven.²⁰

²⁰Another popular, and similarly flexible, choice is the Beta lag polynomial (see Ghysels et al., 2007 for details and more lag polynomial choices).

Equation (3.6) can be extended to include multiple regressors by adding separate weight functions for each one. If lags of the regressors beyond one low-frequency period are excluded, $K_h = m - 1$ is sufficient. If, however, high-frequency observations corresponding to, e.g., $t - 1$ enter the regression, two approaches are possible. First, one may simply set $K_h = 2m - 1$ and estimate one weight function per regressor. Second, separate weight functions for each set of m high-frequency observations (once corresponding to period t and once to period $t - 1$) may be estimated. The latter has the advantage of allowing more flexibility in the weights, but the disadvantage of a larger set of parameters that need to be estimated. Obviously, (3.6) changes to

$$\alpha(L)\Delta y_t^v = c + \delta z_{t-1}^v + \sum_{k=0}^{K_l} \beta_k \sum_{i=0}^{m_j-1} w_{i+1}^{(k)}(\theta) \Delta^{(1/m_j)} x_{t-k-i/m_j}^v + u_t, \quad (3.7)$$

where K_l denotes the number of low-frequency lag periods for x and where $w^{(0)}, \dots, w^{(K_l)}$ are the $K_l + 1$ distinct weight functions. In this paper we opt for the second approach, i.e., Equation (3.7), due its greater flexibility in determining the weights. The two-dimensional Almon lag polynomial then becomes

$$w_{i+1}^{(k)}(\theta) = w_{i+1}(\theta_1^{(k)}, \theta_2^{(k)}) = \frac{\exp(\theta_1^{(k)}(i+1) + \theta_2^{(k)}(i+1)^2)}{\sum_{i=0}^{m_j-1} \exp(\theta_1^{(k)}(i+1) + \theta_2^{(k)}(i+1)^2)}.$$

As far as the long-run term in the ECM-MIDAS model is concerned, Götz et al. (2014) discuss the choice of which high-frequency observation to include in z_{t-1}^v . Actually, it does not make a significant difference in terms of forecasting performance whether one includes the end-of-period observation or an earlier high-frequency observation as long as the structure of short-run dynamics terms is adapted accordingly (see Götz et al., 2014). This result should not come as a surprise given that cointegration is a long-run property (Marcellino, 1999). Due to this argument and for its simplicity, we use the end-of-period observations, x_t^v , to enter the disequilibrium error (as also done in Miller, 2011 or Götz et al., 2014) whenever it is included in the model. For each vintage the cointegrating coefficient will be estimated by OLS implying that we follow the standard two-step Engle-Granger approach (Engle and Granger, 1987).

Note that ECM-MIDAS models including a cointegrating relationship are labeled (1a), whereas the ones without a long-run term are denoted by (1b).

- (2)-(3) **Averaging and Point-in-Time sampling:** Alternative to ECM-MIDAS models one can temporally aggregate the high-frequency series into low-frequency ones as discussed in Marcellino (1999) among others. Averaging, i.e., models (2a)-(2b), or Point-in-Time sampling, i.e., models (3a)-(3b), are the common aggregation schemes here. The former approach takes the average of the high-frequency observations in a certain t -period as the corresponding low-frequency observation, while the latter takes one specific high-frequency observation (often the last one). Both approaches are in fact special cases of the ECM-MIDAS model. Using the notation in (3.7) we have:

$$w_{i+1}^{(k)}(\theta) = w_{i+1} = \begin{cases} 1 & \text{if } i = 0 \\ 0 & \text{else} \end{cases} \quad (\text{Point-in-Time sampling}),$$

$$w_{i+1}^{(k)}(\theta) = w_{i+1} = \frac{1}{m_j} \quad \forall i = 1, \dots, m_j \quad (\text{Average sampling}).$$

- (4) **ARIMA(4, 1, 0):** As a benchmark we consider an *ARIMA*(4, 1, 0) as is done almost throughout the whole investigation in Stark and Croushore (2002).

We estimate models (1a)-(3b) with one low-frequency lag for the regressand and the two regressors, i.e., $\alpha(L) = 1 - \alpha L$ and $K_l = 1$ in Equation (3.7), which needs to be adjusted for the presence of two regressors with different frequencies, m_1 and m_2 .

Chapter 4

Testing for Common Cycles in Non-Stationary VARs with Varied Frequency Data

This chapter proposes a new approach to detecting the presence of common cyclical features when several time series are sampled at different frequencies. We generalize the common-frequency approach introduced by Engle and Kozicki (1993) and Vahid and Engle (1993).

We start with the mixed-frequency VAR representation investigated in Ghysels (2012) for stationary time series. For non-stationary time series in levels, we show that one has to account for the presence of two sets of long-run relationships. The first set is implied by identities stemming from the fact that the differences of the high-frequency $I(1)$ regressors are stationary. The second set comes from possible additional long-run relationships between one of the high-frequency series and the low-frequency variables. Our transformed vector error-correction model (VECM) representations extend the results of Ghysels (2012) and are important for determining the correct set of variables to be used in a subsequent common cycle investigation. This fact has implications for the distribution of test statistics and for forecasting. Empirical analyses with quarterly real gross national product (GNP) and monthly industrial production indices for, respectively, the U.S. and Germany illustrate our new approach. We also conduct a Monte Carlo study

which compares our proposed mixed-frequency models with models stemming from classical temporal aggregation methods.¹

4.1 Introduction

Time series are collected and observed at different sampling frequencies. For instance, national account variables such as the gross domestic product (GDP) are often available on a quarterly basis whilst many other indicators such as the industrial production index, the unemployment rate or price indices are supplied at a monthly frequency. Several variables are only reported annually (e.g. CO₂ emission, health outlook, etc.) whilst most financial time series (stock prices, interest rates, exchange rates, etc.) are recorded almost continuously and are easily obtained at a daily frequency. In this situation it is common in macroeconomics to focus on the lower frequency unless a good proxy for the variable of interest is available at higher frequency. For instance, if one wants to estimate a vector autoregressive model (VAR) with GDP included, one would probably run a VAR model on quarterly time series. The higher frequency variables (monthly, daily,...) are either averaged over quarters (for flow variables) or matched at the end of each quarter (for stock variables).

More recently, researchers have developed techniques for modeling, in a unified framework, variables sampled at different frequencies (see notably the MIDAS approach in Andreou et al., 2010). Most of the mixed-frequency literature which relies on what Ghysels (2012) calls 'observation-driven' models,² as introduced in Cox et al. (1981), is concerned with a regression framework where the dependent variable is the low-frequency variable and the high-frequency variables are in the regressors. Only few papers have extended the analysis towards a multivariate setting. Examples are Qian (2013) and Ghysels (2012).

Qian (2013) deals with a mixed-frequency VAR in which some variables are possibly unobserved to account for the fact that some series may be observed at lower frequencies. Ghysels (2012), in contrast, stacks the high- and low-frequency variables into a vector and formulates a VAR in terms of this vector. Interest-

¹This chapter is based on the paper Götz et al. (2013), published in *Advances in Econometrics*. The authors thank Lutz Kilian, Tom Fomby and the participants of the 12th Annual Advances in Econometrics Conference in Dallas, the 8th Netherlands Econometric Study Group Meeting in Amsterdam 2013 and the 67th European Meeting of the Econometric Society in Gothenburg 2013 for useful suggestions and comments on earlier versions of the paper.

²In contrast to 'parameter-driven' models such as state space models à la Zdrozny (1988). See Seong et al. (2013) for VECMs in state-space forms.

ingly, the model does not involve any latent variables, and thereby no latent shocks, but is formulated exclusively in terms of observable data explaining the term 'observation-driven' model. Note that due to its 'observation-driven' nature, the model's related impulse response functions are entirely driven by observable shocks. Indeed, the mixed-frequency VAR à la Ghysels (2012) jointly specifies the dynamics of all variables concerned without imposing any a-priori restrictions. Furthermore, the ability to apply standard tools of the VAR literature, e.g., impulse response functions, to a mixed-frequency setting gives it a significant edge over single-equation dynamic regression models such as the aforementioned mixed data sampling, or MIDAS, approach.

We follow Ghysels (2012) in this paper in specifying an 'observation-driven' rather than a 'parameter-driven' model.³ We deviate from Ghysels (2012) in that we allow for non-stationary and possibly cointegrated variables. An extension of a mixed-frequency VAR model along these lines is important. We show that for non-stationary time series in levels, one has to account for the presence of two sets of long-run relationships. The first set is characterized by identities stemming from the fact that the differences of the high-frequency $I(1)$ regressors are stationary. The second set comes from additional long-run relationships between one of the high-frequency series and the low-frequency variables. Without this latter set of long-run relationships, the VAR in first differences looks very much like in Ghysels (2012). We propose a VECM representation for these models that allows us to extend the common cyclical feature analysis (see Engle and Kozicki, 1993 or Vahid and Engle, 1993) to a mixed-frequency setup. As such, these mixed-frequency VARs present an alternative to temporally aggregating the high-frequency variable in order to obtain a VAR in a common frequency (Marcellino, 1999 or Hecq, 1998).

The rest of the paper is organized as follows. We start with the stacked VAR representation proposed in Ghysels (2012) and show how to extend it to possibly cointegrated $I(1)$ variables. In Section 4.3 we propose an alternative VAR representation that is useful for testing for the presence of common cyclical features. A suitable likelihood ratio type statistic is proposed in Section 4.4. Our new approach is intended for models involving a small number of variables and low observation frequency. Section 4.5 contains empirical illustrations of our approach

³Note that state space models are unrivalled only under ideal conditions including correct model specification and no parameter estimation errors. These ideal conditions are difficult to satisfy in practice. For example, correct specification is difficult to achieve due to missing observations for the regressand. Moreover, specification errors in one equation will spread to the entire system.

based on the quarterly real gross national product and industrial production for Germany and the United States. Section 4.5 also contains a small forecasting exercise emphasizing the potential usefulness of our methods compared with classical temporal aggregation methods. Section 4.6 presents a Monte Carlo study investigating the rejection frequencies of the aforementioned likelihood ratio type test for common features for both, our proposed mixed-frequency and temporal aggregation based models. Section 4.7 concludes.

4.2 Mixed-Frequency VECMs

Let us start from a two variable mixed-frequency system,⁴ where $y_t, t = 1 \dots T$ is the low-frequency variable and $x_{t-i/m}^{(m)}$ the high-frequency variables with m high-frequency observations per low-frequency period t . In a year/month-example $m = 12$ and the value of i indicates the specific month under consideration, ranging from January ($x_{t-11/12}^{(12)}$) until December ($x_t^{(12)}$). Obviously, $x_{t-m/m}^{(m)} = x_{t-1}^{(m)}$. Note that as far as lag operators, L , or difference operators, Δ , are concerned, we make a distinction between high- and low-frequency operators. To be more precise, L denotes the low-frequency lag operator, i.e., $Ly_t = y_{t-1}$ or $Lx_{t-i/m}^{(m)} = x_{t-1-i/m}^{(m)}$, whereas $L^{j/m}$ denotes the high-frequency counterpart which is only applicable to the high-frequency variables, i.e., $L^{j/m}x_{t-i/m}^{(m)} = x_{t-i/m-j/m}^{(m)} = x_{t-(i+j)/m}^{(m)}$. Similar rules apply for the difference operators, Δ and $\Delta^{(j/m)}$. Finally, note that $L^{1/m}x_{t-(m-1)/m}^{(m)} = x_{t-1}^{(m)}$ and, by the same logic, $\Delta^{(1/m)}x_{t-(m-1)/m}^{(m)} = x_{t-(m-1)/m}^{(m)} - x_{t-1}^{(m)}$. These notational conventions are similar to the ones used in Clements and Galvão (2008, 2009) or Götz et al. (2012, 2014). The following table illustrates the notation for a year/month-example:

Notation	$t = 2011, m = 12$
$x_{t+1-(m-1)/m}^{(m)} = x_{t+1/12}^{(12)}$	$x_{2012, Jan}^{(12)}$
$x_t^{(m)}$	$x_{2011, Dec}^{(12)}$
$x_{t-1/m}^{(m)}$	$x_{2011, Nov}^{(12)}$
$\vdots$	$\vdots$
$x_{t-(m-1)/m}^{(m)} = x_{t-11/12}^{(12)}$	$x_{2011, Jan}^{(12)}$
$x_{t-m/m}^{(m)} = L^{1/m}x_{t-11/12}^{(12)} = x_{t-1}^{(12)}$	$x_{2010, Dec}^{(12)}$
$x_{t-1-1/m}^{(m)}$	$x_{2010, Nov}^{(12)}$

⁴An extension to higher dimensional multivariate systems is straightforward.

Considering each high-frequency variable such that

$$X_t^{(m)} = (x_t^{(m)}, x_{t-1/m}^{(m)}, \dots, x_{t-(m-1)/m}^{(m)})',$$

Ghysels (2012) proposes to difference the $I(1)$ data (e.g., quarterly growth rate of output and the monthly growth rate of the production index) and then construct the $\text{VAR}(p)$:

$$Z_t = \Gamma_1 Z_{t-1} + \dots + \Gamma_p Z_{t-p} + u_t, \quad (4.1)$$

where $Z_t = (y_t, X_t^{(m)})'$ and $u_t = (u_{1t}, u_{2t}, \dots, u_{(m+1)t})' \sim \text{i.i.d. } N(\mathbf{0}, \Sigma_u)$.⁵ In other words, the high-frequency observations are stacked in a vector, together with the low-frequency variable y_t . Note that compared to Ghysels (2012) we place the low-frequency variable first to facilitate the interpretation of the presence of common cycles we develop in Section 4.4.

Equation (4.1) is easy to estimate for small systems but relies on the assumption that there is no cointegration. Assuming that series in Z_t are $I(1)$ and that there exists some cointegration between the variables, (4.1) in first differences is not correctly specified. This can be seen by computing the VECM representation of (4.1) when both the high- and low-frequency variables are $I(1)$ such that

$$\Delta Z_t = \tilde{\Gamma}_1 \Delta Z_{t-1} + \dots + \tilde{\Gamma}_{p-1} \Delta Z_{t-p+1} + \Pi Z_{t-1} + u_t, \quad (4.2)$$

with $\tilde{\Gamma}_i = -\sum_{k=i+1}^p \Gamma_k$, $i = 1, \dots, p-1$ and $\Pi = -(I - \sum_{j=1}^p \Gamma_j) = \alpha\beta'$ with $\text{rank}(\Pi) = (r_0 + r_1) < m + 1$.

Let us focus on the rank of Π in the presence and absence of cointegration between the low- and high-frequency variables. r_0 denotes a first set of known or prespecified cointegrating vectors stemming from the fact that first differences of the high-frequency $I(1)$ variables are stationary. r_1 will denote the potential additional long-run relationship between the two series. Note that in case of cointegration it is sufficient to consider only one relationship between y and one of the various x 's. Indeed, if cointegration between the two variables is present, y_t is cointegrated with each $x_{t-i/m}^{(m)}$, $i = 0, \dots, m-1$. Hence, it does not matter which of these high-frequency candidates is used to capture the cointegrating relation-

⁵It is possible to include deterministic components such as a constant, trends or seasonal dummies here as well. For illustrative purposes we omit the deterministic regressors in the theoretical sections of the paper. In the application and Monte Carlo study, however, a vector of intercepts is included.

ship as long as one is used (see Götz et al., 2014 for details). For simplicity, we model cointegration employing the end-of-period observation of the high-frequency variable, i.e., $x_t^{(m)}$.

Anticipating the results of the empirical section, when r_0 and the associated cointegrating vectors are known, a test on r_1 will be possible using the approach of Horvath and Watson (1995) for testing for cointegration when some of the cointegrating vectors are prespecified.⁶ In particular, we will test the null hypothesis $H_0 : \text{rank}(\Pi) = r_0$ against $H_A : \text{rank}(\Pi) = r_0 + r_1$, where the cointegrating vectors imposed under the null are precisely the prespecified ones discussed above. Under the alternative, r_1 unknown cointegrating vectors are added to the r_0 prespecified ones. The test itself is then a likelihood ratio statistic. In our case, only one unknown cointegrating vector to capture potential cointegration between the low- and high-frequency variables is added. It is of the form $(1 : -\theta : 0_{1 \times (m-1)})$, where the coefficient of y is normalized to one and θ is the cointegrating coefficient.

Our first results can be summarized in the following observation and example.

Observation 1. Suppose that the high-frequency variables, $x_{t-i/m}^{(m)}$, and the low-frequency variable, y_t , are $I(1)$. In the case of no cointegration between the low- and high-frequency variables, $\text{rank}(\Pi) = r_0$. $r_0 = m - 1$ in a bivariate analysis. In the case of cointegration between the low- and high-frequency variables, $\text{rank}(\Pi) = r_0 + r_1$. $r_1 = 1$ in a bivariate analysis.

Indeed, each high-frequency variable in the subvector X_t of Z_t is trivially cointegrated with the other high-frequency variables. Since there are m such variables in Z_t in the bivariate case, we have $m - 1$ prespecified cointegrating relationships r_0 . We model any additional cointegrating relationships between the low- and high-frequency variables by considering one relationship between y_t and $x_t^{(m)}$. This increases the cointegrating rank by one in the bivariate example. These arguments can be extended to a higher dimensional framework.

Example 1. Let us consider a very simple example where the low- and high-frequencies are assumed to be quarterly and monthly, respectively, and where y_t and the elements of $X_t^{(m)}$ are defined as

$$\begin{aligned} y_t &= \theta x_t^{(3)} + u_{1t} \text{ with } u_{1t} = \rho u_{1t-1} + \nu_t \\ x_{t-i/3}^{(3)} &= x_{t-(i+1)/3}^{(3)} + u_{(2+i)t} \text{ for } i = 0, 1, 2, \end{aligned}$$

⁶Horvath and Watson (1995) show that accounting for the fact that some cointegrating vectors are prespecified leads to improved power results in terms of testing for potential further, unknown, cointegrating relationships.

where $|\rho| < 1$, $\nu_t \sim \text{i.i.d. } N(0, 1)$ and all $u_{(2+i)t} \sim \text{i.i.d. } N(0, 1)$, implying that cointegration between the low- and high-frequency variables is present. Representing this in a VAR framework yields

$$\begin{pmatrix} y_t \\ x_t^{(3)} \\ x_{t-1/3}^{(3)} \\ x_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} \rho & (1-\rho)\theta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(3)} \\ x_{t-1-1/3}^{(3)} \\ x_{t-1-2/3}^{(3)} \end{pmatrix} + \begin{pmatrix} \theta \sum_{i=0}^2 u_{(2+i)t} + \nu_t \\ \sum_{i=0}^2 u_{(2+i)t} \\ u_{3t} + u_{4t} \\ u_{4t} \end{pmatrix}.$$

The corresponding VECM representation looks as follows:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t^{(3)} \\ \Delta x_{t-1/3}^{(3)} \\ \Delta x_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} \rho - 1 & (1-\rho)\theta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(3)} \\ x_{t-1-1/3}^{(3)} \\ x_{t-1-2/3}^{(3)} \end{pmatrix} + \begin{pmatrix} \theta \sum_{i=0}^2 u_{(2+i)t} + \nu_t \\ \sum_{i=0}^2 u_{(2+i)t} \\ u_{3t} + u_{4t} \\ u_{4t} \end{pmatrix},$$

such that $\text{rank}(\Pi) = r_0 + r_1 = 2 + 1 = 3$ with long-run relationships, e.g.,

$$\beta' Z_{t-1} = \begin{pmatrix} \rho - 1 & -(\rho - 1)\theta & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(3)} \\ x_{t-1-1/3}^{(3)} \\ x_{t-1-2/3}^{(3)} \end{pmatrix}.$$

If, however, no cointegration existed and both high- and low-frequency variables followed a random walk,

$$\begin{aligned} y_t &= y_{t-1} + v_t \\ x_{t-i/3}^{(3)} &= x_{t-(i+1)/3}^{(3)} + u_{(2+i)t} \text{ for } i = 0, 1, 2, \end{aligned}$$

where v_t and the $u_{(2+i)t}$'s are i.i.d. standard normally distributed, we would obtain the VECM representation:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t^{(3)} \\ \Delta x_{t-1/3}^{(3)} \\ \Delta x_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(3)} \\ x_{t-1-1/3}^{(3)} \\ x_{t-1-2/3}^{(3)} \end{pmatrix} + \begin{pmatrix} v_t \\ \sum_{i=0}^2 u_{(2+i)t} \\ u_{3t} + u_{4t} \\ u_{4t} \end{pmatrix}$$

such that $\text{rank}(\Pi) = r_0 = m - 1 = 2$ and the (prespecified) long-run cointegrating relationships are

$$\beta' Z_{t-1} = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(3)} \\ x_{t-1-1/3}^{(3)} \\ x_{t-1-2/3}^{(3)} \end{pmatrix}.$$

4.3 Alternative VAR/VECM Representations

Starting with the representation (4.2) an alternative representation of this VAR model maybe obtained as illustrated in Hecq and Jacobs (2009). This representation will be used in deriving tests for common cycles. For this purpose we introduce the $(m+1) \times (m+1)$ matrix $C = (M_y : \beta : M_x)'$, with selection matrices $M_y = (1 \ \mathbf{0}_{1 \times m})' \mathbb{1}_{(r=m-1)}$ and $M_x = (\mathbf{0}_{1 \times m} \ 1)'$. The indicator function in M_y ensures that we add a corresponding selection row in case y and x are not cointegrated. Premultiplying both sides of (4.2) by C gives

$$\begin{aligned} C \Delta Z_t &= C \tilde{\Gamma}_1 C^{-1} C \Delta Z_{t-1} + \dots + C \tilde{\Gamma}_{p-1} C^{-1} C \Delta Z_{t-p+1} + \\ &\quad C \Pi C^{-1} C Z_{t-1} + C u_t \\ \Leftrightarrow C Z_t &= (I + C \tilde{\Gamma}_1 C^{-1} + C \Pi C^{-1}) C Z_{t-1} + (C \tilde{\Gamma}_2 C^{-1} - C \tilde{\Gamma}_1 C^{-1}) C Z_{t-2} \\ &\quad + \dots + (-C \tilde{\Gamma}_{p-1} C^{-1}) C Z_{t-p} + C u_t. \end{aligned} \tag{4.3}$$

Whether or not there is cointegration between y and x , it is possible to rewrite (4.3) such that we finally obtain

$$\tilde{Z}_t^D = \Phi_1 \tilde{Z}_{t-1}^D + \Phi_2 \tilde{Z}_{t-2}^D + \dots + \Phi_p \tilde{Z}_{t-p}^D + \tilde{u}_t, \tag{4.4}$$

where

$$\tilde{Z}_t^D = \begin{pmatrix} \Delta y_t \\ \Delta^{(1/m)} x_t^{(m)} \\ \vdots \\ \Delta^{(1/m)} x_{t-(m-2)/m}^{(m)} \\ \Delta^{(1/m)} x_{t-(m-1)/m}^{(m)} \end{pmatrix}, \quad \tilde{Z}_t = \begin{pmatrix} y_t \\ \Delta^{(1/m)} x_t^{(m)} \\ \vdots \\ \Delta^{(1/m)} x_{t-(m-2)/m}^{(m)} \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix}$$

$$\text{and } \tilde{u}_t = \begin{pmatrix} u_{1t} \\ u_{2t} - u_{3t} \\ \vdots \\ u_{mt} - u_{(m+1)t} \\ u_{(m+1)t} \end{pmatrix},$$

i.e., $\tilde{Z}_t^D$ is obtained from $\tilde{Z}_t$ by applying the low-frequency first difference operator Δ to the first element and the high-frequency counterpart $\Delta^{(1/m)}$ to the last element.

We can write (4.4) as

$$\tilde{Z}_t^D = \Phi_1 \tilde{Z}_{t-1}^D + \Phi_2 \Delta \tilde{Z}_{t-2} + \dots + \Phi_p^0 \tilde{Z}_{t-p} + \Phi_p^1 \tilde{Z}_{t-p} + \tilde{u}_t, \quad (4.5)$$

where Φ_p^0 is the matrix with the central part of Φ_p and with zero-coefficients on the first and the last column. In the presence of additional (unknown) r_1 cointegrating vectors $\Phi_p^1 = \Phi_p - \Phi_p^0$ is of rank $r_1 > 0$. In the absence of such cointegrating vectors $\Phi_p = \Phi_p^0$ such that $\Phi_p^1 = \mathbf{0}_{(m+1) \times (m+1)}$.

Remark 4.1. As far as the error term is concerned, it was necessary to assume that u_t is i.i.d. and standard normally distributed in order for (4.2) not to be misspecified. Premultiplying u_t by C in 4.3, however, causes $\tilde{u}_t$ to contain differences of $u_{2t}, \dots, u_{mt}$. Note, however, that these are high-frequency differences in the sense that these are differences of error terms corresponding to two 'neighboring' high-frequency observations. This implies that all terms contained in $\tilde{u}_t$ correspond to observations within one low-frequency period t . Hence, as the transformed VAR, (4.5), is written in terms of the low frequency, premultiplying u_t by C does not lead to a misspecification of the model.

Example 2. Let us again assume that the high- and low-frequency variables are monthly and quarterly, respectively. Furthermore, suppose first that y and x are not cointegrated and let us consider a VAR(1), $Z_t = \Gamma_1 Z_{t-1} + u_t$ (without

deterministic components for simplicity), from which we derive a low-frequency VECM representation, $\Delta Z_t = \alpha\beta'Z_{t-1} + u_t$ as:

$$\Leftrightarrow \begin{pmatrix} \Delta y_t \\ \Delta x_t^{(3)} \\ \Delta x_{t-1/3}^{(3)} \\ \Delta x_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \\ \alpha_{41} & \alpha_{42} \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(3)} \\ x_{t-1-1/3}^{(3)} \\ x_{t-1-2/3}^{(3)} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{pmatrix}.$$

Now, we premultiply both sides of the equation by

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

to obtain

$$\begin{aligned} C\Delta Z_t &= C\Pi Z_{t-1} + Cu_t \\ \Leftrightarrow CZ_t &= CZ_{t-1} + C\Pi C^{-1}CZ_{t-1} + Cu_t \\ \Leftrightarrow \tilde{Z}_t &= (I + C\Pi C^{-1})\tilde{Z}_{t-1} + \tilde{u}_t. \end{aligned}$$

Note that

$$\begin{aligned} \tilde{Z}_t &= \begin{pmatrix} y_t \\ \Delta^{(1/m)}x_t^{(3)} \\ \Delta^{(1/m)}x_{t-1/3}^{(3)} \\ x_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} \Delta y_t \\ \Delta^{(1/m)}x_t^{(3)} \\ \Delta^{(1/m)}x_{t-1/3}^{(3)} \\ \Delta^{(1/m)}x_{t-2/3}^{(3)} \end{pmatrix} + \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}}_A \begin{pmatrix} y_{t-1} \\ \Delta^{(1/m)}x_{t-1}^{(3)} \\ \Delta^{(1/m)}x_{t-1-1/3}^{(3)} \\ x_{t-1-2/3}^{(3)} \end{pmatrix} \\ &= \tilde{Z}_t^D + A\tilde{Z}_{t-1}. \end{aligned}$$

Hence,

$$\begin{aligned}
 \tilde{Z}_t &= (I + C\Pi C^{-1})\tilde{Z}_{t-1} + \tilde{u}_t \\
 \Leftrightarrow \tilde{Z}_t^D &= \underbrace{(I - A + C\Pi C^{-1})}_{\Phi_1 = \Phi_p^0} \tilde{Z}_{t-1} + \tilde{u}_t \\
 \Leftrightarrow \begin{pmatrix} \Delta y_t \\ \Delta^{(1/m)} x_t^{(3)} \\ \Delta^{(1/m)} x_{t-1/3}^{(3)} \\ \Delta^{(1/m)} x_{t-2/3}^{(3)} \end{pmatrix} &= \underbrace{\begin{pmatrix} 0 & \alpha_{11} & \alpha_{12} & 0 \\ 0 & 1 + \alpha_{21} - \alpha_{31} & \alpha_{22} - \alpha_{32} & 0 \\ 0 & \alpha_{31} - \alpha_{41} & 1 + \alpha_{32} - \alpha_{42} & 0 \\ 0 & \alpha_{41} - 1 & \alpha_{42} - 1 & 0 \end{pmatrix}}_{\Phi_1} \times \\
 &\quad \begin{pmatrix} y_{t-1} \\ \Delta^{(1/m)} x_{t-1}^{(3)} \\ \Delta^{(1/m)} x_{t-1-1/3}^{(3)} \\ x_{t-1-2/3}^{(3)} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} - u_{3t} \\ u_{3t} - u_{4t} \\ u_{4t} \end{pmatrix}.
 \end{aligned}$$

This shows that in the presence of prespecified cointegrating vectors only, the mixed-frequency stationary VAR specification of Ghysels (2012) does not impose the full set of restrictions on the dynamics. While consistency of the estimators is still achieved, efficiency might be affected when the VAR is not restricted accordingly. In the framework of this paper, this changes the set of variables that has to be used in the subsequent common feature analysis.

Example 3. More importantly now, if cointegration between y and x is present, as in

$$\Pi = \alpha\beta' = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} \end{pmatrix} \begin{pmatrix} 1 & -\theta & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix},$$

we select $C = (\beta : M_x)'$ with $m = 3$. Note that $CZ_t \neq \tilde{Z}_t$. However, $A_\theta CZ_t = \tilde{Z}_t$, where

$$A_\theta = \begin{pmatrix} 1 & \theta & \theta & \theta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Hence, we have the following:

$$\begin{aligned}
 CZ_t &= (I + C\Pi C^{-1})CZ_{t-1} + Cu_t \\
 \Leftrightarrow A_\theta^{-1}\tilde{Z}_t &= (I + C\Pi C^{-1})A_\theta^{-1}\tilde{Z}_{t-1} + A_\theta^{-1}\tilde{u}_t \\
 \Leftrightarrow \tilde{Z}_t &= A_\theta(I + C\Pi C^{-1})A_\theta^{-1}\tilde{Z}_{t-1} + \tilde{u}_t \\
 \Leftrightarrow \tilde{Z}_t^D &= (A_\theta(I + C\Pi C^{-1})A_\theta^{-1} - A)\tilde{Z}_{t-1} + \tilde{u}_t \\
 &= \underbrace{(I - A + A_\theta C\Pi C^{-1}A_\theta^{-1})}_{\Phi_1}\tilde{Z}_{t-1} + \tilde{u}_t,
 \end{aligned}$$

where $\Phi_1 = \Phi_1^0 + \Phi_1^1$ with

$$\Phi_1^0 = \begin{pmatrix} 0 & \alpha_{12} - \alpha_{11} & \alpha_{13} - \theta\alpha_{11} & 0 \\ 0 & 1 + \alpha_{22} - \alpha_{32} - \theta\alpha_{21} + \theta\alpha_{31} & \alpha_{23} - \alpha_{33} - \theta\alpha_{21} + \theta\alpha_{31} & 0 \\ 0 & \alpha_{32} - \alpha_{42} - \theta\alpha_{31} + \theta\alpha_{41} & 1 + \alpha_{33} - \alpha_{43} - \theta\alpha_{31} + \theta\alpha_{41} & 0 \\ 0 & \alpha_{42} - \theta\alpha_{41} - 1 & \alpha_{43} - \theta\alpha_{41} - 1 & 0 \end{pmatrix}$$

and

$$\Phi_1^1 = \begin{pmatrix} \alpha_{11} \\ \alpha_{21} - \alpha_{31} \\ \alpha_{31} - \alpha_{41} \\ \alpha_{41} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -\theta \end{pmatrix}.$$

using a particular choice of normalization. Note that we can decompose Φ_1 as done above, because its first and fourth columns are linearly dependent, i.e., $\Phi_1 e_1' = -\theta \Phi_1 e_4'$, where e_i is a row-vector of size $m + 1$ with a value of one at the i^{th} position and zeroes everywhere else.

4.4 Common Cycles in Mixed-Frequency VARs

The popular common cycle approach (Vahid and Engle, 1993) aims at testing for the presence of common dynamics in VAR/VECM multivariate systems. Let us test for the presence of common cyclical features when two or more series are observed at different frequencies.

Let us point out that this paper focuses on the particular mixed-frequency case when m is relatively small, say 3 or 12. Indeed, we propose a method based on canonical correlations as well as Full Information Maximum Likelihood (FIML)

techniques. It is well known that the latter has shortcomings when the number of variables is large compared to the number of observations T .

4.4.1 Definitions

Let Ω_t be the information set generated by the collection of sigma-fields $F_t = \sigma(Z_s, s \leq t), t \geq 0$, and let Z_t be adapted to that filtration (Mikosch, 1998). In other words, Ω_t contains all the information present at moment t , implying that y_t and $x_t^{(m)}$ would be given, but $x_{t+1/m}^{(m)}$ would not.

We define the presence of common cycles in the mixed-frequency VAR (4.5) as follows:

Definition 4.1. In a mixed-frequency VAR for the $m + 1$ series $\tilde{Z}_t^D$,

$$\tilde{Z}_t^D = \Phi_1 \tilde{Z}_{t-1}^D + \Phi_2 \Delta \tilde{Z}_{t-2} + \dots + \Phi_p^0 \tilde{Z}_{t-p} + \Phi_p^1 \tilde{Z}_{t-p} + \tilde{u}_t,$$

with

$$\tilde{Z}_t^D = (\Delta y_t, \Delta^{(1/m)} x_t^{(m)}, \Delta^{(1/m)} x_{t-1/m}^{(m)}, \dots, \Delta^{(1/m)} x_{t-(m-1)/m}^{(m)})',$$

a Mixed-Frequency Serial Correlation Common Features (MF-SCCF) arises iff there exists a matrix δ spanning the common feature space such that

$$\mathbb{E}(\delta' \tilde{Z}_t^D | \Omega_{t-1}) = 0.$$

This implies that there exists a δ such that $\delta' \Phi_1 = \dots = \delta' \Phi_{p-1} = \delta' \Phi_p^0 = \delta' \Phi_p^1 = \mathbf{0}$, i.e., δ annihilates the short-run dynamics in the system. In other words, δ lies in the intersection of the left null spaces of all Φ -matrices.

In particular, when δ is a $(m + 1)$ -dimensional column vector normalized on the first element, we say that two series Δy_t and $\Delta^{(1/m)} x_t^{(m)}$ have a MF-SCCF relationship iff there exists a high-frequency-lag polynomial $\gamma(L^{1/m})$ such that $\mathbb{E}([\Delta y_t - \gamma(L^{1/m}) \Delta^{(1/m)} x_t^{(m)}] | \Omega_{t-1}) = 0$. Hence, two series Δy_t and $\Delta^{(1/m)} x_t^{(m)}$ having a MF-SCCF imply that

$$\mathbb{E}([\Delta y_t - \gamma_1 \Delta^{(1/m)} x_t^{(m)}] | \Omega_{t-1}) = \gamma_2 \Delta^{(1/m)} x_{t-1/m}^{(m)} + \dots + \gamma_m \Delta^{(1/m)} x_{t-(m-1)/m}^{(m)}$$

and there exists a (restricted) polynomial serial correlation common feature relationship of order m , PSCCF(m), as defined in Cubadda and Hecq (2001) between the two series.

4.4.2 Testing Procedure

Now that we have defined MF-SCCF, we can formulate the steps we use to test for mixed-frequency common cycles. Note that as far as testing for cointegration and testing for common cyclical features is concerned we follow a two-step procedure (Vahid and Engle, 1993 or Hecq et al., 2006), the validity of which has been established in Paruolo (2006). The steps are as follows:

1. We determine the lag length order in the level (or the log level) of the series for the mixed-frequency VAR in (4.1) to obtain p . We rely on usual information criteria, i.e., Akaike information criterion (AIC), Schwarz information criterion (SC) and Hannan-Quinn information criterion (HQ).
2. Given p , we test for the cointegrating rank $r = r_0 + r_1$. In doing so we rely on the Horvath and Watson (1995) approach because the $m - 1 = r_0$ cointegrating vectors are prespecified.
3. In the absence of additional r_1 cointegrating vectors we perform a reduced rank regression in the transformed VAR(p) using a test on the zero eigenvalues computed from the canonical-correlation analysis

$$CanCor \left\{ \tilde{Z}_t^D, \begin{pmatrix} \tilde{Z}_{t-1}^D \\ \vdots \\ \tilde{Z}_{t-p+1}^D \\ \tilde{Z}_{t-p}^0 \end{pmatrix} \right\}, \quad (4.6)$$

where $\tilde{Z}_{t-p}^0$ is equivalent to $\tilde{Z}_{t-p}$ except for the first and last elements:

$$\tilde{Z}_{t-p}^0 = \begin{pmatrix} 0 \\ \Delta^{(1/m)} x_{t-p}^{(m)} \\ \vdots \\ \Delta^{(1/m)} x_{t-p-(m-2)/m}^{(m)} \\ 0 \end{pmatrix}.$$

Furthermore, $CanCor \{V_t, W_t\}$ denotes the computation of canonical correlations between the two sets of variables V_t and W_t . The likelihood ratio type test, denoted by ξ_{LR} , considers the null hypothesis that there exist at most s common feature vectors against the alternative that there are more

than s , i.e., $H_0 : \text{rank}(\delta) \leq s$ versus $H_A : \text{rank}(\delta) > s$. The test is obtained as

$$\xi_{LR} = -T \sum_{i=1}^s \ln(1 - \hat{\lambda}_i), \quad s = 1, 2, \quad (4.7)$$

where $\hat{\lambda}_i$ is the i^{th} smallest squared canonical correlation, i.e., the i^{th} smallest eigenvalue, computed from (4.6) above, i.e., from

$$\hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{VW} \hat{\Sigma}_{WW}^{-1} \hat{\Sigma}_{WV}, \quad (4.8)$$

or similarly from the symmetric matrix $\hat{\Sigma}_{VV}^{-1/2} \hat{\Sigma}_{VW} \hat{\Sigma}_{WW}^{-1} \hat{\Sigma}_{WV} \hat{\Sigma}_{VV}^{-1/2}$, where $\hat{\Sigma}_{ij}$ are the empirical covariance matrices with i and j being equal to V and/or W . In the mixed-frequency bivariate case and without counting the constant terms, the unrestricted transformed VAR has $(m+1)^2(p-1) + (m+1-2)(m+1)$ parameters. Note that $s = 1, 2$ because our $(m+1)$ -dimensional transformed VAR already contains $m-1$ known cointegrating relationships such that s is maximum 2.

4. In the presence of additional r_1 cointegrating vectors, eigenvalues and eigenvectors are computed from

$$\text{CanCor} \left\{ \tilde{Z}_t^D, \begin{pmatrix} \tilde{Z}_{t-1}^D \\ \vdots \\ \tilde{Z}_{t-p+1}^D \\ \tilde{Z}_{t-p}^D \end{pmatrix} \right\}, \quad (4.9)$$

where

$$\tilde{Z}_{t-p} = \begin{pmatrix} y_{t-p} \\ \Delta^{(1/m)} x_{t-p}^{(m)} \\ \vdots \\ \Delta^{(1/m)} x_{t-p-(m-2)/m}^{(m)} \\ x_{t-p-(m-1)/m}^{(m)} \end{pmatrix},$$

and we focus on the test for $s = 1$. Indeed, due to the additional cointegrating relationship between the low- and high-frequency variables, there is only room for at most one common feature vector in this bivariate example. In

the mixed-frequency bivariate case and without counting the constant terms, the unrestricted transformed VAR has $(m+1)^2(p-1) + (m+1-2+1)(m+1)$ parameters.

5. Alternatively and numerically equivalent is the estimation of the whole system by FIML under the restriction that $\text{rank}(\delta) = 2$.⁷ To be more precise, Vahid and Engle (1993) illustrate that the system can then be written as s so-called pseudo-structural form equations and $n - s = m + 1 - s$ unconstrained reduced-form equations. In the present context, this representation might be easier for counting the number of restrictions as well as for considering additional restrictions on the common feature vectors. Let us write the pseudo-structural model in the presence of both sets of cointegrating vectors (prespecified and not prespecified) as

$$\begin{pmatrix} I & -\gamma \\ 0 & I \end{pmatrix} \tilde{Z}_t^D = \begin{pmatrix} 0 \\ \Lambda_1 \end{pmatrix} \tilde{Z}_{t-1}^D + \dots + \begin{pmatrix} 0 \\ \Lambda_{p-1} \end{pmatrix} \tilde{Z}_{t-(p-1)}^D + \begin{pmatrix} 0 \\ \Lambda_p \end{pmatrix} C^* \tilde{Z}_{t-p} + e_t$$

or, more specifically,

$$\begin{pmatrix} I_s & -\gamma_{s \times (m+1-s)} \\ 0_{(m+1-s) \times s} & I_{(m+1-s)} \end{pmatrix} \begin{pmatrix} \Delta y_t \\ \Delta^{(1/m)} x_t^{(m)} \\ \Delta^{(1/m)} x_{t-1/m}^{(m)} \\ \vdots \\ \Delta^{(1/m)} x_{t-(m-1)/m}^{(m)} \end{pmatrix} = \begin{pmatrix} \underbrace{0}_{s \times (m+1)} \\ \underbrace{\Lambda_1}_{(m+1-s) \times (m+1)} \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta^{(1/m)} x_{t-1}^{(m)} \\ \Delta^{(1/m)} x_{t-1-1/m}^{(m)} \\ \vdots \\ \Delta^{(1/m)} x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \dots +$$

⁷This can be easily accomplished in Eviews or in OxMetrics, for instance.

$$\begin{aligned}
 & \begin{pmatrix} \underbrace{\mathbf{0}}_{s \times (m+1)} \\ \underbrace{\Lambda_{p-1}}_{(m+1-s) \times (m+1)} \end{pmatrix} \begin{pmatrix} \Delta y_{t-p+1} \\ \Delta^{(1/m)} x_{t-(p-1)}^{(m)} \\ \Delta^{(1/m)} x_{t-(p-1)-1/m}^{(m)} \\ \vdots \\ \Delta^{(1/m)} x_{t-(p-1)-(m-1)/m}^{(m)} \end{pmatrix} \\
 & + \begin{pmatrix} \underbrace{\mathbf{0}}_{s \times (r_0+r_1)} \\ \underbrace{\Lambda_p}_{(m+1-s) \times (r_0+r_1)} \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 & -\hat{\theta} \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 & 0 \end{pmatrix}}_{C^*} \times \\
 & \begin{pmatrix} y_{t-p} \\ \Delta^{(1/m)} x_{t-p}^{(m)} \\ \Delta^{(1/m)} x_{t-p-1/m}^{(m)} \\ \vdots \\ \Delta^{(1/m)} x_{t-p-(m-2)/m}^{(m)} \\ x_{t-p-(m-1)/m}^{(m)} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{(m+1)t} \end{pmatrix},
 \end{aligned}$$

where $\Lambda_1, \dots, \Lambda_p$ are the $(m+1-s)$ lower rectangular parts of the corresponding Φ -matrices that are going to be estimated by FIML. Note that the first row of C^* disappears in case $r_1 = 0$. In terms of the terminology introduced by Vahid and Engle (1993), $(I_s - \gamma_{s \times (m+1-s)})\tilde{Z}_t^D$ are the s pseudo-structural form equations. The system is then completed by adding the $m+1-s$ unconstrained reduced-form equations, i.e., $(0_{(m+1-s) \times s} I_{(m+1-s)})\tilde{Z}_t^D$. This shows that we have to estimate $(m+1-s) \times (m+1)$ parameters in each of the Λ_i , $i = 1 \dots p-1$. There are, in the presence of one additional cointegrating vector, $(m+1-s) \times m$ parameters in Λ_p . Finally, there are $s \times (m+1-s)$ parameters in γ . Without additional cointegrating vectors, Λ_p has only $(m+1-s) \times (m-1)$ parameters.

6. The null and alternative hypotheses of the common feature test in terms of the pseudo-structural form become

$$\begin{aligned}
 H_0 : \delta' \Phi_1 &= \dots = \delta' \Phi_p = 0 \\
 H_A : \delta' \Phi_1 &= \dots = \delta' \Phi_p \neq 0,
 \end{aligned}$$

where the rank of δ is implicitly equal to either 1 or 2 as argued before. The test is performed using the following likelihood ratio type test:

$$LR_s = 2(l_{H_A} - l_{H_0}) \sim \chi^2(v),$$

where l_{H_A} and l_{H_0} denote the value of the log-likelihood function evaluated under the alternative and the null hypothesis, respectively. The degrees of freedom v are the difference between the number of parameters in the unrestricted mixed-frequency VAR (with cointegration) and the number of coefficients in the pseudo-structural system.

7. Finally, additional restrictions on the vectors of γ can be imposed if necessary for extracting a meaningful relationship between y and x .

The next section applies all these steps in the context of some empirical applications.

4.5 Applications

To illustrate the methods discussed in the previous sections, we extract two pairs of time series, once for the U.S. and once for Germany.⁸ In both cases y is the quarterly real gross national product (ref. GNPC96 and B1GM for the U.S. and Germany, respectively), seasonally adjusted, *gnp* from here on.⁹ For the U.S. the series is observed from 1948Q1 until 2011Q4, whereas for Germany it is observed from 1991Q1 until 2012Q1. For x we consider, for the same time span, the monthly seasonally adjusted industrial production index (ref. INDPRO and PROD, respectively), or *ipi* from here on. Since *gnp* is a quarterly variable and *ipi* is observed on a monthly basis we have that $m = 3$. Hence, $IP I_t^{(3)} = (ipi_t^{(3)}, ipi_{t-1/3}^{(3)}, ipi_{t-2/3}^{(3)})'$.

4.5.1 Common Feature Analysis

As a first step, we determine the VAR order in the log-level of the series by relying on information criteria. That implies that we compute the mixed-frequency VAR à la Ghysels (2012) disregarding the possibility that the $I(1)$ series may be

⁸From respectively <http://alfred.stlouisfed.org/> and <http://ec.europa.eu/eurostat/>.

⁹Consistent with the distinction between y , x and $X^{(m)}$, we change notation as far as empirical applications are concerned. From now on, lower case letters denote the low-frequency variable y and high-frequency observations x , whereas capital letters are used to describe the high-frequency observations in period t altogether, i.e., $X^{(m)}$.

cointegrated. After assuming the $I(1)$ -ness of both series involved (we indeed do not reject the null of a unit root for all series under consideration), we may test for the cointegrating rank r . As discussed above, due to the presence of r_0 prespecified cointegrating vectors, we follow the approach of Horvath and Watson (1995). In particular, since $m = 3$ we test the null hypothesis that merely the two known cointegrating vectors are present against the alternative of one additional cointegrating vector to capture the long-run relationship between gnp and $IPI^{(3)}$ (Götz et al., 2014). We can formulate the likelihood ratio test statistic as

$$LR_{r_0, r_1} = 2(l_{H_A} - l_{H_0}) = 2(l_{r=r_0+r_1} - l_{r=r_0}),$$

where l_{H_0} and l_{H_A} denote the value of the log-likelihood function evaluated under the null and the alternative hypothesis, respectively, i.e., when once only the prespecified cointegrating vectors are imposed and once the prespecified ones and one capturing potential cointegration between y and x . In terms of our empirical applications, this approach involves the log-likelihood of the following two regressions:

$$\text{Under } H_0 : \begin{pmatrix} \Delta gnp_t \\ \Delta ipi_t^{(3)} \\ \Delta ipi_{t-1/3}^{(3)} \\ \Delta ipi_{t-2/3}^{(3)} \end{pmatrix} \text{ on } \underbrace{\begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} gnp_{t-1} \\ ipi_{t-1}^{(3)} \\ ipi_{t-1-1/3}^{(3)} \\ ipi_{t-1-2/3}^{(3)} \end{pmatrix}}_{\begin{pmatrix} \Delta^{(1/3)} ipi_{t-1}^{(3)} \\ \Delta^{(1/3)} ipi_{t-1-1/3}^{(3)} \end{pmatrix}}, \begin{pmatrix} \Delta gnp_{t-1} \\ \Delta ipi_{t-1}^{(3)} \\ \Delta ipi_{t-1-1/3}^{(3)} \\ \Delta ipi_{t-1-2/3}^{(3)} \end{pmatrix}$$

and

$$\text{Under } H_A : \begin{pmatrix} \Delta gnp_t \\ \Delta ipi_t^{(3)} \\ \Delta ipi_{t-1/3}^{(3)} \\ \Delta ipi_{t-2/3}^{(3)} \end{pmatrix} \text{ on } \underbrace{\begin{pmatrix} 1 & -\theta & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} gnp_{t-1} \\ ipi_{t-1}^{(3)} \\ ipi_{t-1-1/3}^{(3)} \\ ipi_{t-1-2/3}^{(3)} \end{pmatrix}}_{\begin{pmatrix} gnp_{t-1} - \theta ipi_{t-1}^{(3)} \\ \Delta^{(1/3)} ipi_{t-1}^{(3)} \\ \Delta^{(1/3)} ipi_{t-1-1/3}^{(3)} \end{pmatrix}}, \begin{pmatrix} \Delta gnp_{t-1} \\ \Delta ipi_{t-1}^{(3)} \\ \Delta ipi_{t-1-1/3}^{(3)} \\ \Delta ipi_{t-1-2/3}^{(3)} \end{pmatrix}.$$

According to the SC and HQ a lag length of 1 is chosen for both countries under consideration. A VAR(2) is preferred for the U.S. using AIC.¹⁰ Consequently, we carry out the analysis with both lag lengths for the US in order to evaluate the robustness of the results.

It turns out that $LR_{r_0, r_1}^{US} = 8.62$ in the VAR(2) and 8.88 for the VAR(1) whereas $LR_{r_0, r_1}^{GER} = 25.41$ in the VAR(1) case such that with a Horvath and Watson (1995)-critical value of 18.37 we do not reject the null hypothesis of no cointegration between gnp and $IPI^{(3)}$ for the US, but do reject it for Germany. The cointegrating vector from the restricted VECM in the Johansen (1991)'s approach corresponding to r_1 is

$$\text{Germany : } gnp_t - \underset{(0.048)}{0.751} ipi_{t-2/3}^{(3)} \sim I(0).$$

This result will have an impact on how to set up the transformed VAR for both applications which, in turn, has an implication on how to test for common cyclical features thereafter (see Examples 2 and 3 from Section 4.3).

Let us now look at the presence of additional common cyclical feature restrictions. For the U.S., we first look at the VAR(1) in the log-level of the series which

¹⁰Note that 0 was specified as lower and 8 as upper bound on the lag order.

corresponds to the following transformed mixed-frequency VAR:

$$\begin{pmatrix} \Delta gnp_t \\ \Delta^{(1/3)} ipi_t^{(3)} \\ \Delta^{(1/3)} ipi_{t-1/3}^{(3)} \\ \Delta^{(1/3)} ipi_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} 0.005 \\ 0.001 \\ 0.002 \\ 0.001 \end{pmatrix} + \begin{pmatrix} 0 & 0.488 & 0.265 & 0 \\ 0 & 0.247 & 0.023 & 0 \\ 0 & 0.281 & 0.081 & 0 \\ 0 & 0.362 & 0.131 & 0 \end{pmatrix} \times \begin{pmatrix} gnp_{t-1} \\ \Delta^{(1/3)} ipi_{t-1}^{(3)} \\ \Delta^{(1/3)} ipi_{t-1-1/3}^{(3)} \\ ipi_{t-1-2/3}^{(3)} \end{pmatrix} + e_t$$

with a value of the log-likelihood of 3477.474. Because the rank of the transformed VAR coefficient matrix is 2, we know that, by construction, its orthogonal complement and then the common feature space δ is also of rank 2 in this four-dimensional system. Hence, a normalized common feature matrix that takes the form

$$\delta' = \begin{pmatrix} 1 & 0 & \times & \times \\ 0 & 1 & \times & \times \end{pmatrix},$$

where $\times$ denotes a parameter to be estimated, always exists. Premultiplying the VAR by

$$\begin{pmatrix} \delta' \\ 0_{(2 \times 2)} \quad I_{(2)} \end{pmatrix}$$

leads to the pseudo-structural form with the same likelihood because no restrictions are added and hence no test is possible.

We may further restrict δ and investigate whether there exists a common cyclical feature relationship between the growth rates of gnp and $IPI^{(3)}$. Restricting for instance the first possible relationship as in

$$\delta' = \begin{pmatrix} 1 & 0 & 0 & \times \\ 0 & 1 & \times & \times \end{pmatrix}$$

yields a log-likelihood of 3477.067 compared to 3477.474 from the previous mixed-frequency VAR imposing no common feature restrictions. This obviously means a non-rejection of the additional null hypothesis, $H_0 : \delta' \Phi_1 = \dots, \delta' \Phi_p = 0$ (see

section 4.5.1). The MF-SCCF relationship is now given by:

$$\Delta gnp_t = \underset{(0.0001)}{0.005} + \underset{(0.125)}{1.509} \Delta^{(1/3)} ipi_{t-2/3}^{(3)} + w_n,$$

i.e., a relationship between the quarterly growth rate of GNP and the monthly difference of IPI for the first month of the corresponding quarter. Still for the U.S., but with a VAR(2), as favoured by the AIC, we have a log-likelihood of 3480.36 in

$$\begin{pmatrix} \Delta gnp_t \\ \Delta^{(1/3)} ipi_t^{(3)} \\ \Delta^{(1/3)} ipi_{t-1/3}^{(3)} \\ \Delta^{(1/3)} ipi_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} 0.005 \\ 0.001 \\ 0.001 \\ 0.000 \end{pmatrix} + \begin{pmatrix} 0.043 & 0.478 & 0.255 & 0.004 \\ 0.353 & 0.175 & -0.010 & -0.311 \\ 0.259 & 0.227 & 0.045 & -0.031 \\ 0.113 & 0.330 & 0.100 & 0.021 \end{pmatrix} \times \begin{pmatrix} \Delta gnp_{t-1} \\ \Delta^{(1/3)} ipi_{t-1}^{(3)} \\ \Delta^{(1/3)} ipi_{t-1-1/3}^{(3)} \\ \Delta^{(1/3)} ipi_{t-1-2/3}^{(3)} \end{pmatrix} + \begin{pmatrix} 0 & -0.016 & 0.027 & 0 \\ 0 & -0.085 & 0.011 & 0 \\ 0 & -0.096 & -0.046 & 0 \\ 0 & 0.017 & -0.035 & 0 \end{pmatrix} \begin{pmatrix} gnp_{t-2} \\ \Delta^{(1/3)} ipi_{t-2}^{(3)} \\ \Delta^{(1/3)} ipi_{t-2-1/3}^{(3)} \\ ipi_{t-2-2/3}^{(3)} \end{pmatrix} + e_t.$$

In this case there does not necessarily exist a matrix that annihilates all of the dynamics, namely the two matrices. For $s = 1$ we have a log-likelihood value of 3479.877, for $s = 2$ we have a log-likelihood value of 3474.863 in the pseudo-structural systems estimated by FIML. Consequently, we do not reject the null of $s = 2$ common feature vectors with a likelihood ratio test of $2(3480.36 - 3474.86) = 11 \sim \chi^2(8)$. In addition to $s = 2$, we are also able to restrict the first common feature vector as in the ($p = 1$)-case; The log-likelihood with this additional restriction is 3474.732. The relationship is

$$\Delta gnp_t = \underset{(0.0001)}{0.005} + \underset{(0.128)}{1.488} \Delta^{(1/3)} ipi_{t-2/3}^{(3)} + v_n,$$

such that very similar results as before emerge, but now with a formal test for MF-SCCF.

For Germany with $r_0 + r_1$ cointegrating vectors we have:

$$\begin{pmatrix} \Delta gnp_t \\ \Delta^{(1/3)} ipi_t^{(3)} \\ \Delta^{(1/3)} ipi_{t-1/3}^{(3)} \\ \Delta^{(1/3)} ipi_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} 0.102 \\ -1.594 \\ -1.008 \\ 0.166 \end{pmatrix} + \begin{pmatrix} 0 & 0.297 & 0.147 & 0 \\ 0 & 0.167 & -0.093 & 0 \\ 0 & 0.143 & 0.067 & 0 \\ 0 & 0.043 & 0.137 & 0 \end{pmatrix} \begin{pmatrix} gnp_{t-1} \\ \Delta^{(1/3)} ipi_{t-1}^{(3)} \\ \Delta^{(1/3)} ipi_{t-1-1/3}^{(3)} \\ ipi_{t-1-2/3}^{(3)} \end{pmatrix} + \begin{pmatrix} -0.010 \\ 0.162 \\ 0.103 \\ -0.017 \end{pmatrix} (gnp_{t-1} - 0.751 ipi_{t-1-2/3}^{(3)}) + v_t.$$

Due to the ranks of these matrices, we know that there always exists a vector that annihilates the dynamics. This is an application of the mixed form reduced rank approach of Hecq et al. (2006). The common feature relationship and the standard errors (in brackets beneath the estimators) obtained by FIML, however, show that there are no cofeature relationships between the gnp and $IPI^{(3)}$ series alone:

$$\Delta gnp_t = \begin{matrix} 0.006 \\ (0.030) \end{matrix} + \begin{matrix} 2.596 \\ (16.435) \end{matrix} \Delta^{(1/3)} ipi_t^{(3)} - \begin{matrix} 3.411 \\ (24.381) \end{matrix} \Delta^{(1/3)} ipi_{t-1/3}^{(3)} + \begin{matrix} 4.912 \\ (27.25) \end{matrix} \Delta^{(1/3)} ipi_{t-2/3}^{(3)} + w_n,$$

such that the only common cyclical feature relationship involves the growth rate of the gross national product which is consequently a white noise process.

Alternative to a common feature analysis in a mixed-frequency setup, the high-frequency variables can be temporally aggregated in order to work with a common (low-) frequency VAR. Note, however, that common features are not invariant to temporal aggregation. Time aggregation of high-frequency variables may not only cause common feature relationships, that are present in mixed frequencies, to disappear in a common frequency framework, it may also create additional cofeature vectors (Marcellino, 1999).

Two temporal aggregation schemes are considered here, Point-in-Time and Average sampling. The former takes a specific high-frequency observation as the corresponding low-frequency one and is usually applied to stock variables. The

latter simply uses the average of the high-frequency observations as the corresponding t -observation and is usually of interest when flow variables are dealt with (see Marcellino, 1999 or Silvestrini and Veredas, 2008). We only consider the U.S. data for the common feature analysis in these low-frequency VARs. The resulting elasticities (standard errors in brackets) of Δgnp with respect to $\Delta ipi_t^{(3)}$, $\Delta ipi_{t-1/3}^{(3)}$, $\Delta ipi_{t-2/3}^{(3)}$ and $\overline{\Delta IPI_t^{(3)}}$ are 0.7187 (0.0696), 0.4033 (0.0364), 0.279 (0.016) and 0.4097 (0.0328), respectively, where $\overline{IPI_t^{(3)}}$ refers to the average of the monthly $IPI^{(3)}$ -observations in period t . The corresponding common feature test statistics computed from the pseudo-structural forms of the low-frequency VARs are 10.09, 7.9, 31 and 4.48 for the respective temporal aggregation approaches. At the 5% significance level, the null hypothesis of one common feature is rejected in all cases, implying that there exist two independent, i.e., non-common and non-synchronous, cycles for the two variables under consideration. In other words, the commonalities that are present in the mixed-frequency setup are not visible anymore after time aggregation. Finally, note that the potential cofeature elasticities are also very different to the mixed-frequency setup.

4.5.2 Forecasting

In the remainder of this section a small forecasting exercise is presented to emphasize the potential benefits of employing a mixed-frequency approach when dealing with common cycles. Potential gains of incorporating common feature restrictions for forecasting already exist in the literature. Vahid and Issler (2002) show that the costs in terms of forecast accuracy associated with ignoring common feature relationships in VARs may be considerable. Athanasopoulos et al. (2011) also show that models imposing restrictions associated with common cycles outperform unrestricted VAR models for different sample sizes, especially if the data are generated by comparably severe rank restrictions. Anderson and Vahid (2010) argue that VARs "with cointegration and common cycles can be [...] viewed as observable factor models [...] that [...] have potential for forecasting" (page 5). In a set of Monte Carlo studies the authors illustrate that over- and underspecification of the associated rank restrictions deteriorate the forecast performance in almost all cases. While the cointegration structure seems to be more crucial in general, mis-specification of the common cycle rank restrictions particularly worsens the forecast accuracy in the short-run. In their conclusion, Anderson and Vahid (2010)

advise to consider reduced rank error-correction models for forecasting if there are compelling reasons to consider common trends or cycles in a particular data set.

We consider 28 out-of-sample one-step-ahead forecasts of the real U.S. gross national product growth rate corresponding to the period from 2005Q1 until 2011Q4. We recursively update the estimation period and select the lag length at each moment in time. Furthermore, we test for the number of common feature vectors at each moment in time. Note that we do not use real-time data, but final-vintage data. Indeed, real-time data would require a vintage-specific analysis of which models to employ, which lag length to choose, cointegration properties and so on (see Götz et al., 2012 for an example of mixed-frequency models in a real-time data set). As already investigated above, the null hypothesis of no cointegration cannot be rejected for the U.S. such that we assume no cointegration to be present for each estimation period under consideration.

We are interested in the forecast accuracy of three mixed-frequency stationary VAR models. First, the model without restrictions on the short-run parameters, denoted by MF_VAR1.FULL. As pointed out in Sections 4.2 and 4.3, this is the model suggested by Ghysels (2012). Second, the model imposing zero-restrictions corresponding to the absence of cointegration between *gnp* and $IPI^{(3)}$ in the transformed VAR, denoted by MF_VAR1.ZEROS. Finally, the model imposing, on top of the just mentioned zero-restrictions, a set of common feature vectors, denoted by MF_VAR1.CF. In order to investigate whether the development of such mixed-frequency VAR models leads to an improved forecast performance compared to VARs stemming from traditional aggregation methods, we confront our three mixed-frequency VARs with two low-frequency VARs, resulting from applying Average sampling to $IPI^{(3)}$ on the one hand and Point-in-Time sampling using the end-of-period observations, $ipi_t^{(3)}$, on the other hand. In both cases a VAR in first differences, denoted by AV_VAR1.DIFF and PIT_VAR1.DIFF, respectively, are considered. RMSFE errors are computed for each model.

As far as the lag length selection is concerned, it turns out that for both, the mixed-frequency and time aggregated models, a VAR(1) model in first differences is chosen by SC or HQ at each moment in time. When testing for common features in the mixed-frequency VARs, the null hypothesis of $s = 2$ common feature vectors could not be rejected for all estimation periods considered. Hence, MF_VAR1.CF imposes two common feature vectors at each step of the forecasting exercise, whereby the first vector is further restricted as done in the previous subsection. Even though no common feature could have been detected for the low-

frequency VAR at the 5% significance level at each moment in time, we include the models imposing one common feature vector into the analysis, i.e., AV_VAR1_1CF and PIT_VAR1_1CF. As a justification, although no common feature was detected for average sampled *IPI*⁽³⁾ at the 5% significance level, the null hypothesis of one common feature could not be rejected at the 1% significance level.

Figure 4.1 shows the forecasts of the real U.S. gross national product growth rate from 2005Q1 to 2011Q4 corresponding to each of the seven models just introduced as well as the actual data (solid black line). Note that the black dotted or dashed lines correspond to the mixed-frequency transformed VARs, the grey dotted or dashed ones to the VARs based on averaging the high-frequency variable and, finally, the grey solid lines to the VARs based on taking the last monthly observation of IPI per quarter (Point-in-Time).

Figure 4.1: Out-of-sample forecasts

Note: The figure displays out-of-sample one-step-ahead static forecasts of real gross national product growth in the U.S. from 2005Q1 to 2011Q4.

It becomes clear from Figure 4.1 that the forecasts based on the mixed-frequency models are, compared to the competing time aggregated models, able to track the actual data rather well, especially the trough associated with the financial crisis in 2008. The forecasts based on VARs stemming from temporally aggregating the high-frequency variables severely understate this drop in the growth rate of the gross national product. This is reflected in the RMSFEs summarized in Table 4.1.

Table 4.1: RMSFEs of the seven competing models

Model	RMSFE
MF_VAR1_ZEROS	0.005671
MF_VAR1_CF	0.005672
MF_VAR1_FULL	0.005766
PIT_VAR1_DIFF	0.006527
PIT_VAR1_1CF	0.006591
AV_VAR1_1CF	0.007589
AV_VAR1_DIFF	0.00766

Note: The figures represent the RMSFEs of the seven competing models: The model 1) involving no restrictions, 2) imposing zero-restrictions, 3) imposing zero-restrictions and two common features, 4)-5) based on Point-in-Time sampling with and without a common feature imposed, respectively, and 5)-6) based on averaging with and without a common feature imposed.

As is evident in the figures, the forecast accuracy of the mixed-frequency transformed VARs is significantly better than the one of the VARs based on classical temporal aggregation methods. Hence, in the case of predicting the real U.S. gross national product growth rate from 2005Q1 to 2011Q1 it seems advantageous to consider mixed-frequency VAR models instead of temporally aggregating the high-frequency variable.

4.6 Monte Carlo Simulations

Motivated by the results of the forecasting exercise in Section 4.5, we now investigate more closely the behavior of mixed-frequency VAR models in this section. Given a certain lag length, we are interested in the size and power of the LR type test in (4.7) or the one based on (4.9), depending on whether there are additional cointegrating vectors above the prespecified ones.

To this end we conduct a Monte Carlo study, in which we simulate a mixed-frequency VAR that matches some of the characteristics that we have obtained using U.S. data, namely a VAR of order $p = 2$ in the level of the $I(1)$ series and no additional cointegrating vectors above the known ones. Hence, it is a two-variable system with $m = 3$ for the high-frequency series, in which the following parameter

values in the transformed VAR representation are chosen:

$$\begin{pmatrix} \Delta y_t \\ \Delta^{(1/3)} x_t^{(3)} \\ \Delta^{(1/3)} x_{t-1/3}^{(3)} \\ \Delta^{(1/3)} x_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.05 \\ 0.05 \\ 0.05 \end{pmatrix} + \begin{pmatrix} 0.2 & 0.5 \\ -0.2 & 0.2 \\ 0.4 & 0.5 \\ 0.5 & 0.25 \end{pmatrix} \begin{pmatrix} 0.5 & 0.1 \\ 0.6 & 0.5 \\ 0.2 & -0.5 \\ 0.1 & 0.25 \end{pmatrix}' \\
 + \begin{pmatrix} \Delta y_{t-1} \\ \Delta^{(1/3)} x_{t-1}^{(3)} \\ \Delta^{(1/3)} x_{t-1-1/3}^{(3)} \\ \Delta^{(1/3)} x_{t-1-2/3}^{(3)} \end{pmatrix} + \begin{pmatrix} 0 & 0.2 & 0.5 & 0 \\ 0 & -0.2 & 0.2 & 0 \\ 0 & 0.4 & 0.5 & 0 \\ 0 & 0.5 & 0.25 & 0 \end{pmatrix} \times \\
 + \begin{pmatrix} y_{t-2} \\ \Delta^{(1/3)} x_{t-2}^{(3)} \\ \Delta^{(1/3)} x_{t-2-1/3}^{(3)} \\ x_{t-2-2/3}^{(3)} \end{pmatrix} + \begin{pmatrix} \tilde{u}_{1t} \\ \tilde{u}_{2t} \\ \tilde{u}_{3t} \\ \tilde{u}_{4t} \end{pmatrix},$$

where

$$\tilde{u}_t \sim \text{i.i.d. } N(\mathbf{0}, \Sigma_{\tilde{u}}) \text{ with } \Sigma_{\tilde{u}} = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.9 & 0.9 \\ 0.5 & 0.9 & 1 & 0.9 \\ 0.5 & 0.9 & 0.9 & 1 \end{pmatrix}.$$

The normalized cofeature matrix that annihilates the two coefficient matrices of this DGP is fixed to:

$$\delta' = \begin{pmatrix} 1 & 0 & -\frac{4}{3} & \frac{2}{3} \\ 0 & 1 & -1 & 1.2 \end{pmatrix}.$$

Note that because the DGP is generated as a mixed-frequency transformed VAR, we know the number of common feature vectors to be $s = 2$. Consequently, given a lag length p , whether it is the correct one ($p = 2$) or an under- or overestimated one, we can analyze size ($s = 2$) and power ($s = 3$ or $s = 4$) of the LR type test in (4.7).

As already mentioned before, an alternative to considering a mixed-frequency VAR is temporally aggregating the high-frequency variables in order to work in a common (low-)frequency setup. As a consequence, we mimic the scenario in which practitioners are unaware of the mixed-frequency nature of the underlying data and work with time aggregated data for $X^{(m)}$. Assuming they follow the

same testing procedure, i.e., use (4.7) to test for common feature relationships, it is of interest to analyze in how much the use of time aggregated data instead of high-frequency ones changes the results.

Note that as soon as data are temporally aggregated, the dynamics of the underlying system change completely (see Marcellino, 1999, Silvestrini and Veredas, 2008 or Lütkepohl, 2005). Indeed, an n -dimensional VAR(p) can be written in a final equation system (Zellner and Palm, 1974) such that each component is at most an ARMA($np, (n-1)p$) process. In the presence of s SCCF-restrictions the n -dimensional VAR(p) yields at most ARMA($(n-s)p, (n-s)p$) models (Cubadda et al., 2009). In our case, the VAR(2) becomes a VARMA(4,4) after temporal aggregation of $X^{(m)}$ (because $n = 4$ and $s = 2$), that can be written as VAR(∞). Note that for practical reasons, the VARMA(4,4) is usually approximated by a finite order VAR. Consequently, as soon as variables are temporally aggregated, we cannot refer to a certain lag length as correct, under- or overestimated anymore. Note that we do not derive the final equation system corresponding to the mixed-frequency transformed VAR because it depends not only on the particular parameter values chosen, but also on n, m, r and s , making it a tedious and challenging task that could be tackled in future work.

As already mentioned before, time aggregation also has an impact on the analysis of common cycles (Marcellino, 1999). For example, it may cause common feature relationships, that are present in mixed frequencies, to disappear in a common frequency framework as illustrated before. In our case, temporal aggregation of $X^{(m)}$ leads to a 2-dimensional low-frequency VAR, for which the corresponding common feature test can be performed and the rejection frequencies for $s = 1$ or $s = 2$ can be computed. The rejection frequencies for $s = 2$ represent power against a bivariate white noise system. The ones for $s = 1$, however, cannot clearly be interpreted as size or power, because we do not know whether one synchronous common feature is present after time aggregation (in which case it would measure size) or none (in which case it would correspond to power). Again, the precise impact of temporal aggregation in a mixed-frequency VAR on common cycles analyses is left for further research.

For both, the mixed-frequency and time aggregated models, we consider $T = 40$ and 100 corresponding to respectively 10 and 25 years of quarterly data. An additional burning-in period of 100 observations is used to initialize the process and are discarded after generating the data. All figures are based on 10,000 replications.

For the mixed-frequency models, Table 4.2 reports the rejection frequencies (in percent) associated with the likelihood ratio type test in (4.7). Concerning the common feature test statistics, we rely on the estimation of the squared canonical correlation reviewed in Section 4.4. We estimate mixed-frequency VARs of orders from one to six but without imposing the zero-restrictions in the last matrix, i.e., in Φ_p for $p = 1, \dots, 6$. These models are denoted by MF_VAR. Models imposing zero-restrictions are denoted by MF_VAR_0. As argued before, the size of the test should be interpreted as the rejection frequency for $s = 2$, power can be observed for $s = 3$ and $s = 4$.

Table 4.2: Rejection frequencies of the LR type test in (4.7); Mixed-frequency models

		$T = 40$					
	p	1	2	3	4	5	6
MF_VAR	$s = 1$	0.2	0.6	1	2.3	5.7	15.4
	$s = 2$	1.4	8.2	17.7	25.4	44.3	68.5
	$s = 3$	20.4	75.7	79.1	87.3	94.2	98.7
	$s = 4$	100	100	100	100	100	100
MF_VAR_0	$s = 1$	-	0.4	0.7	1.4	3.8	9.4
	$s = 2$	-	6.5	10.3	18.9	34.2	55.9
	$s = 3$	18	77.5	76.6	82.8	90.8	97.3
	$s = 4$	99.8	100	100	100	100	100
		$T = 100$					
	p	1	2	3	4	5	6
MF_VAR	$s = 1$	0.2	0.3	0.3	0.4	0.5	0.8
	$s = 2$	2	6.6	8.2	9.9	12.4	16.6
	$s = 3$	37	99.4	98.5	97.4	96.9	96.9
	$s = 4$	100	100	100	100	100	100
MF_VAR_0	$s = 1$	-	0.3	0.3	0.4	0.5	0.7
	$s = 2$	-	5.7	7.5	9.2	11.1	14.6
	$s = 3$	45.6	99.7	99	97.9	97.1	96.8
	$s = 4$	100	100	100	100	100	100

Note: The figures represent rejection frequencies (in percent) associated with the likelihood ratio type test in equation (4.7). The DGP is a VAR(2) in the level of two $I(1)$ series, where no additional cointegrating relationships above the known ones and two common feature vectors are present. Models with VAR orders from 1 to 6 are estimated where zero-restrictions, corresponding to the absence of additional cointegrating relationships, are once imposed (denoted MF_VAR_0) and once not imposed (denoted MF_VAR). The sample size is $T = 40$ (top) or $T = 100$ (bottom). Size should be interpreted as the rejection frequency for $s = 2$, power can be observed for $s = 3$ and $s = 4$.

Using the correct lag length $p = 2$, the size distortion is small, especially if one correctly accounts for the zeros in the last coefficient matrix. Size distortions are amplified when the lag length is overestimated. This is reflected by the figures in the columns corresponding to $p > 2$. The test is undersized when p is underestimated (see also Hecq et al., 2006). With respect to size unadjusted power, note that for $T = 40$ we reject the presence of 3 common feature vectors 75% of the cases when the lag length is chosen correctly and up to almost 100% of the cases when it is overestimated. When $p = 1$, however, the power deteriorates. The presence of 4 common feature vectors is rejected in all cases. As soon as the sample size increases, we reject both $s = 3$ and $s = 4$ in almost all cases (except for $p = 1$ and $s = 3$).

Table 4.3 reports the rejection frequencies of the low-frequency counterpart to the common feature test statistic in (4.7). The two time aggregation methods already introduced before are considered, i.e., Average and Point-in-Time sampling. Note that as far as the latter is concerned, we only report the results for Point-in-Time sampling on the end-of-period ($\text{PiT-}x_t^{(3)}$) or start-of-period ($\text{PiT-}x_{t-2/3}^{(3)}$) observations to save space (results available upon request). As argued before, note that the frequencies corresponding to $s = 2$ represent power against a bivariate white noise system, whereas the ones for $s = 1$ cannot clearly be interpreted as size or power. Furthermore, we cannot classify the lag lengths as correct, under- or overestimated anymore, because the dynamics of the underlying system change due to temporal aggregation of the high-frequency variable.

For $T = 40$, power against a bivariate white noise system is lower when Average sampling is employed (68-87%, depending on p) than when Point-in-Time sampling is used (81-95%, depending on p and whether the end- or start-of-period observation is concerned). When the sample size increases, this power is equal to 100% in almost all cases. The rejection frequencies corresponding to one common feature relationship, i.e., $s = 1$, depend on the temporal aggregation method chosen, p and T . For a given T , the frequencies increase with the lag length chosen for all three time aggregation approaches. Given p , the rejection frequencies also increase with the sample size (except for Average sampling and $p = 6$). Finally, given p and T , the rejection frequencies are almost always highest when Point-in-Time sampling on the start-of-period observation is performed, i.e., $\text{PiT-}x_{t-2/3}^{(3)}$ (the exception being $p = 1$ and $T = 100$).

Table 4.3: Rejection frequencies of the LR type test in (4.7); Low-frequency models

		$T = 40$					
	p	1	2	3	4	5	6
Average	$s = 1$	3.8	7.1	8.1	10.3	11.1	13.3
	$s = 2$	68.4	88.5	87.4	87.4	86.6	86.8
$\text{PiT}_t^{(3)}$	$s = 1$	4.5	9	9.8	14.1	14.4	16.5
	$s = 2$	81.3	97.2	96.2	97.6	97.1	96.8
$\text{PiT}_{t-2/3}^{(3)}$	$s = 1$	5	22.5	23.8	25.4	25.6	28.3
	$s = 2$	89.2	92.8	95.2	95	94.3	94.8
		$T = 100$					
	p	1	2	3	4	5	6
Average	$s = 1$	4.4	8.8	10.7	12.2	12.1	12.4
	$s = 2$	97.7	100	100	100	100	100
$\text{PiT}_t^{(3)}$	$s = 1$	11.9	21.3	22.1	30.8	28	27.5
	$s = 2$	100	100	100	100	100	100
$\text{PiT}_{t-2/3}^{(3)}$	$s = 1$	5.8	55.7	58.5	56.3	54.2	53.3
	$s = 2$	100	100	100	100	100	100

Note: The figures represent rejection frequencies (in percent) associated with the likelihood ratio type test in equation (4.7). The DGP is a VAR(2) in the level of two $I(1)$ series, where no additional cointegrating relationships above the known ones and two common feature vectors are present. Common low-frequency models obtained via three different temporal aggregation methods are considered: (a) Average sampling (denoted Average), (b) Point-in-Time sampling using the end-of-period observation (denoted $\text{PiT}_t^{(3)}$) and (c) Point-in-Time sampling using the start-of-period observation (denoted $\text{PiT}_{t-2/3}^{(3)}$). The sample size is $T = 40$ (top) or $T = 100$ (bottom). Frequencies for $s = 2$ represent power against a bivariate white noise. system.

4.7 Conclusion

In this paper a new approach to common cycle analyses in multivariate time series processes was proposed when variables are sampled at mixed frequencies. This test allows users to examine commonalities in the business cycle of different countries when the frequency of observations is not the same. It also helps to detect co-movements when the economic indicators are not available at the same sampling frequency as in the applications that were given in this paper.

It was shown that, in the presence of non-stationary variables, the user needs to account for two kinds of possible long-run relationships in a VAR; one stemming from the long-run relationship between the low- and high-frequency variables

and the others. A transformed VAR representation was introduced, from which the set of variables was derived that was used to construct an LR type test for the presence of common cyclical features. An empirical analysis and a small forecasting exercise with the quarterly *gnp* and *IPI*⁽³⁾ illustrated this new approach. Finally, a Monte Carlo study was presented in order to investigate the rejection frequencies of the previously mentioned LR type common feature test. Using a mixed-frequency VAR as DGP, this analysis was done for our mixed-frequency models and several common low-frequency models obtained by temporal aggregation of the high-frequency variables.

From the empirical section as well as the Monte Carlo experiments it became clear that if the underlying data are of a mixed-frequency nature, relying on mixed-frequency models instead of temporally aggregating the high-frequency variables may have advantageous effects. Note that, given the correct lag length, the common feature test behaved very well. Also, common cycles, whose existence was not necessarily visible when working with a low-frequency model, were detected. Consequently, accounting for them in the model lead to gains in terms of forecasting. In other words, if there are useful restrictions present in the variables, there is some gain in making use of them and incorporating them adequately in a mixed-frequency VAR model.

Some issues were not treated in this paper and are currently under study. A common trend/common cycle decomposition could be easily obtained given that the MF-SCCF can be seen as a special case of the PSCCF modeling. Naturally, the development of a strategy for higher dimensional systems is of importance. Finally, as mentioned in the Monte Carlo study, an extension of Cubadda et al. (2009) and Marcellino (1999) to a mixed-frequency transformed VAR is of interest in order to investigate and generalize the effects of temporal aggregation on the lag length selection and the analysis of common cycles in a mixed-frequency setup. A detailed study of the latter would also allow us to state what difference time aggregation of stock or flow variables makes in our framework. Furthermore, it could help answering the question whether mixed-frequency data help to detect the state of the business cycle earlier than with low-frequency data only.

Chapter 5

Nowcasting Causality in Mixed-Frequency Vector Autoregressive Models

This chapter introduces *nowcasting causality* as the mixed-frequency version of instantaneous causality. We analyze the relationship between nowcasting and Granger causality in a mixed-frequency VAR (MF-VAR hereafter) and illustrate its impact on the significance of high-frequency variables in mixed-frequency conditional models.¹

5.1 Mixed-Frequency Structural VARs

Economic time series are available at different frequencies. For two variables, let us denote by y_t , $t = 1, \dots, T$ the low-frequency variable and by $x_{t-i/m}^{(m)}$ the high-frequency variables with m high-frequency observations per low-frequency period t . The value of i indicates how many high-frequency observations, starting from the end of the low-frequency period, i.e., $i = 0$, we take into account. For the quarter/month-scenario $m = 3$ and for $t = 2012Q4$, $i = 1$ refers to November 2012; $x_{2012Q4-2/3}^{(3)}$ to October 2012 etc. While it has been the standard approach to address this issue by simply aggregating the high-frequency series, it potentially

¹This chapter is based on the paper Götz and Hecq (2014a) published in *Economics Letters*. The authors thank Michael Eichler, Daniela Osterrieder, Jean-Pierre Urbain and an anonymous referee, for useful suggestions and comments on earlier versions of the paper.

leads to a loss of information. Mi(xed) Da(ta) S(ampling) regressions (Ghysels et al., 2004) have been developed as means to preserve the information embedded in the higher frequencies without sacrificing parsimony of the model. Until recently, the MIDAS literature was limited to a simple regression framework, in which one of the low-frequency variables is chosen as the dependent variable. Since the work of Ghysels (2012) for stationary series and the extension of Götz et al. (2013) and Ghysels and Miller (2013) for the non-stationary and possibly cointegrated case, we can analyze the links between high- and low-frequency series in an 'observation-driven' (Cox et al., 1981) VAR framework treating all variables as endogenous. Note that the term 'observation-driven' refers to the fact that the variables entering the VAR are observable, contrary to the 'parameter-driven' state space model approach (Zadrozny, 1988, Mariano and Murasawa, 2003 or Seong et al., 2013), in which some variables are latent.

Letting $X_t^{(m)} = (x_t^{(m)}, x_{t-1/m}^{(m)}, \dots, x_{t-(m-1)/m}^{(m)})'$, an unobserved dynamic structural equations model for $Z_t = (y_t, X_t^{(m)})'$ is given by $A_c Z_t = A_1 Z_{t-1} + \dots + A_p Z_{t-p} + \varepsilon_t$, with a diagonal covariance matrix Σ_ε and where A_c pertains to contemporaneous relationships between the series.² We treat this model as a starting point to our analysis; it enables us to trace the impact of particular specifications on reduced-form VAR models. For a quarter/month-example and with $p = 1$ the model reads as

$$\begin{pmatrix} 1 & \beta_1 & \beta_2 & \beta_3 \\ \delta & 1 & -\rho_1 & -\rho_2 \\ 0 & 0 & 1 & -\rho_1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t^{(3)} \\ x_{t-1/3}^{(3)} \\ x_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} \rho_y & \phi_1 & \phi_2 & \phi_3 \\ \pi_1 & \rho_3 & 0 & 0 \\ \pi_2 & \rho_2 & \rho_3 & 0 \\ \pi_3 & \rho_1 & \rho_2 & \rho_3 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(3)} \\ x_{t-1-1/3}^{(3)} \\ x_{t-1-2/3}^{(3)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{pmatrix}, \quad (5.1)$$

where an AR(3) structure for the high-frequency process is assumed in this example;³ the relationship between matrices A_c and A_1 (Ghysels, 2012) is seen in (5.1).

²The analysis can be adapted to the presence of deterministic terms.

³Note that an AR(m) structure for the high-frequency process is often considered for finance applications, where, e.g., a daily volatility variable may depend on its value on the previous day, during the previous week and the entire previous month (Corsi, 2009).

The matrix A_c links contemporaneous values of y and $X^{(m)}$: $\beta_j \neq 0$ implies that y_t is affected by incoming observations of $X_t^{(m)}$, whereas $\delta \neq 0$ implies that the last monthly observation per quarter, e.g., March, of $X_t^{(m)}$ is influenced by y_t . The latter becomes interesting for studying policy analysis, where the high-frequency policy variable may react to current low-frequency conditions. Note that the entire matrix A_c could be non-zero (Ghysels, 2012); the various restrictions leading to (5.1) are imposed to simplify the discussion to come.

5.2 Nowcasting Causality in MF-VARS

Starting from (5.1) we get the unrestricted reduced VAR model $Z_t = A_1^* Z_{t-1} + u_t$

$$Z_t = \begin{pmatrix} \rho_y^* & \phi_1^* & \phi_2^* & \phi_3^* \\ \pi_1^* & a_{2,2}^* & a_{2,3}^* & a_{2,4}^* \\ \pi_2^* & a_{3,2}^* & a_{3,3}^* & a_{3,4}^* \\ \pi_3^* & a_{4,2}^* & a_{4,3}^* & a_{4,4}^* \end{pmatrix} Z_{t-1} + u_t, \quad (5.2)$$

where $A_1^* = A_c^{-1} A_1$, $u_t = A_c^{-1} \varepsilon_t$ and $\Sigma_u = A_c^{-1} \Sigma_\varepsilon A_c^{-1'}$. Equations (5.1) and (5.2) can also be thought of as approximations to a 'true' underlying high-frequency VAR, where y is of the same high frequency as x (Ghysels et al., 2013, p.13). It is clear that testing for the absence of Granger causality (GC hereafter) from $X^{(m)}$ to y corresponds to $\phi_1^* = \phi_2^* = \phi_3^* = 0$ jointly (Granger, 1969). Similarly, testing for Granger non-causality from y to $X^{(m)}$ is associated with the null hypothesis $\pi_1^* = \pi_2^* = \pi_3^* = 0$ (Ghysels et al., 2013 or Götz and Hecq, 2014b).

Importantly, GC is defined in terms of the low frequency, i.e., in terms of index t . Given the mixed-frequency nature of the variables under consideration, it may be of interest to analyze whether knowing the values of $X_t^{(m)}$ (respectively y_t) helps to predict y_t (respectively $X_t^{(m)}$). In the common-frequency setup this is referred to as instantaneous causality (Renault et al., 1998, or Lütkepohl, 2005, p.42), because t refers to the common frequency under consideration. We refer to instantaneous causality in the mixed-frequency case as *nowcasting causality*.

Formally, let Ω_t represent the set of information available at moment t (Mikosch, 1998) such that $x_t^{(m)} \in \Omega_t$, but, e.g., $x_{t+1/m}^{(m)} \notin \Omega_t$, and let Ω_t^W be the corresponding information set containing the information for all stochastic processes except W . Now, we denote by $P[y_{t+1} | \Omega_t^W]$ the best linear forecast of y_{t+1} based on Ω_t^W .

Likewise for $P[X_{t+1}^{(m)}|\Omega_t^W]$. We can now define the absence of nowcasting causality (NC hereafter) in a similar way as GC is defined in Eichler and Didelez (2009):

Definition 5.1. There exists no NC between y and $X^{(m)}$ if

$$P[y_{t+1}|\Omega_t \cup \Omega_{t+1}^y] = P[y_{t+1}|\Omega_t] \text{ and } P[X_{t+1}^{(m)}|\Omega_t \cup \Omega_{t+1}^{X^{(m)}}] = P[X_{t+1}^{(m)}|\Omega_t].$$

In other words, adding information about $X_{t+1}^{(m)}$ to the information set available at period t helps in predicting y_{t+1} or vice versa or both. This concept is fully symmetric, which is why we avoid notions pointing towards a direction of NC as, e.g., ' $X^{(m)}$ is not nowcasting causal for y '. Considering the 'unobserved' system (5.1), the absence of NC corresponds to the restrictions $\beta_1 = \beta_2 = \beta_3$ and $\delta = 0$ jointly. However, as explained below, in the reduced VAR model (5.2), testing for the absence of NC corresponds to checking restrictions on the covariance matrix Σ_u .

The question arises in what sense the presence or absence of NC in Σ_u has an impact on GC in (5.2). Let us use our quarter/month-example and consider various specifications of the matrices A_c and A_1 in (5.1) to investigate the impact of NC on GC from $X^{(m)}$ to y first:

- A) If $\phi_j \neq 0$ for some j we will end up with GC from $X^{(m)}$ to y in the VAR in (5.2), no matter whether NC is present or absent.
- B) Assume $\phi_j = 0 \forall j$.
 - i) In the absence of NC, i.e., $\beta_j = \delta = 0 \forall j$, we have that $\phi_j^* = 0 \forall j$ in (5.2).
 - ii) If $\beta_j \neq 0$ for some j , it turns out that $\phi_j^* \neq 0 \forall j$ in (5.2). In other words, the presence of NC through $\beta_j \neq 0$ implies GC from $X^{(m)}$ to y .
 - iii) If, however, $\beta_j = 0 \forall j$ and NC comes $\delta \neq 0$, then there is no GC from $X^{(m)}$ to y in the reduced-form VAR parameters:

$$A_1^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \pi_1 - \delta + \pi_2\rho_1 + \pi_3\tilde{\rho} & \rho_3 + \rho_1\rho_2 + \rho_1\tilde{\rho} & \rho_1\rho_3 + \rho_2\tilde{\rho} & \rho_3\tilde{\rho} \\ \pi_2 + \pi_3\rho_1 & \tilde{\rho} & \rho_3 + \rho_1\rho_2 & \rho_1\rho_3 \\ \pi_3 & \rho_1 & \rho_2 & \rho_3 \end{pmatrix}, \quad (5.3)$$

where $\tilde{\rho} = \rho_1^2 + \rho_2$. The converse is also true, i.e., whenever $\phi_j^* = 0 \forall j$ we have that $\beta_j = 0 \forall j$.

Adapting Cases A) and B)i)-iii) accordingly, i.e., switching the roles of δ and β_j and replacing ϕ_j^* by π_j^* , reveals that the same conclusions hold for GC from y to $X^{(m)}$.

However, as mentioned before and pointed out by Lütkepohl for instantaneous causality (2005, p. 42), our concept of NC is fully symmetric in the MF-VAR, i.e., if y_t is nowcasting causal for $X_t^{(m)}$, then $X_t^{(m)}$ is nowcasting causal for y_t . This becomes clear from examining $\Sigma_u = A_c^{-1} \Sigma_\varepsilon A_c^{-1'}$. Actually, in the VAR, there is no NC between $X_t^{(m)}$ and y_t if and only if the corresponding errors are uncorrelated (Lütkepohl, 2005, p.47). In the absence of NC, Σ_u is indeed block diagonal:

$$A_c = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -\rho_1 & -\rho_2 \\ 0 & 0 & 1 & -\rho_1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\Rightarrow A_c^{-1} \Sigma_\varepsilon A_c^{-1'} = \begin{pmatrix} \sigma_L & 0 & 0 & 0 \\ 0 & \sigma_H(1 + \rho_1^2 + \tilde{\rho}^2) & \rho_1 \sigma_H(1 + \tilde{\rho}) & \sigma_H \tilde{\rho} \\ 0 & \rho_1 \sigma_H(1 + \tilde{\rho}) & \sigma_H(\rho_1^2 + 1) & \rho_1 \sigma_H \\ 0 & \sigma_H \tilde{\rho} & \rho_1 \sigma_H & \sigma_H \end{pmatrix},$$

where σ_L and σ_H are the variances of ε_{1t} and ε_{jt} , $j = 2, 3, 4$, respectively, i.e., the entries of the diagonal matrix Σ_ε . Hence, the term NC merely describes non-zero correlations between a set of variables, and thereby does not allow the interpretation of causation in a certain direction between $X^{(m)}$ and y .⁴ Stated differently, in the reduced VAR we cannot test whether $\delta \neq 0$ or $\beta_j \neq 0$, we can merely check for the presence or absence of NC.

The preceding discussion implies that despite the presence of NC (due to $\delta \neq 0$), GC from $X^{(m)}$ to y may be absent in a reduced-form VAR. Consider again case B)iii) from above, in which $\delta \neq 0$ and $\beta_j = \phi_j = 0 \forall j$. Equation (5.3) illustrates the absence of GC in the direction of interest, whereas inspecting the corresponding

⁴Lütkepohl (2005, p.48) stresses that the "direction of [nowcasting] causation must be obtained from further knowledge [e.g., economic theory,] on the relationship between the variables".

covariance matrix for this case,

$$\Sigma_u = A_c^{-1} \Sigma_\varepsilon A_c^{-1'} = \begin{pmatrix} \sigma_L & -\delta\sigma_L & 0 & 0 \\ -\delta\sigma_L & \delta^2\sigma_L + \sigma_H(1 + \rho_1^2 + \tilde{\rho}^2) & \rho_1\sigma_H(1 + \tilde{\rho}) & \sigma_H\tilde{\rho} \\ 0 & \rho_1\sigma_H(1 + \tilde{\rho}) & \sigma_H(\rho_1^2 + 1) & \rho_1\sigma_H \\ 0 & \sigma_H\tilde{\rho} & \rho_1\sigma_H & \sigma_H \end{pmatrix} \quad (5.4)$$

with $\mathbb{E}[u_{1t}u_{2t}] = -\delta\sigma_L \neq 0$, reveals the presence of NC between $X^{(m)}$ and y . Similarly, y may not Granger cause $X^{(m)}$ although NC between the two series exists (through $\beta_j \neq 0$ for some j).

Testing for NC can be done using a standard Wald test on the $(1, j)$ -elements of $\widehat{\Sigma}_u$, i.e., $\widehat{\sigma}_{1,j} = \frac{1}{T} \sum_{t=1}^T \widehat{u}_{1t}\widehat{u}_{jt}$, $j = 2, 3, 4$, where $\widehat{u}_{jt}$ corresponds to the residual of equation j in the corresponding VAR. The test is asymptotically distributed as χ_m^2 (see, e.g., Hamilton, 1994, p. 301 or Lütkepohl, 2005, p. 104).

5.3 Granger and Nowcasting Causality in Mixed-Frequency Conditional Models and VARs

Let us now investigate the relationship between the MF-VAR and a single-equation dynamic mixed-frequency regression model. Partitioning Σ_u in the quarter/month VAR(1) example as

$$\Sigma_u = \begin{pmatrix} \sigma_{1,1} & \sigma'_{1,1} \\ \sigma_{\cdot 1} & \Sigma_{2:4} \end{pmatrix},$$

where

$$\sigma_{\cdot 1} = (\sigma_{2,1}, \sigma_{3,1}, \sigma_{4,1})' \text{ and } \Sigma_{2:4} = \begin{pmatrix} \sigma_{2,2} & \cdot & \cdot \\ \sigma_{3,2} & \sigma_{3,3} & \cdot \\ \sigma_{4,2} & \sigma_{4,2} & \sigma_{4,4} \end{pmatrix},$$

we can factorize the VAR in (5.2) into the conditional model for y_t given $X_t^{(m)}$ and the remaining marginal models:

$$\begin{aligned}
 y_t &= \sigma'_{\cdot 1} \Sigma_{2:4}^{-1} X_t^{(m)} + \left[\rho_y^* - \sigma'_{\cdot 1} \Sigma_{2:4}^{-1} \begin{pmatrix} \pi_1^* \\ \pi_2^* \\ \pi_3^* \end{pmatrix} \right] y_{t-1} \\
 &+ \left[\underbrace{\begin{pmatrix} \phi_1^* \\ \phi_2^* \\ \phi_3^* \end{pmatrix}'}_{\phi^*} - \sigma'_{\cdot 1} \Sigma_{2:4}^{-1} \underbrace{\begin{pmatrix} a_{2,2}^* & a_{2,3}^* & a_{2,4}^* \\ a_{3,2}^* & a_{3,3}^* & a_{3,4}^* \\ a_{4,2}^* & a_{4,3}^* & a_{4,4}^* \end{pmatrix}}_{A_{2:4}^*} \right] X_{t-1}^{(m)} + \\
 &\left[u_{1t} - \sigma'_{\cdot 1} \Sigma_{2:4}^{-1} \begin{pmatrix} u_{2t} \\ u_{3t} \\ u_{4t} \end{pmatrix} \right] \\
 &= \pi^* y_{t-1} + \underbrace{\sum_{i=0}^2 \theta_{i,0} x_{t-i/3}^{(3)}}_{\sigma'_{\cdot 1} \Sigma_{2:4}^{-1} X_t^{(m)}} + \underbrace{\sum_{i=0}^2 \theta_{i,1} x_{t-1-i/3}^{(3)}}_{(\phi^* - \sigma'_{\cdot 1} \Sigma_{2:4}^{-1} A_{2:4}^*) X_{t-1}^{(m)}} + u_t^*.
 \end{aligned} \tag{5.5}$$

Equation (5.5) is a standard mixed-frequency regression, the parameters of which being estimated unrestrictedly (what Foroni et al., 2012 call U-MIDAS) or using MIDAS restrictions (e.g., Ghysels et al., 2004). Equation (5.5) reveals that with a MF-VAR as the underlying data generating process, the conditional mixed-frequency model can merge the implications of GC and NC.

Indeed, Granger non-causality from $X^{(m)}$ to y in the MF-VAR implies $\phi^* = \mathbf{0}$, whereas nowcasting non-causality corresponds to $\sigma'_{\cdot 1} = \mathbf{0}$. Assuming that the VAR in (5.2), derived from (5.1), generates the data, let us consider an analyst who works with the mixed-frequency regression model in (5.5) instead. Three cases, in which the analyst may draw different conclusions, i.e., potentially spurious causality (Renault et al., 1998), from the parameter estimates of (5.5) than from the ones of (5.2), can be distinguished:

5.3.1 Case 1

In the absence of GC from $X^{(m)}$ to y in the MF-VAR (i.e., $\phi^* = \mathbf{0}$) it is still possible to obtain non-zero coefficients on $X_{t-1}^{(m)}$ in (5.5), namely if NC between y and $X^{(m)}$ is present (and hence $\sigma'_{\cdot 1} \neq \mathbf{0}$). This case corresponds to the autoregressive matrix

in (5.3) and Σ_u in (5.4) and is analyzed in detail via Monte Carlo simulations in Section 5.4.

As an illustration,⁵ consider the relationship between the monthly growth rate of the seasonally adjusted industrial production index (*ipi* hereafter), and the weekly growth rate of the seasonally adjusted stock of money M2 (*stm* hereafter). Because $m = 4$, we have $Z_t = (ipi_t, STM_t^{(4)})'$, where $STM_t^{(4)} = (stm_t^{(4)}, stm_{t-1/4}^{(4)}, stm_{t-2/4}^{(4)}, stm_{t-3/4}^{(4)})'$. Estimated on the period from January 1991 to December 2010 the MF-VAR(1) model gives

$$\widehat{Z}_t = \widehat{\mu} + \begin{pmatrix} 0.284 & 0.092 & -0.105 & -0.469 & 0.213 \\ (0.069) & (0.205) & (0.194) & (0.266) & (0.182) \\ -0.06 & -0.019 & 0.008 & -0.09 & -0.168 \\ (0.025) & (0.073) & (0.069) & (0.094) & (0.064) \\ -0.021 & 0.153 & 0.017 & -0.014 & 0.021 \\ (0.026) & (0.076) & (0.071) & (0.098) & (0.067) \\ 0.013 & 0.137 & -0.044 & -0.014 & 0.077 \\ (0.017) & (0.05) & (0.047) & (0.065) & (0.044) \\ -0.015 & 0.01 & -0.171 & 0.033 & 0.107 \\ (0.025) & (0.073) & (0.069) & (0.095) & (0.065) \end{pmatrix} Z_{t-1}$$

with standard errors being displayed in brackets. The Wald test for $H_0 : \phi_{1,1}^* = \phi_{1,2}^* = \phi_{1,3}^* = \phi_{1,4}^* = 0$ has a p -value of 0.2484. The Wald statistic for testing nowcasting non-causality à la Hamilton (1994), i.e., $\sigma'_{\cdot 1} = \mathbf{0}$, turns out to be 36.55 with a p -value of practically 0%. Thus, we do not reject Granger non-causality from $STM_t^{(4)}$ to *ipi* in the MF-VAR although NC between the two series is detected. A U-MIDAS regression as in (5.5) yields:

$$\begin{aligned} \widehat{ipi}_t = & 0.0038 + 0.196 ipi_{t-1} - 1.069 stm_t^{(4)} - 0.667 stm_{t-1/4}^{(4)} \\ & + 0.27 stm_{t-2/4}^{(4)} - 0.441 stm_{t-3/4}^{(4)} + 0.142 stm_{t-1}^{(4)} - 0.148 stm_{t-5/4}^{(4)} \\ & - 0.557 stm_{t-6/4}^{(4)} + 0.073 stm_{t-7/4}^{(4)}. \end{aligned}$$

Now, the p -value for testing whether $\theta_{0,1} = \theta_{1,1} = \theta_{2,1} = \theta_{3,1} = 0$ is 0.0821. Hence, in the conditional mixed-frequency regression and at the 10% significance level we would conclude the set of lagged money-variables to be statistically different from zero.

⁵All applications deal with data from the Federal Reserve Bank of St. Louis. Whenever non-stationarity in the time series is detected, we compute growth rates in order to achieve stationarity. Cointegration issues are neglected for simplicity, but the methodology can be extended to the non-stationary, possibly cointegrated, case (see Götz et al., 2013).

5.3.2 Case 2

The parameters in (5.5) are often estimated after MIDAS restrictions (e.g., Ghysels et al., 2004) have been imposed. Without going into too much detail, instead of estimating all θ -coefficients unrestrictedly, which could lead to parameter proliferation issues (especially for large m), the polynomial lag structure is hyper-parameterized to yield

$$y_t = \pi^* y_{t-1} + \beta \sum_{j=0}^1 \sum_{i=0}^2 w_{3j+i+1}(\gamma) x_{t-j-i/3}^{(3)} + \epsilon_t, \quad (5.6)$$

where $w(\gamma)$ is a weight function, for which different specifications are proposed in the literature (see, e.g., Ghysels et al., 2007). In the sequel we employ the two-dimensional, i.e., $\gamma = (\gamma_1, \gamma_2)'$, exponential Almon lag polynomial. Instead of estimating one weight function for all high-frequency variables as done in (5.6), one can also estimate a separate weight function for each set of high-frequency variables per t -period. Note that each weight function is characterized by a two-dimensional vector, i.e., $\gamma^j = (\gamma_1^j, \gamma_2^j)'$ for $j = 1, 2$:

$$y_t = \pi^* y_{t-1} + \beta_1 \sum_{i=0}^2 w_{i+1}^1(\gamma^1) x_{t-i/3}^{(3)} + \beta_2 \sum_{i=0}^2 w_{i+1}^2(\gamma^2) x_{t-1-i/3}^{(3)} + v_t. \quad (5.7)$$

Let us come back to the estimation of (5.5) instead of (5.2). In the absence of NC, but presence of GC from $X^{(m)}$ to y , i.e., $\sigma'_{\cdot 1} = \mathbf{0}$ and $\phi^* \neq \mathbf{0}$, (5.7) is probably more appropriate than (5.6): Due to nowcasting non-causality, coefficients on $X_t^{(m)}$ should equal zero, whereas the ones on the respective observations in period $t-1$ should be non-zero due to existing GC. This implies an abrupt jump in the coefficients from zero to a non-zero value. In (5.7) this is easily achieved when $\widehat{\beta}_1 = 0$, while (5.6) probably yields non-zero coefficients for some observations in $X_t^{(m)}$.

For an application of this case the data consists of the monthly variation of the seasonally adjusted civilian unemployment rate, as regressand, and the aforementioned *stm*, as regressor. The time period considered as well as m are the same as for the first set of data, i.e., $T = 240$ and $m = 4$. Testing for Granger non-causality in the MF-VAR(1) yields a p -value equal to 0.004, testing for nowcasting non-causality gives a p -value of 0.678, and checking for no lags in the mixed-frequency regression results in a p -value of 0.0059. In other words, there is

no NC between the two series, but $STM^{(4)}$ continues to be statistically significant when conditioning on the high-frequency variables.

Here it is of interest to analyze what happens if MIDAS restrictions are imposed and, in particular, if we estimate only one or two weight functions, i.e., estimate (5.6) or (5.7). Given the estimates of $\gamma = (\gamma_1, \gamma_2)'$ in (5.6) with $m = 4$, we can compute the corresponding aggregate version of M as $\sum_{j=0}^1 \sum_{i=0}^3 w_{4j+i+1}(\hat{\gamma}) stm_{t-j-i/4}^{(4)}$. This allows us to run a least squares regression and compute the standard error of the scale coefficient $\hat{\beta} = 31.05$ as well as its t -statistic.⁶ With $t_{\hat{\beta}} = 4.01$ the corresponding coefficients on the whole set of $STM^{(4)}$ -variables ($STM_t^{(4)}$ and $STM_{t-1}^{(4)}$) are jointly statistically different from zero. These coefficients can be obtained by multiplying the γ -dependent weights with the scale coefficient and are plotted as solid line in Figure 1. Crucially and contrary to the missing NC feature, some $STM^{(4)}$ -variables possess a non-zero coefficient ($stm_{t-3/4}^{(4)}$ and to a lesser extent $stm_{t-2/4}^{(4)}$). If, however, (5.7) with $m = 4$ is estimated, $t_{\hat{\beta}_1} = 1.34$ and $t_{\hat{\beta}_2} = 3.84$ implying that $\hat{\beta}_1$, corresponding to the instantaneous $STM^{(4)}$ -variables, is not significantly different from zero. In other words, the nowcasting non-causality feature is preserved when estimating two separate weight functions. This is reflected by the associated coefficients plotted as dotted line in Figure 1.

5.3.3 Case 3

Suppose $X^{(m)}$ Granger causes y in the MF-VAR, i.e., $\phi^* \neq \mathbf{0}$, and NC is present as well, i.e., $\sigma'_{.1} \neq \mathbf{0}$, with the additional restriction that $\phi^* = \sigma'_{.1} \Sigma_{2:4}^{-1} A_{2:4}^*$. In this unlikely case $\widehat{\theta}_{i,1}$, $i = 0, 1, 2$, from (5.5) could be statistically indistinguishable from zero.

We consider the quarterly growth rate of the seasonally adjusted real gross national product (*gnp* hereafter) as the dependent variable; the growth rate of *ipi* is the independent variable such that $m = 3$. This pair of time series is observed from 1948Q3 to 2011Q4. After estimating a VAR(1) we find that both GC in the MF-VAR and NC are detected (p -values ≈ 0). Estimation of the single-equation

⁶Due to non-identification of the weight-specifying parameter γ under the null hypothesis, we would have to follow the approach of Davies (1987) and compute Hansen's (1996) p -value. Although not presented and reported here, the results do not differ qualitatively. The same holds for the estimation of (5.7).

Figure 5.1: Coefficients on $STM_t^{(4)}$ and $STM_{t-1}^{(4)}$ with one or two separate weight functions

Note: This figure shows the coefficients, i.e., their respective weights $\times$ scale coefficient, of $stm_t^{(4)}, \dots, stm_{t-3/4}^{(4)}$ and $stm_{t-1}^{(4)}, \dots, stm_{t-7/4}^{(4)}$, once when only one function is computed on the whole set of variables (solid line) and once when two separate weight functions, one for each set of 4 variables (dotted line), are computed. The weight functions are obtained using the two-dimensional Almon Lag Polynomial (see, e.g., Ghysels et al., 2007). The estimates are $\hat{\beta} = 31.05$, $\hat{\gamma} = (9.34, -0.75)'$ for the solid line and $(\hat{\beta}_1, \hat{\beta}_2)' = (12.6, 27.2)'$, $\hat{\gamma}^1 = (1.4, -0.27)'$ and $\hat{\gamma}^2 = (3.75, -1.18)'$ for the dotted line. Due to the fact that the scale coefficient corresponding to $stm_t^{(4)}, \dots, stm_{t-3/4}^{(4)}$, i.e., $\hat{\beta}_1$, is statistically not different from zero, its estimate is set to zero.

mixed-frequency regression leads to

$$\begin{aligned} \widehat{gnp}_t = & \frac{0.0053}{(0.0005)} - \frac{0.12}{(0.064)} gnp_{t-1} + \frac{0.211}{(0.04)} ipi_t^{(3)} - \frac{0.045}{(0.062)} ipi_{t-1/3}^{(3)} + \\ & \frac{0.253}{(0.047)} ipi_{t-2/3}^{(3)} + \frac{0.042}{(0.06)} ipi_{t-1}^{(3)} - \frac{0.067}{(0.062)} ipi_{t-4/3}^{(3)} + \frac{0.07}{(0.039)} ipi_{t-5/3}^{(3)}. \end{aligned}$$

Testing for $\theta_{0,1} = \theta_{1,1} = \theta_{2,1} = 0$, however, amounts to a p -value of 0.2997. Consequently, estimating the conditional mixed-frequency regression instead of the mixed-frequency VAR leads to $IPI_{t-1}^{(3)} = (ipi_t^{(3)}, ipi_{t-1/3}^{(3)}, ipi_{t-2/3}^{(3)})'$ being statistically insignificant.

5.4 Monte Carlo Simulations

We analyze Case 1 presented in the previous section by generating the VAR(1) in (5.2) from the following dynamic structural equations model:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \delta & 1 & -0.4 & 0.2 \\ 0 & 0 & 1 & -0.4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t^{(3)} \\ x_{t-1/3}^{(3)} \\ x_{t-2/3}^{(3)} \end{pmatrix} = \begin{pmatrix} 0.5 & 0 & 0 & 0 \\ 0.3 & 0.6 & 0 & 0 \\ 0.3 & -0.2 & 0.6 & 0 \\ 0.3 & 0.4 & -0.2 & 0.6 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(3)} \\ x_{t-1-1/3}^{(3)} \\ x_{t-1-2/3}^{(3)} \end{pmatrix} + \varepsilon_t, \quad (5.8)$$

where $\sigma_L = \sigma_H = 1$. Since $\beta_j = \phi_j = 0 \forall j$ and $\delta \neq 0$, A_1^* has the same structure as in (5.3). We test for Granger and nowcasting non-causality in the MF-VAR using a standard Wald test on $\hat{\phi}_j^* \forall j$ for the former and the one described at the end of Section 5.2 for the latter. Furthermore, we test $\theta_{0,1} = \theta_{1,1} = \theta_{2,1} = 0$ in (5.5), i.e., the significance of $X_{t-1}^{(3)}$ in the single-equation mixed-frequency regression. We do so for $\delta \in [-0.95, 0.95]$ and $T = 50, 250$. Figure 2 depicts the various rejection frequencies at a 5% significance level and are all based on 10,000 replications using GAUSS12.

Due to the absence of GC independent from δ the corresponding rejection frequencies for $T = 250$ almost equal the nominal size of 5%, whereas the test is oversized for $T = 50$. For $\delta = 0$, the test for NC has an actual size almost identical to the nominal one for both sample sizes. The rejection frequencies increase quickly (more rapidly for large T) as we turn away from the null. As expected, the test for $\theta_{0,1} = \theta_{1,1} = \theta_{2,1} = 0$ in (5.5) behaves similarly to the one for NC. It rejects close to 5% of the times for $\delta = 0$ and more often as $|\delta|$ increases. However, its rejection frequencies are lower than the ones for NC for both sample sizes: From (5.5), we have that $\sum_{i=0}^2 \theta_{i,1} x_{t-1-i/3}^{(3)} = (\phi^* - \sigma'_{\cdot 1} \Sigma_{2:4}^{-1} A_{2:4}^*) X_{t-1}^{(3)}$, which, for a DGP characterized by an autoregressive matrix as in (5.3), boils down to

Figure 5.2: Rejection frequencies of the various tests

Note: This figure shows the rejection frequencies for Granger non-causality (solid lines) and nowcasting non-causality (dashed lines) in the MF-VAR as well as for the insignificance of $X_{t-1}^{(m)}$ in the univariate mixed-frequency regression (dotted lines) at the 5% level. The data are generated from (5.8) with $\delta \in [-0.95, 0.95]$ (horizontal axis). The sample sizes considered are $T = 50$ (black lines) and 250 (grey lines).

$\left(\delta \frac{\rho_3}{\delta^2 + 1} \quad 0 \quad 0 \right) X_{t-1}^{(3)}$. With $\rho_3 = 0.6$ fixed, we know that for reasonable values of δ the coefficient of $x_{t-1}^{(3)}$ is going to be different from zero.⁷

⁷ $\lim_{|\delta| \rightarrow \infty} \delta \frac{\rho_3}{\delta^2 + 1} = 0$ such that for δ large enough the corresponding coefficient converges to zero. However, for $|\delta| \leq 5$ we have that $\theta_{0,1} > 0.1$.

Chapter 6

Testing for Granger Causality in Large Mixed-Frequency VARs

In this chapter we analyze Granger causality (Granger, 1969) testing in a mixed-frequency VAR, as originally proposed by Ghysels (2012), where the difference in sampling frequencies of the variables is large.

Given a realistic sample size, the number of high-frequency observations per low-frequency period leads to parameter proliferation problems in case we attempt to estimate the model unrestrictedly. We propose two approaches to solve this problem, reduced rank restrictions and a Bayesian mixed-frequency VAR. For the latter, we extend the approach in Banbura et al. (2010) to a mixed-frequency setup. We compare these methods to a common aggregated low-frequency model as well as to the unrestricted VAR in terms of their Granger non-causality testing behavior using Monte Carlo simulations. The techniques are illustrated in an empirical application involving daily realized volatility and monthly business cycle fluctuations.¹

¹This chapter is based on the paper Götz and Hecq (2014b). The authors particularly thank Eric Ghysels for many fruitful discussions, comments and suggestions on earlier versions of the paper. Moreover, the authors thank Daniela Osterrieder, Lenard Lieb, Stephan Smeekes, Jörg Breitung, Roman Liesenfeld and participants of the Workshop on Advances in Quantitative Economics in Maastricht 2013, the CEIS Seminar in Rome 2013, the 33rd International Symposium on Forecasting in Seoul 2013, the 24th (EC)² conference in Cyprus 2013, the 22th SNDE Annual Symposium in New York 2014, the 8th ECB Workshop on Forecasting Techniques, the Cologne Graduate School Research Seminar in June 2014 and the 68th European Meeting of the Econo-

6.1 Introduction

Economic time series are published at various frequencies. While higher frequency variables used to be aggregated (e.g., Silvestrini and Veredas, 2008), it has become more and more popular to consider models that take into account the difference in frequencies of the processes under consideration. As argued extensively in the mixed-frequency literature (e.g., Ghysels et al., 2007), working in a mixed-frequency setup instead of a common low-frequency one is advantageous due to the potential loss of information in the latter scenario and feasibility of the former through MI(xed) DA(ta) S(ampling) regressions (Ghysels et al., 2004), even in the presence of many high-frequency variables compared to the number of observations.

Until recently, the MIDAS literature was limited to the single-equation framework, in which one of the low-frequency variables is chosen as the dependent variable and the high-frequency ones are in the regressors. Since the work of Ghysels (2012) for stationary series and the extension of Götz et al. (2013) and Ghysels and Miller (2013) for the non-stationary and possibly cointegrated case, we can analyze the link between high- and low-frequency series in a VAR system treating all variables as endogenous. Ghysels et al. (2013) define causality in such a mixed-frequency VAR and develop a corresponding test statistic. The authors, however, make an implicit assumption on the variables involved, which does not hold in many economic applications: The number of high-frequency observations within a low-frequency period is rather small, e.g., as in a year/quarter- or quarter/month-example.

In this paper we analyze the finite sample behavior of Granger non-causality tests when the number of high-frequency observations per low-frequency period is large, e.g., as in a month/working day-example (like the application presented in this chapter). To potentially avoid the proliferation of parameters we consider two parameter reduction techniques, reduced rank restrictions and a Bayesian mixed-frequency VAR. With respect to the latter we show how to properly extend the dummy observation approach of Banbura et al. (2010) to a mixed-frequency setting.² Importantly, due to stacking the high-frequency variables in the mixed-frequency VAR (Ghysels, 2012), their approach cannot be applied directly such

metric Society in Toulouse 2014 for useful suggestions and comments on earlier versions of the paper.

²Banbura et al. (2010) refer to these variables as dummy observations. To avoid confusion between high- and low-frequency *observations* and auxiliary *variables*, we use the term 'auxiliary dummy variables'.

that a properly adapted choice of auxiliary dummy variables corresponding to the prior moments is required. The aforementioned approaches are compared to a common low-frequency VAR, obtained by temporally aggregating the high-frequency variable (Breitung and Swanson, 2002), and the unrestricted model, which is expected to suffer from a parameter proliferation problem.

The rest of the paper is organized as follows. In Section 6.2 notations are introduced, the mixed-frequency VAR (MF-VAR hereafter) for our specific case at hand is presented and Granger (non-)causality is defined formally. Section 6.3 discusses the approaches to reduce the number of parameters to be estimated, whereby reduced rank restrictions (Section 6.3.1) as well as Bayesian MF-VARs (Section 6.3.2) are presented in detail. The finite sample performances of these tests are analyzed via a Monte Carlo experiment in Section 6.4. An empirical example with U.S. data on the monthly industrial production index and daily volatility in Section 6.5 illustrates the results. Section 6.6 gives concluding remarks.

6.2 Causality in MF-VARs

6.2.1 Notation

Let us start from a two variable mixed-frequency system, where y_t , $t = 1, \dots, T$, is the low-frequency variable and $x_{t-i/m}^{(m)}$ are the high-frequency variables with m high-frequency observations per low-frequency period t . Contrary to Ghysels et al. (2013), we assume m to be rather large as in a year/month- or month/working day-example. We also assume m to be constant rather than varying with t .³ The value of i indicates the specific high-frequency observation under consideration, ranging from the beginning of each t -period ($x_{t-(m-1)/m}^{(m)}$) until the end ($x_t^{(m)}$ with $i = 0$). These notational conventions have become standard in the mixed-frequency literature and are similar to the ones in Götz et al. (2014), Clements and Galvão (2008, 2009) or Miller (2012).

Furthermore, let $\underline{W}_t = (W'_{t-1}, W'_{t-2}, \dots, W'_{t-p})'$, i.e., it stacks the last p low-frequency lags of any process W . Finally, $\mathbf{0}_{i \times j}$ ($\mathbf{1}_{i \times j}$) denotes an $(i \times j)$ -matrix of zeros (ones), I_i is an identity matrix of dimension i , $\otimes$ represents the Kronecker product and vec corresponds to the operator stacking the columns of a matrix.⁴

³As long as m is deterministic, even time-varying frequency discrepancies do not pose a problem on a theoretical level. However, the assumption of constant m simplifies the notation greatly (Ghysels, 2012).

⁴Extensions towards representations of higher dimensional multivariate systems as in Ghysels et al. (2013) can be considered, but are left for further research here. Analyzing Granger causality

6.2.2 MF-VARS

Considering each high-frequency variable such that

$$X_t^{(m)} = (x_t^{(m)}, x_{t-1/m}^{(m)}, \dots, x_{t-(m-1)/m}^{(m)})',$$

a dynamic structural equations model for the stationary multivariate process $Z_t = (y_t, X_t^{(m)})'$ is given by $A_c Z_t = c + A_1 Z_{t-1} + \dots + A_p Z_{t-p} + \varepsilon_t$,⁵ where Σ_ε is a diagonal covariance matrix and A_c pertains to contemporaneous relationships between the series. Note that the parameters in A_c are related to the ones in A_1 due to stacking the high-frequency observations $X_t^{(m)}$ in Z_t (Ghysels, 2012). Explicitly, for a lag length of $p = 1$ the model reads as:

$$\begin{pmatrix} 1 & \beta_1 & \beta_2 & \dots & \beta_m \\ \delta_1 & 1 & -\rho_1 & \dots & -\rho_{m-1} \\ \delta_2 & 0 & 1 & \dots & -\rho_{m-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_m & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{m+1} \end{pmatrix} + \begin{pmatrix} \rho_y & \theta_1 & \theta_2 & \dots & \theta_m \\ \psi_2 & \rho_m & \dots & \dots & 0 \\ \psi_3 & \rho_{m-1} & \rho_m & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_{m+1} & \rho_1 & \rho_2 & \dots & \rho_m \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{(m+1)t} \end{pmatrix}. \quad (6.1)$$

The matrix A_c links contemporaneous values of y and x : $\beta_j \neq 0$ implies that y_t is affected by incoming observations of $X_t^{(m)}$, whereas $\delta_j \neq 0$ implies that the high-frequency observations are influenced by y_t (see Götz and Hecq, 2014a). The latter

among more than two variables inherently leads to multi-horizon causality (see Lütkepohl, 1993 among others). The latter implies the potential presence of a causal chain: For example, in a trivariate system, X may cause Y through an auxiliary variable Z . To abstract from that scenario, Ghysels et al. (2013) often consider cases, in which high- and low-frequency variables are grouped and causality patterns between these groups, viewed as a bivariate system, are analyzed. They study the presence of a causal chain and multi-horizon causality in a Monte Carlo analysis though.

⁵Compared to Ghysels (2012) we put the low-frequency variable first.

becomes interesting for studying policy analysis, where the high-frequency policy variable(s) may react to current low-frequency conditions (see Ghysels, 2012 for details). Note that an $AR(m)$ model is assumed for the high-frequency process. If instead it is assumed to follow an $AR(q)$ model with $q < m$, one should set $\rho_{q+1} = \dots = \rho_m = 0$ in (6.1).

Pre-multiplying (6.1) by A_c^{-1} we get to the mixed-frequency reduced-form $VAR(p)$, i.e., the aforementioned MF-VAR, model:

$$\begin{aligned} Z_t &= A_c^{-1}c + A_c^{-1}A_1Z_{t-1} + \dots + A_c^{-1}A_pZ_{t-p} + A_c^{-1}\varepsilon_t \\ &= \mu + \Gamma_1Z_{t-1} + \dots + \Gamma_pZ_{t-p} + u_t \end{aligned}$$

or

$$Z_t = \mu + B'\underline{Z}_t + u_t, \quad (6.2)$$

where $B = (\Gamma_1, \Gamma_2, \dots, \Gamma_p)'$. Equation (6.2) is easy to estimate on small systems. As an example, let us explicitly write a $VAR(1)$ as

$$\begin{aligned} \begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} &= \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{m+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,m+1} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m+1,1} & \gamma_{m+1,2} & \cdots & \gamma_{m+1,m+1} \end{pmatrix}}_{\Gamma_1} \\ &\quad \times \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{(m+1)t} \end{pmatrix}, \end{aligned} \quad (6.3)$$

$$u_t \sim N(\mathbf{0}_{((m+1) \times 1)}, \Sigma_u), \quad \Sigma_u = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,m+1} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m+1,1} & \cdots & \cdots & \sigma_{m+1,m+1} \end{pmatrix}.$$

6.2.3 Granger Causality in MF-VARS

Let the information set generated by the collection of sigma-fields $F_t = \sigma(Z_s, s \leq t), t \geq 0$ be denoted by Ω_t . Furthermore, let Z_t be adapted to that filtration such that Ω_t represents the set of information available at moment t (Mikosch, 1998). Ω_t^W denotes the information set containing the information for all stochastic process except W up to moment t . $P[X_{t+h}^{(m)}|\Omega_t^W]$ is the best linear forecast of $X_{t+h}^{(m)}$ based on Ω_t^W and likewise for $P[y_{t+h}|\Omega_t^W]$. Granger non-causality is defined as follows (Eichler and Didelez, 2009):

Definition 6.1. y does not Granger cause $X^{(m)}$ if

$$P[X_{t+1}^{(m)}|\Omega_t^y] = P[X_{t+1}^{(m)}|\Omega_t].$$

Similarly, $X^{(m)}$ does not Granger cause y if

$$P[y_{t+1}|\Omega_t^{X^{(m)}}] = P[y_{t+1}|\Omega_t].$$

In other words, y does not Granger cause $X^{(m)}$ if past information of the low-frequency variable do not help in predicting current (or future) values of the high-frequency variables and vice versa (Granger, 1969). In terms of (6.3), testing for Granger non-causality implies the following null (and alternative) hypotheses:

- y does not Granger cause $X^{(m)}$

$$\begin{aligned} H_0 : \gamma_{2,1} = \gamma_{3,1} = \dots = \gamma_{m+1,1} &= 0, \\ H_A : \gamma_{i,1} &\neq 0 \text{ for at least one } i = 2, \dots, m+1. \end{aligned}$$

- $X^{(m)}$ does not Granger cause y

$$\begin{aligned} H_0 : \gamma_{1,2} = \gamma_{1,3} = \dots = \gamma_{1,m+1} &= 0, \\ H_A : \gamma_{1,i} &\neq 0 \text{ for at least one } i = 2, \dots, m+1. \end{aligned}$$

Ghysels et al. (2013) implicitly assume m to be rather small ($m = 3$ or 4 in their case) in order to estimate the MF-VAR and then test for Granger non-causality. With m large (say $m \geq 12$), we would need a much larger sample than is usually applicable for macroeconomic data sets to estimate the parameters and test for causality properly. Additional lags ($p > 1$) would further complicate the issue.

6.2.4 Remark on Nowcasting Causality

It becomes clear from Section 6.2.3 that Granger non-causality in both directions is defined in terms of the low frequency index t . Given the mixed-frequency nature of the variables under consideration, it may be of interest to analyze whether knowing the values of $x_{t-i/m}^{(m)}, i = 0, \dots, m-1$, helps to predict y_t , which is referred to as instantaneous causality in the common-frequency case (Renault et al., 1998, Lütkepohl, 2005, p.42, Hamilton, 1994, p. 301, Gianetto and Raïssi, 2012).

Götz and Hecq (2014a) have introduced the term nowcasting causality as the mixed-frequency analogue to instantaneous causality, because it amounts to predicting y_t using values of the high-frequency variables *within* period t . In terms of testing for nowcasting non-causality, Götz and Hecq (2014a) argue that $X_t^{(m)}$ is not nowcasting causal for y_t if and only if the corresponding errors of the data generating process of Z_t are uncorrelated (Lütkepohl, 2005). Indeed, testing for nowcasting non-causality corresponds to $\beta_j = \delta_j = 0 \forall j$ in (6.1). In that case, though, Σ_u is block diagonal with $\sigma_{1,j} = \sigma_{j,1} = 0, j = 2, \dots, m+1$, because $\Sigma_u = A_c^{-1} \Sigma_\varepsilon A_c^{-1'}$. Thus, we can test for nowcasting non-causality by using a Wald test on the $(1, j)$ -elements of $\widehat{\Sigma}_u$, i.e., $\widehat{\sigma}_{1,j} = \frac{1}{T} \sum_{t=1}^T \widehat{u}_{1t} \widehat{u}_{jt}, j = 2, \dots, m+1$, where $\widehat{u}_{jt}$ corresponds to the residuals of equation j .⁶

6.3 Parameter Reduction

This section presents techniques that we have considered, and evaluated through a Monte Carlo exercise, with the aim to reduce the amount of parameters to be estimated in the MF-VAR model. Two approaches are discussed in detail, reduced rank restrictions and a Bayesian MF-VAR.

There are many alternative approaches to reduce the number of parameters among which are principal components, Lasso (Tibshirani, 1996) or ridge regressions.⁷ The reason not to consider principal components is that it does not necessarily preserve the dynamics of the VAR under the null. To be more precise,

⁶In this paper we do not include degree of freedom corrections when computing estimates of covariance matrices due to their invariance asymptotically.

⁷Cubadda and Guardabascio (2012) analyze a so-called 'medium- N ' approach arguing that many of the results in the literature favor a number of predictors (N) that is considerably larger than in usual small-scale forecasting problems, but not too large for being forced to rely on double (T and N) asymptotic methods. The authors find, using Monte Carlo simulations, that, under the so-called Helland and Almoy (1994) condition, both principal components and partial least squares (to be discussed later) provide consistent estimates in such a medium- N framework as only the sample size diverges.

nothing prevents, e.g., the first and only principal component to be loading only on y implying that the remaining dynamics enter the error term. In other words, principal components may and will most likely not preserve the autoregressive matrices in (6.2). This affects, in a possibly drastic manner, the block of parameters we test on for Granger non-causality.

6.3.1 Reduced Rank Restrictions

Reduced Rank Regression Model

In order to reduce the number of parameters to estimate in the MF-VAR, we propose the following reduced rank regression model, in which we assume that $rk(B'_{X^{(m)}}) = r < m$ (Carriero et al., 2011) with $B'_{X^{(m)}}$ being obtained from B' by excluding the first columns of $\Gamma_1, \Gamma_2, \dots, \Gamma_p$. The model reads as

$$\begin{aligned} Z_t &= \mu + \gamma_{\cdot,1} y_t + \alpha \sum_{i=1}^p \delta'_i X_{t-i}^{(m)} + \nu_t \\ &= \mu + \gamma_{\cdot,1} y_t + \alpha \delta' \underline{X}_t^{(m)} + \nu_t, \end{aligned} \quad (6.4)$$

where $\gamma_{\cdot,1}$ is the $(m+1) \times p$ matrix containing the first columns of $\Gamma_1, \Gamma_2, \dots, \Gamma_p$ and α and $\delta = (\delta'_1, \dots, \delta'_p)'$ are $(m+1) \times r$ and $pm \times r$ matrices, respectively. Note that (6.4) can also be written in terms of $\underline{Z}_t$. Let us define $\underline{\Gamma}_i \equiv (\gamma_{\cdot,1}^{(i)}, \alpha \delta'_i)$, where $\gamma_{\cdot,1}^{(i)}, i = 1, \dots, p$, corresponds to the i^{th} column of $\gamma_{\cdot,1}$. Then, (6.4) is equivalent to

$$Z_t = \mu + \underline{B}' \underline{Z}_t + \nu_t, \quad (6.5)$$

where $\underline{B} = (\underline{\Gamma}_1, \dots, \underline{\Gamma}_p)'$. For $p = 1$ the models becomes

$$\begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} = \mu + \begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \\ \vdots \\ \gamma_{m+1,1} \end{pmatrix} y_{t-1} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m+1} \end{pmatrix} \delta'_1 \begin{pmatrix} x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \nu_t$$

or

$$\begin{aligned}
 &= \mu + \underbrace{\left[\begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \\ \vdots \\ \gamma_{m+1,1} \end{pmatrix}, \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m+1} \end{pmatrix} \right] \delta'_1}_{\underline{\Gamma}_1} \times \\
 &\quad \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \nu_t,
 \end{aligned}$$

where each $\alpha_i, i = 1, \dots, m+1$, is of dimension $1 \times r$ and where δ_1 is an $m \times r$ matrix. Hence, we could call $\delta' \underline{X}_t^{(m)}$ a vector of r observable high-frequency factors. Note that $r = m - s$, where s represents the rank reduction we are able to achieve within $\underline{X}_t^{(m)}$.⁸

In terms of parameter reduction, the unrestricted VAR in (6.2) requires $p(m+1)^2$ coefficients to be estimated in the autoregressive matrices, whereas the VAR under reduced rank restrictions in (6.4) or (6.5) needs $p(m+1) + r(m+1) + prm$ parameter estimates. As an example, assume $p = 1$ and $m = 20$. Then, if $r = 1, 2$ or 3 , there are, respectively, 62, 103 or 144 coefficients to be estimated in $\underline{\Gamma}_1$ instead of 441 in Γ_1 in (6.3). Note that we do not require y_{t-1} to be included in the same transmission mechanism as the x variables.

Obtaining Observable Factors

We use three ways to obtain the observable factors, canonical correlation analysis (CCA hereafter), partial least squares (PLS hereafter) and heterogeneous autoregressive (HAR hereafter) type restrictions.

⁸There are two ways to justify the reduced rank feature of the autoregressive matrix. First, at the model representation level we may assume that x follows an $AR(q)$ process with $q < m$ and that the last $m - q$ elements of each $X_{t-i}^{(m)}, i = 1, \dots, p$, have a zero coefficient in the equation for y_t . Plugging these restrictions into (6.2) results in a reduced rank of $B'_{X^{(m)}}$. Second, at the empirical level we can interpret the MF-VAR as an approximation of the VARMA obtained after the block marginalization of a high-frequency VAR for each variable. In this situation, reduced rank matrices may empirically not be rejected by the data.

CCA is based on analyzing the eigenvalues and corresponding eigenvectors of

$$\hat{\Sigma}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{X}}^{(m)}}^{-1} \hat{\Sigma}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{Z}}} \hat{\Sigma}_{\underline{\tilde{Z}} \underline{\tilde{Z}}}^{-1} \hat{\Sigma}_{\underline{\tilde{Z}} \underline{\tilde{X}}^{(m)}} \quad (6.6)$$

or, similarly, of the symmetric matrix

$$\hat{\Sigma}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{X}}^{(m)}}^{-1/2} \hat{\Sigma}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{Z}}} \hat{\Sigma}_{\underline{\tilde{Z}} \underline{\tilde{Z}}}^{-1} \hat{\Sigma}_{\underline{\tilde{Z}} \underline{\tilde{X}}^{(m)}} \hat{\Sigma}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{X}}^{(m)}}^{-1/2}. \quad (6.7)$$

For a detailed discussion we refer the reader to Anderson (1951) or Vahid and Engle (1993) for the application to common dynamics. Note that $\hat{\Sigma}_{ij}$ represents the empirical covariance matrix of processes i and j . Furthermore, $\tilde{Z}$ and $\underline{\tilde{X}}^{(m)}$ indicate Z_t and $\underline{X}_t^{(m)}$, respectively, to be concentrated out by the variables that do not enter in the reduced rank regression, i.e., the intercept and y_{t-1} . Denoting by $\hat{V} = (v_1, v_2, \dots, v_r)$, with $v'_i v_j = 1$ for $i = j$ and 0 otherwise, the eigenvectors corresponding to the r largest eigenvalues of the matrix in (6.7), we obtain the estimators of α and δ as:

$$\begin{aligned} \hat{\alpha} &= \hat{\Sigma}_{\underline{\tilde{Z}} \underline{\tilde{Z}}}^{-1} \hat{\Sigma}_{\underline{\tilde{Z}} \underline{\tilde{X}}^{(m)}} \hat{\Sigma}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{X}}^{(m)}}^{-1/2} \hat{V} \\ \hat{\delta} &= \hat{\Sigma}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{X}}^{(m)}}^{-1/2} \hat{V}. \end{aligned}$$

Note that the estimation of the eigenvectors obtained from the canonical correlation analysis in (6.6) or (6.7) may, however, perform poorly with high-dimensional systems, because inversions of the large variance matrices $\hat{\Sigma}_{\underline{\tilde{Z}} \underline{\tilde{Z}}}^{-1}$ and $\hat{\Sigma}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{X}}^{(m)}}^{-1}$ are required.

As an alternative to CCA we use a PLS algorithm similar to the one used in Cubadda and Hecq (2011) or Cubadda and Guardabascio (2012). In order to make the solution of this eigenvalue problem invariant to scale changes of individual elements, we compute the eigenvectors associated with the largest eigenvalues of the matrix $\hat{D}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{X}}^{(m)}}^{-1/2} \hat{\Sigma}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{Z}}} \hat{D}_{\underline{\tilde{Z}} \underline{\tilde{Z}}}^{-1} \hat{\Sigma}_{\underline{\tilde{Z}} \underline{\tilde{X}}^{(m)}} \hat{D}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{X}}^{(m)}}^{-1/2}$ with $\hat{D}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{X}}^{(m)}}$ and $\hat{D}_{\underline{\tilde{Z}} \underline{\tilde{Z}}}$ being diagonal matrices having the diagonal elements of, respectively, $\hat{\Sigma}_{\underline{\tilde{X}}^{(m)} \underline{\tilde{X}}^{(m)}}$ and $\hat{\Sigma}_{\underline{\tilde{Z}} \underline{\tilde{Z}}}$ as their entries. The computation of $\hat{\alpha}$ and $\hat{\delta}$ works in a similar fashion as with CCA-based factors.

Finally, we may *impose* the presence of $r = 3p$ factors,⁹ inspired by the Corsi HAR-model (Corsi, 2009). For $i = 1, \dots, p$:

⁹To ensure that $r < m$ we assume hence that $p < \frac{1}{3}m$ in this case.

$$\delta_i = \begin{pmatrix} \mathbf{0}_{3(i-1) \times m} \\ \mathbf{1} & \mathbf{0}_{1 \times (m-1)} \\ \mathbf{1}_{1 \times (\frac{1}{4}m)} & \mathbf{0}_{(1 \times \frac{3}{4}m)} \\ \mathbf{1}_{1 \times m} \\ \mathbf{0}_{3(p-i) \times m} \end{pmatrix}' \Rightarrow \delta' \underline{X}_t^{(m)} = \begin{pmatrix} x_{t-1}^{(m)} \\ \sum_{i=0}^{\frac{1}{4}m-1} x_{t-1-i/m}^{(m)} \\ \sum_{i=0}^{m-1} x_{t-1-i/m}^{(m)} \\ \vdots \\ x_{t-p}^{(m)} \\ \sum_{i=0}^{\frac{1}{4}m-1} x_{t-p-i/m}^{(m)} \\ \sum_{i=0}^{m-1} x_{t-p-i/m}^{(m)} \end{pmatrix}.$$

For $m = 20$, i.e., the month/working day-example mentioned before, and $p = 1$ this corresponds to

$$\delta' \underline{X}_t^{(m)} = \begin{pmatrix} x_{t-1}^{(20)} \\ \sum_{i=0}^4 x_{t-1-i/20}^{(20)} \\ \sum_{i=0}^{19} x_{t-1-i/20}^{(20)} \end{pmatrix} \equiv \begin{pmatrix} x_{t-1}^D \\ x_{t-1}^W \\ x_{t-1}^M \end{pmatrix},$$

with x_t^D, x_t^W and x_t^M denoting daily, weekly and monthly measures, respectively. As noted in Ghysels and Valkanov (2012) and Ghysels et al. (2007), these HAR-type restrictions are a special case of MIDAS with step functions introduced in Forsberg and Ghysels (2007). Considering partial sums of regressors x as $X_t(K, m) = \sum_{i=0}^K x_{t-i/m}^{(m)}$, a MIDAS regression with M steps reads as

$$y_t = \mu + \sum_{j=1}^M \beta_j X_t(K_j, m) + \epsilon_t,$$

where $K_1 < \dots < K_M$. Alternative to this type of HAR-restrictions, we could use MIDAS restrictions as introduced in Ghysels et al. (2004). Indeed, the difference between HAR and MIDAS models can be small (Ghysels and Valkanov, 2012). However, step functions have the advantage of not requiring non-linear estimation methods, since the distributed lag pattern is approximated by a number of discrete steps. Also, implementing MIDAS restrictions and testing for Granger non-causality implies the well-known Davies (1987) problem, i.e., the parameters

determining the MIDAS weights (see Ghysels et al., 2004 for details) are not identified under the null hypothesis.¹⁰

Testing for Granger Non-Causality

Given r , we can test for Granger non-causality by defining R as the matrix that picks the set of coefficients we want to do inference on, i.e., $R\text{vec}(\hat{B})$, where $\hat{B}$ is the estimator of B . For a general construction of the matrix R in the presence of several low- and high-frequency variables, we refer the reader to Ghysels et al. (2013). The Wald test is then constructed as

$$\tilde{\xi}_W = \left[R\text{vec}(\hat{B}) \right]' (R\hat{\Omega}R')^{-1} \left[R\text{vec}(\hat{B}) \right], \quad (6.8)$$

with

$$\hat{\Omega} = (W'W)^{-1} \otimes \hat{\Sigma},$$

where $\hat{\Sigma} = \frac{1}{T}\hat{u}'\hat{u}$ is the empirical covariance matrix of the disturbance terms and W is the regressor set, i.e., an intercept, y_{t-1} and the high-frequency factors $\delta'X_t^{(m)}$. As illustrated in Ghysels et al. (2013), ξ_W is asymptotically $\chi_{\text{rank}(R)}^2$. A robust version of (6.8) is also implemented in the empirical section.

6.3.2 Bayesian MF-VARs

Restricted MF-VARs

No restrictions are placed on the coefficients in the autoregressive matrices, i.e., B , in (6.2). While this may seem reasonable for the first equation, i.e., the one for y_t , it is less clear for the remaining ones. Ghysels (2012) discusses the issue of parsimony in MF-VAR models by specifying the high-frequency process to allow a number of parameters that is independent from m . In particular, assuming x to

¹⁰To properly test for Granger non-causality in this case, a grid for the weight specifying parameter vector has to be considered and the corresponding Wald tests for each candidate have to be computed. Subsequently, one can calculate the supremum of these tests (Davies, 1987) and obtain an 'asymptotic p -value' using bootstrap techniques (see Hansen, 1996 or Ghysels et al., 2007 for details). While this approach is feasible, it is computationally more demanding due to applying bootstrapping within a relatively large (due to large m) regression. Admittedly, the HAR-type restrictions provide less flexibility than the MIDAS approach, yet rely on linear estimation of the model, which simplifies the analysis greatly and is therefore very appealing from an applied perspective.

follow an ARX(1) process with only one lag of y in the regressor set yields

$$x_{t-i/m}^{(m)} = \mu_{i+2} + \rho x_{t-(i+1)/m}^{(m)} + \pi y_{t-1} + v_{(i+2)t} \text{ for } i = 0, \dots, m-1. \quad (6.9)$$

Completing the system with the first equation of the unrestricted MF-VAR, i.e.,

$$y_t = \mu_1 + \sum_{k=1}^p \left[\gamma_{1,1}^{(k)} y_{t-k} + \sum_{j=2}^{m+1} \gamma_{1,j}^{(k)} x_{t-k-(m-1)/m}^{(m)} \right] + v_{1t}, \quad (6.10)$$

where $\gamma_{i,j}^{(k)}$ corresponds to the (i, j) -element of matrix Γ_k in (6.2), leads to

$$\begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \sum_{i=0}^{m-1} \rho^i \mu_{2+i} \\ \sum_{i=0}^{m-2} \rho^i \mu_{3+i} \\ \vdots \\ \mu_{m+1} \end{pmatrix} + \begin{pmatrix} \gamma_{1,1}^{(1)} & \gamma_{1,2}^{(1)} & \gamma_{1,3}^{(1)} & \cdots & \gamma_{1,m+1}^{(1)} \\ \pi \sum_{i=0}^{m-1} \rho^i & \rho^m & & & \\ \pi \sum_{i=0}^{m-2} \rho^i & \rho^{m-1} & & & \\ \vdots & \vdots & & & \\ \pi & \rho & & & \end{pmatrix} \times \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \sum_{k=2}^p \begin{pmatrix} \gamma_{1,1}^{(k)} & \gamma_{1,2}^{(k)} & \cdots & \gamma_{1,m+1}^{(k)} \\ & \mathbf{0}_{m \times (m+1)} & & \end{pmatrix} \times \begin{pmatrix} y_{t-k} \\ x_{t-k}^{(m)} \\ \vdots \\ x_{t-k-(m-1)/m}^{(m)} \end{pmatrix} + \underbrace{\begin{pmatrix} v_{1t} \\ \sum_{i=0}^{m-1} \rho^i v_{(m+1-i)t} \\ \sum_{i=0}^{m-2} \rho^i v_{(m+1-i)t} \\ \vdots \\ v_{(m+1)t} \end{pmatrix}}_{v_t^*}, \quad (6.11)$$

the restricted MF-VAR. As for the error terms in (6.9) and (6.10), we denote, similarly to Ghysels (2012), that $\mathbb{E}(v_{it}v_{it}) = \sigma_{HH}$, $\mathbb{E}(v_{1t}v_{it}) = \sigma_{HL}$ for $i = 2, \dots, m+1$, and that $\mathbb{E}(v_{1t}v_{1t}) = \sigma_{LL}$. Furthermore, each error term is assumed to possess a zero mean and to be normally distributed. Consequently, $v_t^* \sim N(\mathbf{0}_{(21 \times 1)}, \Sigma_{v^*})$ in (6.11) with

$$\Sigma_{v^*} = \begin{pmatrix} \sigma_{LL} & \sigma'_{.L} \\ \sigma_{.L} & \Sigma_{v^*}^H \end{pmatrix}, \quad (6.12)$$

where

$$\sigma_{.L} = \sigma_{HL} \left(\sum_{i=0}^{m-1} \rho^i, \sum_{i=0}^{m-2} \rho^i, \dots, \sum_{i=0}^1 \rho^i, 1 \right)'$$

and

$$\Sigma_{v^*}^H = \sigma_{HH} \begin{pmatrix} \sum_{i=0}^{m-1} \rho^{2i} & \sum_{i=0}^{m-2} \rho^{2i+1} & \dots & \sum_{i=0}^1 \rho^{2i+m-2} & \rho^{m-1} \\ \sum_{i=0}^{m-2} \rho^{2i+1} & \sum_{i=0}^{m-2} \rho^{2i} & \dots & \sum_{i=0}^1 \rho^{2i+m-3} & \rho^{m-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum_{i=0}^1 \rho^{2i+m-2} & \sum_{i=0}^1 \rho^{2i+m-3} & \dots & \sum_{i=0}^1 \rho^{2i} & \rho \\ \rho^{m-1} & \rho^{m-2} & \dots & \rho & 1 \end{pmatrix}.$$

The Auxiliary Dummy Variable Approach for Mixed-Frequency Data

We will rely heavily on the restricted VAR in (6.11) when formulating prior beliefs for the parameters, which are going to be estimated using Bayesian techniques. Indeed, as pointed out by Carriero et al. (2011), Bayesian methods allow the imposition of such restrictions, while admitting influence of the data as well. Consequently, Bayesian shrinkage has become a standard tool when being faced with large-dimensional estimation problems such as large VARs (e.g., Banbura et al., 2010, Kadiyala and Karlsson, 1997). Within a MF-VAR, Ghysels (2012) describes a way to sample the MIDAS hyperparameters and subsequently formulates prior beliefs for the remaining parameters.¹¹ Once these hyperparameters are taken

¹¹Sekhposyan et al. (2014) use a Sims-Zha shrinkage prior and the algorithm in Waggoner and Zha (2003) to solve the parameter proliferation problem with Bayesian estimation techniques. Bayesian methods within mixed-frequency VARs were also considered by Schorfheide and Song (2012). However, the VAR under consideration differs from the one of Ghysels (2012) in the sense that the high-frequency observations of the low-frequency variables are assumed to be missing. Hence, it is a 'parameter-driven' (Cox et al., 1981) high-frequency VAR with latent variables

care of, the Bayesian analyses of mixed- and common-frequency VAR models are quite similar and, hence, traditional Bayesian VAR techniques (e.g., Kadiyala and Karlsson, 1997, Litterman, 1986) can be applied. This amounts to imposing a normal inverted Wishart prior for the VAR parameters, which retains the principles of the previously specified prior beliefs.

Banbura et al. (2010) show, for the common-frequency case, that adding a set of auxiliary dummy variables to the system is equivalent to imposing the aforementioned normal inverted Wishart prior. In the mixed-frequency scenario this approach can, however, not be applied directly due to the stacked nature of $X_t^{(m)}$ within Z_t . In the sequel, we properly extend this auxiliary dummy variable approach to the mixed-frequency setup and, thereby, show precisely how to choose the moments of the normal inverted Wishart prior when faced with mixed frequencies.

Let us start by considering the MF-VAR in (6.2) and choosing the moments for the prior distributions of the coefficients in B :

$$\begin{aligned} \mathbb{E}[\gamma_{i,j}^{(k)}] &= \begin{cases} \rho^m & \text{if } i = j = k = 1 \\ \rho^{m+j-i} & \text{if } k = 1, j = 2, i > 1 \\ 0 & \text{else} \end{cases} , \\ \text{Var}[\gamma_{i,j}^{(k)}] &= \begin{cases} \phi \frac{\lambda^2}{(km+j-2)^2} \mathbb{S}_{LH} & \text{for } i = 1, j > 1 \\ \phi \frac{\lambda^2}{(km+2-i)^2} \mathbb{S}_{HL} & \text{for } j = 1, i > 1 \\ \frac{\lambda^2}{(km+j-i)^2} & \text{else} \end{cases} , \end{aligned} \quad (6.13)$$

where all $\gamma_{i,j}^{(k)}$ are assumed to be a priori independent and normally distributed. Note that the covariance matrix of the residuals is for now assumed diagonal and fixed, i.e., $\Sigma_u = \Sigma = \Sigma_d$ with $\Sigma_d = \text{diag}(\sigma_L^2, \sigma_H^2, \dots, \sigma_H^2)$ of dimension $m + 1$. Furthermore, $\mathbb{S}_{ij} = \frac{\sigma_i^2}{\sigma_j^2}$, $i, j = L, H$. For μ we take a diffuse prior.

This specification of prior beliefs is derived from the Minnesota prior in Litterman (1986) and was extended to the mixed-frequency case by Ghysels (2012). In short, λ determines the tightness of the prior distributions around the ARX(1) specification in (6.9), the influence of low- on high-frequency data and vice versa is controlled by ϕ and, finally, $\mathbb{S}_{ij}$ governs the difference in scaling between y and the x -variables. Note that $\lambda \approx 0$ results in the posterior coinciding with the prior, whereas $\lambda = \infty$ causes the posterior mean to coincide with the OLS estimate of the unrestricted VAR in (6.2). From (6.13) it becomes clear that the prior variances

instead of the 'observation-driven' VAR considered here. We comment on this difference more elaborately at a later stage.

decrease very quickly as m grows. Hence, in contrast to De Mol et al. (2008), we can already assume that, for a given sample size, λ needs to *grow* with the size of the system in order for the data to have any influence on the estimates also when m is large.

Define $\underline{Z} = (\underline{Z}_1^\mu, \dots, \underline{Z}_T^\mu)'$, where $\underline{Z}_t^\mu = (\underline{Z}_t', 1)'$, and let us re-write the MF-VAR in (6.2) in the following way:

$$\underbrace{vec(Z)}_{(m+1)T \times 1} = \underbrace{\underline{Z}^*}_{(m+1)T \times (m+1)n} \underbrace{vec(B^*)}_{(m+1)n \times 1} + \underbrace{vec(E)}_{(m+1)T \times 1}, \quad (6.14)$$

where $Z = (Z_1, \dots, Z_T)'$, $\underline{Z}^* = I_{m+1} \otimes \underline{Z}$, $E = (u_1, \dots, u_T)'$, $B^* = (B', \mu)'$ and $n = (m+1)p + 1$. In order to drop the undesirable feature of a fixed and diagonal covariance matrix Σ , we impose a normal inverted Wishart prior (at the cost of having to set $\phi = 1$) with the following form (Kadiyala and Karlsson, 1997):

$$\begin{aligned} vec(B^*) | \Sigma &\sim N(vec(B_0^*), [Z_0^{*'}(\Sigma^{-1} \otimes I_T)Z_0^*]^{-1}) \text{ and} \\ \Sigma &\sim iW(V_0, v_0), \end{aligned} \quad (6.15)$$

where $Z_0^* = Z_0 \otimes I_{m+1}$.¹² B_0^* and Z_0 have to be chosen as to let expectations and variances of the elements in B^* coincide with the moments of the prior in (6.13). Likewise, V_0 and v_0 need to be set such that $\mathbb{E}[\Sigma] = \Sigma_d$, i.e., the fixed, diagonal covariance matrix introduced before.

As for the common-frequency case, the priors in (6.13) ensure that more recent lags provide more reliable information than more distant ones. However, due to the stacked nature of $X_t^{(m)}$, the prior variances $Var[\gamma_{i,j}^{(k)}]$ do not only decrease for larger low-frequency lag lengths $k = 1, \dots, p$, but also *within* a low-frequency period. Put differently, given λ , ϕ and $\mathbb{S}_{ij}$, the coefficients are not shrunk according to the low-frequency time difference between the corresponding variables, but according to their *high-frequency* time difference. This is the reason why not only k , but also m, j and i appear in the denominators of the prior variances. As a consequence, and in severe contrast to the common-frequency case in Banbura et al. (2010), the matrix Z_0 , which needs to be chosen to match prior variances with the variance of $vec(B^*)$, changes with each column of B^* .

Still, we can show that adding $T_d = n^{aux} + (m+1)$, with $n^{aux} = m(2p+1)$, auxiliary dummy variables Y_d and X_d to (6.14) is equivalent to imposing the

¹²We can re-write the MF-VAR in (6.2) as in Banbura et al. (2010). In that case, the variance of $vec(B^*)$ simplifies to $\Sigma \otimes (Z_0' Z_0)^{-1} = \Sigma \otimes \Omega_0$. However, it will prove useful to express the MF-VAR as done in (6.14) in our mixed-frequency setup.

normal inverted Wishart prior in (6.15). To this end, let

$$\begin{aligned}
 B_0^{*aux} &= (X_d' X_d)^{-1} X_d' Y_d, \\
 Z_0^{aux} &= X_d \\
 V_0 &= (Y_d - X_d B_0^{*aux})' (Y_d - X_d B_0^{*aux}) \text{ and} \\
 v_0 &= m + 3 = T_d - n^{aux} + 2,
 \end{aligned} \tag{6.16}$$

and subsequently set

$$\begin{aligned}
 \text{vec}(B_0^*) &= S' \text{vec}(B_0^{*aux}) \text{ and} \\
 \text{Var}[\text{vec}(B^*)] &= S' [Z_0^{*aux'} (\Sigma^{-1} \otimes I_T) Z_0^{*aux}]^{-1} S,
 \end{aligned} \tag{6.17}$$

where S is an $[(m+1)n^{aux} \times (m+1)n]$ -dimensional selection matrix filled with ones and zeros, the precise construction of which is given in Appendix 6.A. Intuitively speaking, Z_0^{aux} is an auxiliary matrix constructed as the 'union' of the Z_0 matrices corresponding to the different columns of B^* . The non-random matrix S then selects, for each column of B^* , the corresponding elements of Z_0^{aux} in order to let the variance of each element in B^* match the corresponding prior variance. Likewise, B_0^{*aux} is an auxiliary matrix from which we derive B_0^* . The auxiliary dummy variables that imply a matching of the prior moments turn out to be

$$\underbrace{Y_d}_{T_d \times (m+1)} = \begin{pmatrix} \mathbf{0}_{2(m-1) \times 1} & \vdots & \\ \hline \sigma_L m \rho^m / \lambda & & D_\rho \\ \hline 0 & \vdots & \\ \hline \mathbf{0}_{(m(2p-1)-1) \times (m+1)} & & \\ \text{diag}(\sigma_L, \sigma_H, \dots, \sigma_H) & & \\ \hline \mathbf{0}_{1 \times (m+1)} & & \end{pmatrix}, \tag{6.18}$$

$$\underbrace{X_d}_{T_d \times n^{aux}} = \begin{pmatrix} J_p^1 \otimes \text{diag}(\sigma_L, \sigma_H) / \lambda & \mathbf{0}_{2pm \times (m-1)} & \mathbf{0}_{2pm \times 1} \\ \mathbf{0}_{(m-1) \times 2pm} & J_p^2 \sigma_H / \lambda & \mathbf{0}_{(m-1) \times 1} \\ \mathbf{0}_{(m+1) \times 2pm} & \mathbf{0}_{(m+1) \times (m-1)} & \mathbf{0}_{(m+1) \times 1} \\ \mathbf{0}_{1 \times 2pm} & \mathbf{0}_{1 \times (m-1)} & \epsilon \end{pmatrix}, \tag{6.19}$$

where $D_\rho = \text{diag}^a(m\rho^m, (m-1)\rho^{m-1}, \dots, 2\rho^2, \rho) \frac{\sigma_H}{\lambda} \otimes (0, 1)'$, with $\text{diag}^a(\cdot)$ denoting an anti-diagonal matrix. Furthermore,

$$\begin{aligned} J_p^1 &= \text{diag}(1, 2, \dots, m, m+1, \dots, 2m, \dots, pm), \\ J_p^2 &= \text{diag}(pm+1, pm+2, \dots, pm+m-1). \end{aligned}$$

The last line of both, Y_d and X_d , corresponds to the diffuse prior for the intercept (ϵ is a very small number), the block above imposes the prior for Σ and the remaining blocks set the priors for the coefficients $\gamma_{i,j}^{(k)}$. As Banbura et al. (2010), we set $\sigma_i^2 = s_i^2, i = L, H$, where s_i^2 is the variance of a residual from an $\text{AR}(p)$, respectively an $\text{AR}(mp)$, model for y_t , respectively $x_t^{(m)}$.¹³

Let us now augment the model in (6.14) by the auxiliary dummy variables in (6.18) and (6.19). To this end, let $\underline{Z}_j^d = (\underline{Z}', (X_d S_j)')', j = 1, \dots, m+1$, where S_j is the j^{th} block of the selection matrix S (see Appendix 6.A for details). Here, S_j picks the elements of $Z_0^{aux} = X_d$ corresponding to column j of B^* . Then, the augmented system becomes

$$\underbrace{\text{vec}(Z_*)}_{(m+1)(T+T_d) \times 1} = \underbrace{\underline{Z}_*^*}_{(m+1)(T+T_d) \times (m+1)n} \underbrace{\text{vec}(B^*)}_{(m+1)n \times 1} + \underbrace{\text{vec}(E_*)}_{(m+1)(T+T_d) \times 1}, \quad (6.20)$$

where $Z_* = (Z', Y_d')'$, $E_* = (E', E_d')'$ and $\underline{Z}_*^*$ is block diagonal with $\underline{Z}_*^* = \text{diag}\{\underline{Z}_1^d, \underline{Z}_2^d, \dots, \underline{Z}_{m+1}^d\}$ and off-diagonal elements equal to $\mathbf{0}_{(T+T_d) \times n}$. The posterior then has the form

$$\begin{aligned} \text{vec}(B^*) | \Sigma, Z &\sim N(\text{vec}(\hat{B}^*), [\underline{Z}_*^{*'} (\Sigma^{-1} \otimes I_{T+T_d}) \underline{Z}_*^*]^{-1}) \text{ and} \\ \Sigma | Z &\sim iW(\hat{V}, T+m+3), \end{aligned} \quad (6.21)$$

where

$$\text{vec}(\hat{B}^*) = [\underline{Z}_*^{*'} (\Sigma^{-1} \otimes I_{T+T_d}) \underline{Z}_*^*]^{-1} \underline{Z}_*^{*'} (\Sigma^{-1} \otimes I_{T+T_d}) \text{vec}(Z_*),$$

and where $\hat{V} = \hat{E}_*' \hat{E}_*$ with $\text{vec}(\hat{E}_*) = \text{vec}(Z_*) - \underline{Z}_*^* \text{vec}(\hat{B}^*)$ such that $\hat{E}_*$ is a $(T+T_d) \times (m+1)$ matrix. Note that the posterior mean of the coefficients equals

¹³We propose to choose λ in such a way as to control the size of our Granger causality tests, because their analysis is the ultimate purpose of this work. To this end, we also allow λ to change with T , the sample size. More details on how we choose λ are given in the Monte Carlo section.

the GLS estimate of a SUR regression of $vec(Z_*)$ on $\underline{Z}_*$. As for the common-frequency case, it can be checked that it also coincides with the posterior mean for the prior setup in (6.13). In practice, we estimate Σ in the standard way, i.e., $\hat{\Sigma} = \frac{1}{T+T_d} \hat{E}_*^{ols'} \hat{E}_*^{ols}$, where $\hat{E}_*^{ols}$ denotes the OLS residuals of the system in (6.20).¹⁴

As a final comment, note that (6.10) actually corresponds to a U-MIDAS regression (Forni et al., 2012) without contemporaneous observations of the high-frequency variable. It is worth mentioning that MIDAS restrictions (Ghysels et al., 2004) can be imposed here as an alternative. However, doing so implies leaving the linear framework, which is needed to apply the auxiliary dummy variable approach presented above. Furthermore, even after having drawn the MIDAS hyperparameters, one still faces the aforementioned Davies (1987) problem when attempting to test for Granger non-causality from $X^{(m)}$ to y (see the end of Section 6.3.1). For those reasons we leave the corresponding parameters in (6.10) unrestricted.

Testing for Granger Non-Causality

Ultimately, our goal is to compare our parameter reduction techniques with each other, and two benchmark models to be introduced below, in terms of their Granger non-causality testing behavior. The implementation of the Bayesian MF-VAR via the auxiliary dummy variable approach is, in this sense, preferable to 'usual' Bayesian methods, i.e., computing the posterior distribution from the prior distribution and the data using Bayes' rule (see, e.g., Zellner, 1996), since we obtain a closed-form solution for the posterior mean of the coefficients of interest. As for the parameter estimates, it is desirable to deal with a testing framework that is common to all approaches considered in this paper, i.e., the Wald test. However, construction and especially interpretation of the latter need to be treated with special care, because, unlike in Section 6.3.1, we are not permitted to act as sampling theorists, or frequentists, anymore. To be more precise, the parameter vector $vec(B^*)$ is not assumed fixed, but random with posterior distribution, or posterior density, given by (6.21).

To investigate the presence or absence of Granger causality we consider Bayesian confidence intervals, in particular highest posterior density (HPD) confidence intervals (Bauwens et al., 2000). We express the tightness of each interval by α , which is also used to denote the significance level in the frequentist framework.

¹⁴The i^{th} column of $\hat{E}_*^{ols}$, $i = 1, \dots, m+1$, corresponds to the residuals of a regression of $(Z'_{\cdot,i}, Y'_{d,i})'$ on $\underline{Z}_i^d$, where $Z_{\cdot,i}$ and $Y_{d,i}$ denote the i^{th} columns of Z and Y_d , respectively.

However, for our Bayesian method it is interpreted in such a way that $(1 - \alpha)\%$ of the probability mass is within the respective interval. In other words, the probability that a model parameter falls within the bounds of the interval is equal to $1 - \alpha$. In contrast, with a frequentist confidence interval, whose range needs to be interpreted as the one in which a model parameter would fall into $(1 - \alpha)\%$ of the times with repeated sampling, the aforementioned probability is either 0 or 1. The interval centered at the modal value for unimodal symmetric posterior densities as in (6.21) is then called the HPD (Zellner, 1996 or Bauwens et al., 2000).

The connection to Wald tests is now immediate using the equivalence between confidence intervals and respective test statistics. Once again, however, one needs to be careful when interpreting the outcomes. Unlike the frequentist approach, a large value for the test statistic does not imply rejection of a null hypothesis with a significance level of $\alpha\%$. It rather implies a large probability, $(1 - \alpha)\%$, for the model parameters to fall within regions that do not contain the tested values, i.e., zeros in our case. Put differently, the larger the test statistic the smaller the probability that the model parameters fall into regions containing 0, i.e., the smaller the probability that there is no Granger causality. Consequently, we consider the test statistic computed as

$$\xi_W^* = [Rvec(\hat{B}^*)]' \left\{ R[\underline{Z}_*^* (\hat{\Sigma}^{-1} \otimes I_{T+T_d}) \underline{Z}_*^*]^{-1} R' \right\}^{-1} [Rvec(\hat{B}^*)], \quad (6.22)$$

where R is the suitably adjusted matrix picking the set of coefficients to analyze Granger non-causality and where the auxiliary dummy variables enter ξ_W^* through $vec(\hat{B}^*)$, $\underline{Z}_*^*$ and $\hat{\Sigma}$.¹⁵ To clarify, if the p -value corresponding to the $\chi_{rank(R)}^2$ distributed test statistic above is larger (smaller) than, e.g., 5%, the intervals containing the model parameters with 95% probability (do not) contain zeros, i.e., it is likely for Granger causality to be absent (present).

6.3.3 Common Low-Frequency VARs

Before the introduction of MIDAS regression models, high-frequency variables were usually aggregated to the low frequency in order to obtain a common frequency for all variables appearing in a regression (Silvestrini and Veredas, 2008 or Marcellino,

¹⁵Note also, that, similar to the relationship between the posterior mean of the coefficients to the GLS estimate of an adequately constructed SUR regression, there is large sample correspondence between classical Wald and Bayesian posterior odds tests (Andrews, 1994). For certain choices of the prior distribution, the posterior odds ratio is approximately equal to the Wald statistic. Andrews (1994) shows that for any significance level α there exist priors such that the aforementioned correspondence holds, and vice versa.

1999). Likewise for systems, the observations of a high-frequency, say monthly, variable were usually aggregated to a low frequency, say quarterly, such that the VAR could be estimated in the resulting common frequency.

Temporally aggregating the high-frequency variable also leads to a great reduction in parameters which need to be estimated. After all, each set of m high-frequency variables per t -period is aggregated into one low-frequency observation. In terms of the $(p = 1)$ -example in (6.3), instead of $(m+1)^2$ parameters to estimate in Γ_1 , a common low-frequency VAR as in

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \underbrace{\begin{pmatrix} \gamma_{1,1}^{LF} & \gamma_{1,2}^{LF} \\ \gamma_{2,1}^{LF} & \gamma_{2,2}^{LF} \end{pmatrix}}_{\Gamma_1^{LF}} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \quad (6.23)$$

requires only 4 parameter estimates in Γ_1^{LF} .¹⁶ Note that $x_t = W(L^{1/m})x_t^{(m)}$, where $W(L^{1/m})$ denotes a high-frequency lag polynomial of order A , i.e., $W(L^{1/m})x_t^{(m)} = \sum_{i=0}^A w_i x_{t-i/m}^{(m)}$ (Silvestrini and Veredas, 2008). This generic specification nests the two dominating aggregation schemes in the literature, Point-in-Time ($A = 0, w_0 = 1$) and Average sampling ($A = m - 1, w_i = 1/m \forall i$), where the former is usually applied to stock and the latter to flow variables. In view of the high-frequency variable we consider in our empirical application, the natural logarithm of bipower variation, we focus on Average sampling in this paper.

Of course, the decrease in parameters to estimate comes at the cost of disregarding information embedded in the high-frequency process. As argued in Miller (2011), if the aggregation scheme used to compute x_t is different from the true underlying aggregation scheme, potentially crucial high-frequency information will be forfeited. Additionally, aggregating a high-frequency variable may lead to 'spurious' (non-)causality in the common low-frequency setup (Breitung and Swanson, 2002), since causality is a property which is not invariant to temporal aggregation (Marcellino, 1999 or Sims, 1971).

As far as testing for Granger non-causality is concerned, we can rely on the Wald statistic in (6.8), where the set of regressors W , the matrix R and the coefficient matrix B are suitably adjusted.

¹⁶Of course, a similar argument holds for $p > 1$: Instead of $p(m+1)^2$ parameters to estimate in the autoregressive matrices $\Gamma_1, \dots, \Gamma_p$ in (6.2), the common low-frequency VAR(p) requires only $4p$ parameter estimates in $\Gamma_1^{LF}, \dots, \Gamma_p^{LF}$.

6.3.4 Unrestricted VARs

Finally, we can attempt to estimate the full MF-VAR in (6.2) ignoring the fact that, given the sample sizes usually available, the amount of parameters may be too large to estimate them properly or to test for any causality pattern adequately. Still, it may serve as benchmark case. To this end we estimate the MF-VAR using ordinary least squares disregarding the parameter proliferation problem we inevitably run into.¹⁷ Similar to the previous section, we can test for Granger non-causality using the Wald statistic in (6.8).¹⁸

6.4 Monte Carlo Simulations

In order to assess the finite sample performance of our different parameter reduction techniques, we conduct a Monte Carlo experiment. Unlike Ghysels et al. (2013), we do not start with a common high-frequency data generating process (DGP hereafter), but rather a mixed-frequency one. The former implies the assumption that the low-frequency variable has high-frequency observations that are missing. Given such a situation it seems natural to cast the model in state space form and estimate the parameters using the Kalman filter. However, this amounts to a 'parameter-driven' model (Cox et al., 1981), which contains latent processes, by construction. The latter is a feature we try to avoid in our MF-VAR. Say we are interested in the impact of shocks to one or several variables on the whole system. Using a high-frequency DGP with missing observations implies that shocks to these latent processes are also latent and unobservable. This is undesirable given that, e.g., policy shocks are, of course, observable (Foroni and Marcellino, 2013).

Recall from Section 6.2.1 that $Z_t = (y_t, x_t^{(m)}, \dots, x_{t-(m-1)/m}^{(m)})' = (y_t, X_t^{(m)})'$. In light of our empirical investigation (Section 6.5) we set $m = 20$, i.e., a month/working day-example. Furthermore, we start by investigating the case where $p = 1$ and keep the analysis of higher lag orders for future research.

As far as investigating the size of our Granger non-causality tests is concerned, we assume that the data are generated as a mixed-frequency white noise process,

¹⁷In this sense the comparison is related to the one of U-MIDAS (Foroni et al., 2012) and MIDAS regression models.

¹⁸Again, W , R and B have to be adjusted adequately.

i.e.,

$$Z_t = u_t. \quad (6.24)$$

To analyze the power of the tests, we generate two different DGPs that are closely related to the restricted VAR in (6.11):

$$Z_t = \Gamma_P Z_{t-1} + u_t \quad (6.25)$$

with

$$\Gamma_P = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} & \cdots & \gamma_{1,21} \\ \gamma_{2,1}^* & \rho^{20} & 0 & \cdots & 0 \\ \gamma_{3,1}^* & \rho^{19} & \vdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{21,1}^* & \rho & 0 & \cdots & 0 \end{pmatrix},$$

where $\gamma_{1,1} = 0.5$, $\rho = 0.6$ and $\gamma_{1,j} = 2w_{j-1}^*(-0.01)$ for $j > 1$.¹⁹ The parameter values for w_j^* and $\gamma_{j,1}^*$, $j = 2, \dots, 21$, are depicted in Figure 6.1.²⁰

The reason to consider two different DGPs for power is that due to the zero-mean feature of $w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*$, $j = 2, \dots, 21$, in the second power DGP, we expect the presence of Granger causality to be 'hidden' when Average sampling the high-frequency variable as discussed in Section 6.3.3. If this expectation can be confirmed, the situation underlying the second power DGP would be one, in which classical temporal aggregation fails (Marcellino, 1999) and an alternative way of dealing with high-frequency observations would be called for. The first

¹⁹ Additionally, we have considered three alternative DGPs for size, all based on the restricted VAR in (6.11): (i) $\gamma_{1,i} = \gamma_{i,1}^* = 0$, (ii) $\gamma_{1,i} = 0$ and $\gamma_{i,1}^* \neq 0$ or (iii) $\gamma_{i,1}^* = 0$ and $\gamma_{1,i} \neq 0$ $\forall i = 2, \dots, 21$. Results are available upon request.

²⁰ The parameter values have been chosen to mimic part of the structure of the restricted VAR in (6.11) and to ensure stability of the system. The first DGP we consider for investigating power is characterized by positive, but decreasing coefficients $w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*$. In other words, the later the observation of x in $t-1$, the larger its impact on y_t . Furthermore, positive values of y in the preceding period will increase x towards the end of the current period, but hardly have an impact on it at the beginning. One can think of a delayed impact of y on x here. For the second power DGP the general pattern of the coefficients $w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*$ does not change, except that they now possess a mean of zero. This implies that large values of x early in $t-1$ have a negative effect on y_t , whereas large x -values late in $t-1$ have a positive effect. Similarly, positive values of y in period $t-1$ have a negative impact on early x -values in period t , but a positive impact on x -values towards the end of the period.

Figure 6.1: Parameter values for w^* and $\gamma_{j,1}^*$

Note: This figure graphs the parameter values for $w_{j-1}^*(\psi)$ and $\gamma_{j,1}$, $j = 2, \dots, 21$. In the first power DGP $w_i^*(\psi) = w_i(\psi)$ with $w_i(\psi) = \frac{\exp(\psi i^2)}{\sum_{i=1}^{20} \exp(\psi i^2)}$, whereas in the second power DGP $w_i^*(\psi) = w_i(\psi) - \overline{w(\psi)}$, where the horizontal bar symbolizes the arithmetic mean. This specification actually corresponds to the two-dimensional exponential Almon lag polynomial with the first parameter set to zero (Ghysels et al., 2007). Likewise, $\gamma_{j,1}^* = \gamma_{j,1}$ (first power DGP) and $\gamma_{j,1}^* = \gamma_{j,1} - \overline{\gamma_{\cdot,1}}$ (second power DGP) for $j = 2, \dots, 21$ with $\gamma_{2,1} = 0.25$ and $\gamma_{j,1} = (\gamma_{j-1,1})^{1.11}$ for $j = 3, \dots, 21$. Note that $\overline{\gamma_{\cdot,1}} = \frac{1}{20} \sum_{j=2}^{21} \gamma_{j,1}$.

power DGP serves as a benchmark in the sense that we do not a priori expect one approach to be better or worse than the others.

For the methods in Section 6.3 we analyze size and size-adjusted power for $T = 50, 250$ and 500 , corresponding to roughly 4, 21 and 42 years of monthly data. Note that an additional 100 monthly observations are used to initialize the process. As far as the error term is concerned, we assume $u_t \sim N(\mathbf{0}_{(21 \times 1)}, \Sigma_u)$, where Σ_u has the same structure as Σ_{v^*} in (6.12) with $\sigma_{LL} = 0.5$, $\sigma_{HH} = 1$ and $\sigma_{HL} = -0.1$.²¹

²¹We choose these values since they are close to the data we use in the empirical section to come. In particular, the sample variance of the low-frequency variable, an estimator of σ_{LL} , turns out to be 0.55, the mean of the sample variances of $x_{t-i/20}^{(20)}$, $i = 0, \dots, 19$ is equal to 1.09 and approximates σ_{HH} and, finally, the mean of the covariances between y and $x_{t-i/20}^{(20)}$, $i = 0, \dots, 19$ estimating σ_{HL} is -0.22 . We deviate from the exact values to achieve positive-definiteness of Σ_u . Note also that since $\sigma_{HL} \neq 0$ nowcasting causality is present (see Section 6.2.4). If we set

Similar to the lag length, we restrict the amount of factors (when considering CCA and PLS) to one in this paper and leave the analysis of higher factor dimensions for further research. The figures in the tables below represent the percentage amount of rejections at $\alpha = 5\%$.²² All figures are based on 2,500 replications and are computed using GAUSS12.

6.4.1 Size

Table 6.1 contains the size results for the the common low-frequency VAR (LF-VAR hereafter), the unrestricted VAR (U-VAR hereafter), the VARs after reduced rank restrictions have been imposed and the Bayesian MF-VAR (B-MF-VAR hereafter) implemented via the auxiliary dummy variable approach. With respect to reduced rank restrictions, the factors are obtained via CCA, PLS or HAR-type restrictions.

Table 6.1: Size of Granger non-causality tests

S'le Size	LF-VAR	U-VAR	CCA	PLS	HAR	B-MF-VAR	Direction
T=50	0.062	0.130	0.202	0.346	0.067	0.050	$X^{(m)}$ to y
T=250	0.059	0.067	0.294	0.349	0.052	0.058	
T=500	0.056	0.061	0.300	0.360	0.057	0.054	
T=50	0.052	0.926	0.858	0.634	0.604	0.713	y to $X^{(m)}$
T=250	0.054	0.108	0.185	0.139	0.097	0.271	
T=500	0.045	0.066	0.124	0.105	0.063	0.146	
T=50		0.086	0.113	0.131	0.065	0.435	y to $X^{(m)}$ (Bonferroni)
T=250		0.057	0.082	0.095	0.056	0.097	
T=500		0.041	0.066	0.082	0.040	0.061	

Note: The figures represent the percentage of rejections of the Wald test statistic in (6.22) for B-MF-VAR or the one in (6.8), properly adapted if necessary, for the remaining methods at $\alpha = 5\%$. Prior to estimation, the initial VAR has been transformed into a common low-frequency VAR as discussed in Section 6.3.3, left unrestricted as discussed in Section 6.3.4, reduced rank restrictions, with CCA-, PLS- or HAR-based factors, have been imposed as discussed in Section 6.3.1 or the VAR has been augmented to a Bayesian MF-VAR through the auxiliary dummy variable approach in Section 6.3.2. The number of factors (for CCA and PLS) and the lag length of the estimated VARs are equal to 1. The underlying DGP is found in (6.24), where the variance-covariance matrix of the error term is equal to Σ_u .

it equal to zero instead, nowcasting causality would have been absent. The results are, however, similar in both cases.

²²Outcomes for $\alpha = 10\%$ and $\alpha = 1\%$ are available upon request. Note that as far as for as the methods in Sections 6.3.1, 6.3.3 and 6.3.4 are concerned, α represents the usual significance level. For the Bayesian approach α needs to be interpreted as outlined in Section 6.3.2.

Let us focus on the top two blocks of Table 6.1. While the low-frequency VAR has size close to the nominal one, the unrestricted approach suffers from parameter proliferation for small T . Note that the former result is not surprising because the flat aggregation scheme underlying the low-frequency VAR is correct in this mixed-frequency white noise DGP. For large sample sizes the two approaches perform almost equally well. As far as reduced rank restrictions and testing the direction from $X^{(m)}$ to y is concerned, computing the factors by CCA or PLS delivers size distortions, whereas the use of HAR-type factors leads to very good size results irrespective of T . With respect to testing for causality in the reverse direction, the tests are severely oversized for small T , whereas they perform more satisfactory (especially for HAR) as T grows. Finally, the Bayesian MF-VAR yields optimal size results for the direction from $X^{(m)}$ to y by construction as we choose λ as to control the size of Granger non-causality tests in this direction (see below). For the reverse direction, however, a similar picture as for CCA or PLS arises, i.e., size distortions also for large T .

When testing for Granger non-causality from y to $X^{(m)}$ in all approaches except the low-frequency VAR is the test oversized for small T , but yields better size results for a larger sample size. Note that we compute a joint test on $20p$ parameters in this case, which distorts our size results, especially for small T . In order to address this issue, we re-compute the rejection frequencies using the Bonferroni correction (Dunn, 1961) and report the results in the lowest block of Table 6.1. It turns out that the results improve for all sample sizes, yet only for the reduced rank restrictions with HAR-type factors (and the unrestricted VAR to a lesser degree) are the outcomes satisfactory also for small T .

As far as B-MF-VAR and the choice of λ is concerned, closer inspection of the prior variances of the parameters of interest reveals that $\text{Var}[\gamma_{1,j}^{(k)}] < \text{Var}[\gamma_{i,1}^{(k)}]$ for $i, j = 2, \dots, m+1$ and $k = 1, \dots, p$ (see (6.13)). Consequently, for a given m and λ the parameters corresponding to Granger (non-)causality from $X^{(m)}$ to y get shrunk more than the ones corresponding to the direction from y to $X^{(m)}$. As the Bonferroni correction presents a way to lower size distortions in the latter case, we opted to set λ as to control size in the former direction. Denoting the value of λ that equates the corresponding actual and nominal size for a certain T and m by $\lambda(T, m)$, we obtain $\lambda(50, 20) \approx 3.12$, $\lambda(250, 20) \approx 4.8$ and $\lambda(500, 20) \approx 5.6$.²³

²³For $m = 4, 5, 7$ and 10 we obtain $\lambda(i, 4) = 1.3, 1.45, 1.6, \lambda(i, 5) = 1.55, 2, 2.8, \lambda(i, 7) = 1.8, 2.4, 3.7$ and $\lambda(i, 10) = 2.45, 2.6, 4.5$ for $i = 50, 250$ and 500 , respectively. A detailed analysis of the relationship between m, T and λ as well as the potential introduction of an equation-dependent λ , say λ_L and λ_H , to address the different evolvement of prior variances of the

6.4.2 Power

Table 6.2 reports size-adjusted power of the Granger non-causality tests for the various parameter reduction techniques and the unrestricted VAR. The top half of the table contains the outcomes for the first power DGP ($w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*$, $j = 2, \dots, 21$, possess non-zero means), while the bottom half does so for the second power DGP ($w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*$, $j = 2, \dots, 21$, possess means of zero). Within each half, the top two blocks correspond to the usual Wald statistics and the bottom block reports the results after using the Bonferroni correction.

Table 6.2: Size-adjusted power of Granger non-causality tests

Sample Size	LF-VAR	U-VAR	CCA	PLS	HAR	B-MF-VAR	Direction
T=50	0.992	1	0.909	0.975	1	1	$X^{(m)}$ to y
T=250	1	1	1	1	1	1	
T=500	1	1	1	1	1	1	
T=50	0.182	0.063	0.004	0.098	0.116	0.172	y to $X^{(m)}$
T=250	0.866	0.880	0.857	0.908	0.935	0.931	
T=500	0.995	0.999	0.999	0.999	0.999	0.999	
T=50		0.125	0.181	0.181	0.250	0.094	y to $X^{(m)}$ (Bonferroni)
T=250		0.970	0.976	0.969	0.982	0.980	
T=500		1	1	1	1	1	
T=50	0.046	0.963	0.636	0.865	0.999	1	$X^{(m)}$ to y
T=250	0.054	1	1	1	1	1	
T=500	0.064	1	1	1	1	1	
T=50	0.058	0.054	0.016	0.057	0.065	0.144	y to $X^{(m)}$
T=250	0.049	0.350	0.304	0.420	0.411	0.376	
T=500	0.058	0.779	0.829	0.862	0.826	0.766	
T=50		0.075	0.071	0.063	0.098	0.076	y to $X^{(m)}$ (Bonferroni)
T=250		0.472	0.528	0.489	0.514	0.519	
T=500		0.874	0.928	0.906	0.896	0.888	

Note: The underlying DGP is found in (6.25), where the top half of the Table reports the outcomes for the first power DGP, in which $w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*$, $j = 2, \dots, 21$, contain a non-zero mean, and where the bottom half gives the results for the second power DGP, in which they possess a mean of zero. For the rest see Table 1.

Granger (non)-causality determining parameters, are of great interest, but beyond the scope of this paper and thus left for future work.

Let us start by considering the first power DGP, for which no method is a priori expected to perform better or worse than the others. When testing for Granger non-causality from $X^{(m)}$ to y , all parameter reduction approaches lead to very good power results. For the reverse direction we obtain a similar pattern as for size, i.e., size-adjusted power is low for small T , yet improves quickly as T grows. Although the results improve slightly when using the Bonferroni correction (except for B-MF-VAR), the general pattern remains the same.

Let us now turn to the second power DGP. Most importantly, the results for the low-frequency VAR show that aggregation of the high-frequency variable by averaging annihilates the causality features between the series for all sample sizes (size-adjusted power actually almost coincides with the nominal size of 5%). Put differently, the outcomes confirm that Granger causality is not invariant to temporal aggregation (Marcellino, 1999) if the causal effects get averaged out as in this case. For the other approaches a similar pattern as for the first power DGP emerges: Granger non-causality from $X^{(m)}$ to y is usually rejected, whereby power is slightly lower for small T in case of CCA and, to a lesser extent, PLS.²⁴ For the reverse direction, power is very small for $T = 50$ (whereby B-MF-VAR provides the least bad outcome and CCA the worst) and is satisfactory for $T = 500$. For $T = 250$ size-adjusted power equals about 30-40% for all approaches (except LF-VAR, of course) implying much smaller values than in the top half of the table. Overall, it seems that reduced rank restrictions with HAR-type factors results in best size-adjusted power (except maybe for $T = 50$ and testing the direction from y to $X^{(m)}$). Application of the Bonferroni correction yields similar outcomes.

A final comment on the results of the Monte Carlo study. Note that these results are specific to one particular example, in which $m = 20$, the lag length and number of factors are one and which is based on DGPs that are closely related to the restricted VAR in (6.11). It is well possible that the outcomes change for different situations, i.e., reduced rank restrictions using factors based on CCA and PLS or B-MF-VAR may perform better in other scenarios. In any case, both approaches are valuable contributions to the set of parameter reduction techniques that could be considered in practice as the next section shows.

²⁴PLS has the disadvantage that the loadings corresponding to the first factor are bound to be larger than zero. With several negative coefficients in the underlying DGP, this could explain losses of power for the direction from $X^{(m)}$ to y . This problem would most likely vanish for a larger amount of factors, since negative loadings are possible from the second factor onwards.

6.5 Application

We apply the parameter reduction approaches described in Section 6.3 to a MF-VAR consisting of the monthly growth rate of the industrial production index (*ipi* hereafter), a measure of business cycle fluctuations, and the logarithm of daily bipower variation (*bv* hereafter), a measure of volatility. While the degree to which macroeconomic variables can help to predict volatility movements has been investigated widely in the literature (see Schwert, 1989b, Hamilton and Gang, 1996, or Engle and Rangel, 2008, among others), the reverse, i.e., whether the future path of the economy can be forecasted using return volatility, has been granted comparably few attention (examples are Schwert, 1989a, Mele, 2007, and Andreou et al., 2000). Instead of using an aggregate measure of volatility measure (Chauvet et al., 2013) such as, e.g., realized volatility or GARCH, we use daily bipower variation computed on 5-minute returns.²⁵ With the *bv*-series being available at a higher frequency than most indicators of business cycle fluctuations, we obtain the mixed-frequency framework analyzed in this paper.

The sample covers the period from January 2000 to June 2012, yielding a sample size of $T = 150$, and $m = 20$. Figure 6.2 plots the data. We take $m = 20$ as it is the maximum amount of working days that is available in *every* month throughout the sample we deal with. Consider the following table illustrating the notation for our month/working day-example, where $t = 2012M06$, i.e., June 2012.

Notation	Meaning
ipi_t	<i>ipi</i> in June, 2012
No working day	<i>bv</i> on June 30, 2012
$bv_t^{(20)}$	<i>bv</i> on June 29, 2012
$bv_{t-1/20}^{(20)}$	<i>bv</i> on June 28, 2012
$\vdots$	$\vdots$
$bv_{t-19/20}^{(20)}$	<i>bv</i> on June 4, 2012
ipi_{t-1}	<i>ipi</i> in May, 2012
$bv_{t-1}^{(20)}$	<i>bv</i> on May 31, 2012
$bv_{t-1-1/20}^{(20)}$	<i>bv</i> on May 30, 2012

²⁵Data are obtained from the database in Heber, Gerd, Asger Lunde, Neil Shephard and Kevin Sheppard (2009), Oxford-Man Institute's realized library (version 0.2), Oxford-Man Institute, University of Oxford. Note that *bv* is considered as a measure that is less sensitive to jumps compared to realized volatility.

In this particular case, it means that although there are 21 open days in June 2012, we do not consider the first day, i.e., June 1. For May 2012 we do not consider the first three days. An alternative (balanced) strategy would have been to take the maximum number of days in a *particular* month (i.e., 23, usually in July, August or October) and to create additional values for non-existing days in other months whenever necessary. As far as the treatment of daily data is concerned we have also taken $bv_t^{(20)} = bv_{t-1/20}^{(20)}$ when there are no quotations for $bv_t^{(20)}$. Because $m = 20$ we have $BV_t^{(20)} = (bv_t^{(m)}, bv_{t-1/m}^{(m)}, \dots, bv_{t-(m-1)/m}^{(m)})'$.

Figure 6.2: Growth rate of the Industrial Production Index and the logarithm of bipower variation

Note: This figure shows the monthly growth rate of the industrial production index (lower line), i.e., ipi , and the logarithm of daily bipower variation (top lines), i.e., $BV^{(20)}$, for the time period from January 2000 to June 2012

Table 6.3 contains the outcomes of Granger non-causality tests for all approaches discussed in Section 6.3. Note that a length of $p = 1$ and 2 is considered and that the figures represent p -values. For all approaches, except the one described in Section 6.3.2, we also consider a heteroscedasticity consistent variant

of (6.8) by computing a robust estimator of Ω (see Ravikumar et al., 2000) to account for the potential presence of a time varying multivariate process:

$$\hat{\Omega}_R = T((W'W)^{-1} \otimes I_{m+1})\hat{S}_0((W'W)^{-1} \otimes I_{m+1}), \quad (6.26)$$

where

$$\hat{S}_0 = \frac{1}{T} \sum_{t=1}^T (W_t \otimes \hat{u}_t)'(W_t \otimes \hat{u}_t).$$

The top part of the table shows the outcomes after the initial MF-VAR has been transformed into a common low-frequency VAR by temporally aggregating the high-frequency variable: $bv_t^{(aggr)} = \ln(\frac{1}{20} \sum_{i=0}^{19} \exp(bv_{t-i/20}^{(20)}))$. Clearly, the figures suggest bi-directional Granger causality.

The second block from the top presents the outcomes when leaving the MF-VAR unrestricted. The outcomes partly differ if we consider the standard Wald test or its robust version. Based on the latter we usually reject Granger non-causality in both directions. Based on the former, however, we only do so for $p = 2$; for $p = 1$ no Granger causality is found in any direction. However, although not displayed here,²⁶ the Bonferroni correction points towards Granger causality from ipi to $BV^{(20)}$. Hence, apart from the Wald test for $p = 1$, we conclude again bi-directional Granger causality.

As far as parameter reduction by reduced rank restrictions is concerned, canonical correlation tests²⁷ favor $r = 2$ factors whereas the approach of Cubadda and Hecq (2011) selects $r = 1$. Consequently, we present the results for 1, 2 as well as 3 factors, i.e., $r = 1, 2, 3$. Of course, for the HAR-type factors, the number of factors is $r = 3p$. The third block of Table 6.3 displays the results when the factors are computed using CCA, whereas the subsequent two blocks show the outcomes when they are based on PLS and HAR-type restrictions, respectively.

²⁶Results available upon request.

²⁷For *i.i.d.* normally distributed disturbances, the likelihood ratio test statistic for the null hypothesis that there exist at most $r = 20 - s$ common feature combinations within $X_t^{(m)}$ is given by $\zeta_{LR(s)} = -T \sum_{j=1}^s \ln(1 - \hat{\lambda}_j)$, where $\hat{\lambda}_j$ denotes the $s = 1, \dots, 20$, smallest eigenvalues associated with $\hat{\Sigma}_{\tilde{Z}\tilde{Z}}^{-1/2} \hat{\Sigma}_{\tilde{Z}\tilde{X}^{(m)}} \hat{\Sigma}_{\tilde{X}^{(m)}\tilde{X}^{(m)}}^{-1} \hat{\Sigma}_{\tilde{X}^{(m)}\tilde{Z}} \hat{\Sigma}_{\tilde{Z}\tilde{Z}}^{-1/2}$, i.e., the matrix in (6.7) with the roles of $\tilde{Z}$ and $\tilde{X}^{(m)}$ interchanged.

Table 6.3: Testing for Granger non-causality between $\Delta \ln IPI$ and $\ln BV$

	Factors	$p = 1$		$p = 2$		Direction
		Wald	White	Wald	White	
LF-VAR		< 0.001	0.002	< 0.001	0.021	<i>bv to ipi</i>
		0.001	0.015	0.004	0.05	<i>ipi to bv</i>
U-VAR		0.128	0.069	0.028	0.005	<i>bv to ipi</i>
		0.127	0.005	0.04	< 0.001	<i>ipi to bv</i>
CCA	$r = 1$	0.02	0.039	0.064	0.118	<i>bv to ipi</i>
	$r = 2$	0.05	0.09	0.004	0.031	
	$r = 3$	0.003	0.009	0.01	0.061	
	$r = 1$	0.09	< 0.001	0.011	< 0.001	<i>ipi to bv</i>
	$r = 2$	0.135	0.006	0.002	< 0.001	
	$r = 3$	0.213	0.012	0.002	< 0.001	
PLS	$r = 1$	< 0.001	0.002	0.001	0.013	<i>bv to ipi</i>
	$r = 2$	0.002	0.006	< 0.001	0.011	
	$r = 3$	< 0.01	0.003	0.001	0.028	
	$r = 1$	0.057	< 0.001	0.015	< 0.001	<i>ipi to bv</i>
	$r = 2$	0.063	< 0.001	0.011	< 0.001	
	$r = 3$	0.055	< 0.001	0.011	< 0.001	
HAR	$r = 3$	0.007	0.019	0.011	0.072	<i>bv to ipi</i>
	$r = 3$	0.081	< 0.001	0.016	< 0.001	<i>ipi to bv</i>
B-MF-VAR		0.006		0.001		<i>bv to ipi</i>
		0.022		0.003		<i>ipi to bv</i>

Note: The figures represent the p -values associated with the Wald test statistic in (6.22) for B-MF-VAR, the one in (6.8), properly adapted if necessary, for the remaining methods, labeled 'Wald', and the one derived from (6.26), labeled 'White'. The models considered are the common low-frequency VAR as discussed in Section 6.3.3, the unrestricted VAR as discussed in Section 6.3.4, the VAR with reduced rank restrictions, using CCA-, PLS- or HAR-based factors, as discussed in Section 6.3.1 and the B-MF-VAR implemented via the auxiliary dummy variable approach as discussed in Section 6.3.2. The lag order p is equal to 1 or 2 and the number of factors (for CCA and PLS) is equal to 1, 2 or 3.

For all three approaches based on reduced rank restrictions, Granger causality from $BV^{(20)}$ to ipi is detected for almost all combinations of p , f , 'Wald' and 'White'. For the reverse direction and $p = 1$, 'White' also leads to a rejection of Granger non-causality, whereas 'Wald' either suggests a rejection of Granger non-causality at roughly 6% (PLS) and 9% (HAR or CCA with $r = 1$) or the absence of Granger causality altogether (CCA with $r > 1$). For a lag length of 2, both tests imply a rejection of the null hypothesis in almost cases. These results are generally confirmed by the use of the Bonferroni correction.

Finally, the bottom part of Table 6.3 contains the results for the Bayesian MF-VAR implemented through the auxiliary dummy variable approach.²⁸ Here, the figures clearly suggest bi-directional Granger causality between the two variables of interest. Again, this conclusion is confirmed by the use of the Bonferroni correction.

Overall, the outcomes point towards bi-directional Granger causality between uncertainty in financial markets, as measured by bipower variation, and business cycle fluctuations, as measured by growth in the industrial production index.

6.6 Conclusion

Using the example of analyzing the link between uncertainty in financial markets and economic fluctuations, we investigated Granger non-causality testing in a MF-VAR, where the mismatch between the sampling frequencies of the variables under consideration is large. Indeed, the observable MF-VAR by Ghysels (2012) quickly becomes very large when the mismatch between sampling frequencies grows, causing estimation and inference to be potentially problematic due to parameter proliferation.

To potentially avoid this issue we discussed two parameter reduction techniques in detail, reduced rank restrictions and a Bayesian MF-VAR approach. The former achieves a reduction in parameters by looking for a reduced rank structure within the coefficient matrix pertaining to the high-frequency variables (Carriero et al., 2011). The corresponding observable high-frequency factors were computed either by CCA or PLS, or imposed as HAR-type factors. The Bayesian MF-VAR was implemented through the mixed-frequency extension of the approach in Banbura et al. (2010). Using a Minnesota prior for the coefficients (Ghysels, 2012), it was shown that augmenting the unrestricted MF-VAR by a set of auxiliary dummy variables and estimating the resulting system by GLS is equivalent to considering the mean of the corresponding posterior distribution. Finally, we also transformed the MF-VAR into a common low-frequency one by a simple aggregation scheme. Subsequently, we compared these approaches with each other and an unrestricted VAR in terms of their Granger non-causality testing behavior using Monte Carlo simulations.

In our particular case we found that while the low-frequency VAR may lead to very poor results, the unrestricted VAR resulted in size distortions for small sample

²⁸Note that we do not compute a heteroscedasticity consistent variant of (6.22) for this case.

sizes due to parameter proliferation. Reduced rank restrictions based on CCA and PLS yielded relatively poor size and power outcomes, whereas the ones based on HAR-type factors lead to very good size and the overall best power results. Despite power being nearly as high as for the latter approach, the Bayesian MF-VAR approach lead to size distortions when testing for causality in the direction from low- to high-frequency variables. Size for the reverse direction was nearly optimal by construction. Reduced rank restrictions and Bayesian MF-VARs are, in any case, valuable contributions to the available set of techniques to address the proliferation of parameters in large MF-VAR models.

An application investigating the presence of a causal link between business cycle fluctuations and volatility illustrated the practical usefulness of these approaches. In most cases, bi-directional Granger causality between uncertainty in financial markets and the growth rate of the industrial production index was detected.

A number of avenues should be pursued in future work: A formal comparison of the canonical correlation approach with alternative options; An analysis of the relationship between the tightness of the prior distribution around the restricted MF-VAR, the sample size and the frequency discrepancy between the variables (λ , T and m); A potential generalization of the the paper's findings to different data generating processes, especially a common high-frequency DGP;²⁹ The consideration of higher lag orders and a larger number of factors for CCA and PLS; Reduced-rank Bayesian VARs; The assessment of the various models' forecasting performances; An extension of the Monte Carlo study to investigate the presence of causal chains and their impact on Granger causality tests.

Nevertheless, the approaches presented in this paper indicate that many empirical research questions, where the sampling frequencies of the variables involved differ considerably and where a link between these variables is of interest, can be addressed in the future. More importantly, these problems can be tackled without the need to temporally aggregate the high-frequency variable, thus avoiding the risk to create non-existing or conceal existing causality patterns.

²⁹It would be useful to explore whether causality patterns in a common high-frequency VAR can be preserved in the MF-VARs after the various parameter reduction approaches have been applied (Ghysels et al., 2013).

6.A Appendix: Construction of the Selection Matrix S

Let us investigate the variance of $vec(B^*)|\Sigma$ more closely. Noting that we have to choose Z_0 , or Z_0^* in (6.15), in such a way as to let the variances of the corresponding coefficients coincide with the prior variances in (6.13), it turns out that we need to set

$$Var[vec(B^*)|\Sigma] = \begin{pmatrix} \sigma_L^2 \Omega_0^{(2)} & \mathbf{0}_{n \times n} & \dots & \dots & \mathbf{0}_{n \times n} \\ \mathbf{0}_{n \times n} & \sigma_H^2 \Omega_0^{(2)} & \mathbf{0}_{n \times n} & \dots & \mathbf{0}_{n \times n} \\ \vdots & \mathbf{0}_{n \times n} & \sigma_H^2 \Omega_0^{(3)} & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0}_{n \times n} \\ \mathbf{0}_{n \times n} & \mathbf{0}_{n \times n} & \dots & \mathbf{0}_{n \times n} & \sigma_H^2 \Omega_0^{(m+1)} \end{pmatrix},$$

where

$$\Omega_0^{(i)} = diag\left(\frac{\lambda^2}{(m+2-i)^2 \sigma_L^2}, \frac{\lambda^2}{(m+2-i)^2 \sigma_H^2}, \frac{\lambda^2}{(m+3-i)^2 \sigma_H^2}, \dots, \frac{\lambda^2}{(m+m+1-i)^2 \sigma_H^2}, \dots, \right. \\ \left. \dots, \frac{\lambda^2}{(pm+2-i)^2 \sigma_L^2}, \frac{\lambda^2}{(pm+2-i)^2 \sigma_H^2}, \dots, \frac{\lambda^2}{(pm+m+1-i)^2 \sigma_H^2}, \frac{1}{\epsilon^2} \right)$$

for $i = 2, \dots, m+1$.

Hence, unlike in the common-frequency case, where $\Omega_0^{(i)} = \Omega_0 \ \forall i$ (Banbura et al., 2010), the set of variances changes due to the stacked nature of the vector Z_t and the specific lag structure for each coefficient (see the variances in (6.13)). Let us form an auxiliary matrix Ω_0^{aux} , which contains the union of all elements in $\Omega_0^{(2)}, \dots, \Omega_0^{(m+1)}$. Each matrix $\Omega_0^{(i)}$ contains $p+1$ new elements compared to $\Omega_0^{(i-1)}$ for $i > 2$. As there are n elements in $\Omega_0^{(2)}$, we end up with a dimension of $m(2p+1) = n^{aux}$ for the square matrix Ω_0^{aux} :

$$\Omega_0^{aux} = diag\left(\frac{\lambda^2}{1^2 \sigma_L^2}, \frac{\lambda^2}{1^2 \sigma_H^2}, \dots, \frac{\lambda^2}{m^2 \sigma_L^2}, \frac{\lambda^2}{m^2 \sigma_H^2}, \frac{\lambda^2}{(m+1)^2 \sigma_L^2}, \frac{\lambda^2}{(m+1)^2 \sigma_H^2}, \dots, \frac{\lambda^2}{(2m)^2 \sigma_H^2}, \dots, \right. \\ \left. \dots, \frac{\lambda^2}{(pm)^2 \sigma_L^2}, \frac{\lambda^2}{(pm)^2 \sigma_H^2}, \frac{\lambda^2}{(pm+1)^2 \sigma_H^2}, \dots, \frac{\lambda^2}{(pm+m-1)^2 \sigma_H^2}, \frac{1}{\epsilon^2} \right).$$

All that remains is to define a $[(m+1)n^{aux} \times (m+1)n]$ -dimensional selection matrix S such that $S'(\Sigma_d \otimes \Omega_0^{aux})S = Var[vec(B^*)|\Sigma]$. Note that from Ω_0^{aux} we can then derive $Z_0^{aux} = X_d$ by using $\Omega_0^{aux} = Z_0^{aux'} Z_0^{aux}$.

Let us denote by $\underline{1}_i$ an n^{aux} -dimensional column vector with a one in row i and zeros elsewhere. It turns out that S is a block-diagonal matrix, i.e., $S =$

$diag(S_1, S_2, \dots, S_{m+1})$, where each off-diagonal block is $\mathbf{0}_{n^{aux} \times n}$. Each $S_j, j = 1, \dots, m+1$, can be described as follows:

$$S_j = (S_j^1, S_j^2, \mathbf{1}_{n^{aux}}) \text{ for } j = 2, \dots, m+1,$$

where

$$S_j^1 = (\quad \mathbf{1}_{2m-(2j-3)}, \mathbf{1}_{2m-(2j-4)}, \mathbf{1}_{2m-(2j-6)}, \dots, \mathbf{1}_{4m-(2j-2)}, \mathbf{1}_{4m-(2j-3)}, \dots, \\ \mathbf{1}_{6m-(2j-2)}, \dots, \dots, \mathbf{1}_{2pm-(2j-3)}, \mathbf{1}_{2pm-(2j-4)}, \dots, \mathbf{1}_{2pm}),$$

$$S_j^2 = \begin{cases} \emptyset & \text{for } j = m+1 \\ (\mathbf{1}_{2pm+1}, \mathbf{1}_{2pm+2}, \dots, \mathbf{1}_{2pm+m-(j-1)}) & \text{else} \end{cases}$$

and

$$S_1 = S_2.$$

Each of the indices in S_j reveals which row element of the j^{th} column of B_0^{aux} gets selected by S . Put differently, the indices that are missing in S_j correspond to elements of B_0^{aux} (in column j) that will not get chosen by S . This implies that the values of these elements do not play a role for the computation of $vec(B_0^*)$. Consequently, when constructing B_0^{aux} , and subsequently also Y_d , we assign these elements a value of zero for simplicity.

Chapter 7

Conclusion

This chapter provides concluding remarks for this thesis. Since each chapter contains its own conclusion, including possible extensions of the respective work, I refrain from going into too much detail. However, main overall conclusions as well as general limitations of the thesis as a whole will be identified.

This thesis dealt with modeling non-stationary and stationary time series using single-equation dynamic mixed-frequency regression and VAR models. Thereby, each chapter contributed to this topic in a different way. While Chapters 2 and 3 were concerned with mixed-frequency models in a single-equation format, Chapters 4 to 6 focused on VAR specifications. In terms of covariance-stationarity of the time series under consideration, Chapters 2 to 4 covered the non-stationary case, whereas Chapters 5 and 6 were restricted to the stationary setup.

An error-correction model for mixed-frequency non-stationary time series was introduced in Chapter 2, whereby special attention was given to the composition of the cointegrating relationship and its impact on the short-run dynamics structure. In particular, it was shown that by choosing MIDAS models a practitioner can easily guard himself against dynamic misspecification, while still preserving information contained in the high-frequency process. This feature makes MIDAS models an attractive choice compared to competing approaches when it comes to forecasting non-stationary mixed-frequency data.

The ECM-MIDAS models of Chapter 2 came into action in the subsequent chapter, in which the analyses of mixed-frequency time series and real-time data sets were combined. To be more precise, the repeated observation forecasting approach of Stark and Croushore (2002) was extended to an ADL setting, in which

the regressors are sampled at higher frequencies than the regressand. Furthermore, the time series involved were allowed to be non-stationary and possibly cointegrated. The results of Stark and Croushore (2002) were extended in the sense that model selection is vintage-specific also in the presence of additional regressors. Moreover, using a weighting scheme derived from the one in Amisano and Geweke (2013), it was illustrated how to incorporate revision process information into real-time studies.

Chapter 4 represented the multivariate counterpart of Chapter 2 by introducing a mixed-frequency VECM, thereby extending the mixed-frequency VAR of Ghysels (2012) to the non-stationary case. It was shown that a researcher needs to account for two sets of cointegrating relationships and that even in the absence of cointegration between the low- and high-frequency variables, the mixed-frequency VAR of Ghysels (2012) may lead to efficiency losses. Using the mixed-frequency VECM representation we extended the common cyclical feature analysis of, e.g., Engle and Kozicki (1993) to the mixed-frequency setting. Compared to temporally aggregated models, mixed-frequency models may not only have advantages in terms of forecasting, but also as far as the detection of common cycles is concerned.

Chapters 5 and 6 dealt with causality testing within the aforementioned mixed-frequency VAR of Ghysels (2012), implying a restriction to covariance-stationary time series. While nowcasting causality and its impact on causality testing within conditional mixed-frequency models was discussed in the former, Granger causality testing within large mixed-frequency VARs was focused on in the latter chapter. In particular, due to large frequency discrepancies among the variables under consideration, two approaches to reduce the number of parameters to be estimated were considered in Chapter 6, reduced rank restrictions and Bayesian techniques. As to the latter, we showed how to extend the dummy observations approach of Banbura et al. (2010). Thereby, it was illustrated precisely how to incorporate prior beliefs into a prior distribution, clearing the road towards more detailed analyses of Bayesian mixed-frequency VAR models in the future.

Overall, this thesis illustrates not only that mixed-frequency models are an attractive choice compared to their common-frequency counterparts, but also that many settings, previously unexplored for mixed-frequency time series, can be tackled in the future. Among others, mixed-frequency models for non-stationary and possibly cointegrated variables, a testing procedure for common cycles and ways to investigate causal links between mixed-frequency processes in (large) systems are now available. The detailed analysis of Bayesian mixed-frequency VAR models

is yet another example of where this thesis has opened possibly new, or relatively unexplored, branches in the respective literature. The novelties presented in this thesis are not only interesting from a theoretical perspective, but are potentially very useful to practitioners who see themselves confronted with (non-stationary) mixed-frequency time series data.

Of course, the analyses presented here are far from complete. Indeed, the conclusions at the end of each chapter contained several potential avenues for future research. Without listing and commenting on these issues once again at this stage, let me address a rather general issue.

As mentioned already in the introduction, the traditional approach to deal with mixed-frequency data has been to assume a common high frequency to be underlying the data and subsequently consider the high-frequency observations of the low-frequency variables to be missing (see Zadrozny, 1988 or Mariano and Murasawa, 2003 among others). In this thesis, however, the DGPs were 'observation-driven', i.e., of a mixed-frequency nature (Andreou et al., 2010). This strategy releases the researcher from having to deal with latent processes, which possibly exacerbates the interpretation of results, e.g., latent shocks in a VAR model (Ghysels, 2012). Additionally, it is reasonable to assume that economic agents make decisions at a finer frequency than when time series are released (Forni and Marcellino, 2013), e.g., central banks make decisions mostly on a monthly basis, whereas estimation often occurs at the quarterly frequency. While this strategy obviously simplifies the analysis greatly, due to increased control over the parameters, it also biases the outcomes towards the mixed-frequency models under consideration. Furthermore, the assumption of data being truly sampled at mixed frequencies is, at least, debatable.

Consequently, a key extension of almost all chapters in this thesis is to check whether the respective results and conclusions are robust to a high-frequency DGP, e.g., as done in Chapter 2. Furthermore, though, it may be of interest how the model specifications introduced in this thesis compare to the traditional approach of dealing with mixed frequencies, i.e., using a state-space model and the Kalman filter (as done in Kuzin et al., 2009).

However, on a more philosophical level and in my humble opinion, a thorough discussion about the true underlying frequency of time series processes is necessary. After all, the arguments in favor of a common high-frequency, say monthly, instead of a mixed-frequency, say quarterly/monthly, DGP are equally valid when comparing the very same high-frequency DGP with one at an even higher frequency, say

daily. Continuing this thought experiment leads us to a DGP in which all variables are sampled continuously and in which mixed frequencies arise from observing a different amount of discrete time realizations of the variables under consideration (see Chambers, 2003 or Renault et al., 1998). Given the availability of tractable continuous time models, though, it is doubtful whether such an investigation will appear in the near future or whether it is possible at all.

In conclusion, apart from a variety of open questions and possible future research options, this thesis has enlarged the set of situations, in which mixed-frequency time series can be considered without the need to temporally aggregate the high-frequency variable and thereby risk forfeiting important information. Hopefully, the results presented in this thesis can contribute, theoretically and empirically, to the development of mixed-frequency time series analyses.

Bibliography

- Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud (2011). Nowcasting gdp in real-time: A density combination approach. Technical report.
- Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica* 33(1), 178–196.
- Amemiya, T. and R. Y. Wu (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association* 67(339), pp. 628–632.
- Amisano, G. and J. Geweke (2013). Prediction using several macroeconomic models. Technical report.
- Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* 25, 177–190.
- Anderson, H. M. and F. Vahid (2010). Vars, cointegration and common cycle restrictions. Technical report.
- Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics* 22(3), 327–351.
- Andreou, E., E. Ghysels, and A. Kourtellis (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics* 158(2), 246–261.
- Andreou, E., E. Ghysels, and A. Kourtellis (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics* 31, 240–251.

- Andreou, E., D. R. Osborn, and M. Sensier (2000). A comparison of the statistical properties of financial variables in the usa, uk and germany over the business cycle. *Manchester School* 68(4), 396–418.
- Andrews, D. W. K. (1994). The large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests. *Econometrica* 62(5), 1207–1232.
- Athanasopoulos, G., O. T. de Carvalho Guilln, J. V. Issler, and F. Vahid (2011). Model selection, estimation and forecasting in var models with short-run and long-run restrictions. *Journal of Econometrics* 164(1), 116–129.
- Bai, J., E. Ghysels, and J. H. Wright (2013). State Space Models and MIDAS Regressions. *Econometric Reviews* 32(7), 779–813.
- Banbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics* 25(1), 71–92.
- Bauwens, L., M. Lubrano, and J.-F. Richard (2000). *Bayesian inference in dynamic econometric models*. Oxford University Press.
- Boswijk, P. H. (1994). Testing for an unstable root in conditional and structural error correction models. *Journal of Econometrics* 63(1), 37–60.
- Breitung, J. and N. R. Swanson (2002). Temporal aggregation and spurious instantaneous causality in multiple time series models. *Journal of Time Series Analysis* 23, 651–666.
- Carriero, A., G. Kapetanios, and M. Marcellino (2011). Forecasting large datasets with bayesian reduced rank multivariate models. *Journal of Applied Econometrics* 26(5), 735–761.
- Chambers, M. J. (2003). The asymptotic efficiency of cointegration estimators under temporal aggregation. *Econometric Theory* 19(1), 49–77.
- Chauvet, M., Z. Senyuz, and E. Yoldas (2013). What does financial volatility tell us about macroeconomic fluctuations? Technical report.
- Clark, T. E. and M. W. McCracken (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* 105(1), 85–110.
- Clements, M. and A. B. Galvão (2008). Macroeconomic forecasting with mixed-frequency data. *Journal of Business & Economic Statistics* 26, 546–554.

- Clements, M. and A. B. Galvão (2009). Forecasting u.s. output growth using leading indicators: An appraisal using midas models. *Journal of Applied Econometrics* 24(7), 1187–1206.
- Clements, M. and D. Hendry (1998). *Forecasting economic time series*. Number 9780521634809 in Cambridge Books. Cambridge University Press.
- Clements, M. P. and A. B. Galvão (2011). Forecasting with vector autoregressive models of data vintages: U.s. output growth and inflation. *International Journal of Forecasting* forthcoming.
- Colacito, R., R. Engle, and E. Ghysels (2011). A component model of dynamic correlations. *Journal of Econometrics* 164, 45–59.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7(2), 174–196.
- Cox, D. R., G. Gudmundsson, G. Lindgren, L. Bondesson, E. Harsaae, P. Laake, K. Juselius, and S. L. Lauritzen (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics* 8(2), 93–115.
- Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature* 49(1), 72–100.
- Croushore, D. and T. Stark (2001). A real-time data set for macroeconomists. *Journal of Econometrics* 105(1), 111–130.
- Cubadda, G. and B. Guardabascio (2012). A medium-n approach to macroeconomic forecasting. *Economic Modelling* 29(4), 1099–1105.
- Cubadda, G. and A. Hecq (2001). On non-contemporaneous short-run co-movements. *Economics Letters* 73(3), 389–397.
- Cubadda, G. and A. Hecq (2011). Testing for common autocorrelation in data rich environments. *Journal of Forecasting* 30(3), 325–335.
- Cubadda, G., A. Hecq, and F. C. Palm (2009). Studying co-movements in large multivariate data prior to multivariate modelling. *Journal of Econometrics* 148(1), 25–35.
- Davidson, J. (2000). *Econometric theory*. Blackwell Publishers Ltd.

- Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternatives. *Biometrika* 74(1), 33–43.
- de Brouwer, G. and N. R. Ericsson (1998). Modeling inflation in australia. *Journal of Business & Economic Statistics* 16(4), 433–449.
- De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics* 146(2), 318–328.
- Diebold, F. X. (2012). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold-mariano tests. Working Paper 18391, National Bureau of Economic Research.
- Diebold, F. X. and J. A. Lopez (1996). *Forecast evaluation and combination*, Volume 1 of *Handbook of Statistics*, Chapter 14, pp. 241–268. Amsterdam: North-Holland.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13(3), 253–263.
- Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association* 56(293), 52–64.
- Eichler, M. and V. Didelez (2009). On Granger-causality and the effect of interventions in time series. Technical report.
- Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55(2), 251–276.
- Engle, R. F. and S. Kozicki (1993). Testing for common features. *Journal of Business & Economic Statistics* 11(4), 369–380.
- Engle, R. F. and J. G. Rangel (2008). The spline-garch model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies* 21(3), 1187–1222.
- Engle, R. F. and B. S. Yoo (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics* 35(1), 143–159.
- Foroni, C. (2012). Econometric models for mixed-frequency data. Technical report.

- Foroni, C. and M. Marcellino (2013). Mixed frequency structural models: estimation, and policy analysis. Technical report.
- Foroni, C., M. Marcellino, and C. Schumacher (2012). U-midas: Midas regressions with unrestricted lag polynomials. Technical report.
- Forsberg, L. and E. Ghysels (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics* 5(1), 31–67.
- Galvão, A. B. (2012). Changes in predictive ability with mixed frequency data. *International Journal of Forecasting* forthcoming.
- Ghysels, E. (2012). Macroeconomics and the reality of mixed frequency data. Technical report.
- Ghysels, E. and J. I. Miller (2013). Testing for Cointegration with Temporally Aggregated and Mixed-frequency Time Series. Technical report.
- Ghysels, E., K. Motegi, and J. Hill (2013). Granger causality tests with mixed data frequencies. Discussion Paper.
- Ghysels, E., A. Plazzi, and R. Valkanov (2007). Valuation in the us commercial real estate. *European Financial Management* 13, 472–497.
- Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The midas touch: Mixed data sampling regression models. CIRANO Working Papers 2004s-20, CIRANO.
- Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131(1-2), 59–95.
- Ghysels, E., A. Sinko, and R. Valkanov (2007). Midas regressions: Further results and new directions. *Econometric Reviews* 26(1), 53–90.
- Ghysels, E. and R. Valkanov (2006). Linear time series processes with mixed data sampling and midas regression models. Discussion Paper.
- Ghysels, E. and R. Valkanov (2012). *Forecasting volatility with MIDAS*, Chapter 16, pp. 383–401. Handbook of Volatility Models and Their Applications. John Wiley & Sons, Inc., Hoboken, NJ, USA.
- Gianetto, Q. G. and H. Raïssi (2012). Testing instantaneous causality in presence of non constant unconditional variance. Technical report.

- Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics* 55(4), 665–676.
- Götz, T. B. (2010). Mixed data sampling for nonstationary time series. Master’s thesis, Maastricht University, School of Business and Economics.
- Götz, T. B. and A. Hecq (2014a). Nowcasting causality in mixed frequency vector autoregressive models. *Economics Letters* 122(1), 74–78.
- Götz, T. B. and A. Hecq (2014b). Testing for granger causality in large mixed-frequency vars. Working Paper.
- Götz, T. B., A. Hecq, and J.-P. Urbain (2012). Real-time forecast density combinations (forecasting u.s. gdp growth using mixed-frequency data). Research Memoranda 021, Maastricht : METEOR, Maastricht Research School of Economics of Technology and Organization.
- Götz, T. B., A. Hecq, and J.-P. Urbain (2013). *Testing for common cycles in non-stationary VARs with varied frequency data*, Volume 32, pp. 361–393. Emerald Group Publishing Limited.
- Götz, T. B., A. Hecq, and J.-P. Urbain (2014). Forecasting mixed frequency time series with ecm-midas models. *Journal of Forecasting* 33, 198–213.
- Gourieroux, C. and A. Monfort (1990). *Time Series and Dynamic Models*. Cambridge University Press.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37(3), 424–438.
- Greene, W. H. (2008). *Econometric analysis* (6th ed.). Pearson Prentice Hall.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton Univ. Press.
- Hamilton, J. D. and L. Gang (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics* 11(5), 573–93.
- Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64(2), 413–30.

- Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13(2), 281–291.
- Hecq, A. (1998). Does seasonal adjustment induce common cycles? *Economics Letters* 59(3), 289–297.
- Hecq, A. and J. P. A. M. Jacobs (2009). On the var-vecm representation of real time data. Discussion Paper.
- Hecq, A., F. C. Palm, and J.-P. Urbain (2006). Common cyclical features analysis in var models with cointegration. *Journal of Econometrics* 132(1), 117–141.
- Helland, I. S. and T. Almoy (1994). Comparison of prediction methods when only a few components are relevant. *Journal of the American Statistical Association* 89(426), 583–591.
- Horvath, M. T. and M. W. Watson (1995). Testing for cointegration when some of the cointegrating vectors are prespecified. *Econometric Theory* 11(05), 984–1014.
- Jacobs, J. P. and S. van Norden (2011). Modeling data revisions: Measurement error and dynamics of ‘true’ values. *Journal of Econometrics* 161(2), 101–109.
- Jacobs, J. P. A. M., S. Sarferaz, J.-E. Sturm, and S. van Norden (2013). Modeling Multivariate Data Revisions. Technical report.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* 59(6), 1551–1580.
- Kadiyala, K. R. and S. Karlsson (1997). Numerical methods for estimation and inference in bayesian var-models. *Journal of Applied Econometrics* 12(2), 99–132.
- Kuzin, V. N., M. Marcellino, and C. Schumacher (2009). MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area. Technical report.
- Kvedaras, V. and A. Rackauskas (2010). Regression models with variables of different frequencies: The case of a fixed frequency ratio. *Oxford Bulletin of Economics and Statistics* 72(5), 600–620.
- Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions - five years of experience. *Journal of Business & Economic Statistics* 4(1), 25–38.

- Lütkepohl, H. (1993). *Testing for causation between two variables in higher dimensional VAR models*, pp. 75–91. Studies in Applied Econometrics. Springer-Verlag, Heidelberg.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Publishing Company, Incorporated.
- Mankiw, N. and M. Shapiro (1986). News or noise: An analysis of gnp revisions. *Survey of Current Business* 66, 20–25.
- Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics* 17(1), 129–136.
- Mariano, R. S. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics* 18(4), 427–443.
- Mele, A. (2007). Asymmetric stock market volatility and the cyclical behavior of expected returns. *Journal of Financial Economics* 86(2), 446–478.
- Mikosch, T. (1998). *Elementary stochastic calculus with finance in view*. World Scientific Singapore.
- Miller, J. I. (2011). Conditionally efficient estimation of long-run relationships using mixed-frequency time series. Working Papers 1103, Department of Economics, University of Missouri.
- Miller, J. I. (2012). Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. Technical report.
- Nijman, T. E. and F. C. Palm (1990). Predictive Accuracy Gain from Disaggregate Sampling in ARIMA Models. *Journal of Business & Economic Statistics* 8(4), 405–15.
- Paruolo, P. (2006). Common trends and cycles in $i(2)$ var systems. *Journal of Econometrics* 132(1), 143–168.
- Pons, G. and A. Sansó (2005). Estimation of cointegrating vectors with time series measured at different periodcity. *Econometric Theory* 21(4), 735–756.
- Qian, H. (2013). Vector autoregression with mixed frequency data. Technical report.

- Ravikumar, B., S. Ray, and N. E. Savin (2000). Robust wald tests in sur systems with adding-up restrictions. *Econometrica* 68(3), 715–720.
- Renault, E., K. Sekkat, and A. Szafarz (1998). Testing for spurious causality in exchange rates. *Journal of Empirical Finance* 5(1), 47–66.
- Rodriguez, A. and G. Puggioni (2010). Mixed frequency models: Bayesian approaches to estimation and prediction. *International Journal of Forecasting* 26(2), 293–311.
- Schorfheide, F. and D. Song (2012). Real-time forecasting with a mixed-frequency VAR. Technical report.
- Schwert, G. W. (1989a). Business cycles, financial crises, and stock volatility. *Carnegie-Rochester Conference Series on Public Policy* 31(1), 83–125.
- Schwert, G. W. (1989b). Why does stock market volatility change over time? *Journal of Finance* 44(5), 1115–53.
- Sekhposyan, T., M. W. McCracken, and M. T. Owyang (2014). Real-time forecasting with a large, mixed frequency, bayesian var. Working Paper.
- Seong, B., S. K. Ahn, and P. A. Zadrozny (2013). Estimation of vector error correction models with mixed-frequency data. *Journal of Time Series Analysis* 34(2), 194–205.
- Silvestrini, A. and D. Veredas (2008). Temporal aggregation of univariate and multivariate time series models: A survey. *Journal of Economic Surveys* 22(3), 458–497.
- Sims, C. A. (1971). Discrete approximations to continuous time distributed lags in econometrics. *Econometrica* 39(3), 545–563.
- Stark, T. and D. Croushore (2002). Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics* 24(4), 507–531.
- Stock, J. H. and M. W. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* 41(3), 788–829.
- Tay, A. and K. F. Wallis (2000). Density forecasting: A survey. Technical report.
- Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika* 59(3), pp. 525–531.

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)* 58, 267–288.
- Timmermann, A. (2006). *Forecast Combinations*, Volume 1 of *Handbook of Economic Forecasting*, Chapter 4, pp. 135–196. Elsevier.
- Urbain, J.-P. (1992). On weak exogeneity in error correction models. *Oxford Bulletin of Economics and Statistics* 54(2), 187–207.
- Vahid, F. and R. F. Engle (1993). Common trends and common cycles. *Journal of Applied Econometrics* 8(4), 341–360.
- Vahid, F. and J. V. Issler (2002). The importance of common cyclical features in var analysis: a monte-carlo study. *Journal of Econometrics* 109(2), 341–363.
- Waggoner, D. F. and T. Zha (2003). A Gibbs sampler for structural vector autoregressions. *Journal of Economic Dynamics and Control* 28(2), 349–366.
- Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics* 67(s1), 983–994.
- Wohlrabe, K. (2009). Forecasting with mixed-frequency time series models. Technical report.
- Zadrozny, P. (1988). Gaussian likelihood of continuous-time armax models when data are stocks and flows at different frequencies. *Econometric Theory* 4(1), 108–124.
- Zellner, A. (1996). *An introduction to bayesian inference in econometrics*. John Wiley and Sons, Inc.
- Zellner, A. and F. Palm (1974). Time series analysis and simultaneous equation econometric models. *Journal of Econometrics* 2(1), 17–54.

Nederlandse samenvatting

De frequentie waarmee economische tijdreeksen worden waargenomen verschilt per variabele. Zo wordt het bruto binnenlands product ieder kwartaal gepubliceerd, terwijl veel economische indicatoren iedere maand of nog vaker verschijnen. Deze variatie in verschijningsfrequentie zorgt voor problemen voor econometristen, die moeten besluiten hoe gebruik te maken van de variabelen die vaker waargenomen worden. Een mogelijkheid is om alle data met hoge frequentie direct te gebruiken zonder beperkingen, waarbij de verschillen in frequentie tussen de data effectief genegeerd worden. Deze strategie wordt echter onaantrekkelijk als de verschillen in frequentie groot worden of als er veel variabelen zijn, omdat in dit geval erg veel parameters geschat moeten worden terwijl men daarvoor slechts een beperkt aantal waarnemingen ter beschikking heeft. Als alternatief kan men de data met hoge frequentie aggregeren, ofwel samenvoegen, danwel de data met lage frequentie splitsen, om tot een gemeenschappelijke frequentie voor alle variabelen te komen. Binnen de eerste aanpak kan onderscheid gemaakt worden tussen een vaststaande aggregatie-methode of een die van de data afhangt, waarbij deze tweede optie bekend is geworden als *MI(xed) DA(ta) S(ampling)*.

Ondanks dat de literatuur over dit onderwerp sinds de introductie van MIDAS modellen snel gegroeid is, zijn er maar weinig artikelen die er niet van uitgaan dat de betreffende variabelen *stationair* zijn, wat wil zeggen dat de eigenschappen van de tijdreeksen niet veranderen over de tijd. Aangezien het van veel (macro-)economische variabelen, waargenomen met verschillende frequenties, bekend is dat ze niet stationair zijn, is dit een dubieze aanname. Tijdreeksen zoals bruto binnenlands product, inflatie, nationaal inkomen en wisselkoersen zijn niet stationair en kunnen enkel getransformeerd worden tot stationair door het verschil

tussen hun huidige waarde en de waarde in de vorige periode te berekenen. Ook zijn er veel tijdreeksen die dezelfde veranderlijke eigenschappen delen zoals een gezamenlijke trend, wat *coïntegratie* genoemd wordt en betekent dat er een bepaalde (lange termijn) relatie is tussen deze variabelen die wel stationair is.

Naast de bovengenoemde focus op stationaire tijdreeksen was de literatuur over modellen met variërende frequenties voornamelijk beperkt tot enkelvoudige regressie-modellen. Recentelijk is echter meer aandacht gekomen voor het ontwikkelen van systeem-modellen, in het bijzonder *vector autoregressive (VAR)* modellen, voor processen van variërende frequentie. Echter, net als in de enkelvoudige modellen worden de variabelen nog steeds verondersteld stationair te zijn. Daarnaast zijn veel van de problemen die in meervoudige modellen voorkomen nog niet onderzocht in de context van modellen met variërende frequenties, zoals bijvoorbeeld de causale verbanden tussen variabelen wiens frequentie flink verschilt.

Dit proefschrift complementeert de bestaande literatuur over tijdreeksen met variërende frequenties op de manier zoals beschreven in de vorige twee alinea's. In het bijzonder behandelt dit proefschrift het voorspellen van mogelijk niet-stationaire en gecoïntegreerde tijdreeksen gebruik makende van modellen met variërende frequentie in zowel een enkelvoudig regressie- als een systeem-kader. Wat betreft stationariteit van de variabelen kan onderscheid gemaakt worden tussen hoofdstukken 2 tot 4 die niet-stationaire en mogelijk gecoïntegreerde gevallen bekijken en hoofdstukken 5 en 6 die enkel stationaire reeksen toelaten. Wat betreft het enkelvoudige model ten opzichte van het systeem model, kan onderscheid gemaakt worden tussen hoofdstukken 2 en 3 die het eerste geval en hoofdstukken 4 tot 6 die het tweede geval beschouwen.

Hoofdstuk 2 introduceert een *error-correction* model (ECM) voor niet-stationaire tijdreeksen met variërende frequenties, waar speciale aandacht uitgaat naar de vorm van de lange termijn relatie en diens invloed op de overige, korte termijn, variabelen. In het bijzonder wordt aangetoond dat door een MIDAS model te gebruiken de onderzoeker zich gemakkelijk kan weren tegen dynamische misspecificatie, terwijl de informatie die in het proces met hoge frequentie besloten ligt bewaard blijft.

Deze ECM-MIDAS modellen worden gebruikt in hoofdstuk 3, waarin de analyse van tijdreeksen met variërende frequenties en van *real-time* data worden gecombineerd. Specifiek wordt de *repeated observation forecasting* methode van Stark and Croushore (2002) uitgebreid om verklarende variabelen mee te nemen, die mogelijk met een hogere frequentie worden waargenomen dan de afhankelijke variabele.

Bovendien wordt geïllustreerd door middel van een specifiek schema van veranderlijke gewichten, hoe informatie van het revisie-proces kan worden meegenomen in real-time toepassingen.

Hoofdstuk 4 kan worden gezien als de multivariate tegenhanger van hoofdstuk 2 door het introduceren van een *vector error-correction models (VECM)* met variërende frequenties, waarmee het VAR model van Ghysels (2012) wordt uitgebreid naar niet-stationaire tijdreeksen. Er wordt aangetoond dat een onderzoeker rekening moet houden met twee groepen van lange termijn relaties, en dat zelfs in afwezigheid van cointegratie tussen de variabelen met lage en hoge frequentie het VAR model met variërende frequenties kan leiden tot verliezen in efficiëntie. Gebruik makend van de VECM representatie met variërende frequenties breiden we de analyse van gemeenschappelijke cyclische patronen van, bijvoorbeeld, Engle and Kozicki (1993) uit naar de situatie met variërende frequenties. Vergeleken met temporeel geaggregeerde modellen, hebben modellen met variërende frequenties niet alleen voordelen wat betreft het doen van voorspellingen, maar ook wat betreft de detectie van gemeenschappelijke cycli.

Hoofdstukken 5 en 6 bekijken het toetsen op causaliteit in een VAR model met variërende frequenties. Terwijl *nowcasting* causaliteit en diens impact op het toetsen op causaliteit in enkelvoudige regressie-modellen met variërende frequentie wordt besproken in het eerste hoofdstuk, is het toetsen op Granger causaliteit in grote VAR modellen met variërende frequenties onderwerp van het laatste hoofdstuk. Meer specifiek, doordat er grote verschillen zijn in frequenties van de variabelen, worden er twee methodes bekeken om het aantal parameters dat geschat moet worden te verminderen, te weten beperkingen op de gereduceerde rang en Bayesiaanse technieken. Voor deze laatste aanpak wordt getoond hoe a priori opvattingen kunnen worden opgenomen in een a priori distributie in de setup met variërende frequenties, waarmee de weg wordt vrijgemaakt voor diepere analyse van Bayesiaanse VAR modellen met variërende frequenties in de toekomst.

Valorization

As defined by the National Valorization Committee, knowledge valorization represents "the process of creating value from knowledge, by making knowledge suitable [...] and available for social [...] and economic use and by making knowledge suitable for translation into competitive products, services, processes and new commercial activities". In the following addendum I outline the knowledge valorization of modeling non-stationary and stationary mixed-frequency time series, i.e. of this dissertation. I do so by discussing the (social and economic) relevance of the topic and the respective results, potential target groups, translated services, processes and activities, the innovativeness of the research and the implementation of the valorization plans.

As stated in the opening sentence of this thesis, time series of different kinds are available at mixed frequencies. The global sea level¹ or alcohol consumption per capita are examples of annual time series, national account variables, such as the gross domestic product (GDP), appear each quarter, customer surveys are often repeatedly conducted on a monthly basis, whereas stock prices or an individual's weight are measured daily or even intra-daily. The aforementioned list reflects the diversity in time series' frequencies across different fields of research, i.e., the environment, health, economics, marketing and finance. This list can obviously be extended by countless variables originating from different fields. Many socially and economically relevant question concern relationships between different time series: Are rising sea levels (annual) really due to global warming, i.e., rising average temperatures (daily or monthly)? Is a country's alcohol consumption per capita (annual) an indicator for the share of obese habitants measured by, say,

¹As measured by the KNMI climate explorer.

each individual's body mass index² (daily)? What is the effect of stock prices (daily) or the industrial production index (monthly) on the growth of an economy, as measured by GDP (quarterly)?

Questions such as the ones above demand researchers to take the mixed-frequency characteristic of the time series involved into account. More precisely, when developing regression models to study the relationship between the variables in question, they face the situation that many variables are available at different frequencies. Two straightforward approaches to deal with that issue come to mind: (i) Simply use all observations at one's disposal or (ii) temporally aggregate (e.g., average) the high-frequency observations in order to reduce all time series involved to the same frequency. However, both strategies incur serious drawbacks. Intuitively speaking, with respect to the former, one may end up with a model that is too large (due to including all observations) to draw any meaningful conclusions from it, whereas, with respect to the latter, one may apply an inappropriate temporal aggregation scheme.

In recent years, researchers have developed a battery of models to address the mixed-frequency nature of time series, while circumventing the aforementioned drawbacks. Notable here is the MI(xed) DA(ta) S(ampling) approach of Ghysels et al. (2004), which allows for a data-driven temporal aggregation scheme, while keeping the 'size' of the model tractable. Since their introduction, MIDAS models have been analyzed and applied extensively, mostly to economic or financial variables. However, many time series, especially from the former category³ have a characteristic, whose effect on the performance of mixed-frequency models was still relatively unexplored: They are non-stationary. Intuitively speaking this means that the average (mean) and the variability (variance) of these series change over time rather than staying constant. Failing to properly account for this characteristic may have a damaging effect on the underlying model, i.e., it may not capture the relationship between the time series of interest correctly. Say we attempt to build a model for GDP using indicators of higher frequencies. Assume further that our goal is to predict the values of GDP in the near future. The latter is an activity central banks and many research institutions focus on and that is vital to policy makers, because anticipating a country's economic outlook may lead to changes in government spending as well as monetary and fiscal policy. Not taking non-

²Weight divided by height squared.

³For example GDP, the industrial production index, disposable income, consumption and many more.

stationarities of the relevant time series into account may lead to a significantly worsened forecasting performance.

Another gap in the literature is that, until recently, MIDAS regression models were limited to the single-regression framework, i.e., in which one variable (usually the low-frequency one) was 'explained' by the other (usually high-frequency) variables. Extensions to a system approach, in which we can simultaneously analyze the effect of, e.g., alcohol consumption per capita on obesity and the other way around, were either not available or demanded the researcher to treat some variables, e.g., daily observations of a yearly variable, as missing. Ghysels (2012) directly extended the MIDAS approach to a system by introducing an 'observable' mixed-frequency vector autoregressive model (VAR). One notion of interest in such multivariate systems is causality, so, e.g., do stock prices last month influence the industrial production index this month and the other way around? However, similar to problem (i) above, the mixed-frequency VAR of Ghysels (2012) may quickly get too large if the differences in frequencies between the variables are rather large. Despite its importance (due to the previously outlined vast amount of time series that are available at differing frequencies), there did not exist a way to deal with that issue.

This dissertation fills the two aforementioned gaps in the mixed-frequency literature by firstly (Chapters 2 to 4) introducing, applying and analyzing extensions of the MIDAS and mixed-frequency VAR models to the non-stationary case and by secondly (Chapters 5 and 6) investigating causality testing within (large-dimensional) mixed-frequency VAR models. The methods and results presented in this dissertation allow researchers, central bankers, policy makers, professional forecasters and analysts, on the one hand, to investigate relationships between non-stationary mixed-frequency time series⁴ and, on the other hand, to test for causal links between several mixed-frequency time series in (potentially large-dimensional) VAR models.

Generally speaking, the results of this thesis are useful to anyone who is interested in the relationship between time series that are sampled at differing frequencies and that are potentially non-stationary. Due to the fact that interest in matters such as causal links and forecasting is very pronounced among macroeconomists, who constantly deal with time series that are, on the one hand, sampled at varying frequencies and, on the other hand, non-stationary, the outcomes of this

⁴These series may even share a common stochastic trend, i.e., they are allowed to be characterized by a long-run relationship, i.e., they are allowed to cointegrate.

thesis are predominantly interesting for central bankers and macroeconomic forecasters. The general public is continuously made aware of rising or falling future GDP growth rates, industrial production indices, unemployment rates, real wages and so on. The construction of these predictions, which are done by central banks or research institutions, is usually based on the combination of various forecasts obtained using a large set of models. This battery of models consists of different specifications that try to mimic the true underlying relationship between the variables of interest.

This 'model toolbox' should be extended by the models and methods introduced in this thesis in order to incorporate mixed-frequency non-stationary regression models and (potentially large-dimensional) mixed-frequency VARs. Consistent with the conclusion of this dissertation (Chapter 7), this allows researchers to make more precise predictions of non-stationary time series such as GDP, test for common cycles among several mixed-frequency series and investigate causal links between mixed-frequency processes in (large) systems, e.g., between stock prices (daily) and GDP (quarterly). Note that this 'model toolbox' extension comes at little costs, since the methods are easily programmable (in standard econometric software packages used by the institutions in question) and implementable in a user-friendly and generic way, implying that applicants only need to 'feed in' the relevant data to ultimately obtain (and interpret) the respective outcomes.

Despite their immediate interest for macroeconomists, the innovations presented in this thesis are, obviously, not restricted to that particular field. As mentioned above, research questions from other fields require the investigation of relationships between several time series, which are potentially sampled at various frequencies. In fact, any setup involving either mixed-frequency non-stationary time series or a (large) system of mixed-frequency variables, where causal links between these series are of interest, benefits from this dissertation. Environmental economists assessing the causality pattern between, e.g., global sea levels (annual) and average (daily) temperatures, could make use of the methods as well as a health economist investigating the high-frequency determinants of a country's alcohol consumption per capita, given that one or more of the series qualify as non-stationary processes. As for macroeconomists, these models may present a valuable addition to the available set of specifications and is equally easy and cheap to implement.

Apart from the direct applicability of the approaches developed in this thesis, they also provide a range of options for potentially valuable future research.

The construction of mixed-frequency panels, allowing also for non-stationary time series, is only one important avenue for future work that could be interesting to the target groups outlined above. For further extensions of the methods presented here, I refer the reader to the conclusions of the respective chapters. I hope that the process of knowledge valorization outlined in this addendum gets, in fact, realized by an application of the respective methods outside the academic world. Of course, I also look forward to any extensions on the work of non-stationary and stationary mixed-frequency regression models, both on a theoretical and empirical level.

Curriculum Vitae

Thomas Götz was born on August 14, 1986 in Cologne, Germany. He attended high school (Gymnasium) between 1996 and 2005 at Irmgardis Gymnasium in Cologne. Subsequently, he studied Econometrics at Maastricht University. In August 2010 he obtained his M.Sc. with distinction (cum laude). His master thesis is entitled "Mixed Data Sampling with Non-Stationary Time series".

After graduation, Thomas joined the Department of Quantitative Economics as a Ph.D. candidate in September 2010, under the supervision of Dr. Alain Hecq and Prof. dr. Jean-Pierre Urbain. From January to April 2014 Thomas spent three months at the University of North Carolina as a visiting Ph.D. student under the supervision of Prof. dr. Eric Ghysels. The results of his research are presented in this thesis. Thomas presented his work at various international conferences and parts of this thesis are published or forthcoming in international refereed academic journals.