Macroeconomics, Nonlinearities, and the Business Cycle

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Magnus Reif, M.Sc.
aus Kiel

Kiel, 2019

Erstbegutachtung:	Prof. Dr. Maik Wolters
Zweitbegutachtung:	Prof. Dr. Matei Demestrescu
Drittbegutachtung:	Prof. Dr. Tino Berger

Tag der mündlichen Prüfung: 14.11.2019

Acknowledgments

Since I have started working on this thesis in April 2014, five "hungry and foolish" years have passed by. During these years, I learned a lot, made new friends, and met many interesting people. Completing this thesis would not have been possible without the help and support of these people, of which four stand out. First, I am indebted to Maik Wolters for being my first supervisor. He came up with the idea for our first research project and provided extremely helpful comments that greatly assisted my further research. His guidance, support, and cooperation since the final stage of my master studies were invaluable. Besides, I owe thanks to Maik Wolters for encouraging me to apply to the ifo institute in Munich. Second, almost equally important for the completion of this thesis was the support of Kai Carstensen, who accepted to be the chair of my examination committee. His rigorous research attitude and his astute way of thinking about economic issues has shaped each of my projects. Third, I am grateful for the very efficient, enjoyable, and open-minded collaboration with Markus Heinrich that vastly improved the quality of this thesis. Finally, I also thank Matei Demetrescu, who was unhesitatingly taking the effort of being my second supervisor.

Moreover, I am grateful to Timo Wollmershäuser for giving me the opportunity to work in a highly motivated and stimulating environment. I thank Timo Wollmershäuser for many insightful discussions about the German business cycle and ways to model its characteristics. Being part of the ifo center for macroeconomics and surveys allowed me to engage in applied policy consulting that strongly influenced my subsequent research. I appreciate the support of the entire ifo team and want to thank in particular Tim Oliver Berg, Christian Grimme, Steffen Henzel, Stefan Lautenbacher, Robert Lehmann, Felix Schröter, Marc Stöckli, and Klaus Wohlrabe for numerous helpful comments, discussions, and plenty of laughter. Further, I am grateful to the ifo institute—especially Andreas Peichl—for providing excellent conditions for my research. In particular, I am thankful for the financial support the institute provided to present my work at international conferences. I also owe thanks to Svetlana Rujin, who was willing to read large parts of this thesis. Her comments have improved its scientific level tremendously.

Last, but certainly not least, I thank my mother, Dorothee Reif, and my brother, Falko Reif, for backing me throughout the last years and pointing at more important things than economic research at the right moments.

Contents

List of Figures

List of Tables

1 | Outline

This thesis is a collection of four academic articles, each constituting one chapter. These articles are interconnected by two major themes. First, each essay aims at describing and explaining business cycle dynamics. Second, each article resorts to nonlinear, macroeconometric models. By this means the articles address the large criticism the economic profession has received for their inability in predicting the extraordinary strong downturn in 2008/2009—coined by Stock and Watson (2017) the "Mother of All Forecast Errors". As put forward by studies like Ng and Wright (2013) or Chauvet and Potter (2013) using these kinds of models that allow for shifts in the structure of the economy is one possibility to engage in this criticism.

This thesis should be regarded as a contribution to applied econometrics and can be clustered into two categories. The articles in chapter one and two take an applied stance by exploring different modeling strategies for the German economy. The articles in chapter three and four cover methodological contributions, especially with regard to economic forecasting. In the following, I provide a more detailed summary of these articles and a description of my contribution to each of these studies.

The first article, entitled "Predicting Ordinary and Severe Recessions – An Application to the German Business Cycle", is joint work with Kai Carstensen, Markus Heinrich, and Maik H. Wolters (see Carstensen, Heinrich, Reif, and Wolters, 2017). This study consists of two parts. The first part demonstrates that a Markov-switching dynamic factor model (MS-DFM) with three states provides a better ex post characterization of the German business cycle than commonly applied two-state models. We show that adding a third state helps to effectively distinguish between ordinary and severe recessions. Another novelty of this article is that we use a flexible indicator selection procedure to overcome the curse of dimensionality pervasive in MS-DFMs. In fact, we use machine learning techniques to select a subset of the most informative indicators from a larger dataset. In the second part, we conduct a real-time forecast experiment to address the question raised by Ng and Wright (2013), that is, can we detect early signals for economic turning points in real-time, particularly in the presence of the Great Recession? We illustrate that combining the three state MS-DFM with the automatic variable selection procedure provides timely information on business cycle turning points. Notably, the model is able to predict the chronology of the Great Recession one quarter in advance. The key idea for

this article was developed by Maik Wolters. Markus Heinrich and I refined it by using factor models instead of univariate models and came up with the idea of a third state. Furthermore, we were responsible for the technical implementation and the construction of the dataset. I wrote the first draft of the article, which was enormously improved by Kai Carstensen and Maik Wolters. Disentangling the individual contributions of Markus Heinrich and myself with regard to this project is—due to the project's complexity and length—almost impossible.

The second article is entitled "Time-Varying Dynamics of the German Business Cycle" and provides both a description of how the German business cycle has evolved over time and an explanation of the driving forces shaping this evolution (see Reif, 2019). I extend the literature by estimating a time-varying parameter vector autoregression with stochastic volatility (TVP-SV-VAR) and provide results based on both reduced-form estimates and a structural identification of the model. The reduced-form analysis reveals substantial time-variation in the variables' trends, volatilities, and persistences. Most importantly, I find a strong reduction of German output growth volatility. Regarding the question, whether this reduction is rather caused by good luck or good policy, the results favor the good luck hypothesis as the major driving force. In fact, the findings from the structural analysis show that the systematic response of the economy to the identified shocks is fairly stable over time, while the shocks' magnitude has strongly declined. This article's research question was motivated by the insights gained during the work on Carstensen et al. (2017), suggesting important structural change in the German business cycle. The article strongly benefited from various discussions with and comments of Timo Wollmershäuser and Robert Lehmann regarding the volatility of German gross domestic product and analysis in the context of the ifo economic projections.[1]

The third article is entitled "Macroeconomic Uncertainty and Forecasting Macroeconomic Aggregates" (see Reif, 2018). This research project was motivated by the increasing interest of practitioners and researchers in the effects of economic uncertainty on macroeconomic developments. While various studies focus on structural analysis of fluctuations in uncertainty, my analysis mainly concerns the impact of economic uncertainty on macroeconomic forecasts. Since previous research has demonstrated that the link between economic uncertainty and the real economy is subject to nonlinearities, I employ both linear and nonlinear Bayesian VARs. Moreover, I propose a new approach of estimating Bayesian threshold VARs by combining two methods. Using an out-of-sample forecast exercise, I examine the models' forecast performance with regard to point and density forecasts. I find that accounting for nonlinearities is beneficial with regard to forecasting, especially concerning density forecasts and in the presence of high uncertainty.

The fourth article, entitled "Forecasting using Mixed-Frequency VARs with Time-Varying Parameters", is a joint research project with Markus Heinrich (see Heinrich and Reif, 2018). The

[1] See, for example, Wollmershäuser et al. (2017) for a univariate analysis of the volatility of German gross domestic product growth.

main research question of this study is whether accounting for parameter instability improves the accuracy of nowcasts and short-term forecasts. We combine two strands of literature, namely studies on mixed-frequency models and studies on structural change. By implementing the state-space approach for mixed-frequency models in a time-varying parameter framework, this article introduces a new forecasting model—a mixed-frequency TVP-SV-VAR. Moreover, we extend the literature by employing a hyperparameter optimization routine in a mixed-frequency set-up and assessing its impact on both point and density forecasts. We compare the forecast accuracy of this model with several other linear and nonlinear models and demonstrate that the combination of mixed-frequencies and time-variation in the models' coefficients is particularly helpful with regard to inflation forecasts. I came up with the key idea (forecasting with time-varying parameter VARs) for this research project. After an extensive literature research and several discussions, Markus Heinrich and I, jointly developed the final research question. The writing and the technical implementation was carried out by both of us, with almost equal shares.

References

Carstensen, K., M. Heinrich, M. Reif, and M. H. Wolters (2017). Predicting Ordinary and Severe Recessions with a Three-State Markov Switching Dynamic Factor Model. An Application to the German Business Cycle. CESifo Working Paper 6457, Center for Economic Studies and Ifo Institute (CESifo), Munich.

Chauvet, M. and S. Potter (2013). Forecasting Output. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Chapter 3, 141–194. Elsevier.

Heinrich, M. and M. Reif (2018). Forecasting using Mixed-Frequency VARs with Time-Varying Parameters. ifo Working Paper Series 273, ifo Institute - Leibniz Institute for Economic Research at the University of Munich.

Ng, S. and J. H. Wright (2013). Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling. *Journal of Economic Literature 51*(4), 1120–54.

Reif, M. (2018). Macroeconomic Uncertainty and Forecasting Macroeconomic Aggregates. ifo Working Paper Series 265, ifo Institute - Leibniz Institute for Economic Research at the University of Munich.

Reif, M. (2019). Time-Varying Dynamics of the German Business Cycle. Mimeo.

Stock, J. H. and M. W. Watson (2017). Twenty Years of Time Series Econometrics in Ten Pictures. *Journal of Economic Perspectives 31*(2), 59–86.

Wollmershäuser, T., W. Nierhaus, N. Hristov, D. Boumans, M. Göttert, C. Grimme, S. Lauterbacher, R. Lehmann, W. Meister, A. Peichl, M. Reif, F. Schröter, T. Schuler, M. Stöckli, K. Wohlrabe, A. Wolf, and C. Zeiner (2017). ifo Konjunkturprognose 2017/2018: Deutsche Wirtschaft stark und stabil. *ifo Schnelldienst 70*(12), 30–83.

2 | Predicting Ordinary and Severe Recessions - An Application to the German Business Cycle

(**with Kai Carstensen, Markus Heinrich, and Maik H. Wolters**)

Abstract We estimate a Markow-switching dynamic factor model with three states based on six leading business cycle indicators for Germany preselected from a broader set using the Elastic Net soft-thresholding rule. The three states represent expansions, normal recessions and severe recessions. We show that a two-state model is not sensitive enough to reliably detect relatively mild recessions when the Great Recession of 2008/2009 is included in the sample. Adding a third state helps to clearly distinguish normal and severe recessions, so that the model identifies reliably all business cycle turning points in our sample. In a real-time exercise the model detects recessions timely. Combining the estimated factor and the recession probabilities with a simple GDP forecasting model yields an accurate nowcast for the steepest decline in GDP in 2009Q1 and a correct prediction of the timing of the Great Recession and its recovery one quarter in advance.

Keywords: Markov-Switching Dynamic Factor Model, Great Recession, Turning Points, GDP Nowcasting, GDP Forecasting
JEL-Codes: C53, E32, E37

2.1 Introduction

The failure of macroeconomists to predict the Great Recession of 2008 and 2009 has evoked much public criticism. While the debate mostly focuses on the state of macroeconomic modeling, it has also raised the question why professional forecasters even at the onset of the Great Recession did not foresee the steep output contraction that loomed around the corner. The case of Germany illustrates this failure. It was not until November 2008 that professional forecasters started predicting a recession despite clear warning signals accumulating throughout the year 2008.[1] For example, the expectation component of the Ifo business climate index—viewed by professional forecasters as one of the most important early indicators for German GDP—began its descent already in June 2007 and plunged heavily in July 2008, well before GDP plummeted in the fourth quarter of 2008 and the first quarter of 2009.

In this chapter, we take up the debate and ask whether it is possible to reliably predict in real time both business cycle turning points and GDP growth rates around these turning points, particularly during the Great Recession episode. We focus on Germany as a representative of the group of countries that show little persistence in GDP growth (other countries with this characteristic are, *inter alia*, Italy, Japan, Australia, and Norway). The lack of persistence is important because the usual approach to predict GDP growth by augmenting an autoregressive distributed lag model with a business cycle measure derived from coincident indicators (see, e.g., Chauvet and Potter, 2013) works well only for countries like the US that exhibit significant sample autocorrelations.[2] As a more promising approach for low-persistence countries we suggest to directly exploit the information of leading indicators for future GDP. For Germany we show that this yields very competitive one-quarter ahead forecasts of business cycle turning points and GDP growth.

To extract information from leading indicators of the German business cycle, we use the Markov-switching dynamic factor model (MS-DFM) proposed by Diebold and Rudebusch (1996) and Kim and Yoo (1995) because it has been shown to be a valuable device for assessing the state of an economy (Chauvet, 1998; Kim and Nelson, 1998; Camacho, Pérez-Quirós, and Poncela, 2014) and its results are much more timely available than those of simple benchmark approaches such as the Bry-Boschan algorithm. However, unlike the previous literature we specify the MS-DFM with three states. Specifically, we add to the conventional expansion and (ordinary) recession states a third state which reflects a severe recession.[3] This is motivated both by the

[1] See Drechsel and Scheufele (2012) for an analysis of the performance of leading indicators during the financial crisis and Heilemann and Schnorr-Bäcker (2017) for a detailed documentation of the chronological sequence of data releases and publications of professional forecasts in 2008.

[2] The cross-country difference in the persistence of GDP growth and its implications for forecasting are hardly discussed in the literature. One exception is Stock and Watson (2005) who document that, among the G7 countries, Germany, Italy and Japan have negligible persistence in the post-1984 period.

[3] Three-state Markov-switching models have been applied mainly to the US (Boldin, 1996; Layton and Smith, 2000; Krolzig and Toro, 2001; Ferrara, 2003; Nalewaik, 2011; Ho and Yetman, 2014) but also to the euro area

general perception that the Great Recession was different from previous post-war recessions and may thus require a special econometric treatment, and by our empirical finding documented below that an MS-DFM with two states becomes instable in 2008.[4]

We also address the question of how to determine the number of states in real time. This is highly relevant as the severe recession state is only weakly identified before the Great Recession which is probably why studies analyzing pre-2008 data report that the German business cycle can well be represented with two states (Bandholz and Funke, 2003; Artis et al., 2004; Kholodilin and Siliverstovs, 2006). We propose to choose—at each point in real time—the number of states that optimizes the quadratic probability score which measures how well the MS-DFM fits the Bry-Boschan algorithm. Thereby, we effectively train the MS-DFM to yield results close to a simple benchmark but at the same time exploit its advantage to detect turning points instantaneously at the sample end.

Another methodological contribution to the literature is to prepend a flexible indicator selection procedure to the MS-DFM. This is important because there are many potentially useful business cycle indicators available for an economy to be fed into the MS-DFM, while the nonlinear one-step estimation approach by Kim and Yoo (1995), which simultaneously determines the factor and the state probabilities, is subject to numerical problems if the number of parameters is large.[5] We use a soft thresholding procedure that accounts for multivariate correlations among the variables to extract a small number of variables from a medium-sized set of pre-selected indicators because Bai and Ng (2008) show that hard thresholding, i.e., using statistical tests to ensure that a predictor is significant irrespective of other predictors, might be inadequate in such situations. Specifically, we use the elastic net (EN) algorithm of Zou and Hastie (2005), which is a convex combination of a ridge regression and a Least Absolute Shrinkage and Selection

(McAdam, 2007; Artis, Krolzig, and Toro, 2004; Anas et al., 2008). However, they have been implemented in univariate and vectorautoregressive contexts but not in a dynamic factor model. In addition, these papers intend to identify a recession, a normal growth regime, and a high growth regime, the latter being typically interpreted as a recovery in the line of Sichel (1994) and Morley and Piger (2012). The only exception is Hamilton (2005) who identifies a severe recession regime in a univariate model of the US.

[4]Another way to approach this problem is to stay with a two-regime model but make the regime-dependent growth rates follow a random walk as in Eo and Kim (2016). However, their setting differs is several important respects from ours. First, they analyze US GDP in a univariate approach. Extending it to a factor model is computationally very demanding. Second, they model the full postwar sample which is characterized by a secular decline in US growth rates while we model only the most recent 25 years of German data for which a similar decline is much less obvious. Third, their focus is on extracting in-sample features of the business cycle while we are mainly concerned with real-time forecasting for which too much parameter flexibility typically reduces forecast accuracy.

[5]While this problem can be circumvented by a two-step approach which first extracts a linear factor from the data set and subsequently uses this factor to estimate a univariate Markov-switching model, Camacho, Pérez-Quirós, and Poncela (2015) argue that the one-step method is—although it involves a higher computational burden—more robust against misspecification. Furthermore, Doz and Petronevich (2016) compare the performance of both methods on dating French business cycle turning points and find that one-step estimation is more precise in indicating the beginning and end of recessions.

Operator (LASSO). It is suited particularly for data sets with highly correlated variables like business cycle indicators.

We structure our empirical analysis in two parts. We first study whether the MS-DFM reasonably describes the German business cycle ex post using revised data for the period January 1991 to June 2016. Subsequently, we examine how well the MS-DFM is suited to timely detect, and predict, business cycle turning points in real time. In both parts, we compare the properties of models with two and three states, emphasizing the Great Recession period.

In the ex post analysis presented in Section 2.4, we apply the EN algorithm to select three out of 16 hard indicators such as new orders and three out of 19 survey indicators, all of which have been considered as early indicators in the literature on German business cycle dynamics. Using six indicators has been proven to capture business cycle dynamics quite well for several countries (see, e.g., Chauvet, 2001; Camacho and Martinez-Martin, 2015; Aastveit, Jore, and Ravazzolo, 2016).[6] We then feed these six indicators in one-factor MS-DFMs with two and three states, estimate the parameters, and smooth out the factors, which can be interpreted as composite leading indicators, and the conditional state probabilities. It turns out that the three-state model is superior in several dimensions. Its factor correlates more strongly with GDP growth (if aggregated to the quarterly frequency) and its states can be interpreted nicely as expansion, ordinary recession, and severe recession, while the two-state model seems to identify a low-growth regime and a medium-severe recession regime that is too fierce for any pre-2008 downturn and too mild for the Great Recession. The three-state model also dates recessions in general, and the Great Recession in particular, much more in line with conventional wisdom and the Bry-Boschan algorithm.[7] In contrast, the two-state model is less sensitive and thus typically comes a bit late because the business cycle needs to deteriorate considerably before it is classified as medium-severe recession.

In Section 2.5 we present the second part of our empirical analysis. We ask whether the superiority of the three-state model carries over to a forecasting situation in real time in which the data exhibit a ragged-edge structure and the Bry-Boschan algorithm is not suited because its standard version requires a lag of at least 5 months until it is able to signal a turning point. To this end, we set up a recursive nowcasting exercise from January 2001 to June 2016 that in each month selects six indicators by means of the EN algorithm and estimates one-factor MS-DFMs

[6]The results are similar when we only select four indicators or when we increase the number of selected indicators up to ten. Using more than ten indicators leads to noisy and more and more unreliable recession signals.

[7]We apply the Bry-Boschan algorithm because there is no widely accepted monthly business cycle chronology for the German economy available against which we can assess the results of our MS-DFM. The chronology published by the Economic Cycle Research Institute (ECRI) is based on both an unknown data set and an unknown method and is provided with a lag of approximately one year. The business cycle dates published by the OECD are determined by applying the Bry-Boschan algorithm on the OECD's composite leading indicator on a quarterly basis. A useful proposal is made by Schirwitz (2009) who suggests a consensus business cycle chronology based on the results of different methods. However, it is again on the quarterly frequency. Hence, we use the Bry-Boschan algorithm applied to monthly industrial production as a benchmark.

with two and three states. We find that the two-state model signals turning points fairly well but becomes instable during the Great Recession, while the three-state model appears poorly identified before the Great Recession but works properly thereafter. These results suggest that a forecaster would have dismissed the two-state model after the Great Recession and moved towards the three-state model. To operationalize this, we use real-time model selection based on the quadratic probability score and the BIC which yields a combined two-state/three-state model. It produces precise and timely nowcasts of business cycle turning points.

Using a recursive out-of-sample forecasting exercise we even demonstrate that the combined model is able to provide excellent 3-month ahead turning point predictions that would have been extremely useful for policy makers during the Great Recession. In particular, it predicts an upcoming recession with almost 100 percent probability already in July 2008 and thus four months ahead of most professional forecasters. Moreover, in March 2009 it correctly predicts that the recession comes to an end soon, one month before the German public started to discuss a third stimulus package.

Finally, we assess whether point forecasts of German GDP growth rates benefit from including the information provided by the MS-DFMs. Specifically, we augment an autoregressive forecasting model with the dynamic factor and the recession probabilities extracted from the early indicators. Specifically, augmenting an autoregressive forecasting model with the dynamic factor and the recession probabilities extracted from the early indicators considerably improves nowcasts and short-term forecasts, especially during recessions. In particular, it yields an accurate nowcast for the steepest decline in GDP in 2009Q1.

This chapter adds to the literature that applies Markov-switching models to the German business cycle. Ivanova, Lahiri, and Seitz (2000) estimate univariate Markov-switching models for various interest rate spreads and examine their predictive power for business cycle turning points. Bandholz and Funke (2003) use an MS-DFM model with a bivariate data set to construct a leading indicator for the German business cycle. Kholodilin (2005) augment that model with a second factor and interpret it as a coincidence indicator. Abberger and Nierhaus (2010) demonstrate the predictive power of the Ifo business climate index with regard to business cycle turning points in a univariate framework. Proaño and Theobald (2014) use Probit models rather than a Markov-switching approach to predict German recessions. None of these contributions considers a flexible data selection approach based on a large data set or a distinction between severe and ordinary recessions. Moreover, they are based on revised data, while we analyse the predictive ability of the model in a real-time setting.

The remainder of the chapter is structured as follows: Section 2.2 outlines our baseline MS-DFM model and the estimation method. In Section 2.3 we describe our data set and the variable selection procedure. Section 2.4 and 2.5 present our estimation results as described above. Finally, Section 2.6 concludes.

2.2 The Markov-switching dynamic factor model

We use a Markov-switching dynamic factor model (MS-DFM) to extract common nonlinear business cycle dynamics from a set of leading indicators. We distinguish between n_h hard indicators, $y_{it}^{(h)}$, such as new orders, interest rates, and oil prices, which typically account for rather short-term fluctuations, and n_s survey indicators, $y_{it}^{(s)}$, such as the Ifo business climate index and the ISM purchasing managers index which capture primarily medium-term business cycle dynamics. The distinction is important because quarterly growth rates of hard indicators generally correlate well with quarter-on-quarter GDP growth (and monthly growth rates of hard indicators correlate with monthly business cycle indicators like industrial production), while business surveys typically rather fit year-on-year GDP growth. We model these differences along the lines of Camacho et al. (2014): For the hard indicators we assume a standard factor structure,

$$y_{i,t-l_{h,i}}^{(h)} = \gamma_i^{(h)} f_t + z_{it}^{(h)}, \qquad i = 1, \ldots, n_h, \tag{2.1}$$

where $y_{i,t-l_{h,i}}^{(h)}$ is a hard indicator in monthly growth rates, $z_{it}^{(h)}$ is an idiosyncratic component, f_t is a scalar dynamic factor that leads the month-on-month business cycle dynamics by three months, and $l_{h,i}$ is the lag with which the hard indicator i enters the model. For the survey indicators we assume a slightly different specification,

$$y_{i,t-l_{s,i}}^{(s)} = \gamma_i^{(s)} \sum_{k=0}^{11} f_{t-k} + z_{it}^{(s)}, \qquad i = 1, \ldots, n_s, \tag{2.2}$$

where $y_{i,t-l_{s,i}}^{(s)}$ is a soft indicator in levels, $z_{it}^{(s)}$ is an idiosyncratic component and $l_{s,i}$ is the lag with which the soft indicator i enters the model. We include the sum of lags 0 to 11 of the factor as a parsimonious way to incorporate the phase shift associated with a year-on-year growth cycle that correlates with the survey indicators.[8]

For all indicators, we take into account that they lead the cycle to different extents and thus should enter the factor model with different lags $l_{h,i}$ and $l_{s,i}$. To make the factor lead the business cycle by 3 months, we include indicators that lead GDP by 1, 2, and 3 quarters with a lag of 0, 3, and 6 months, respectively (in Section 2.3 below we describe in detail how we choose the indicators and their lags).

[8] As a robustness check we apply an Almon lag structure as a more flexible weighting scheme. Specifically, we model the survey indicators as $y_{i,t-l_{s,i}}^{(s)} = \gamma_i^{(s)} g(\boldsymbol{\delta}, L) f_t + z_{it}^{(s)}$ where $g(\boldsymbol{\delta}, L) = \sum_{k=0}^{11} c(\boldsymbol{\delta}, k) L^k$ is a lag polynomial, L denotes the lag operator, and $\boldsymbol{\delta} = [\delta_0, \delta_1]$. We specify $c(\boldsymbol{\delta}, k)$ as an exponential Almon lag $c(\boldsymbol{\delta}, k) = \frac{exp(\delta_0 k + \delta_1 k^2)}{\sum_{k=0}^{11} exp(\delta_0 k + \delta_1 k^2)}$. Since results do not improve, we keep the parsimonious specification. Results are available upon request.

Following Doz and Petronevich (2016), we model the vector of idiosyncratic components, $z_t = [z_{1t}^{(h)}, \ldots, z_{n_h t}^{(h)}, z_{1t}^{(s)}, \ldots, z_{n_s t}^{(s)}]'$, as a diagonal VAR process of lag order q,

$$z_t = \psi_1 z_{t-1} + \cdots + \psi_q z_{t-q} + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d\, N(0, \Sigma_z), \tag{2.3}$$

where $\psi_1, \ldots, \psi_q$ and Σ_z are diagonal matrices, and ε_t is a vector of independent Gaussian shocks. We specify the common factor as an autoregressive process of lag order p with regime-dependent intercept,

$$f_t = \beta_{S_t} + \phi_1 f_{t-1} + \cdots + \phi_p f_{t-p} + \eta_t, \qquad \eta_t \sim i.i.d\, N(0, 1), \tag{2.4}$$

where η_t is an independent Gaussian shock. The intercept, β_{S_t}, depends on the state variable $S_t \in \{1, ...m\}$ as follows:

$$\beta_{S_t} = \beta_1 S_{1,t} + \beta_2 S_{2,t} + \cdots + \beta_m S_{m,t},$$

where $S_{m,t}$ is equal to unity if S_t equals m and zero otherwise. We assume that S_t follows a first-order ergodic Markov chain. The corresponding $m \times m$ transition matrix, Π, has elements p_{ij} defining the probability to switch from regime i to regime j, with $\sum_{j=1}^{m} p_{ij} = 1$ for every $i = 1, \ldots, m$. We do not impose restrictions on the duration of any regime. We consider models with two regimes ($m = 2$) that represent expansions and recessions and with three regimes ($m = 3$) with the aim to distinguish in addition between ordinary and severe recessions.

Defining the vector $y_t = [y_{1,t-l_{h,1}}^{(h)}, \ldots, y_{n_h,t-l_{h,n_h}}^{(h)}, y_{1,t-l_{s,1}}^{(s)}, \ldots, y_{n_s,t-l_{s,n_s}}^{(s)}]'$ of dimension $n = n_h + n_s$, we cast the model into state-space form,

$$y_t = B a_t, \tag{2.5}$$

$$a_t = \mu_{S_t} + F a_{t-1} + R\omega_t, \tag{2.6}$$

where a_t is the state vector, ω_t is a vector of independent Gaussian shocks with mean zero and covariance matrix Q, B, F and R are coefficient matrices, and μ_{S_t} is a state-dependent intercept. For details, we refer to Appendix A.1.

We estimate the MS-DFM by numerically maximizing the highly nonlinear likelihood function.[9] To this end, we employ the filter proposed by Kim (1994), see Appendix A.2 for details. It yields the latent dynamic regime dependent factor as well as the Markov-switching probabilities.

We use the following starting values. In a first step, we approximate f_t by a static principal components analysis and plug it into (2.4) with invariant intercept to estimate starting values for ϕ_1 to ϕ_p by OLS. We also plug f_t into (2.1) and (2.2), and run OLS regressions to obtain starting values for $\gamma^{(h)}$ and $\gamma^{(s)}$. The residuals of these regressions approximate the idiosyncratic components z_t. We use them to estimate a diagonal VAR model of lag order q to find starting

[9] We use the Matlab globalsearch class based on the routine fmincon to obtain a global maximum.

values for ψ_1 to ψ_q, and Σ_z. In the next step, we take all these values to initialize and estimate a dynamic factor model with a single regime. This yields starting values for $\gamma^{(h)}$, $\gamma^{(s)}$, $\phi_1, \ldots, \phi_p$, $\psi_1, \ldots, \psi_q$ and Σ_z. Finally, combining the results of the single-regime model with starting values for the transition matrix and the regime dependent means completes the set of required parameters. Specifically, we initialize the transition matrix by assuming persistent regimes (high values on the main diagonal and small values on the off-diagonal). We construct starting values for the regime dependent means as follows. In case of the two-state model we take the average over all positive factor values and the average of all negative factor values for the expansion and recession regime, respectively. For the three state model we use the same approach and take in addition the smallest factor value in the sample as starting value for the mean of the severe recession regime.

2.3 Indicator selection

While there are many business cycle indicators available for the German economy, the challenge is to reduce their number such that they carry all necessary cyclical information without overburdening the nonlinear maximum likelihood technique described above with estimating too many parameters. Boivin and Ng (2006) demonstrate that even linear factor models do not always benefit from adding more and more variables in particular in the context of forecasting. Camacho et al. (2015) focus specifically on MS-DFMs and show that, once a small number of high quality indicators is included, adding more indicators yields only minor improvements in terms of the identification of business cycle turning points. Finally, Schumacher (2010) shows that feeding only a set of targeted predictors into an otherwise standard factor model can improve prediction accuracy of German GDP. Hence, we first pre-select a medium-sized set of potentially useful indicators based on previous results in the literature and then apply to it a variable selection algorithm that chooses only a few final indicators to be fed into the MS-DFM.

Our pre-selection is primarily based on previous results of the literature (Fritsche and Stephan, 2002; Kholodilin and Siliverstovs, 2006; Drechsel and Scheufele, 2012; Lehmann and Wohlrabe, 2016) on the German business cycle. As hard indicators we choose 6 industrial order inflow series, 2 commodity prices, 3 interest rates, the German contribution to the EMU M2, and the DAX index which have all been found to give early business cycle signals. To take into account Germany's dependence on foreign markets, we also include US industrial production as a simple indicator for world market fluctuations. We finally add German consumer prices and employment as important economic state variables even though they are typically thought to lag the business cycle. We leave it to the indicator selection algorithm below to decide whether they are promising candidates. As survey indicators we pre-select 9 series published by the European commission and 7 series published by the Ifo institute. These series cover a broad range of economic activity, with a specific focus on expectations. We add the purchasing manager index

for the US, the Belgium business confidence indicator—which is sometimes found to lead the EU cycle (see Vanhaelen, Dresse, and De Mulder, 2000)—and the Euro-coin index to reflect the importance of major foreign markets.[10] Altogether, we pre-select a set of 35 monthly business cycle indicators, of which 16 are categorized as hard and 19 as survey indicators. To ensure stationarity, we apply log differencing to all hard indicators—except for interest rates and spreads where we compute differences without taking logs—while the survey indicators are stationary by construction. The indicators are then standardized to mean zero and variance one. A complete description is provided in Appendix A.5. Our sample starts in January 1991 in order to avoid any issues associated with the German reunification break, and runs until June 2016.

Based on the pre-selected data set, we employ an automatic indicator selection algorithm. As our goal is to provide early signals for business cycle turning points, the algorithm should select only those hard indicators that exhibit a strong lead correlation with quarter-on-quarter GDP growth rates, $\Delta \log(GDP_t)$, and only those survey indicators that exhibit a strong lead correlation with year-on-year GDP growth rates, $\Delta_4 \log(GDP_t)$. To this end, we transform our monthly indicators to quarterly frequency by averaging over the respective quarter and estimate the predictive regressions

$$\Delta \log(GDP_t) = \sum_{i=1}^{16} \sum_{l=1}^{3} b_{i,l}^{(h)} y_{i,t-l}^{(h)} + u_t^{(h)}, \tag{2.7}$$

for hard indicators and

$$\Delta_4 \log(GDP_t) = \sum_{i=1}^{19} \sum_{l=1}^{3} b_{i,l}^{(s)} y_{i,t-l}^{(s)} + u_t^{(s)}, \tag{2.8}$$

for survey indicators, where l denotes quarterly lags. The parameters $b^{(h)}$ and $b^{(s)}$ are estimated using the elastic net (EN) proposed by Zou and Hastie (2005) and successfully used by Bai and Ng (2008) for indicator selection.[11] The elastic net is a convex combination of a ridge regression and a LASSO and yields nonzero parameter estimates only for a few important indicators. It solves the following optimization problem:

$$\mathsf{L} = (\lambda_1, \lambda_2, \mathsf{b}) = |\mathsf{y} - \mathsf{X}\mathsf{b}|^2 + \lambda_1 |\mathsf{b}|_1 + \lambda_2 |\mathsf{b}|^2, \tag{2.9}$$

[10] Although it consists of both hard and survey indicators, the Euro-coin index is assigned to the survey category because it exhibits, as the other survey indicators, the highest correlation with year-on-year GDP growth (Altissimo et al., 2010).

[11] Another method to identify the relevant indicators in the context of predicting recessions are boosted regression trees (see Ng, 2014; Döpke, Fritsche, and Pierdzioch, 2017), which complement the probit approach and thus are not applicable in our case.

where $\mathsf{b} = (\mathsf{b}_1, \ldots, \mathsf{b}_N)'$ is a $N \times 1$ dimensional coefficient vector, and

$$|\mathsf{b}|_1 = \sum_j |\mathsf{b}_j| \quad \text{and} \quad |\mathsf{b}|^2 = \sum_j \mathsf{b}_j^2.$$

$\mathsf{y} = (\mathsf{y}_1, \ldots, \mathsf{y}_T)'$ denotes a centered response variable—in our setting either $\Delta \log(GDP_t)$ or $\Delta_4 \log(GDP_t)$—and $\mathsf{X} = (\mathsf{x}_1, \ldots, \mathsf{x}_N)$ is a set of N standardized predictors $\mathsf{x}_i = (\mathsf{x}_{1i}, \ldots, \mathsf{x}_{Ti})'$—in our setting either the hard indicators $y_{i,t-l}^{(h)}$, $i = 1, \ldots, 16$, $l = 1, \ldots, 3$, or the survey indicators $y_{i,t-l}^{(s)}$, $i = 1, \ldots, 19$, $l = 1, \ldots, 3$.

The tuning parameters λ_1 and λ_2 control the weight on the L_1 and L_2-norm penalty, respectively. For increasing relative weight λ_1 the EN approaches the LASSO which is known to shrink coefficients to zero due to the non-smoothness of its objective function, while for increasing relative weight λ_2 the EN approaches the ridge regression which is capable of handling highly correlated predictors. Zou and Hastie (2005) show that the EN inherits both properties and is thus particularly suited for our purpose. They also demonstrate that the EN can be transformed into a LASSO problem which can be estimated by the Least Angle Regression (LARS) of Efron, Hastie, Johnstone, and Tibshirani (2004). This algorithm, called LARS-EN, is a forward stepwise additive fitting procedure. The number of steps, k, equal the number of included variables and corresponds, for given λ_2, to a specific value of λ_1. Hence, instead of choosing λ_1 and λ_2, one may equivalently choose λ_2 and k which is what we do in the following.[12] For a detailed description of the LARS-EN algorithm along with the estimated coefficients, we refer to Appendix A.3.

We apply the LARS-EN algorithm to both (2.7) and (2.8) and choose in both cases $\lambda_2 = 100$ which is a fairly large value and allows high correlation between the selected indicators.[13] We select $n_h = 3$ hard indicators and $n_s = 3$ survey indicators in order to avoid predominance of one category and so to balance their relative merits: Hard indicators are often thought to give more reliable signals ex post but suffer from publication lags and strong revisions in real time, while soft indicators are timely available and remain largely unrevised but might be more loosely

[12] For given λ_2 this works as follows. Since LASSO shrinks coefficients to zero, start with a sufficiently large λ_1 (which yields zero estimates of all coefficients) and iteratively lower λ_1 until the prespecified number, k, of nonzero coefficient estimates is obtained.

[13] Higher values for λ_2 do not change the selection. Smaller values for λ_2 cause LARS-EN to select only one of a set of correlated indicators which is problematic in our setting because we rather select similar indicators with high correlation and good forecasting power for GDP than very different indicators of which some are only loosely related to GDP. We also tried to choose λ_2 according to cross validation based on the MSE but this method leads to inferior results which is why we do not report them here.

connected to the "hard" outcome variables such as GDP we are interested in.[14] In both cases, we thus set the elastic net parameter k to 3.[15]

2.4 Ex post business cycle dating for Germany

In the following, we apply our dynamic factor model combined with the LARS-EN indicator selection to identify the German business cycle turning points in the full sample. Such an ex post business cycle dating based on revised data is of its own interest as it complements simple but purely univariate dating algorithms like Bry-Boschan and undisclosed multivariate procedures like the one published by the Economic Cycle Research Institute (ECRI). Our main interest is, however, to show that our empirical approach produces reasonable results in-sample before we subsequently use it to predict turning points out-of-sample in a real-time forecasting setting.

2.4.1 Selected indicators

We first apply the LARS-EN algorithm with the aforementioned settings to the pre-selected set of indicators. We obtain the following results. The selected hard indicators comprise—in the order of selection—foreign orders of capital goods, domestic orders of intermediate goods, and domestic orders of capital goods. The selected survey indicators include—again in the order of selection—overall production expectations, Ifo business expectations, and Ifo export expectations. All six indicators are selected with a lag of one quarter implying that they lead the business cycle by three months. To obtain a factor with the same lead property, we include the indicators contemporaneously in the monthly factor model, i.e., set $l_{h,i} = l_{s,i} = 0$ in equations (2.1) and (2.2) for all i. Altogether the selection reflects common knowledge that orders of production inputs and business expectations are valuable early indicators. It also highlights the openness of the German economy as foreign trade plays a role in both indicator sets.

2.4.2 Factor estimate for MS(2)-DFM

Based on the selected indicators, we first estimate a "classical" two-state model, MS(2)-DFM, that distinguishes between expansions and recessions. Before estimation, we have to determine the lag orders of the factor and the idiosyncratic components. Camacho and Pérez-Quirós (2007) and Aastveit et al. (2016) argue that the main dynamics of a business cycle can be captured solely by a switching intercept, and Boldin (1996) shows for univariate Markov-switching models

[14] A series of robustness checks showed that our specification is in fact optimal to produce reliable real-time recession signals. It clearly dominates the alternative specifications $n_h = 1$ and $n_s = 5$, $n_h = 2$ and $n_s = 4$, and $n_h = 5$ and $n_s = 1$ and is slightly better than the specification $n_h = 4$ and $n_s = 2$. The results are available upon request.

[15] In some instances, the LARS-EN algorithm selects two different lags of the same indicator. In such a case, we include in our factor model the lag selected first and increase k by one to select another indicator to be included.

that overparameterization can lead to severe problems. Therefore, we set the lag order, p, of the factor to zero.[16] This allows us to treat our intercept as a switching mean. The autocorrelation functions of the idiosyncratic components indicate a lag order of $q = 2$.

The estimated means, probabilities and factor loadings of the MS(2)-DFM are reported in Table 2.1, while the autoregressive parameters of the idiosyncratic components are presented in Table 2.8 in Appendix A.4. State 1 features a positive mean of $\beta_1 = 0.32$, a high persistence probability, and occurs 87 percent of the time unconditionally. It can thus be interpreted as an expansionary regime. State 2 exhibits a negative mean of $\beta_2 = -2.12$, is less persistent, and takes place 13 percent of the time which is why it appears like a standard recession regime. However, the estimated means have a strong implication. To see this, recall that the factor is constructed from standardized indicators and has a sample mean of approximately zero. Therefore, the expansionary (recessionary) mean describes the average positive (negative) deviation from "normal times". While the scale is arbitrary, the relative sizes are not. Hence, the estimates imply that a recession is, in absolute terms, about 6.5 times stronger than an expansion. This appears very large and is a consequence of effectively treating the Great Recession as a normal recession.

Table 2.1: Estimated parameters of the MS(2)-DFM

Parameter	β_1	β_2	p_{11}	p_{22}	P_1	P_2	$\gamma_1^{(h)}$	$\gamma_2^{(h)}$	$\gamma_3^{(h)}$	$\gamma_1^{(s)}$	$\gamma_2^{(s)}$	$\gamma_3^{(s)}$
Estimate	0.32	−2.12	0.97	0.79	0.87	0.13	0.23	0.38	0.20	0.12	0.11	0.11
	(0.10)	(0.34)	(0.02)	(0.13)			(0.03)	(0.05)	(0.03)	(0.01)	(0.01)	(0.01)

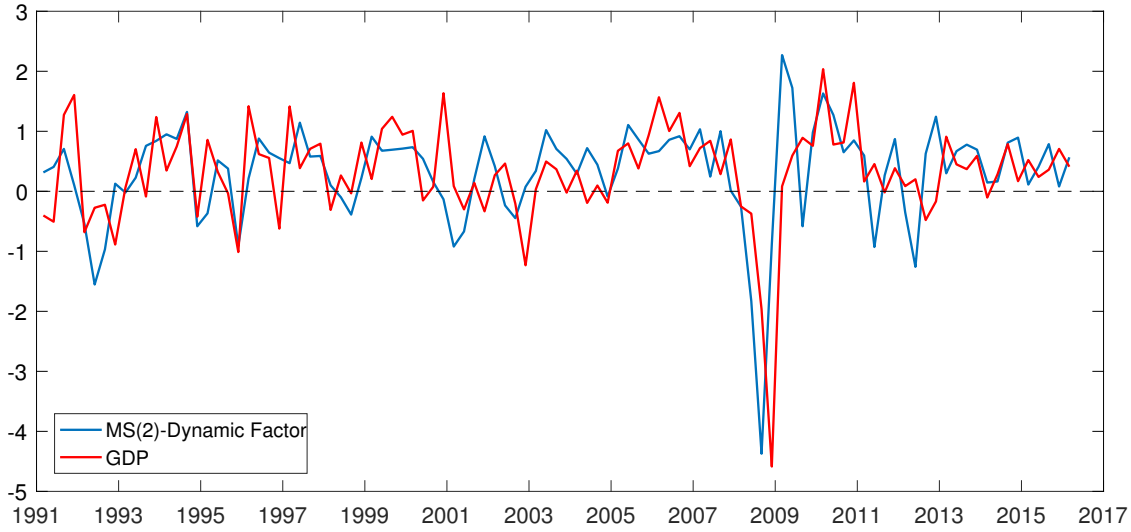
Notes: Estimated standard errors are reported in parentheses below the estimates. $\mathrm{P}_1 = \mathrm{P}(S_t = 1)$ and $\mathrm{P}_2 = \mathrm{P}(S_t = 2)$ are the unconditional probabilities of being in the expansionary and recessionary states, respectively. $\gamma_1^{(h)}$ to $\gamma_3^{(h)}$ and $\gamma_1^{(s)}$ to $\gamma_3^{(s)}$ refer to the factor loadings of the hard indicators (new foreign orders of capital goods, new domestic orders of intermediate goods, and new domestic orders of capital goods) and soft indicators (overall production expectations, overall business expectations, and export expectations), respectively.

Nevertheless, the factor corresponds closely to GDP growth, see Figure 2.1 where we display quarterly averages of the filtered factor along with quarterly German GDP growth rates. Even though the factor solely summarizes the fluctuations of the six leading indicators identified above, it tracks GDP growth remarkably well. In several instances it appears to lead GDP growth as intended by construction. In fact, it exhibits the strongest correlation of 0.64 to GDP growth with a lead of one quarter which suggests that already the MS(2)-DFM may be well suited to forecast business cycle turning points.

Finally, note that the estimated factor loadings are all positive and significantly different from zero, implying procyclicality of the selected indicators. As in previous studies (Camacho et al., 2014; Camacho and Pérez-Quirós, 2010), the soft indicators load more weakly on the factor than the hard indicators.

[16] We also estimated models with $p = 1$ and $p = 2$, but obtained inferior results for the in-sample fit, thereby confirming the results of Camacho and Pérez-Quirós (2007) and Aastveit et al. (2016).

Figure 2.1: Filtered factor of the MS(2)-DFM and GDP growth



Notes: The figure displays quarterly averages of the filtered factor estimated from an MS(2)-DFM, and quarterly German GDP growth rates, 1991Q1-2016Q2. The factor is re-scaled to fit mean and variance of GDP.

2.4.3 Factor estimate for MS(3)-DFM

Now we introduce a third state. The idea is to account for, and predict, extraordinary strong output contractions like the Great Recession. The majority of the literature only considers two regimes. The few exceptions that consider three regimes rather aim at identifying weak growth phases (sometimes called stall phases) in addition to recessions and expansions (Boldin, 1996; Ferrara, 2003; Artis et al., 2004; Nalewaik, 2011). Instead, we aim at identifying regime 1 as expansionary, regime 2 as ordinary recession and regime 3 as severe recession as in Hamilton (2005).[17]

To identify the three regimes and obtain numerically stable results of the numerical estimation procedure, we impose two economically sensible restrictions on the 3×3 transition matrix. Specifically, as in Hamilton (2005) we do not allow to directly switch from regime 1 (expansion) to regime 3 (severe recession) or vice versa. This is motivated by the observation that the Great Recession started off like an ordinary recession at the beginning of 2008, became severe after the Lehman collapse (industrial production dropped by more than 3 percent in each of the four

[17] Proaño (2017) also identifies three business cycle states for Germany. He distinguishes above trend growth, trend growth, and recessions, rather than the expansionary, recessionary and severe recessionary regime that we are interested in.

months between November 2008 and February 2009), and phased out in the subsequent months.[18] The restricted transition matrix reads as follows:

$$\Pi_r = \begin{bmatrix} p_{11} & (1-p_{11}) & 0 \\ p_{21} & p_{22} & (1-p_{21}-p_{22}) \\ 0 & (1-p_{33}) & p_{33} \end{bmatrix}. \tag{2.10}$$

Except for adding a third state, we apply the same specification choices as before. In particular, we include the same six indicators as in the two-state model, set the lag order, p, of the factor to zero and the lag order, q, of the idiosyncratic components to two. The estimated parameters of the MS(3)-DFM are reported in Table 2.2. They compare favorably to the results of the two-state approach because the relative size of the means is more in line with what one would expect. The first regime has a positive mean implying that an expansion is characterized by a positive deviation from average times. The second regime has a negative mean of an absolute size that is 2.5 times the mean of the first regime. Hence, a normal recession is characterized by a negative deviation from average times, and it is 2.5 times as strong as an expansion. The third regime has a much lower mean and can thus safely be interpreted as a severe recession. The estimate implies that a severe recession is more than five times worse than a normal recession. Not much surprisingly given the development of the Great Recession, a severe recession is estimated to be much less persistent than normal recessions and expansions. In addition, the probability to switch from the ordinary recession to the severe recession is much lower ($1-\hat{p}_{21}-\hat{p}_{22} = 0.01$) than to switch back ($1-\hat{p}_{33} = 0.34$), and the unconditional probability of being in a severe recession is much lower than that of being in an ordinary recession. The factor loadings are significantly positive and also very similar in magnitude compared to the ones of the MS(2)-DFM.

The factor of the MS(3)-DFM corresponds closely to GDP growth, see Figure 2.2. Again it appears to lead GDP growth. It exhibits the strongest correlation of 0.68 to GDP growth with a lead of one quarter. This correlation is slightly larger than the one of the two-state factor which indicates that the three-state model might be better suited to predict German business cycle turning points.

2.4.4 Which model gives a more realistic characterization of the German business cycle?

In the following, we present the smoothed recession probabilities of the two-state and three-state models and assess whether they give a realistic picture of the German business cycle phases. As a benchmark we would ideally use a generally accepted monthly business cycle chronology for

[18] A likelihood ratio test of the two restrictions was not rejected with a p-value of almost 1. In a model without the two zero restrictions the point estimates of the two transition probabilities are virtually zero with large standard errors which suggests that they are not well identified by the data.

Table 2.2: Estimated parameters of the MS(3)-DFM

Parameter	β_1	β_2	β_3	p_{11}	p_{22}	p_{33}	p_{21}	P_1	P_2	P_3
Estimate	0.61	−1.42	−7.93	0.94	0.83	0.66	0.16	0.73	0.26	0.01
	(0.12)	(0.29)	(1.08)	(0.03)	(0.10)	(0.51)	(0.09)			
Parameter	$\gamma_1^{(h)}$	$\gamma_2^{(h)}$	$\gamma_3^{(h)}$	$\gamma_1^{(s)}$	$\gamma_2^{(s)}$	$\gamma_3^{(s)}$				
Estimate	0.20	0.24	0.18	0.09	0.09	0.09				
	(0.02)	(0.04)	(0.02)	(0.01)	(0.01)	(0.01)				

Notes: Estimated standard errors are reported in parentheses below the estimates. $\mathrm{P}_1 = \mathrm{P}(S_t = 1)$, $\mathrm{P}_2 = \mathrm{P}(S_t = 2)$, and $\mathrm{P}_3 = \mathrm{P}(S_t = 3)$ are the unconditional probabilities of being in the states of expansion, recession, and severe recession, respectively. $\gamma_1^{(h)}$ to $\gamma_3^{(h)}$ and $\gamma_1^{(s)}$ to $\gamma_3^{(s)}$ refer to the factor loadings of the hard indicators (new foreign orders of capital goods, new domestic orders of intermediate goods, and new domestic orders of capital goods) and soft indicators (overall production expectations, overall business expectations, and export expectations), respectively.

Germany comparably to the one of the NBER for the US. Since this is not available, we construct our own benchmark. To this end, we apply the Bry-Boschan business cycle dating algorithm because it is an often-used method and easily replicable. Given a monthly benchmark series, x_t, the algorithm defines peaks by

$$\wedge_t = \{(x_{t-d}, \cdots, x_{t-1}) < x_t > (x_{t+1}, \cdots, x_{t+d})\},$$

and troughs by

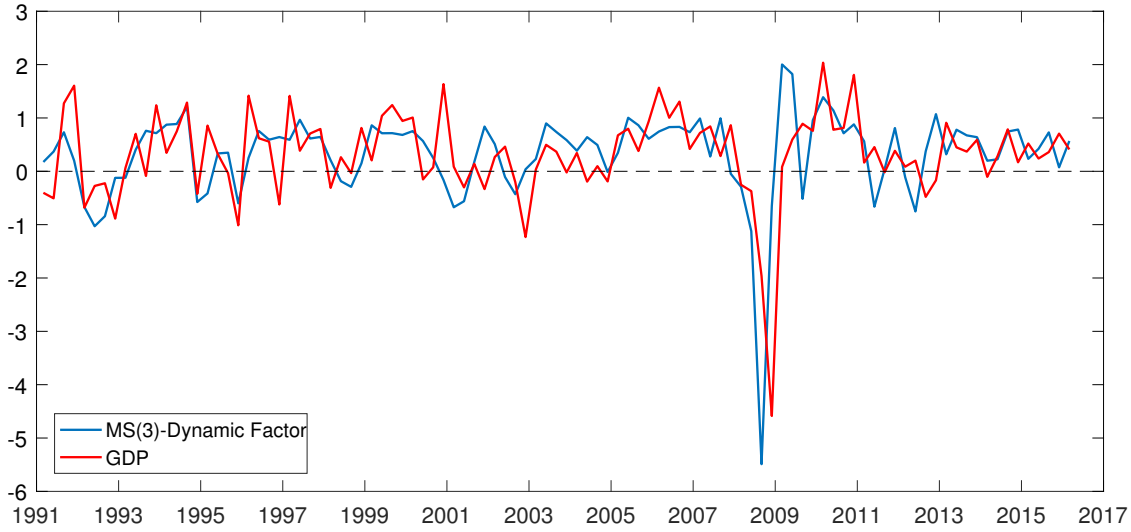
$$\vee_t = \{(x_{t-d}, \cdots x_{t-1}) > x_t < (x_{t+1}, \cdots, x_{t+d})\},$$

where d is the minimum duration which also implies that peak and trough must be at least d periods apart. The definition reveals the major drawback of the algorithm. To identify a turning point it requires at least d subsequent observations. Throughout the literature it has become standard to assume $d = 5$ months (and additionally a minimum length of a full cycle of 15 months). We follow this convention. Thus the algorithm exhibits a lag of at least 5 months until it signals that the state of the business cycle has changed, while the MS-DFM is—if it is applied in a real-time situation—designed to identify turning points instantaneously.

As benchmark series, x_t, to be fed into the Bry-Boschan algorithm we choose industrial production excluding construction.[19] This is motivated by the stylized fact that German industrial and overall activity are so strongly correlated that the industry sector, which exhibits a much more pronounced cyclical behavior than GDP, is generally thought of as the driver of the German business cycle. As industrial production is available at a monthly frequency, it enables us to determine the state of the economy on a monthly basis.

[19] We exclude construction because particularly in the 1990s after German reunification the construction cycle was decoupled from the overall business cycle in Germany.

Figure 2.2: Filtered factor of the MS(3)-DFM and GDP growth



Notes: The figure displays quarterly averages of the filtered factor estimated from an MS(3)-DFM, and quarterly German GDP growth rates, 1991Q1-2016Q2. The factor is re-scaled to fit mean and variance of GDP.

Figure 2.3 presents the smoothed recession probabilities of the MS(2)-DFM (panel a) and the MS(3)-DFM (panel b). Generally, they match the Bry-Boschan classification (indicated by shaded areas) quite well. In particular, they start rising slightly before, or at the beginning of, all benchmark recessions. Further, the MS(3)-DFM model identifies the steepest contraction of GDP during the Great Recession as a severe recession regime, while the probability of a severe recession is close to zero for the rest of the sample.

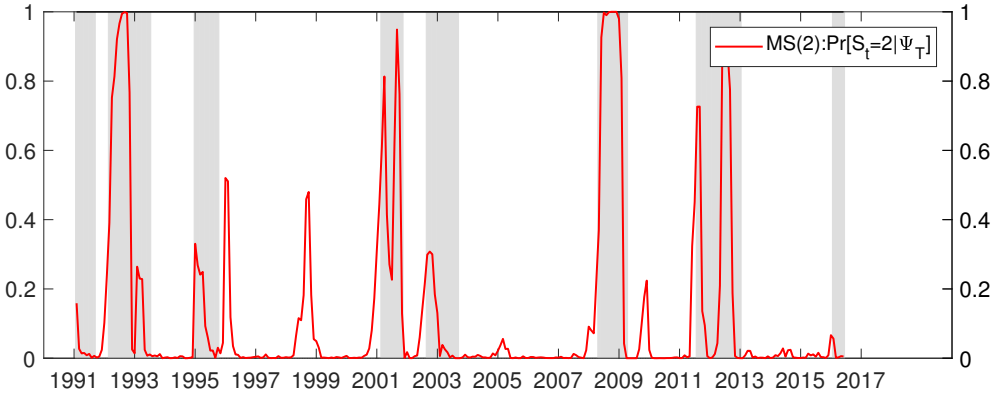
There are some important differences between the recession probabilities of the two models. The two-state model detects the Bry-Boschan recessions starting in January 1995 and September 2002 with probabilities of less than 0.4, while the three-state model identifies them with probabilities of more than 0.9.[20] This finding indicates that the three-state model is much more sensitive than its two-state counterpart because the distinction between ordinary and severe recessions allows it to assign already mildly weak times as (ordinary) recessions. The increased sensitivity can also be inferred from panel (c) of Figure 2.3 which displays the recession probabilities of the two-state model and the joint probabilities of a normal or severe recession of the three-state model. Clearly, the latter are always higher than the former.

As a potential drawback, an increased sensitivity may go hand in hand with a higher risk of false alarms. In fact, the three-state model indicates the existence of a recession in a few cases when both the two-state model and the Bry-Boschan algorithm do not. It is instructive

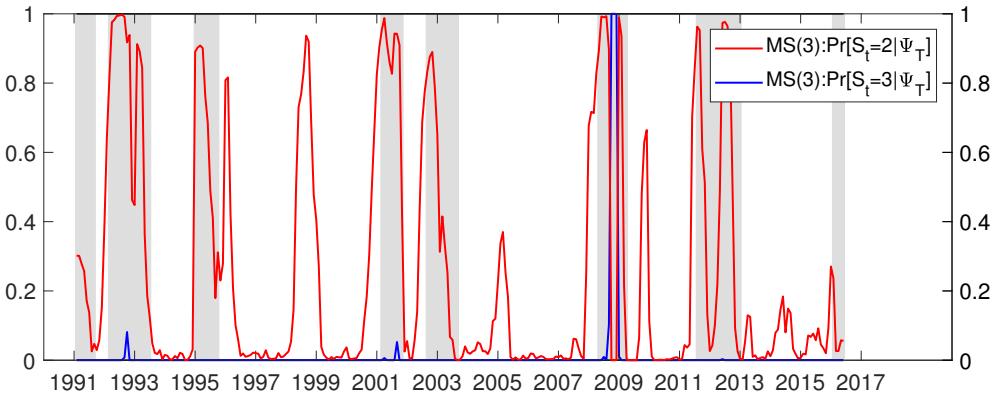
[20] A similar episode is the Bry-Boschan recession of February to June 2016. We do not take it too seriously, however, because the data are still relatively preliminary which may induce divergences in the information content of the early indicators and industrial production that may vanish after future data revisions.

Figure 2.3: Recession probabilities of MS(2)-DFM and MS(3)-DFM

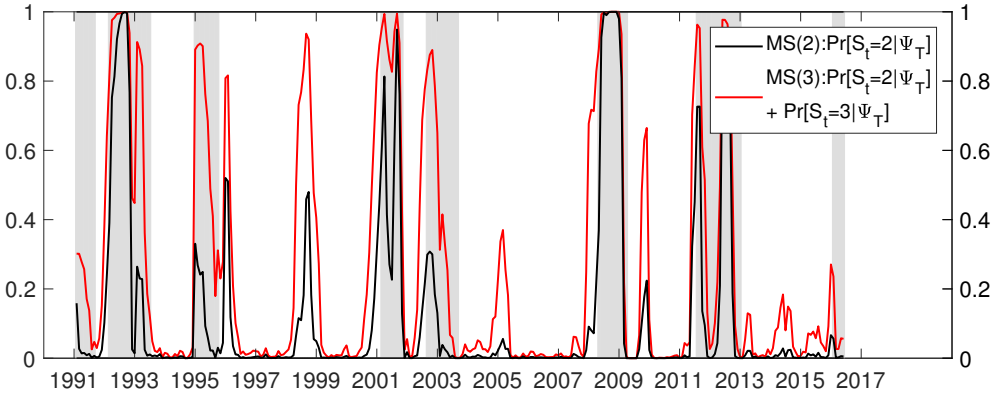
(a) Probability of a recession estimated from an MS(2)-DFM



(b) Probability of a normal and severe recession estimated from an MS(3)-DFM



(c) Comparison of the recession probabilities



Notes: Panel (a) displays smoothed recession probabilities of the MS(2)-DFM. Panel (b) displays smoothed probabilities of an ordinary recession (red line) and severe recession (blue line) of the MS(3)-DFM. Panel (c) compares the probability of a recession from the MS(2)-DFM (red line) with the joint probability of an ordinary or severe recession from the MS(3)-DFM (black line). Shaded areas correspond to the recessions dated by the Bry-Boschan algorithm.

to examine these additional signals in 1998, 2005 and at the end of 2009 in more detail. In September 1998 the recession probability of the three-state model exceeds 0.5 for 7 months in a row while the two-state probability remains slightly below 0.5 and the benchmark does not indicate a recession at all. At that time the German business cycle was temporarily fragile as indicated by a majority of the selected indicators. After a peak in July 1998, industrial production exhibited a weak period of more than 6 months before it picked up again. However, the trough was already in November 1998 which is not more than five months away from the peak. Hence, the Bry-Boschan algorithm neglects this episode.

In 2005 the recession probability of the three-state model rises to a value of just below 0.4. While it thus gives only a weak signal, it does so for good reasons. In mid-2004 the selected soft indicators started a gradual decline that continued until April 2005, and the selected hard indicators (domestic and foreign orders of capital goods and domestic orders of intermediate goods) exhibited two weak months in February and March 2005. As a consequence, the recession probability increases. The model result coincides with the assessment of professional forecasters at that time. For example, according to the Ifo business cycle forecast of June 2005 the German economy "started stuttering" (Flaig et al., 2005). Today we know that the German economy in 2005 was rather stagnating. Industrial production decreased in February, May, August and November 2005 but not in two or more months in a row. Hence, there is no local minimum to be identified by the Bry-Boschan algorithm, which therefore neglects this episode.

From November 2009 to January 2010 the recession probability of the three-state model exhibits a brief hike with a maximum of nearly 0.7. It reflects a decline in domestic orders of capital goods during September 2009 to January 2010 and drops in foreign orders of capital goods in November 2009 and January 2010, indicating a weakening business cycle. In addition, the soft indicators increase only very moderately in these months. In fact, the recovery of the German economy from the Great Recession paused during the winter 2009/10. After industrial production had increased by 11% in the first five months after its trough in April 2009, it stagnated until February 2010, but again there was no clear minimum which is why the Bry-Boschan algorithm does not indicate a recession.

These examples demonstrate that it is ultimately a matter of definition whether an episode should be classified as a recession and that it is important to combine information from hard and soft indicators. It also shows that what might appear as oversensitivity at first sight, may carry useful information that is more nuanced than a 0-1 rule.

To illustrate the leading properties of the two models, Figure 2.4 takes a closer look at the Great Recession. In panel (a) the red line represents the smoothed recession probabilities of the two-state model. Since the factor is designed to lead GDP by one quarter, a recession probability measured in month t refers to month $t + 3$. Specifically, the recession probability first exceeds 0.5 in June 2008 and thus predicts that a recession starts in September 2008, the month of the Lehman collapse. While this appears like a sensible result, it is by now conventional

wisdom that the Great Recession in Germany started earlier that year[21] while the most severe production declines came a few months later. The root of the problem is again the missing distinction between ordinary and severe recessions. As the two-state model identifies a single "average" recession, it comes late when a recession is mild.

In contrast, the three-state model almost perfectly matches the Great Recession. Panel (b) of Figure 2.4 displays the smoothed probabilities of an ordinary recession (red line) and a severe recession (blue line). The probability of an ordinary recession first rises above 0.5 in January 2008 indicating a recession start three months later in April which compares well with the development of output: the second quarter of 2008 saw the first (small) decline in GDP. The probability of a severe recession exceeds 0.5 during October to December 2008 implying that January to March 2009 are the core recession months. In fact, GDP loss in the first quarter of 2009 was by a large margin the steepest of the Great Recession. Also, industrial production fell maximally in January 2009. Altogether, the three-state model indicates that the Great Recession occurred between April 2008 and May 2009 while the Bry-Boschan algorithm identifies May 2008 to April 2009.

To more formally evaluate the two-state and three-state models against the Bry-Boschan benchmark, we employ the quadratic probability score

$$QPS = \frac{1}{T}\sum_{t=1}^{T}[B_{t+k} - \mathrm{P}_t(\text{recession})]^2, \tag{2.11}$$

where B_{t+k} denotes the binary Bry-Boschan benchmark series with lead equal to $k \in \{0, 1, 2, 3\}$ and P_t(recession) is the smoothed probability to be in a recession (two-state model) or in an ordinary or severe recession (three-state model). QPS takes an optimal value of zero if the smoothed probabilities calculated by a model coincide with the benchmark.

In addition, we compute the false positives measure

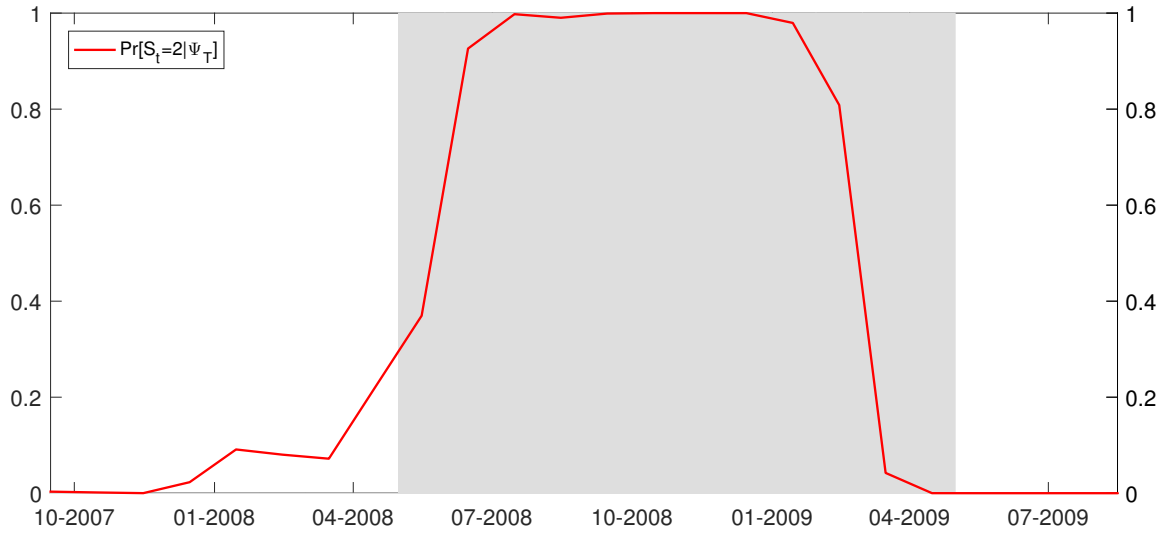
$$FPS = \frac{1}{T}\sum_{t=1}^{T}[B_{t+k} - I\{\mathrm{P}_t(\text{recession}) > 0.5\}]^2, \tag{2.12}$$

where $I\{\mathrm{P}_t(\text{recession}) > 0.5\}$ is an indicator function taking the value of 1 if the smoothed probability of being in a recession is higher than 0.5 and 0 otherwise. Hence, this measure counts the number of false signals, i.e. incorrectly predicted periods, of the model. The lower the FPS is, the better is the model's ability to reliably predict recessions.

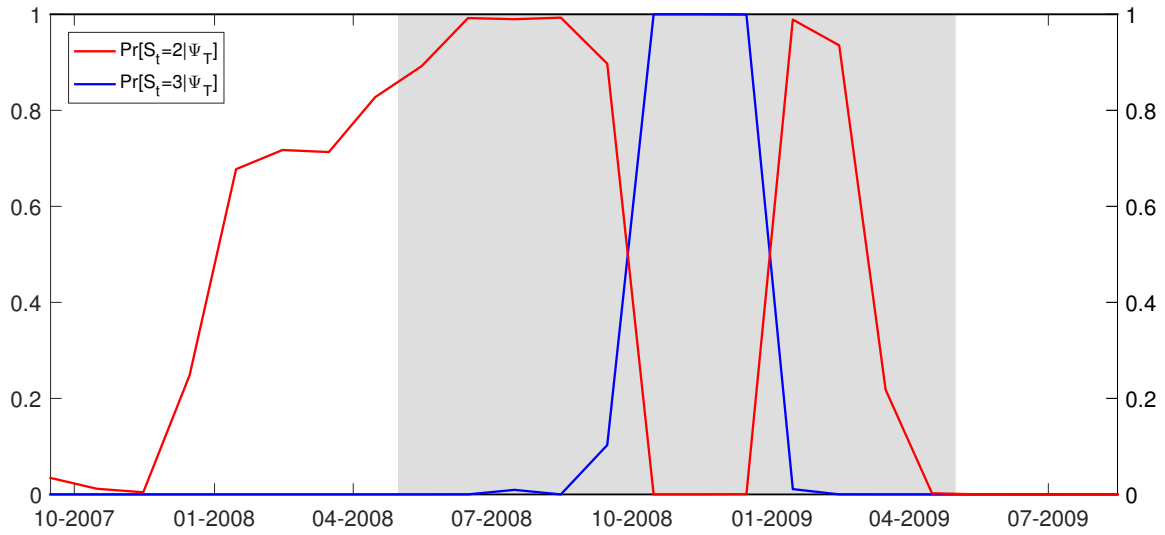
[21] Using a simple rule-of-thumb that defines a recession as at least two consecutive quarters of negative real GDP growth, one would date the start of the recession in the second quarter of 2008. Official business cycle dates from the CEPR Euro Area Business Cycle Dating Committee are only available for the euro area as a whole. According to those the business cycle peak occurred in the first quarter of 2008. Business cycle dates for Germany are released by the Economic Cycle Research Institute (ECRI) which dates the peak of the previous expansion in April 2008.

Figure 2.4: Recession probabilities of MS(2)-DFM and MS(3)-DFM during the Great Recession

(a) MS(2)-DFM



(b) MS(3)-DFM



Notes: Panel (a) displays smoothed recession probabilities of the MS(2)-DFM during the Great Recession. Panel (b) displays smoothed recession probabilities of the MS(3)-DFM during the Great Recession. The solid lines depict the model-based recession probabilities which lead the business cycle by three months. Shaded areas correspond to the recessions dated by the Bry-Boschan algorithm.

Table 2.3 reports the QPS and FPS measures for the two-state and three-state models. According to both quality measures the three-state approach provides a superior in-sample fit for all measures. This suggests that using an MS(3)-DFM gives a more realistic characterization of

the German business cycle than using a more classical two-state model. Since it provides detailed information in terms of regime probabilities we also prefer it over a simple 0-1 classification scheme like the Bry-Boschan algorithm that in addition can only classify downturns that last at least five months as recessions.

Table 2.3: QPS and FPS measures

	QPS				FPS			
k	0	1	2	3	0	1	2	3
MS(2)-DFM	0.1830	0.1702	**0.1661**	0.1725	0.2164	**0.2131**	0.2164	0.2262
MS(3)-DFM	0.1491	0.1240	**0.1089**	0.1121	0.2164	0.1803	0.1574	**0.1541**

Notes: QPS is the quadratic probability measure defined in (2.11). FPS is the false positives measure defined in (2.12). $k \in \{0, 1, 2, 3\}$ refers to the lead of of the Markov-switching models compared to the Bry-Boschan benchmark.

Additionally, the QPS and FPS measures corroborate that the Markov-switching models exhibit a lead compared to the Bry-Boschan benchmark. Specifically, the QPS measure is minimal at $k = 2$ suggesting that both models have a lead of two months, while the FPS measure is lowest at $k = 1$ month for the two-state model and $k = 3$ for the three-state model. Taken together, these results indicate that it is possible to achieve a leading property of almost one quarter by carefully selecting a set of leading indicators and integrating them into a Markov-switching dynamic factor model.

2.4.5 Monthly business cycle chronology for Germany

In some situations it may be valuable to have a dichotomous monthly business cycle chronology (even though recession probabilities are much more informative). Characterizing months with a recession probability greater than 0.5 as recessionary and assuming a lead of three months, we derive such a chronology from our preferred three-state model, see Table 2.4. We also report the chronologies based on the two-state model and the Bry-Boschan algorithm.

According to our three-state model Germany has experienced eight recessionary phases since January 1991. Particularly pronounced episodes are the post-reunification recession (May 1992 to July 1993), the "dot com" recession (March 2001 to January 2002), the Great Recession (April 2008 to May 2009), and the European sovereign debt crisis which consists of two phases (September 2011 to February 2012 and September to December 2012) summarized in columns 8a and 8b of Table 2.4.

The two-state model identifies solely those four pronounced recessions. However, the timing is always a little late and the recession lengths appear a bit underestimated. For example, according to the two-state model the Great Recession lasted only nine months and the European debt crisis as little as four months. In contrast, the Bry-Boschan benchmark indicates eight recessionary

Table 2.4: Benchmark recession dates for Germany

		Recession no.									
		1	2	3	4	5	6	7	8a	8b	9
MS(3)-DFM	start	–	05.92	04.95	08.98	03.01	10.02	04.08	09.11	09.12	–
	end	–	07.93	09.95	02.99	01.02	04.03	05.09	02.12	12.12	–
MS(2)-DFM	start	–	07.92	–	–	06.01	–	09.08	09.12		–
	end	–	02.93	–	–	01.02	–	05.09	12.12		–
Bry-Boschan	start	01.91	03.92	01.95	–	03.01	09.02	05.08	08.11		08.15
	end	09.91	07.93	10.95	–	11.01	09.03	04.09	01.13		12.15

Notes: Recessions are defined as $PR[S_t = 2|\Psi_T] \geq 0.5$ (MS(2)-DFM) and $PR[S_t = 2|\Psi_T] + PR[S_t = 3|\Psi_T] \geq 0.5$ (MS(3)-DFM), where Ψ_T is the information set available at the sample end. Episodes that last less than 4 months are excluded.

phases which in most cases coincide well with the three-state model. Exceptions are the two episodes at the sample beginning and the sample end which may be the result of a sample edge problem (in particular, potential data revisions render the 2015 recession tentative), and the episode between August 1998 and February 1999 already discussed in the previous subsection.

2.5 Real-time business cycle assessment and forecasting

In this section, we apply the Markov-switching dynamic factor models to nowcast and forecast business cycle turning points, as well as GDP growth rates, in real time. In doing so, we exploit the advantage of these models to indicate turning points instantaneously and thereby circumvent the endpoint problem inherent to the Bry-Boschan algorithm which leads to delayed signals.

2.5.1 Nowcasting German business cycle turning points

To assess the nowcasting ability of the two-state and three-state models, we perform a nowcasting experiment over the evaluation period January 2001 until June 2016 using real-time data. We choose this evaluation period because it includes five recessions which allows us to judge the results with some confidence, while the initialization sample of ten years (1991M01-2000M12) is still sufficient to estimate an MS-DFM. In addition, we include equally long periods before and after the Lehman bankruptcy which helps us to understand whether adding a third state—which is hardly identifiable before the Great Recession—would have made a difference in real time.

We construct a real-time data set consisting of the same pre-selected set of 35 indicators as in the previous section. To this end, we take the series of new orders, employed persons, and inflation from the real-time database of Deutsche Bundesbank,[22] and US industrial production

[22]Some releases miss some observations at the beginning of the sample. In such cases, we use growth rates from previous releases to fill the gaps by means of backward chaining.

from the real-time database of the OECD. The remaining hard indicators are determined on financial markets and are not revised.[23] The survey indicators are revised only very marginally, hence we neglect these revisions.

In each step of the nowcasting experiment, we go through the selection and estimation stages described in previous sections. To obtain a nowcast for month $\tau \in \{2001\text{M}01, \ldots, 2016\text{M}06\}$, we first apply the LARS-EN algorithm to the sample available at the end of this month and select a set of three hard and three survey indicators. Subsequently, we feed these indicators into the Markov-switching dynamic factor models with zero lags for the factor and two lags for idiosyncratic component, estimate the parameters, and smooth out the state probabilities. As a result, we not only obtain a series of real-time probabilities but also a time-varying selection of indicators for the period January 2001 until June 2016. Figures 2.12 and 2.13 in the Appendix depict the recursive selection of the hard and survey indicators, respectively.

It turns out that the real-time indicator selection is stable in the sense that changes in the chosen indicator sets occur infrequently. The selection reflects the traditional dependence of the German business cycle on global developments. Of the six indicators, the LARS-EN algorithm always picks two hard indicators (foreign orders of capital goods and, with very few exceptions, one of the two commodity prices) and one survey indicator (the Euro-coin indicator until April 2013 and the Ifo export expectations thereafter) that summarize external information while only two survey indicators (the Ifo business expectations and another Ifo indicator) are more closely related to the domestic situation. Interestingly, the Great Recession does not seem to affect the selection with one exception which may signal an increased relevance of the domestic economy: the sixth indicator is foreign orders of intermediate goods until February 2009 but domestic orders of intermediate goods thereafter.

The real-time nowcasts of the recession probabilities are constructed using all available information at a certain point of time. Since we only select indicators that lead the business cycle by at least 3 months, it would be sufficient to include indicators of period $\tau - 3$ and earlier in order to compute filtered probabilities of period τ. However, such an approach would neglect important information as, at the end of period τ, the realizations of, say, new orders for period $\tau - 2$ and survey indicators for period τ are already known. Therefore, we compute the real-time probabilities by means of backward smoothing taking all observations into account that are known in period τ.[24] Like Chauvet and Hamilton (2006) and Hamilton (2011), we find that these smoothed probabilities are much more stable and reliable than their filtered counterparts.

[23] The only exception is the German contribution to EMU M2. However, it is so rarely and slightly revised that we can safely take it as being unrevised.

[24] Note that this leads to ragged edges in the data structure. We deal with that complication by using the method of Mariano and Murasawa (2003) which is extended to the nonlinear Markov-switching framework by Camacho, Pérez-Quirós, and Poncela (2018). It consists of replacing the missing observations at the end of the sample by random numbers distributed independently of the model's parameters. These random numbers are in turn eliminated by an appropriately defined Kalman filter. As shown by Camacho et al. (2018), neither the maximum of the likelihood function nor the estimated filtered probabilities depend on these random numbers.

The upper panel of Figure 2.5 depicts the smoothed recession probability generated in real time by the MS(2)-DFM. It shows a roughly similar evolution as the one based on full sample estimates discussed in Section 2.4 but deviates from it in two episodes, 2004 and 2005, when it falsely signals recessions. A difference between real-time and ex post analysis can be caused by two factors. First, the ex post model is applied to revised data which is relevant in many cases because revisions of some hard indicators can be huge. Second, the real-time model suffers from the usual sample-end problem while the ex post model knows how the indicators evolve over the whole sample. This affects not only the smoothed probabilities but also the variable selection algorithm. For example, it may take a while until the real-time model replaces an indicator with deteriorating information content by another one that is better suited.

During the first episode, the real-time recession probability rose to slightly below 0.5 in June 2004, mainly because the selected three soft indicators—Ifo business expectations, intermediate goods production expectations, and the EuroCoin index—started to ease off at the beginning of 2004. The ex post recession probability does not react because based on the full sample, the latter two indicators are replaced by the overall production expectations and the Ifo export expectations which evolved more positively. In particular, export expectations tended to increase in the first three quarters of 2004. Nevertheless, industrial production stagnated—in March 2004 it was on the same level as in November 2003—but without a clear local minimum which is why the (ex post) Bry-Boschan algorithm does not indicate a recession.

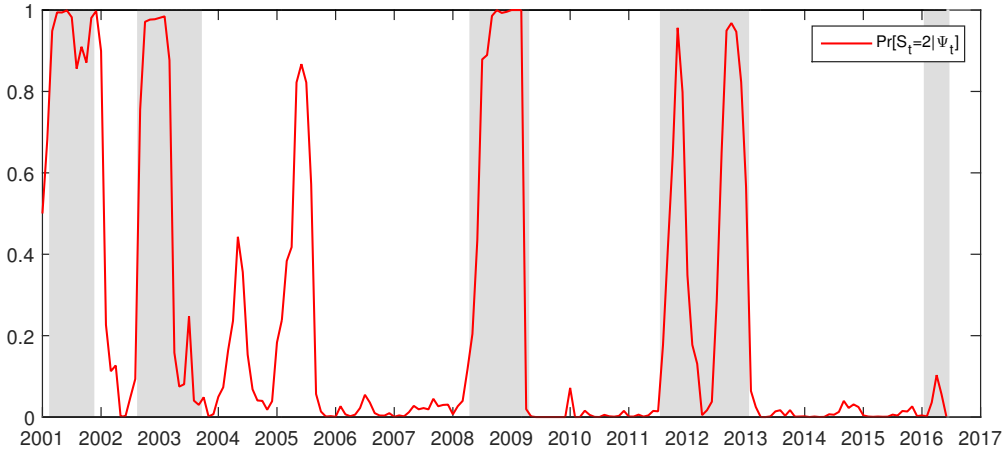
In the second episode of June to August 2005, the real-time recession probability exceeded 0.8. Again, this was primarily due to a temporary decline in the selected soft indicators at the beginning of 2005. In addition, foreign orders of capital goods and foreign orders of intermediate goods exhibited a few weak months. While there are differences to the ex post analysis both due to indicator selection and data revisions, it is remarkable that the three-state model based on revised full-sample data gives a (weak) recession signal at the same time, see the discussion in Section 2.4.4 above. As argued there, the German economy stagnated at that time but did not slip into a recession. Hence, the Bry-Boschan algorithm does not react but there is good reason for an increased recession probability.

The real-time recession probabilities estimated by the MS(3)-DFM are shown in the lower panel of Figure 2.5. They differ substantially from those based on the full sample. In particular, before the Great Recession the third state is not well identified in real time and the probabilities of being in the second or third state exhibit erratic fluctuations. Immediately after both orders and the early indicators have plummeted in the end of 2008, the third state starts to identify a severe recession. Hence, the advantage of having a third state kicks in at this point of time.

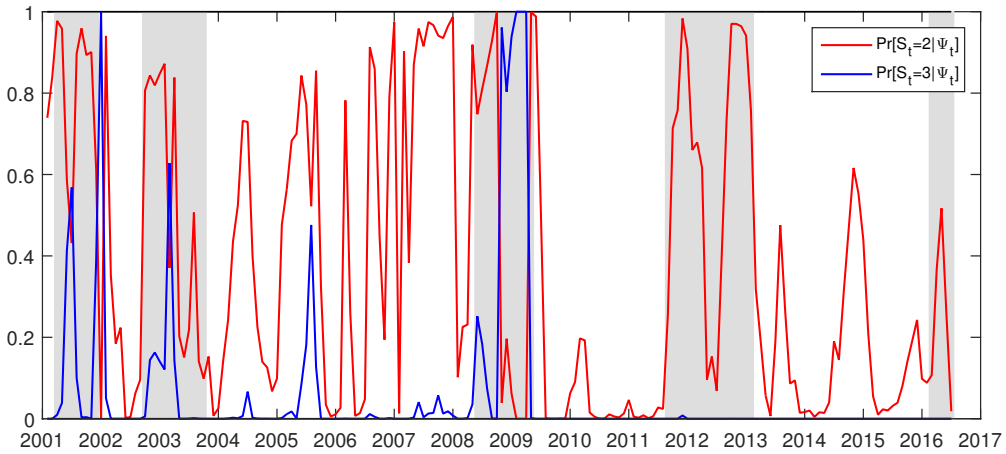
After the Great Recession, the real-time MS(3)-DFM raises two false alarms which do not show up in the ex post analysis. In August 2013 the real-time recession probabilities increased to slightly below 0.5, caused by a temporary weakness of both the selected hard indicators—foreign orders of capital goods, domestic orders of intermediate goods, and the HWWA index—and the

Figure 2.5: Real-time nowcasts of recession probability

(a) MS(2)-DFM



(b) MS(3)-DFM



Notes: Smoothed recession probabilities of (a) MS(2)-DFM and (b) MS(3)-DFM recursively estimated with real-time data. Ψ_t denotes the information set as of period t. Shaded areas correspond to the recessions of the benchmark business cycle chronology from Section 2.4.4.

selected survey indicators. In particular, the Ifo business expectations declines from February until May. In the ex post analysis the recession probabilities do not exceed 0.2 because of data revisions and differences in the selection of the hard indicators. Most notably, a real-time stagnation of foreign orders of capital goods in May is revised into a strong increase by roughly 2.4%. Moreover, the HWWA index, which is selected in real time, is ex post replaced by domestic orders of capital goods, which evolve less negatively. Industrial production in turn shows a very erratic behavior with alternating months of positive and negative growth between June and

November 2013. Therefore, the Bry-Boschan algorithm cannot detect a local minimum and does not signal a recession in autumn 2013.

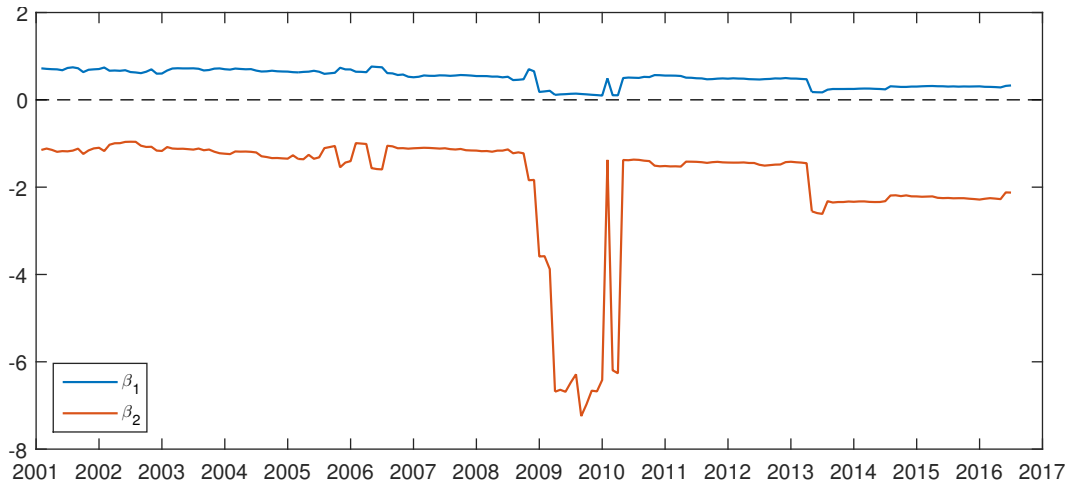
In November and December 2014 there was another false alarm with real-time recession probabilities exceeding 0.5, primarily due to a marked decline in Ifo business expectations and overall production expectations. In addition, domestic orders of intermediate goods and the HWWA commodity price index, which reflects the demand situation on world markets, exhibited weak or even negative growth rates over most of the year. The ex post analysis does not signal a recession mainly for two reasons. First, the indicator selection differs. In particular, instead of the HWWA commodity price index the ex post model selects domestic orders of capital goods which evolve less negatively. Second, the downswing of the domestic order inflow is much more pronounced in real time than using revised data. For instance, in June 2015 domestic orders of capital goods drop by 3.5% in real-time, while the revised decline is only 2.7%. The real-time results also differ from the ex post results of the Bry-Boschan algorithm because industrial production sharply decreased only in August 2014 and in January 2015 while it also saw a few positive months such that a clear local minimum is missing.

To further understand what happens inside the two models, Figure 2.6 takes a closer look at their recursively estimated state-specific means. The two-state model (panel a) exhibits a break at the beginning of the Great Recession. Before, the model is remarkably stable with a first state that has a positive mean and a second state that has a negative mean of similar absolute magnitude. Since the factor is extracted from standardized indicators and thus has a sample mean of approximately zero, the first state can be interpreted as expansion, while the second state represents a recession. During the Great Recession, however, the expansion mean is estimated as approximately zero whereas the recession mean falls dramatically. At that time, a user of this model would have found the model's result unconvincing, both because of its instability and—perhaps more importantly—because of its interpretation: neither an upswing with a growth rate that merely equals the sample average nor an extreme contraction could have been easily reconciled with what was observed as expansions and recessions before the onset of the Great Recession. These findings probably would have been interpreted as a signal that "this time is different" and that a third state is necessary to characterize the German business cycle properly.

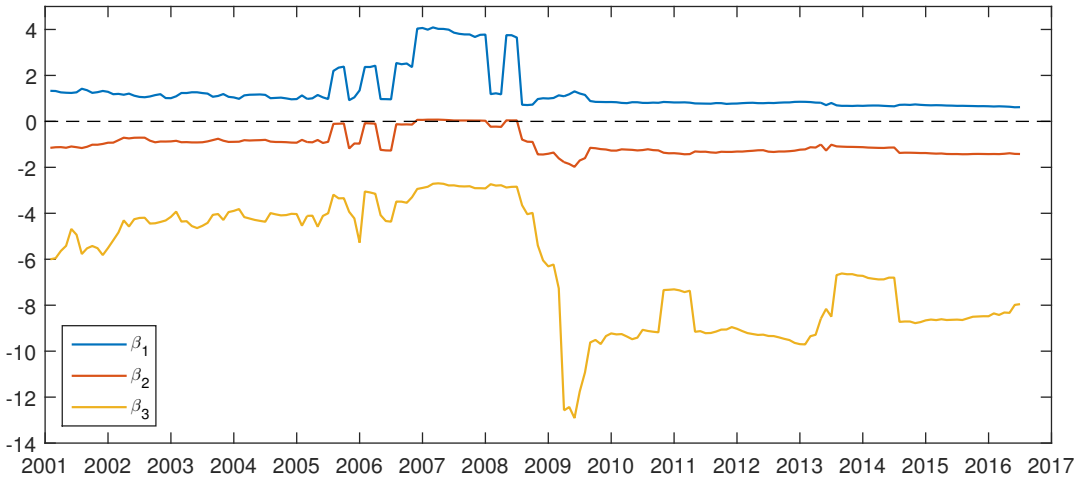
In contrast, the three-state model is instable before the Great Recession because the third state is only weakly identified during this time. Until 2005 the first two states would have been interpreted as expansion and ordinary recession, while the third state having a mean considerably smaller than the second state would have been labeled a severe recession. However, during the boom of 2006 to mid 2008 which preceded the Great Recession, the first state signals a strong boom and the third state a recession of similar absolute magnitude whereas the second state indicates "average times" with mean zero and thus average growth—an interpretation difficult to reconcile with prior experience. This changes again with the beginning of the Great Recession.

Figure 2.6: Recursively estimated means for MS(2)-DFM and MS(3)-DFM

(a) MS(2)-DFM



(b) MS(3)-DFM



Notes: Means of an MS(2)-DFM and MS(3)-DFM, estimated recursively with real-time data. Blue line: first state, red line: second state, yellow line: third state.

As more and more bad news come in, the model starts to extract a severe recession regime with a very negative mean that fluctuates—after a few months of undershooting—in a range that is considerably below the pre-crisis level. In addition, the means of the first and second state stabilize at levels that lend to the interpretation of expansion and mild recession, respectively. This stabilization is also visible in Figure 2.5 where the smoothed real-time recession probabilities largely coincide with those based on the full sample shown in Figure 2.3.

2.5.2 Model selection in real time

The results of the nowcasting experiment directly raise the issue of model selection in real time. We suggest to use either of the following two criteria to compare the two-state and three-state models. The first criterion is the QPS which measures how closely the Markov-switching models match the business cycle turning points identified by the Bry-Boschan algorithm applied to German industrial production in real time. We select the model with the better fit.[25] The second criterion is the BIC which may have the advantage over the QPS that it balances fit against parsimonity.[26] Both criteria are applied exclusively to the information sets available at each point in time to make sure this is in fact a real-time model selection without any benefit of hindsight.

Figure 2.7 plots the differences between the two-state and three-state models in terms of the QPS, $dQPS = QPS(2) - QPS(3)$, and in terms of the BIC, $dBIC = BIC(2) - BIC(3)$, based on the real-time estimates over the period January 2001 until July 2016. In both cases, a positive value indicates an advantage of the MS(3)-DFM. We find that the two-state model is superior up to the end of 2008, while the three-state model is favored thereafter. The exact change dates are very similar: the $dBIC$ selects November 2008 as the first month with an advantage of the three-state model, while the $dQPS$ identifies December 2008. Note that this date coincides with the aforementioned break in the recursively estimated state-specific means of the two-state model, see panel (a) of Figure 2.6. Hence, a user of these models would have noticed by December 2008 that introducing a third state is necessary to obtain a well-specified model.

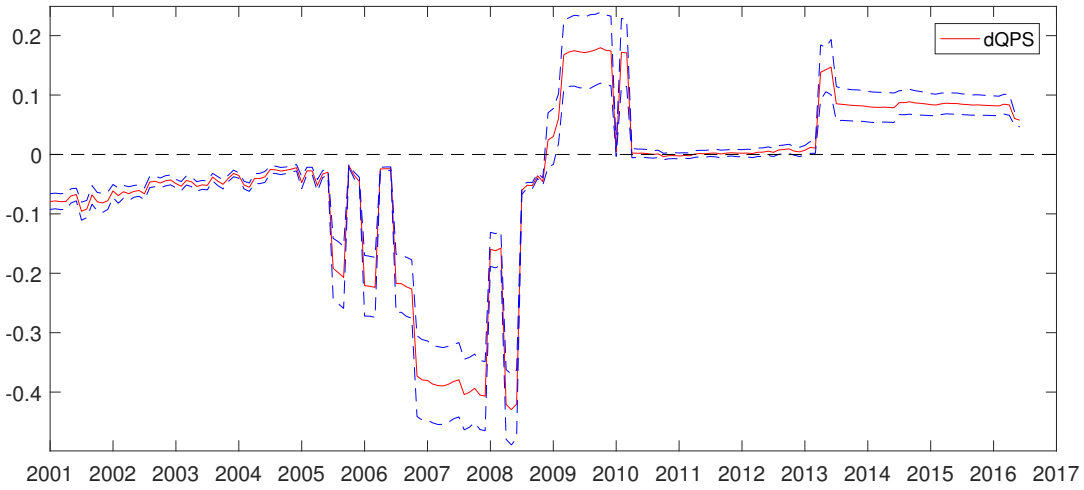
Panel (a) of Figure 2.8 shows the smoothed nowcast probabilities of a combination of the MS(2)/MS(3)-DFM, with the shift implemented in December 2008, for the whole sample, while panel (b) zooms in on the Great Recession period. The shift occurs when the probability for a severe recession reaches one in December 2008. The economy gets back to an ordinary recession in April 2009. This information about the magnitude of the recession might have been extremely helpful at this point in time as it perfectly matches the steepest part of the Great Recession: industrial production dropped by -7.2% in January 2009 and GDP dropped by -4.6% in the first quarter of 2009. Further taking the publication lag of two months for industrial production and one quarter for GDP into account, the model could have given timely information about the

[25] We take industrial production from the real-time database of the Bundesbank and run the Bry-Boschan algorithm on the information set available at each point in time. This implies that the real-time Bry-Boschan algorithm gives different results at sample ends than the ex post Bry-Boschan algorithm because identification of a turning point requires a lag of $d = 5$ months. While this means that model selection may react with a delay in real time, it is probably exactly the way an applied researcher would proceed who does not benefit from hindsight. For the Great Recession we therefore find that the real-time Bry-Boschan algorithm detects the recession start of May 2008 not before using the industrial production data vintage released on 7 November 2008 which includes the first five recession months May to September 2008.

[26] Smith, Naik, and Tsai (2006) propose a specific Markov-switching specific criterion. However, it is designed for models in which all parameters switch and thus does not work with our model. They also show that the Akaike information criterion (AIC) always selects the model with more states. Hence, we prefer the BIC over the AIC.

Figure 2.7: Recursive differences in QPS and BIC between MS(2)-DFM and MS(3)-DFM

(a) Recursive difference in QPS



(b) Recursive difference in BIC

Notes: Panel (a) shows the recursively computed $dQPS = QPS(2) - QPS(3)$ with 95% confidence bands. Panel (b) shows the recursively computed $dBIC = BIC(2) - BIC(3)$. In panel (b) we trim the observation of December 2005 (-72.56) to -40 to make the graph better readable.

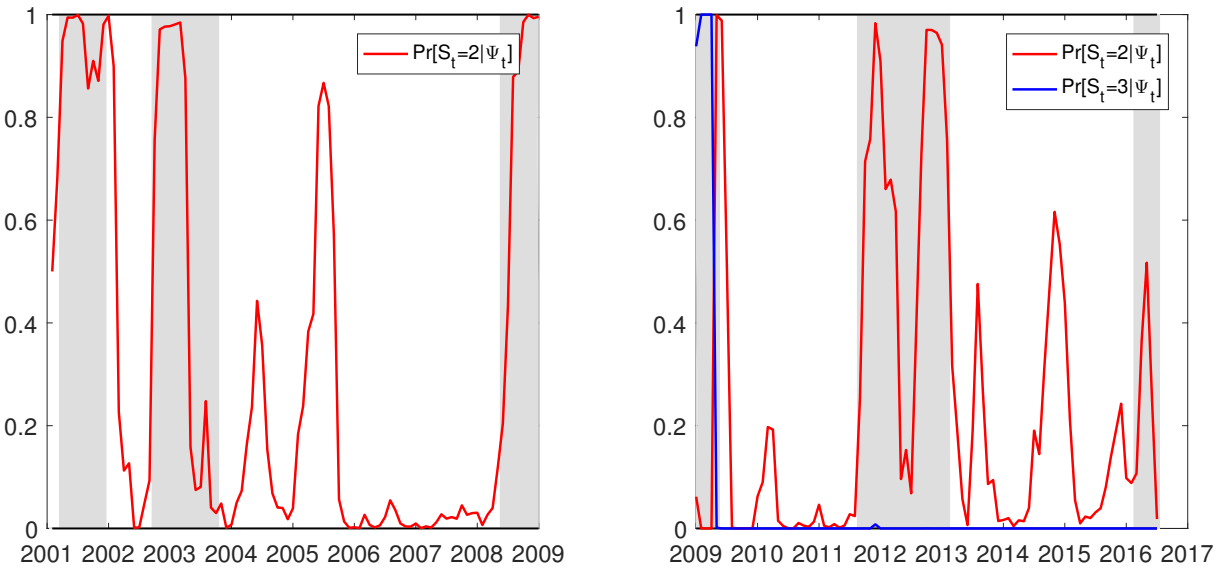
economic situation at that time and thus provided background for policy-makers to counteract the situation before knowing how deep the recession really was.

2.5.3 Forecasting German business cycle turning points

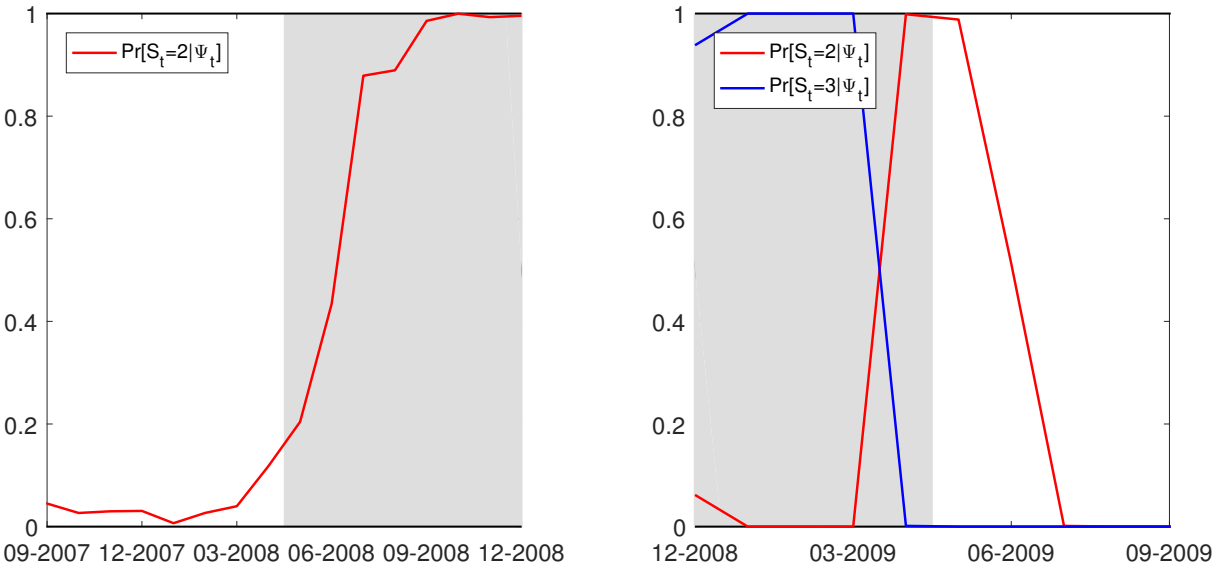
Markov-switching models can also be used to forecast future turning points. While nowcasting business cycle turning points in real time is generally difficult enough and accuracy deteriorates

Figure 2.8: Real-time nowcast of recession probabilities using an MS(2) and an MS(3)-DFM

(a) Full evaluation sample 2001M01 to 2016M06



(b) Great Recession sample



Notes: Smoothed recession probabilities of an MS(2) (left panel) and an MS(3)-DFM (right panel) during the Great Recession recursively estimated with real-time data. The split indicates the shift from MS(2)-DFM to MS(3)-DFM. Shaded areas correspond to the recessions from the benchmark business cycle chronology from Section 2.4.4.

quickly with the forecast horizon (Hamilton, 2011), our selection of early indicators that lead GDP by up to three months enables us to directly filter the probabilities $Pr[S_t = i|\Psi_{t-3}]$ from the data. It turns out that the probability forecasts of both the two-state and three-state models are somewhat more volatile than the corresponding nowcasts. This is not surprising because less information is available. Technically, this is reflected in the fact that the nowcasts are smoothed probabilities while the 3-month ahead forecasts are only filtered.

To save space, we solely report the predicted recession probabilities of the combined MS(2)/ MS(3)-DFM with the shift taking place in December 2008 as discussed above.[27] Figure 2.9 shows that they contain very useful information. For example, in July 2008 the model forecasts a recession with almost 100 percent probability for October which is remarkable as most forecasters identified the recession not before November (Heilemann and Schnorr-Bäcker, 2017). It also predicts the recovery very timely. The forecast made in January 2009 already predicts for April 2009 that the severe recession ends and the economy is back in a normal recession. And in March 2009 the model first predicts that the recession ends three months later in June 2009. We believe that this information would have been valuable at that time. For example, the German parliament passed a large stimulus package known as "Konjunkturpaket II" in February 2009, and in April the German public started to discuss another stimulus package because the end of the recession seemed far away.[28]

2.5.4 Point forecasts of German GDP

Chauvet and Potter (2013) compare a large number of GDP-forecasting models including linear univariate and multivariate time series models, DSGE models and Markov-switching models. They find that MS-DFMs are by a large difference the most successful models in predicting GDP during US recessions in real time and even outperform expert forecasts from the Blue Chip Survey. To check whether they are also useful for predicting German GDP, we conduct an out-of-sample forecast experiment using real-time data. Our MS-DFMs do not include GDP and thus do not provide directly a GDP forecast. Therefore, we augment an autoregressive distributed lag (ADL) model for quarterly GDP growth with the estimated factor and the smoothed recession probabilities,

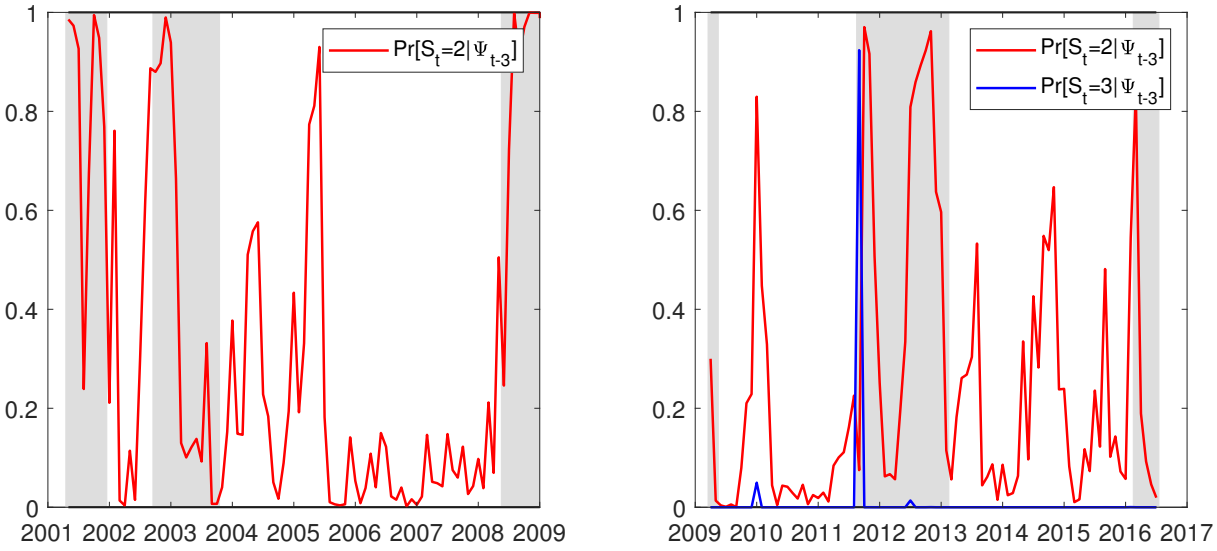
$$\Delta \log(\text{GDP}_{t+h}) = c + \sum_{j=1}^{p} \alpha_j \Delta \log(\text{GDP}_{t-j}) + \sum_{j=0}^{r} \gamma_j f_{t-j} + \sum_{j=0}^{s} \delta_j \Pi_{t-j} + \varepsilon_t, \tag{2.13}$$

where h denotes the forecast horizon, f_t denotes the quarterly average of the monthly factor, and Π_{t-j} is the quarterly average of the smoothed probability that period $t-j$ experiences a

[27] Results for the single MS(2)-DFM and MS(3)-DFM models are available upon request.

[28] The combined MS(2)/MS(3)-DFM forecasts include the same false alarms, and for the same reasons, as the respective nowcasts. Therefore, we do not discuss them here but refer the interested reader to the detailed analysis presented in Section 2.5.1.

Figure 2.9: Real-time forecast recession probabilities of MS(2)/MS(3)-DFM



Notes: One-quarter ahead recession probability forecasts $Pr[S_t = i|\Psi_t]$, $i = 2, 3$, of a n MS(2) (left panel) and MS(3)-DFM (right panel) recursively estimated with real-time data. The split indicates the shift from MS(2)-DFM to MS(3)-DFM which identified in December 2008 and thus effective for a three-month ahead forecast in March 2009. Shaded areas correspond to the recessions from the benchmark business cycle chronology from Section 2.4.4.

recession. Note that we use a direct rather than an iterative forecasting procedure. We compare the performance of the following forecasting models including a nested benchmark AR-model:

- AR: Our benchmark is a purely autoregressive model with p lags ($\gamma_j = \delta_j = 0$).
- ADL-DFM(1): This is a one-state, i.e. linear, dynamic factor model including p lags of GDP growth and r lags of the factor and $\delta_j = 0$ as there are no switches between states. We consider this model in order to check whether including additional information via a linear factor is already sufficient to improve upon the benchmark AR forecasts or whether a Markov-switching framework is essential.
- ADL-DFM(2) and ADL-DFM(3): These are ADL models which include p lags of GDP growth, r lags of a state-dependent factor and s lags of the recession probabilities generated by the MS-DFM(2)and the MS-DFM(3) model, respectively. For the latter it turned out that distinguishing between mild and severe recessions did not improve forecasting power which is why we only report results based on the joint probability $Pr[S_t = 2|\Psi_T] + Pr[S_t = 3|\Psi_T]$.
- ADL-DFM(2&3): This is an ADL model which includes p lags of GDP growth and r lags of the factor and s lags of the recession probabilities generated by the MS-DFM(2) or MS-

DFM(3) depending on which one is preferred by the BIC. The switch from the MS-DFM(2) to the MS-DFM(3) occurs in the fourth quarter of 2008.

We recursively construct real-time nowcasts ($h = 0$) and h-step forecasts for $h = 1, \ldots, 4$ quarters based on an expanding window of vintage data.[29] Since we apply direct-step forecasting, for each model we consider one lag order specification per forecast horizon h. We proceed as follows. It is a well-known feature of German GDP growth that it has almost no autocorrelation (see, for example, Pirschel and Wolters, 2018, for a comparison of autocorrelation functions of German and US GDP). Therefore, we include only one lag of GDP ($p = 1$) in all specifications. The recession probability Π_t is a first-order Markov process and includes by construction all relevant information which is why, at least theoretically, it is not necessary to include distributed lags. Since, in addition, Π_t leads GDP by one quarter, we include solely its first lag in the nowcast specifications ($s = 1$, $\gamma_0 = 0$) and its contemporaneous value in the forecast specifications ($s = 0$).[30] The factor f_t also leads GDP by one quarter. Therefore, we again exclude its current value from the nowcast specifications ($\delta_0 = 0$), but include it in the forecast specifications. At each recursion of our out-of-sample forecasting experiment we then choose the maximum lag order r as the one that minimizes the BIC.

We evaluate nowcasts and forecasts over the sample 2001Q1 to 2016Q2. Since GDP figures are subject to data revisions, we compare each forecast with the realisation published two quarters later. For example, a nowcast of 2001Q1 is compared with the value released by the end of 2001Q3. Exceptions are the major revisions of the German national account in 2005, 2011 and 2014. Here we use the last release before the revision to ensure that we take into account early data revisions but abstract from benchmark revisions which are difficult to forecast.

Before evaluating nowcasts and forecasts systematically based on RMSEs we graphically inspect the main characteristics of the different forecasting models. Figure 2.10 shows nowcasts and one-step ahead forecasts for the AR, ADL-DFM(1), ADL-DFM(2) and the ADL-DFM(3) model. The simple AR model captures the mean of GDP growth well, but mostly misses expansions and recessions. Both nowcast and forecast are basically flat until the Great Recession when the AR model reacts too late and too moderate. Afterwards, the nowcast becomes somewhat more accurate, while the forecast remains always close to the sample mean. Given the weak autocorrelation of German GDP, the result is not surprising.

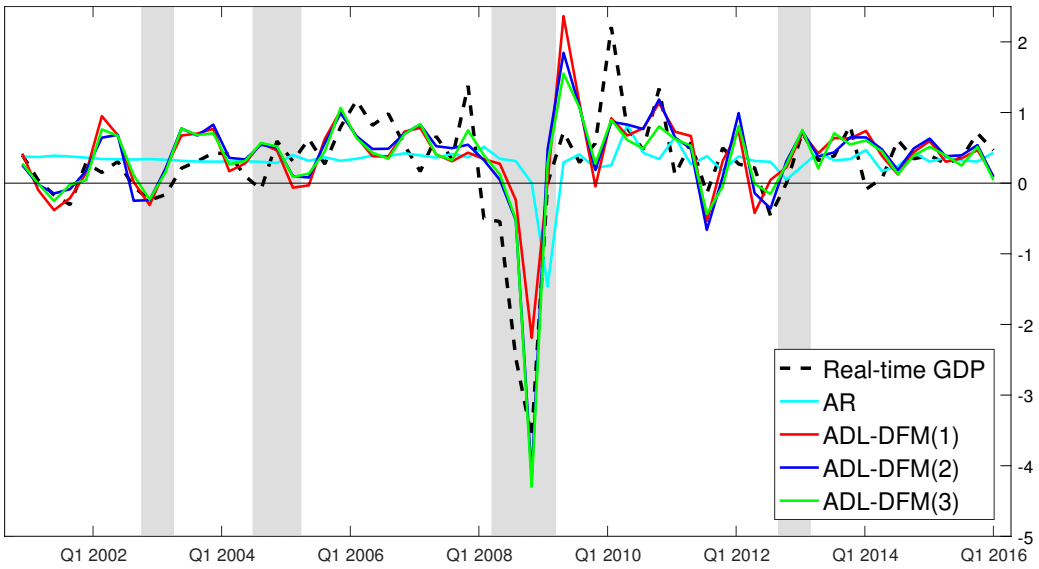
The ADL-DFM(1) includes additional information via the factor estimated from a one-state, and thus linear, dynamic factor model. It turns out that this information and in particular

[29] A real-time nowcast of, say, 2001Q1 uses the data vintage available at the end of this quarter which includes—due to its one-quarter publication lag—GDP until 2000Q4. Correspondingly, an h-step ahead forecast is based on the data vintage available at the end of 2001Q1$-h$ which includes GDP until 2000Q4$-h$. Note that some indicators also have publication lags but the dynamic factor model has not because it filters out the last observation of the vintage based on the information contained in surveys and commodity prices that are published without delay.

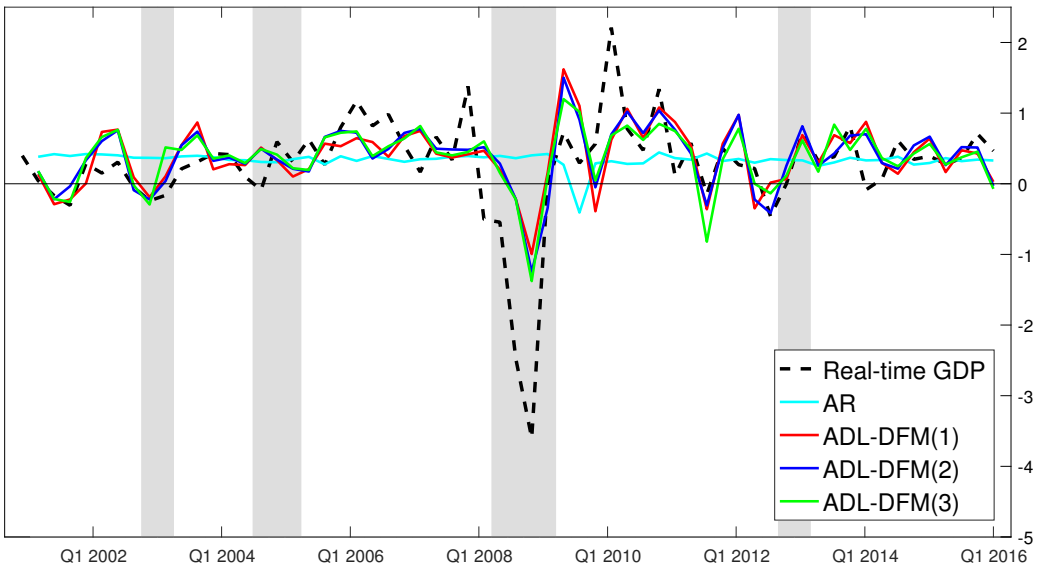
[30] We checked specifications that allowed higher choices for s but got worse RMSEs which supports our argument.

Figure 2.10: Real-time nowcasts and one-quarter ahead forecasts of GDP growth

(a) Nowcasts ($h = 0$)



(b) Forecasts ($h = 1$)



Notes: Nowcasts and forecasts are based on real-time GDP and quarterly averages of monthly recession probabilities and dynamic factors. Shaded areas correspond to recessions according to the Bry-Boschan algorithm.

the leading property of the factor is extremely valuable in generating accurate predictions. The model detects most turning points and misses only some episodes like the strong expansion in 2006 or the spike of GDP growth in 2010. It gets the timing of the largest drop in GDP during

the Great Recession right, even one quarter in advance, though by far not its actual depth. After the Great Recession the model considerably overestimates the strength of the recovery.

Turning to the Markov-switching models, both the ADL-DFM(2) and ADL-DFM(3) models improve during expansions and normal recessions only slightly upon the ADL-DFM(1) model. This changes, however, during the Great Recession and the subsequent recovery when their nowcasts for 2009Q1 are almost exactly correct and their one-quarter ahead forecasts for 2009Q1 outperform the ADL-DFM(1) model by a noticeable amount, even though they still underpredict the actual depth of the recession. During the recovery from the Great Recession, they again make more accurate predictions than the ADL-DFM(1) model. It is further noticeable that except for the Great Recession the differences between the nowcasts and the 1-quarter ahead forecast are surprisingly small for all versions of the ADL-DFM framework. The 1-quarter forecasts are almost as accurate as the nowcast.

Based on the graphical analysis we conclude that in normal times it is sufficient to use the leading information extracted by a linear factor model. In contrast, during highly volatile times like the Great Recession and the subsequent recovery when disagreement among forecasters is usually high (see, for instance, Dovern, 2015), predictions improve substantially when applying the MS-DFM(2) and MS-DFM(3) to account for the potential nonlinearity induced by those extraordinary business cycle movements.

As to the question whether to specify two or three states, we find that the predictions of the ADL-DFM(2) and ADL-DFM(3) are extremely close to each other, both before and after the Great Recession. Hence, using the information provided by the three-state Markov-switching model throughout the entire sample does not worsen GDP forecast accuracy despite the erratic switches between states before the Great Recession documented, inter alia, in Figure 2.5. Consequently, it does not make a difference here when the real-time model selection approach discussed above is applied. Since the shift from two to three states is detected in 2008Q4, the predictions of the ADL-DFM(2&3) model, which uses the information provided by the combined MS(2)/MS(3)-DFM, equal the predictions of the ADL-DFM(2) until the 2008Q4 and the ADL-DFM(3) thereafter. This is why we do not include these predictions in Figure 2.10.

Table 2.5 reports RMSEs relative to the AR model for forecasts up to $h = 4$. To test whether the forecast are significantly different from the benchmark AR-model, we employ the test proposed by Clark and West (2007). We find that all factor models provide significantly better predictions than the AR benchmark up to forecast horizon $h = 2$, with decreasing margin as the forecast horizon h increases. For a horizon of $h = 3$, only the ADL-DFM(1) outperforms the benchmark, and for $h = 4$ the AR model dominates even if not significantly so. These results are not surprising as by construction the factor leads GDP by only one quarter. Hence, for higher forecast horizons, the information provided by the factor models is much less relevant while the additional parameter estimation uncertainty remains unchanged. However, using the information from the MS-DFM is beneficial for forecasting GDP growth in spite of the low persistence

of German GDP. In line with the graphical inspection, we also find that differences between the ADL-DFM(2), ADL-DFM(3), and ADL-DFM(2&3) models are rather small, especially for forecast horizons of up to two quarters.

Table 2.5: Relative RMSEs

Model	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$
ADL-DFM(1)	0.7669***	0.8815***	0.9205***	0.9431***	1.2943
ADL-DFM(2)	0.6565***	0.8215***	0.8839***	1.3114	1.4523
ADL-DFM(3)	0.6472***	0.8616***	0.8983***	1.2471	1.5145
ADL-DFM(2&3)	0.6426***	0.8602***	0.9072***	1.2653	1.5201

Notes: Root mean squared errors relative to an AR-benchmark model. *, **, and *** denote significance on the 10% level, 5% level, and 1% level, respectively, according to the Clark-West test with Newey-West standard errors.

The graphical analysis showed that the Markov-switching models perform particularly well during recessions. Hence, it is of interest to analyze differences in forecast precision between recessions and expansions systematically. To this end, we employ the quarterly version of the Bry-Boschan algorithm (Harding and Pagan, 2002) to GDP. The recession subsample includes 11 quarters (2002Q4-2003Q1, 2004Q3-2005Q1, 2008Q2-2009Q1, and 2012Q4-2013Q1), while the expansion subsample covers the remaining 55 quarters. Table 2.6 reports the corresponding RMSEs relative to the AR-benchmark model. These results confirm the finding by Chauvet and Potter (2013) that the advantage of Markov-switching models is largest during recessions. Interestingly, these models also improve upon the linear factor model during expansions, albeit to a smaller extent.

2.6 Conclusion

We provide evidence that Markov-switching dynamic factor models together with a flexible variable pre-selection algorithm are an appropriate device to predict and date business cycle turning points for the German economy. It turns out that a three-state model is more sensitive than a two-state model and provides a better ex post characterization of the German business cycle, especially because it identifies the Great Recession as a severe recession. Using real-time data we show that nowcasts and one-quarter ahead forecasts capture business cycle dynamics in Germany well even though German GDP growth is characterized by very low persistence.

During the Great Recession the model predicts the timing of events one quarter in advance starting with the initially mild downturn, the severe recessionary phase afterwards, and finally the recovery. Further, a comparison of the two- and three-state model clearly signals that the three-state model would have been preferable in December 2008 right before the biggest downturn of the German economy. Hence, for professional forecasters using this framework during the Great

Table 2.6: Relative RMSEs for recessions

Model	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$
			Recessions		
ADL-DFM(1)	0.6186**	0.8154**	0.8567**	0.9100**	1.3229
ADL-DFM(2)	0.4910**	0.7274***	0.7634**	0.9923	1.5005
ADL-DFM(3)	0.5111**	0.7920**	0.8611**	1.1746	1.4960
ADL-DFM(2&3)	0.5019**	0.7831**	0.8643**	1.1715	1.5006
			Expansions		
ADL-DFM(1)	0.9812***	1.0186***	1.0665***	1.0184	1.2160
ADL-DFM(2)	0.8816***	1.0066***	1.1353***	1.8687	1.3172
ADL-DFM(3)	0.8403***	1.0048***	0.9871**	1.4066	1.5621
ADL-DFM(2&3)	0.8406***	1.0171***	1.0087***	1.4669	1.5705

Notes: Relative root mean squared errors during recessions and expansions. *, **, and *** denote significance on the 10% level, 5% level, and 1% level, respectively, according to the Clark-West test with Newey-West standard errors.

Recession would have been valuable to predict events systematically based on leading indicators. Moreover, the framework would have been highly useful for policymakers in order to plan the timing of policies to mitigate the crisis without the danger of stimulating the economy when the recovery was already on the way.

References

Aastveit, K. A., A. S. Jore, and F. Ravazzolo (2016). Identification and Real-Time Forecasting of Norwegian Business Cycles. *International Journal of Forecasting 32*(2), 283–292.

Abberger, K. and W. Nierhaus (2010). Markov-Switching and the Ifo Business Climate: the Ifo Business Cycle Traffic Lights. *OECD Journal: Journal of Business Cycle Measurement and Analysis 2010*(2), 1–13.

Altissimo, F., R. Cristadoro, M. Forni, M. Lippi, and G. Veronese (2010). New Eurocoin: Tracking Economic Growth in Real Time. *The Review of Economics and Statistics 92*(4), 1024–1034.

Anas, J., M. Billio, L. Ferrara, and G. L. Mazzi (2008). A System for Dating and Detecting Turning Points in the Euro Area. *The Manchester School 76*(5), 549–577.

Artis, M., H.-M. Krolzig, and J. Toro (2004). The European Business Cycle. *Oxford Economic Papers 56*(1), 1–44.

Bai, J. and S. Ng (2008). Forecasting Economic Time Series Using Targeted Predictors. *Journal of Econometrics 146*(2), 304–317.

Bandholz, H. and M. Funke (2003). In Search of Leading Indicators of Economic Activity in Germany. *Journal of Forecasting 22*(4), 277–297.

Boivin, J. and S. Ng (2006). Are More Data Always Better for Factor Analysis? *Journal of Econometrics 132*(1), 169–194.

Boldin, M. D. (1996). A Check on the Robustness of Hamilton's Markov Switching Model Approach to the Economic Analysis of the Business Cycle. *Studies in Nonlinear Dynamics & Econometrics 1*(1), 35–46.

Camacho, M. and J. Martinez-Martin (2015). Monitoring the World Business Cycle. *Economic Modelling 51*, 617–625.

Camacho, M. and G. Pérez-Quirós (2007). Jump-and-Rest Effect of US Business Cycles. *Studies in Nonlinear Dynamics & Econometrics 11*(4), 1–39.

Camacho, M. and G. Pérez-Quirós (2010). Introducing the Euro-STING: Short Term Indicator of Euro Area Growth. *Journal of Applied Econometrics 25*(4), 663–694.

Camacho, M., G. Pérez-Quirós, and P. Poncela (2014). Green Shoots and Double Dips in the Euro Area: A Real Time Measure. *International Journal of Forecasting 30*(3), 520–535.

Camacho, M., G. Pérez-Quirós, and P. Poncela (2015). Extracting Nonlinear Signals from Several Economic Indicators. *Journal of Applied Econometrics 30*(7), 1073–1089.

Camacho, M., G. Pérez-Quirós, and P. Poncela (2018). Markov-Switching Dynamic Factor Models in Real Time. *International Journal of Forecasting 34*(4), 598–611.

Chauvet, M. (1998). An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching. *International Economic Review 39*(4), 969–996.

Chauvet, M. (2001). A Monthly Indicator of Brazilian GDP. *Brazilian Review of Econometrics 21*(1), 1–15.

Chauvet, M. and J. D. Hamilton (2006). Dating Business Cycle Turning Points. In C. Milas, P. Rothman, D. v. Dijk, and D. E. Wildasin (Eds.), *Nonlinear Time Series Analysis of Business Cycles*, Chapter 1, 1–54. Emerald Group Publishing Limited.

Chauvet, M. and S. Potter (2013). Forecasting Output. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Chapter 3, 141–194. Elsevier.

Clark, T. E. and K. D. West (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics 138*(1), 291–311. 50th Anniversary Econometric Institute.

Diebold, F. X. and G. D. Rudebusch (1996). Measuring Business Cycles: A Modern Perspective. *The Review of Economics and Statistics 78*(1), 67–77.

Döpke, J., U. Fritsche, and C. Pierdzioch (2017). Predicting Recessions with Boosted Regression Trees. *International Journal of Forecasting 33*(4), 745–759.

Dovern, J. (2015). A Multivariate Analysis of Forecast Disagreement: Confronting Models of Disagreement with Survey Data. *European Economic Review 80*, 16–35.

Doz, C. and A. Petronevich (2016). Dating Business Cycle Turning Points for the French Economy: An MS-DFM approach. In S. J. K. Eric Hillebrand (Ed.), *Dynamic Factor Models*, Volume 35 of *Advances in Econometrics*, Chapter 12, 481–538. Emerald Group Publishing Limited.

Drechsel, K. and R. Scheufele (2012). The Performance of Short-Term Forecasts of the German Economy Before and During the 2008/2009 Recession. *International Journal of Forecasting 28*(2), 428–445.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least Angle Regression. *The Annals of Statistics 32*(2), 407–499.

Eo, Y. and C.-J. Kim (2016). Markov-Switching Models with Evolving Regime-Specific Parameters: Are Postwar Booms or Recessions All Alike? *Review of Economics and Statistics 98*(5), 940–949.

Ferrara, L. (2003). A Three-Regime Real-Time Indicator for the US Economy. *Economics Letters 81*(3), 373–378.

Flaig, G., W. Nierhaus, O.-E. Kuntze, A. Gebauer, S. Henzel, O. Hülsewig, A. Dehne, E. Langmantel, W. Meister, M. Ruschinski, and B. Schimpfermann (2005). ifo Konjunkturprognose 2005/2006: Nur zögerliche Erholung. *ifo Schnelldienst 58*(12), 29–63.

Fritsche, U. and S. Stephan (2002). Leading Indicators of German Business Cycles. An Assessment of Properties/Frühindikatoren der deutschen Konjunktur. Eine Beurteilung ihrer Eigenschaften. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics 222*(3), 289–315.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica 57*(2), 357–384.

Hamilton, J. D. (2005). What's Real About the Business Cycle? *Federal Reserve Bank of St. Louis Review Jul*(87(4)), 435–452.

Hamilton, J. D. (2011). Calling Recessions in Real Time. *International Journal of Forecasting 27*(4), 1006–1026.

Harding, D. and A. Pagan (2002). A Comparison of Two Business Cycle Dating Methods. *Journal of Economic Dynamics & Control 27*(9), 1681–1690.

Hastie, T., R. Tibshirani, and J. Friedman (2017). *Elements of Statistical Learning*. Springer.

Heilemann, U. and S. Schnorr-Bäcker (2017). Could the Start of the German Recession 2008-2009 Have Been Foreseen? Evidence from Real-Time Data. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics 237*(1), 29–62.

Ho, W.-Y. A. and J. Yetman (2014). Do economies stall? *Applied Economics 46*(35), 4267–4275.

Hoerl, A. E. and R. W. Kennard (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics 12*(1), 55–67.

Ivanova, D., K. Lahiri, and F. Seitz (2000). Interest Rate Spreads as Predictors of German Inflation and Business Cycles. *International Journal of Forecasting 16*(1), 39–58.

Kholodilin, K. A. (2005). Forecasting the German Cyclical Turning Points: Dynamic Bi-Factor Model with Markov Switching. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics 225*(6), 653–674.

Kholodilin, K. A. and B. Siliverstovs (2006). On the Forecasting Properties of the Alternative Leading Indicators for the German GDP: Recent Evidence. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics 226*(3), 234–259.

Kim, C.-J. (1994). Dynamic Linear Models with Markov-switching. *Journal of Econometrics 60*(1-2), 1–22.

Kim, C.-J. and C. R. Nelson (1998). Business Cycle Turning Points, A New Coincident Index, And Tests Of Duration Dependence Based On A Dynamic Factor Model With Regime Switching. *Review of Economics and Statistics 80*(2), 188–201.

Kim, M.-J. and J.-S. Yoo (1995). New Index of Coincident Indicators: A Multivariate Markov Switching Factor Model Approach. *Journal of Monetary Economics 36*(3), 607–630.

Krolzig, H.-M. and J. Toro (2001). A New Approach to the Analysis of Business Cycle Transitions in a Model of Output and Employment. Economics Series Working Papers 59, University of Oxford, Department of Economics.

Layton, A. P. and D. Smith (2000). A Further Note on the Three Phases of the US business Cycle. *Applied Economics 32*(9), 1133–1143.

Lehmann, R. and K. Wohlrabe (2016). Looking Into the Black Box of Boosting: The Case of Germany. *Applied Economics Letters 23*(17), 1229–1233.

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.

McAdam, P. (2007). USA, Japan and the Euro Area: Comparing Business-Cycle Features. *International Review of Applied Economics 21*(1), 135–156.

Morley, J. and J. Piger (2012). The Asymmetric Business Cycle. *Review of Economics and Statistics 94*(1), 208–221.

Nalewaik, J. J. (2011). Forecasting Recessions Using Stall Speeds. In *Finance and Economics Discussion Series 2011-24, Federal Reserve Board*. Citeseer.

Ng, S. (2014). Boosting Recessions. *Canadian Journal of Economics/Revue canadienne d'économique 47*(1), 1–34.

Pirschel, I. and M. H. Wolters (2018). Forecasting with Large Datasets: Compressing Information Before, During or After the Estimation? *Empirical Economics 55*(2), 573–596.

Proaño, C. R. (2017). Detecting and Predicting Economic Accelerations, Recessions, and Normal Growth Periods in Real-Time. *Journal of Forecasting 36*(1), 26–42.

Proaño, C. R. and T. Theobald (2014). Predicting Recessions with a Composite Real-Time Dynamic Probit Model. *International Journal of Forecasting 30*(4), 898–917.

Schirwitz, B. (2009). A Comprehensive German Business Cycle Chronology. *Empirical Economics 37*(2), 287–301.

Schumacher, C. (2010). Factor Forecasting Using International Targeted predictors: The Case of German GDP. *Economics Letters 107*(2), 95–98.

Sichel, D. E. (1994). Inventories and the Three Phases of the Business Cycle. *Journal of Business & Economic Statistics 12*(3), 269–277.

Smith, A., P. A. Naik, and C.-L. Tsai (2006). Markov-Switching Model Selection Using Kullback–Leibler Divergence. *Journal of Econometrics 134*(2), 553–577.

Stock, J. H. and M. W. Watson (2005). Understanding Changes in International Business Cycle Dynamics. *Journal of the European Economic Association 3*(5), 968–1006.

Vanhaelen, J.-J., L. Dresse, and J. De Mulder (2000). The Belgian Industrial Confidence Indicator: Leading Indicator of Economic Activity in the Euro Area? Working Paper 12, National Bank of Belgium.

Zou, H. and T. Hastie (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(2), 301–320.

A Appendix

A.1 Construction of the state space form

We start defining the $(12 + nq)$-dimensional state vector

$$a_t = [f_t, \ldots, f_{t-11}, z_t', \ldots, z_{t-q+1}']'.$$

Now the measurement equations (2.1) and (2.2) can be jointly written as

$$y_t = Ba_t,$$

where

$$B = \begin{bmatrix} \gamma^{(h)} & 0_{n_h \times 1} & \cdots & 0_{n_h \times 1} & I_{n_h} & 0_{n_h \times n_s} & 0_{n_h \times (nq-q)} \\ \gamma^{(s)} & \gamma^{(s)} & \cdots & \gamma^{(s)} & 0_{n_s \times n_h} & I_{n_s} & 0_{n_s \times (nq-q)} \end{bmatrix},$$

and $\gamma^{(h)} = [\gamma_1^{(h)}, \ldots, \gamma_{n_h}^{(h)}]'$ and $\gamma^{(s)} = [\gamma_1^{(s)}, \ldots, \gamma_{n_s}^{(s)}]'$.

The transition equation can be written as

$$a_t = \mu_{S_t} + Fa_{t-1} + R\omega_t,$$

using the following definitions. The system matrix is

$$F = \begin{bmatrix} F_{11} & 0_{12 \times nq} \\ 0_{nq \times 12} & F_{22} \end{bmatrix},$$

where F_{11} is the (12×12)-dimensional companion matrix of an AR(12) process with lag coefficients ϕ_1 to ϕ_{12} of which coefficients 3 to 12 restricted to zero because we only allow a maximum lag order of $p = 2$ for f_t, and F_{22} is the $(nq \times nq)$-dimensional companion matrix of an n-dimensional VAR process with q lags and coefficient matrices ψ_1 to ψ_q. The intercept vector is nonzero only for f_t and thus is

$$\mu_{S_t} = \left[\beta_{S_t}, 0_{1 \times (11+nq)}\right]'.$$

The vector of iid shocks, $\omega_t = [\eta_t, \varepsilon_t']'$, is iid normally distributed with mean zero and diagonal covariance matrix

$$Q \equiv \mathrm{E}(\omega_t \omega_t') = \begin{bmatrix} 1 & 0_{1 \times n} \\ 0_{n \times 1} & \Sigma_z \end{bmatrix}.$$

Finally, we define the coefficient matrix

$$R = \begin{bmatrix} R_{11} & 0_{12 \times n} \\ 0_{nq \times 1} & R_{22} \end{bmatrix},$$

where $R_{11} = [1, 0_{1\times 11}]'$ and $R_{22} = [I_n, 0_{n\times(nq-n)}]'$.

A.2 Estimation of the MS-DFM

We employ the filter proposed by Kim (1994) to estimate the MS-DFM. Based on the initialization $a_{0|0} = (I - F)^{-1}\mu_{S_t}$ and $P_{0|0} = (I - F \otimes F)^{-1}\text{vec}(Q)$, the recursion consists of the usual prediction and updating steps. To this end, let us define $P^{(j,i)}_{t|t-1}$ as the variance of z_t conditional on Ψ_{t-1}, the information available in $t-1$, and on $S_t = j$ and $S_{t-1} = i$, $P^{(i)}_{t|t}$ as the variance of z_{t-1} conditional on Ψ_t and $S_{t-1} = i$, and equivalently $a^{(j,i)}_{t|t-1}$ and $a^{(i)}_{t-1|t-1}$. Then the prediction step is

$$a^{(j,i)}_{t|t-1} = Fa^{(i,k)}_{t-1|t-1} + \mu^{(j)}_{S_t}, \tag{A.1}$$

$$P^{(j,i)}_{t|t-1} = FP^{(i,k)}_{t-1|t-1}F' + RQR', \tag{A.2}$$

and the updating step is

$$a^{(j,i)}_{t|t} = a^{(j,i)}_{t|t-1} + K^{(j,i)}_t(y_t - Ba^{(j,i)}_{t|t-1}), \tag{A.3}$$

$$P^{(j,i)}_{t|t} = (I_{2n+p} - K^{(j,i)}_t B)P^{(j,i)}_{t|t-1}, \tag{A.4}$$

where the Kalman gain is defined by $K^{(j,i)}_t = P^{(j,i)}_{t|t-1}B'(BP^{(j,i)}_{t|t-1}B')^{-1}$. However, each recursion generates an m-fold increase in the number of states to be considered. Therefore, we apply the approximation by Kim (1994),

$$a^{(j)}_{t|t} = \frac{\sum_{i=1}^m Pr[S_{t-1} = i, S_t = j|\Psi_t]a^{(j,i)}_{t|t}}{Pr[S_t = j|\Psi_t]}, \tag{A.5}$$

$$P^{(j)}_{t|t} = \frac{\sum_{i=1}^m Pr[S_{t-1} = i, S_t = j|\Psi_t](P^{(j,i)}_{t|t} + (a^{(j)}_{t|t} - a^{(j,i)}_{t|t})(a^{(j)}_{t|t} - a^{(j,i)}_{t|t})')}{Pr[S_t = j|\Psi_t]}, \tag{A.6}$$

which reduces the number of possible states of $a_{t|t}$ and $P_{t|t}$ to m per period by taking weighted averages over the states and feeding them into the prediction steps (A.1) and(A.2).

The corresponding log likelihood function is obtained by Hamilton (1989):

$$ln\ L = \sum_{t=1}^{T} ln\left(\sum_{S_t=1}^{m}\sum_{S_{t-1}=1}^{m} f(y_t|S_t, S_{t-1}, \Psi_{t-1})Pr(S_t = j, S_{t-1} = i|\Psi_{t-1})\right). \tag{A.7}$$

Evaluating it requires calculating the weights $Pr(S_t=j, S_{t-1}=i|\Psi_{t-1})$, which can be expressed as the product of the probability of being in a certain regime at period $t-1$ and the corresponding transition probability:

$$Pr(S_t=j, S_{t-1}=i|\Psi_{t-1}) = p_{ij}Pr(S_{t-1}=i|\Psi_{t-1}). \tag{A.8}$$

Updating this probability with information up to period t yields the filtered probabilities:

$$\begin{aligned} Pr(S_t=j, S_{t-1}=i|\Psi_t) &= \frac{f(S_t=j, S_{t-1}=i, y_t|\Psi_{t-1})}{f(y_t|I_{t-1})} \\ &= \frac{f(y_t|S_t=j, S_{t-1}=i, \Psi_{t-1})Pr(S_t=j, S_{t-1}=i|\Psi_{t-1})}{\sum_{S_t=1}^{m}\sum_{S_{t-1}=1}^{m} f(y_t|S_t=j, S_{t-1}=i, \Psi_{t-1})Pr(S_t=j, S_{t-1}=i|\Psi_{t-1})}, \end{aligned}$$

and

$$Pr(S_t=j|\Psi_t) = \sum_{i=1}^{m} Pr[S_{t-1}=i, S_t=j|\Psi_t].$$

Based on an initialization—we employ the unconditional probabilities as derived by Hamilton (1989)— the steps can be iterated forward over the sample to obtain the filtered probabilities for each period. Along with the filter recursions, this yields all the information we need to estimate the latent dynamic regime dependent factor as well as the Markov-switching probabilities.

A.3 LARS-EN algorithm

In the following, we explain in more detail how the elastic net works and present results for the full sample. Let us focus on the selection of hard indicators, $y_{it}^{(h)}$, since the selection of the soft indicators works equivalently. The aim is to choose those hard indicators that jointly predict quarterly GDP growth well. We start from the quarterly predictive regression (2.7),

$$\Delta\log(GDP_t) = \sum_{i=1}^{16}\sum_{l=1}^{3} b_{i,l}^{(h)} y_{i,t-l}^{(h)} + u_t^{(h)}, \tag{A.9}$$

where $\Delta\log(GDP_t)$ is centered at zero and all regressors are standardized. Applying OLS would yield, in general, nonzero parameters for all three lags of all 16 indicators. To obtain a sparse solution, i.e., a solution that contains parameter estimates of zero, and thus really selects indicators, we estimate the parameters by means of the elastic net (EN) proposed by Zou and Hastie (2005). To this end, we define the $T \times 1$ vector $\mathsf{y} = (\Delta\log(GDP_1), \ldots, \Delta\log(GDP_T))'$ and the $T \times 48$ matrix X with rows

$$\mathsf{X}_t = (y_{1,t-1}^{(h)}, y_{1,t-2}^{(h)}, y_{1,t-3}^{(h)}, \ldots, y_{16,t-1}^{(h)}, y_{16,t-2}^{(h)}, y_{16,t-3}^{(h)}),$$

and corresponding 48×1 vector of coefficients

$$\mathsf{b} = (b_{1,1}^{(h)}, b_{1,2}^{(h)}, b_{1,3}^{(h)}, \ldots, b_{16,1}^{(h)}, b_{16,2}^{(h)}, b_{16,3}^{(h)})'.$$

Then we solve the elastic net optimization problem

$$\mathsf{L} = (\lambda_1, \lambda_2, \mathsf{b}) = |\mathsf{y} - \mathsf{X}\mathsf{b}|^2 + \lambda_1 |\mathsf{b}|_1 + \lambda_2 |\mathsf{b}|^2, \tag{A.10}$$

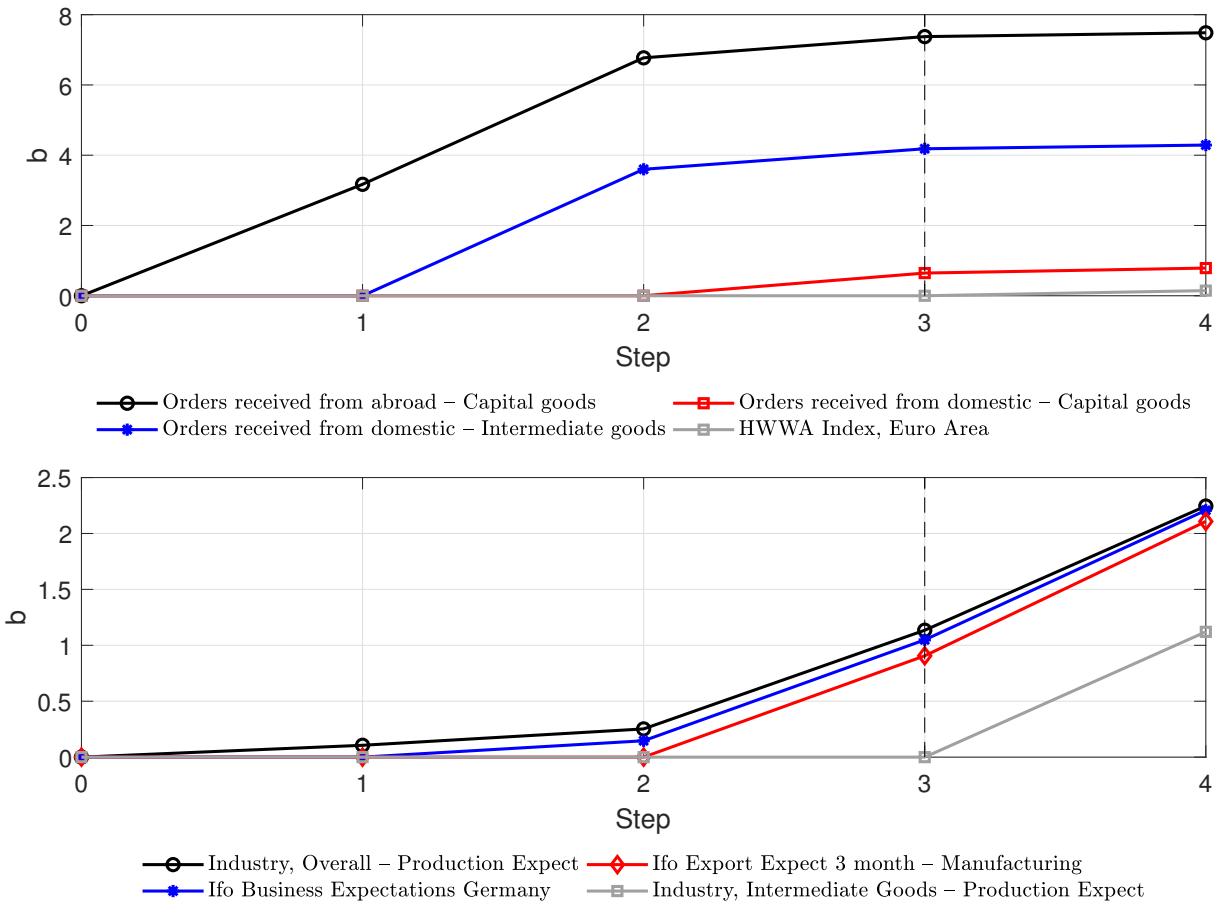
where $|\cdot|_1$ and $|\cdot|^2$ denote the L_1 and L_2 norm, respectively. The specific shape of the L_1 norm induces, for sufficiently large λ_1, a sparse solution that can be interpreted as regressor subset selection, see Hastie, Tibshirani, and Friedman (2017). We follow Zou and Hastie (2005) who show that the elastic net optimization problem can be rewritten as a LASSO optimization problem which can be solved by an adaption of the least angle regression (LARS) originally proposed by Efron et al. (2004). The adaption to the elastic net, called LARS-EN, allows to transform the two tuning parameters λ_1 and λ_2 into the tuple (k, λ_2), where k is the number of regressors to be selected. The intuition behind it is simple: the larger we choose λ_1, the more dominates the L_1 norm which favors a sparse solution. One can think of the LARS-EN algorithm as starting, for fixed λ_2, from a very large value of λ_1 such that b is estimated as a zero vector. By successively lowering λ_1, more and more nonzero parameter estimates show up and thus k increases. Since we intend to select three hard indicators, we set $k = 3$.

We also need to choose a value for the other tuning parameter, λ_2, which determines the weight of the L_2 norm in the optimization problem. To understand how λ_2 affects the estimation problem, note that the elastic net collapses to the LASSO if $\lambda_2 = 0$. The LASSO is known to select almost arbitrarily only one predictor from a subset of highly correlated regressors. This is the so-called grouping effect. By the very nature of our problem — we intend to extract common business cycle information from a set of selected leading indicators — our regressors are potentially strongly correlated and we deliberately want to choose correlated ones for the subsequent factor model to make sense. Further note that the elastic net reduces to the ridge regression if $\lambda_1 = 0$. Ridge regression is able to deal with highly correlated regressors. In fact, it was originally motivated for the extreme case that the cross-product $\mathsf{X}'\mathsf{X}$ is not even invertible (Hoerl and Kennard, 1970). In general, the elastic net is a kind of combination of the LASSO and ridge regression. The larger we choose λ_2, the more dominates the L_2 norm which allows efficient handling of correlated regressors and avoids the grouping effect. We experimented with different choices for λ_2 and found the value of 100 to work well which is in the range of values considered by Zou and Hastie (2005). The results turned out to be robust to choosing higher values but smaller values gave rise to the grouping effect.

To get an idea of how the elastic net works with our data, let us consider the selection of hard indicators in the regression (2.7) for the ex post analysis. We set $\lambda_2 = 100$. The LARS-EN

algorithm starts with a prohibitively large $\lambda_1 = 95.2$ so that all parameters are estimated as zero. Successively lowering λ_1 allows the inclusion of more and more regressors. The upper panel of Figure 2.11 shows how the parameter estimates evolve step by step. In step $k = 1$, λ_1 is lowered to 82.9 which allows to include the first regressor, foreign orders of capital goods (lag 1), with parameter 3.17. In step $k = 2$, λ_1 is lowered to 63.6. Now the first regressor has a larger parameter, 6.77, and a second regressor, domestic orders of intermediate good (lag 1), is added with parameter 3.60. In step $k = 3$, λ_1 is lowered to 59.6 which allows to add domestic orders of capital goods (lag 1) as third regressor. Hence, this choice of λ_1 corresponds to our objective of $k = 3$ and we use the selected three indicators in our Markov-switching models. Of course, it is possible to take more steps and thus add more variables. To illustrate this, step $k = 4$ is also shown.

Figure 2.11: Evolution of the estimated parameters of the LARS-EN



Notes: The colored lines indicate how the parameter estimates for the regressors stated in the legend change step by step. Each step k corresponds to a specific value λ_1 that allows to include another regressor. The dashed line indicates step $k = 3$.

The upper panel of Table 2.7 reports the estimated parameters of the third LARS-EN step applied to the selection regression (2.7). As a comparison we also show the OLS estimates of the same parameters. (Of course, OLS yields nonzero estimates of all parameters but for ease of presentation we leave them out here.) Clearly, the elastic net estimates are absolutely smaller than the unconstrained OLS estimates. The lower panels of Figure 2.11 and Table 2.7 show the results of the analogous selection regression (2.8) for the survey indicators. While the parameter values obviously change, the general procedure remains the same.

Table 2.7: Parameters of the selection regressions estimated by LARS-EN and OLS

Indicator	b	β_{OLS}
Regression of GDP on hard indicators		
Foreign orders of capital goods	7.38	10.92
Domestic orders of intermediate goods	4.18	8.81
Domestic orders of capital goods	0.65	7.59
Regression of GDP on soft indicators		
Overall production expectations	1.13	10.40
Ifo business expectations	1.05	14.42
Ifo export expectations	0.91	16.41

Notes: b in the upper and lower panels denotes the estimated parameters of equations (2.7) and (2.8) that are nonzero based on LARS-EN with $\lambda_2 = 100$ and $k = 3$. β_{OLS} denotes the respective parameter estimates obtained by OLS.

A.4 Detailed estimation results

In this section we report the estimated autoregressive parameters of the idiosyncratic components of both the MS(2)-DFM and the MS(3)-DFM. Recall that the vector of idiosyncratic components, $z_t = [z_{1t}^{(h)}, z_{2t}^{(h)}, z_{3t}^{(h)}, z_{1t}^{(s)}, z_{2t}^{(s)}, z_{3t}^{(s)}]'$ is modeled as a diagonal VAR process of lag order $q = 2$ with diagonal covariance matrix. Hence, each component $i = 1, \ldots, 6$ follows an independent AR(2) process with AR parameters $\psi_{i,1}$ and $\psi_{i,2}$, where $\psi_{i,j}$ is the ith diagonal element of the parameter matrix ψ_j defined in (2.3). Table 2.8 shows these parameters estimated by maximum likelihood. It turns out that, while being stationary by assumption, most idiosyncratic components are fairly persistent.

A.5 Data: indicators, sources, and real-time selection

The majority of the series is downloaded from Thomson Reuters Datastream, while the remaining indicators are directly obtained from the German Bundesbank, the ECB and the OECD. Tables 2.9 and 2.10 list the hard and survey indicators, respectively, together with their sources and the transformations we applied. For the hard indicators we report the sources for both our ex post

Table 2.8: Autoregressive parameters of the idiosyncratic components

Indicator		MS(2)-DFM $\psi_{i,1}$	MS(2)-DFM $\psi_{i,2}$	MS(3)-DFM $\psi_{i,1}$	MS(3)-DFM $\psi_{i,2}$
$z_{1t}^{(h)}$	Foreign orders of capital goods	−0.63 (0.06)	−0.31 (0.06)	−0.66 (0.06)	−0.34 (0.06)
$z_{2t}^{(h)}$	Domestic orders of intermediate goods	−0.32 (0.07)	−0.09 (0.07)	−0.41 (0.07)	−0.16 (0.07)
$z_{3t}^{(h)}$	Domestic orders of capital goods	−0.65 (0.06)	−0.21 (0.06)	−0.68 (0.06)	−0.24 (0.06)
$z_{1t}^{(s)}$	Overall production expectations	0.63 (0.08)	0.14 (0.07)	0.73 (0.07)	0.06 (0.07)
$z_{2t}^{(s)}$	Ifo business expectations	1.03 (0.07)	−0.21 (0.07)	1.09 (0.07)	−0.27 (0.07)
$z_{3t}^{(s)}$	Ifo export expectations	0.83 (0.07)	0.07 (0.07)	0.88 (0.07)	0.01 (0.07)

Notes: $\psi_{i,j}$ denotes the autoregressive parameter of idiosyncratic component i for lag j. In terms of the notation of equation (2.3), it is the ith diagonal element of the (diagonal) parameter matrix ψ_j. Estimated standard errors are reported in parentheses below the estimates.

analysis and our real-time analysis. The survey indicators are stationary by construction and thus left untransformed. They are published without (noticeable) revisions, hence the use of a specific real-time data set is not necessary.

The hard and survey indicators selected by the LARS algorithm in each step of our real-time analysis are reported in Figures 2.12 and 2.13. Note that we exclude from the Figures all variables that are never selected.

Table 2.9: Hard indicators

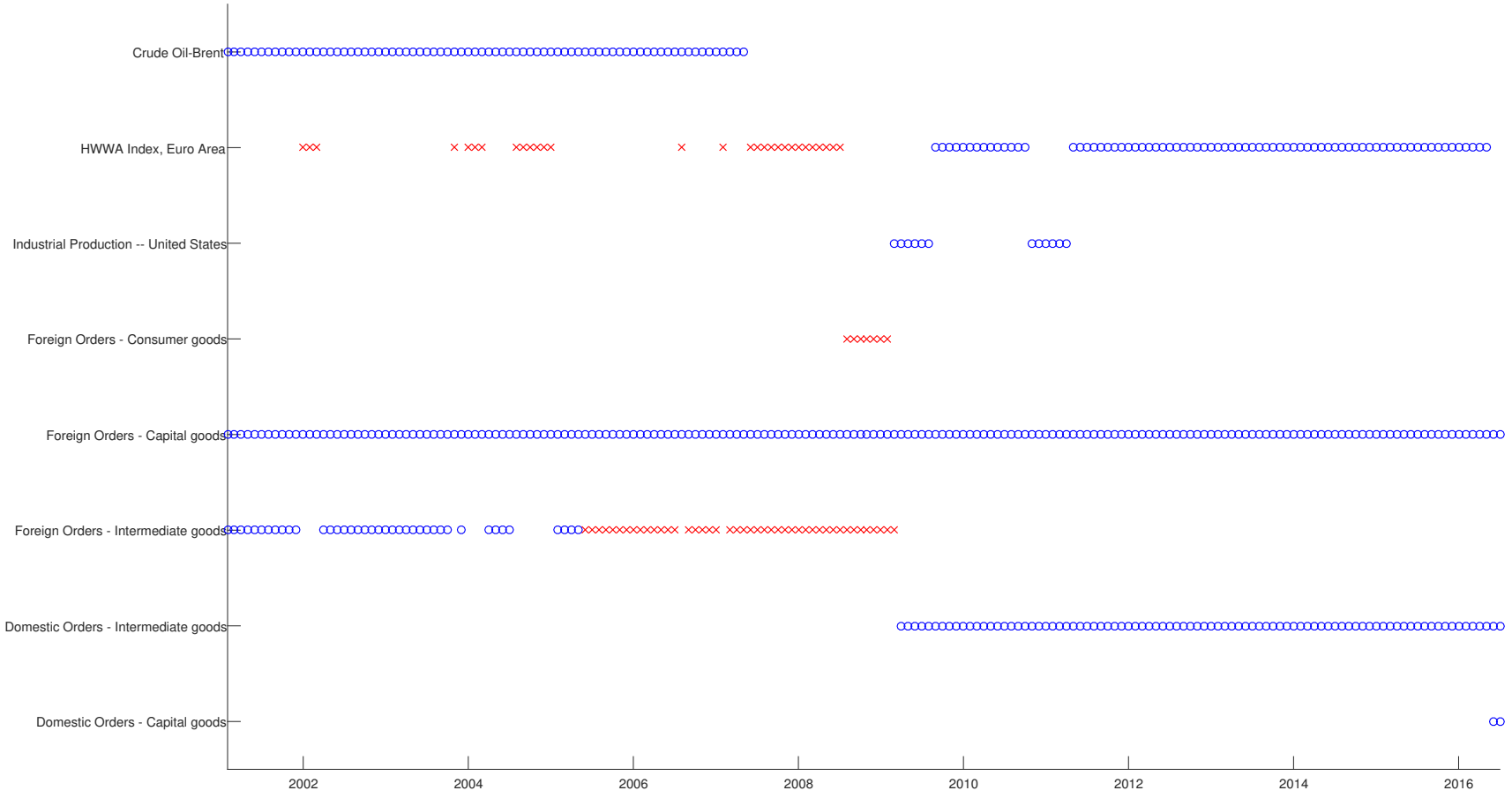
Name of Series	**Datastream code** (ex post data set)	**Real-time code** (real-time data set)	**Transformation** diff	log
New Orders				
Domestic orders of capital goods, Germany	BDDCPORDG	BBKRT.M.DE.Y.I.IO2.ACM03.C.I	yes	yes
Domestic orders of consumer goods, Germany	BDDCNORDG	BBKRT.M.DE.Y.I.IO2.ACM04.C.I	yes	yes
Domestic orders of intermediate goods, Germany	BDDBPORDG	BBKRT.M.DE.Y.I.IO2.ACM02.C.I	yes	yes
Foreign orders of capital goods, Germany	BDOCPORDG	BBKRT.M.DE.Y.I.IO3.ACM03.C.I	yes	yes
Foreign orders of consumer goods, Germany	BDOCNORDG	BBKRT.M.DE.Y.I.IO3.ACM04.C.I	yes	yes
Foreign orders of intermediate goods, Germany	BDOBPORDG	BBKRT.M.DE.Y.I.IO3.ACM02.C.I	yes	yes
Interest rates				
Yield on German federal securities, residual maturity 9 to 10 years	BDT0557	no revisions	yes	no
Fibor – 3 month (monthly average)	BDINTER3	no revisions	yes	no
Term spread on German federal securities - (10y-3m)	BDT0557-BDINTER3	no revisions	yes	no
Commodity prices				
Brent crude oil price, US-Dollar	OILBREN	no revisions	yes	yes
HWWA commodity price index, Euro	BDHWWAINF	no revisions	yes	yes
General economic indicators				
Dax performance index	DAXINDX	no revisions	yes	yes
German contribution to EMU M2	BDTXI302A	no revisions	yes	yes
Employed persons – overall economy, Germany	last vintage	BBKRT.M.DE.S.L.BE1.CA010.P.A	yes	yes
Consumer prices – all categories, Germany	last vintage	BBKRT.M.DE.Y.P.PC1.PC100.R.I	yes	yes
Foreign markets				
US industrial production	last vintage	from OECD.Stat	yes	yes

Notes: "no revisions" indicates that the series is assumed to be published without revisions, "last vintage" indicates that we use the last vintage of the real-time data set as ex post data.

Table 2.10: Survey indicators

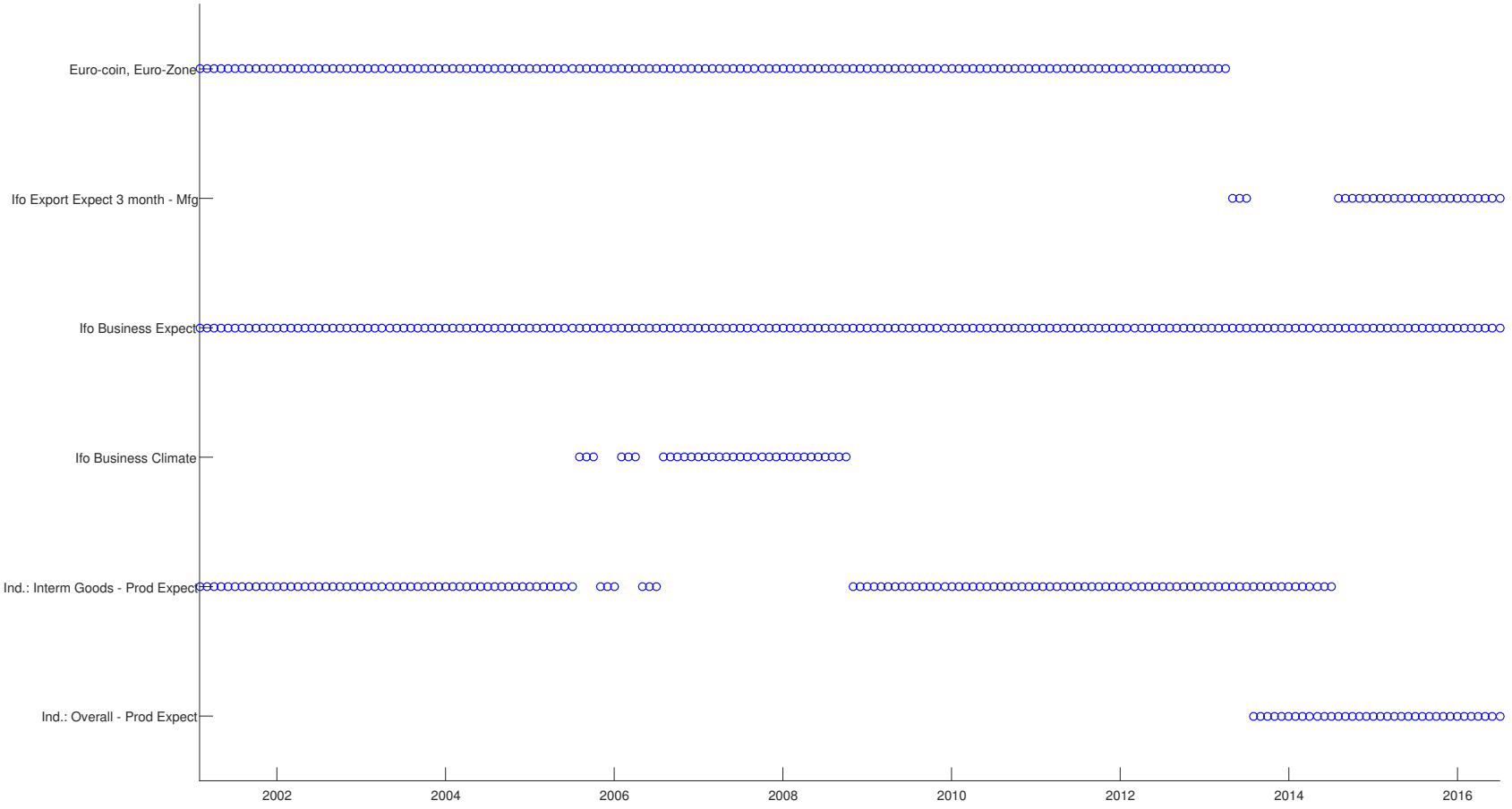
Name of Series	**Datastream-Code**	**Source**
Overall production expectations, Industry, Germany	BDTTA5BSQ	European commission
Intermediate goods production expectations, Industry, Germany	BDITM5.BQ	European commission
Investment goods production expectations, Industry, Germany	BDIVE5.BQ	European commission
Overall employment expectations, Germany	BDTTA7BSQ	European commission
Overall order books, Industry, Germany	BDTTA2BSQ	European commission
Consumer confidence indicator, Germany	BDCNFCONQ	European commission
Consumer survey: economic situation next 12 months, Germany	BDEUSCEYQ	European commission
Economic sentiment indicator, Germany	BDEUSESIG	European commission
Economic sentiment indicator, Euro Zone	EKEUSESIG	European commission
Ifo business climate index, Germany	BDCNFBUSQ	Ifo institute
Ifo business climate manufacturing, Germany	BDIFDMTLQ	Ifo institute
Ifo business climate capital goods, Germany	BDIFDMPLQ	Ifo institute
Ifo business expectations index, Germany	BDCYLEADQ	Ifo institute
Ifo business expectations manufacturing, Germany	BDIFDMTKQ	Ifo institute
Ifo business expectations capital goods, Germany	BDIFDMPKQ	Ifo institute
Ifo export expectations, Germany	BDIFDMTJQ	Ifo institute
ISM purchasing managers index, USA	USCNFBUSQ	ISM institute
Belgium business indicator, Belgium	BGCNFBUSQ	National Bank of Belgium
Euro-coin, Euro-Zone	EMECOIN.Q	CEPR

Figure 2.12: Real-time variable selection—hard indicators



Note: Recursive real time variable selection with LARS-EN. Variables selected at $\circ$ *= lag* 1, $\times$ *= lag* 2, $*$ *= lag* 3. *Variables which do not appear are never selected.*

Figure 2.13: Real-time variable selection—survey indicators



Notes: Recursive real time variable selection with LARS-EN. Variables selected at ∘ = *lag* 1, × = *lag* 2, ∗ = *lag* 3. Variables which do not appear are never selected.

3 | Time-Varying Dynamics of the German Business Cycle

Abstract This chapter investigates whether there have been structural changes in the German business cycle since the 1970s. Using a time-varying parameter VAR with stochastic volatility, I present evidence based on both reduced-form estimates and a structural identification. With regard to the former, I document substantial shifts in the long-run growth rates, shock volatilities, and the persistence of the variables considered. In particular, German GDP growth rates exhibit a strong decrease in volatility and an increase in persistence. Regarding the structural analysis, I use sign restrictions to identify key macroeconomic shocks. My main result is that the impact responses of the variables to these shocks have decreased over time. Finally, to assess the relative importance of these shocks, I conduct a counterfactual analysis and conclude that smaller supply shocks are a major driver of structural changes and output growth stabilization in Germany.

Keywords: Time-varying parameters, Bayesian vector autoregression, counterfactuals, stochastic volatility, Great Moderation
JEL-Codes: E31, E32, E52, E58

3.1 Introduction

During the last five decades, the German economy was subject to enormous structural changes. It consummated its reunification, integrated into the global and in particular European economy, and transferred its monetary authority from the Bundesbank to the European Central Bank. These changes not only came along with a substantial change in the composition of German GDP over time.[1] A number of studies have also documented a decline in German output growth volatility.

However, the timing, the extent, and the sources of the so-called "Great Moderation" in Germany are not beyond dispute. Stock and Watson (2005) document a near monotonic decline of GDP growth volatility since the 1960s, driven by a decrease of the residual variances. Fritsche and Kuzin (2005) confirm this finding, however, attribute it to an increasing persistence of the GDP growth process caused by a change in the conduct of monetary policy. Buch, Doepke, and Pierdzioch (2004) and Aßmann, Hogrefe, and Liesenfeld (2009), by constrast, present evidence in favor of a discrete transition to a lower volatility state happening in the early 1990s. While Buch et al. (2004) also attribute the declining volatility to a change of monetary policy, Aßmann et al. (2009) highlight the importance of shifts in the composition of GDP. Finally, Mills and Wang (2003) and Summers (2005) find a single structural break in the residual variances of the growth process taking place already in the mid 1970s.

The objective of this chapter is twofold. On the one hand, I provide a more comprehensive view on the Great Moderation in Germany by modeling the joint dynamics of four German macroeconomic variables—GDP deflator inflation, GDP growth, a short-term interest rate, and the growth rate of the money stock—using a time-varying parameter VAR with stochastic volatility (TVP-SV-VAR). On the other hand, I employ a structural identification to investigate if the reduction in output growth volatility is rather driven by the reduction of shocks over time (good luck) or by a change in the systematic response of the economy to these shocks (good policy).

Compared to the studies highlighted above, which are either based on linear multivariate models, univariate models with discrete breaks, or univariate models with gradual parameter change, the TVP-SV-VAR has three advantages. First, the researcher can refrain from taking any stance on whether there is abrupt, gradual, or no structural change at all.[2] Second, the TVP-SV-VAR allows for both drifting VAR coefficients and drifting volatilities. Thus, it can capture time-variation in the high- and low-frequency domain of the variables considered. Third, this nonlinear multivariate framework allows to simultaneously identify structural shocks and their evolution over time. The estimation of the model is conducted along the lines of Cogley

[1] On the expenditure side, the share of exports constantly increased from about 20% in the 1970s to more than 40% in 2017. On the production side, the share of the service sector increased from roughly 50% in 1970 to almost 70% in 2017, whereas the share of manufacturing dropped from 37% to 22% in the same period.

[2] For instance, Baumeister and Peersman (2013) or Antolin-Diaz, Drechsel, and Petrella (2017) show that the random walk law of motion, commonly applied in these models, is able to handle each of these situations.

and Sargent (2005) and Primiceri (2005), that is, I use a Gibbs sampler to consecutively draw from the respective conditional posteriors of the coefficients.

The reduced-form analysis investigates how the structural transformations affected the time series properties of the German economy. In particular, I examine whether the variables' trends, volatilities, and persistences are time-varying. I document that each of these statistics is subject to substantial change over time. With regard to the variables' trends, I find that inflation and GDP growth exhibit a steady decline until the mid 2000s, which is followed by an anew rise. In contrast, the trends of the monetary variables constantly decline until the end of the sample, implying that the trend nominal interest rate is close to and the trend real interest is significantly below zero percent. Concerning the variables' volatility, the results suggest that the overall noise hitting the German economy is steadily decreasing over time. I show that this decline can be attributed to a strong decrease of the volatility of (reduced-form) shocks hitting GDP growth and inflation, thus, pointing at good luck as an important driver of output growth stabilization in Germany. However, I also provide evidence in favor of a change in the shock propagation, indicated by a slight increase of persistence of GDP growth over time.

Regarding the structural analysis, I introduce identifying assumptions on the reduced-form innovations. Specifically, I follow Benati (2008) and identify three major aggregate shocks by imposing restrictions on the signs of each shock. Using the procedure proposed by Baumeister and Peersman (2013), which takes into account the nonlinear model structure, I investigate how the propagation of these identified shocks to the economy evolves over time. I show that, although the conduct of monetary policy has substantially changed since the 1970s, its impact on the evolution of inflation and output growth remained fairly stable. In contrast, I find that both variations of the response to and the magnitude of supply shocks account for large parts of the output growth stabilization in Germany.

The remainder of this chapter is as follows. Section 3.2 outlines the model. Section 3.3 provides a brief overview of the dataset. Sections 3.4 and 3.5 present the results from the reduced-form and the structural analysis, respectively. Section 3.6 concludes.

3.2 The model

To investigate whether there are structural changes in the German economy, I resort to a time-varying parameter VAR with stochastic volatility. This model allows for changes in both the shocks' sizes and transmissions. The model reads as follows:

$$y_t = c_t + \sum_{i=1}^{p} B_{i,t} y_{t-i} + u_t \equiv X_t \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \Omega_t), \tag{3.1}$$

where θ_t contains the VAR coefficients stacked in a vector. y_t is a vector of endogenous variables in quaterly frequency, containing observations on a short-term nominal interest rate, GDP deflator inflation, GDP growth, and the growth rate of the money stock. To be comparable with previous studies, I set the lag length to $p = 2$.[3] Following Primiceri (2005), I assume that the time-varying covariance matrix of reduced-form residuals, Ω_t, can be decomposed into a lower-triangular matrix A_t and a diagonal matrix Σ_t according to

$$A_t \Omega_t A_t' = \Sigma_t \Sigma_t', \tag{3.2}$$

where the diagonal elements of Σ_t are the stochastic volatilities and A_t has ones on the main diagonal and nonzero entries for the remaining lower-triangular elements, describing the contemporaneous relationships between the volatilities. Defining σ_t as the vector of the diagonal elements of Σ_t and a_t as the vector of nonzero elements of A_t stacked by rows, allows to formulate the laws of motion for the time-varying parameters as follows:

$$\theta_t = \theta_{t-1} + \nu_t, \qquad \nu_t \sim N(0, Q), \tag{3.3}$$

$$\log \sigma_t = \log \sigma_{t-1} + e_t, \qquad e_t = (e_{1,t}, \ldots, e_{n,t})' \sim N(0, \Psi), \tag{3.4}$$

$$a_t = a_{t-1} + \upsilon_t, \qquad \upsilon_t = (\upsilon_{1,t}', \ldots, \upsilon_{n,t}')' \sim N(0, \Phi). \tag{3.5}$$

To obtain a stable system at each t, I impose a stability constraint on θ (Cogley and Sargent, 2001). Moreover, I postulate that Ψ is diagonal and Φ is block-diagonal where the blocks relate to the equations of the VAR, implying that the contemporaneous relations are correlated within equations, but uncorrelated across equations.[4] The model estimation is conducted along the lines of Primiceri (2005) and Cogley and Sargent (2005). For details regarding the prior distributions, the Markov chain Monte Carlo algorithm, and the convergence of the Markov chains see Appendix B.1.

3.3 Data

The sample contains quarterly observations from 1960:Q2 until 2018:Q2. This facilitates to investigate the effects of the Great Recession and the subsequent turmoil in the European Monetary Union (EMU) on the German economy. Regarding this sample period, two major issues have to be taken into account: First, the German reunification, and second, the construction of the EMU. The first issue mainly affects real GDP and the GDP deflator. I employ the seasonally adjusted series provided by the OECD quarterly national accounts, which refers to West Germany

[3] See, for instance, Cogley and Sargent (2005); Benati (2008); Gambetti, Pappa, and Canova (2008).

[4] This structure increases computational efficiency and simplifies inference by enabling to estimate the covariances equation by equation (Primiceri, 2005).

until 1991 and afterwards to reunified Germany. To address the second issue, I use seasonally adjusted data for M2 provided by the Deutsche Bundesbank, refering to German M2 until 1998 and afterwards to the German contribution to euro area M2. Finally, I use quarterly averages of FIONIA until the end of 1999 and afterwards I switch to EONIA. GDP, GDP deflator, and M2 enter the model in percentage quarter-on-quarter growth rates. In the following, I refer to this specification as model A. Since EONIA approaches the effective lower bound (ELB) in the euro area from 2009 onward, which obviously reduces the volatility of the series and the shock sizes, I also estimate a model including FIONIA until the end of 1999, EONIA until 2004, and the shadow rate for the euro area provided by Wu and Xia (2017) afterwards. The shadow rate—introduced by Black (1995)—is a hypothetical interest rate, which would arise in the absence of a lower bound on interest rates and can capture additional features of monetary policy that do not directly affect the actual short-term interest rate.[5] I label this specification model B. The latter provides a rough gauge of the impact of unconventional monetary policy on the time series properties of the German economy. To make the figures for the interest rates commensurable with the remaining series, I compute the quarterly effective interest rate as $r_t = ((1+r_t^A)^{0.25}-1)\cdot 100$, where r_t^A denotes the annualized quarter-on-quarter interest rate.

3.4 Reduced-form analysis

This section provides reduced-form evidence for changes of the time series properties of the German economy. To this end, I rewrite the VAR in (3.1) in companion form:

$$Y_t = \mu_t + F_t Y_{t-1} + V_t, \qquad V \sim N(0, \Omega^*), \tag{3.6}$$

where F_t is the VAR's companion matrix, containing the AR-coefficients, μ_t contains the VAR intercepts, and the first $n \times n$ elements of Ω^* correspond to Ω. In the following, I examine changes both in the low- and high frequency domain of the variables considered. Regarding the low-frequency domain, the analysis focuses on the variables' long-run trends and persistence, while for the high-frequency domain the variables' volatility is examined.

3.4.1 Long-run means

Following Cogley, Primiceri, and Sargent (2010), I approximate the long-run trends by:

$$z_t \approx (I - F_t)^{-1}\mu_t, \tag{3.7}$$

[5]The series can be downloaded from the website of Jing Cynthia Wu (`https://sites.google.com/view/jingcynthiawu/shadow-rates`).

where I is an identity matrix of conformable size. This approximation is based on Beveridge and Nelson (1981), defining the stochastic trend of a series as the value the series is expected to converge to in the absence of shocks, that is, $z_t = \lim_{h\to\infty} E_t y_{t+h}$. Figure 3.1 graphs the evolution of z_t for model A (solid line) and model B (dashed line) along with 68% probability bands. To ease comparison, the trends are expressed in terms of annualized rates. Overall, the results for both models are similar—each trend features a decline over time. Apparently, the estimates of model B are smoother, especially after the Great Recession.

The long-run trend of the (nominal) short-term interest rate exhibits the lowest amount of time-variation, but shows the well-known decline over time (see, for instance, Summers, 2014). From the 1970s until the mid 2000s, it decreases by roughly one percentage point. However, it drops sharply to zero percent (model A) and even below (model B) afterwards. Moreover, the distribution of the estimated long-run trend widens considerably after the mid 2000s. While the upper bound, depicted by the 84th percentile, do not change much in this period, I find a strong drop in the lower bound (16th percentile). This indicates that both the estimation uncertainty and the posterior probability for negative trend interest rates has strongly increased in the last 10 years of the sample.

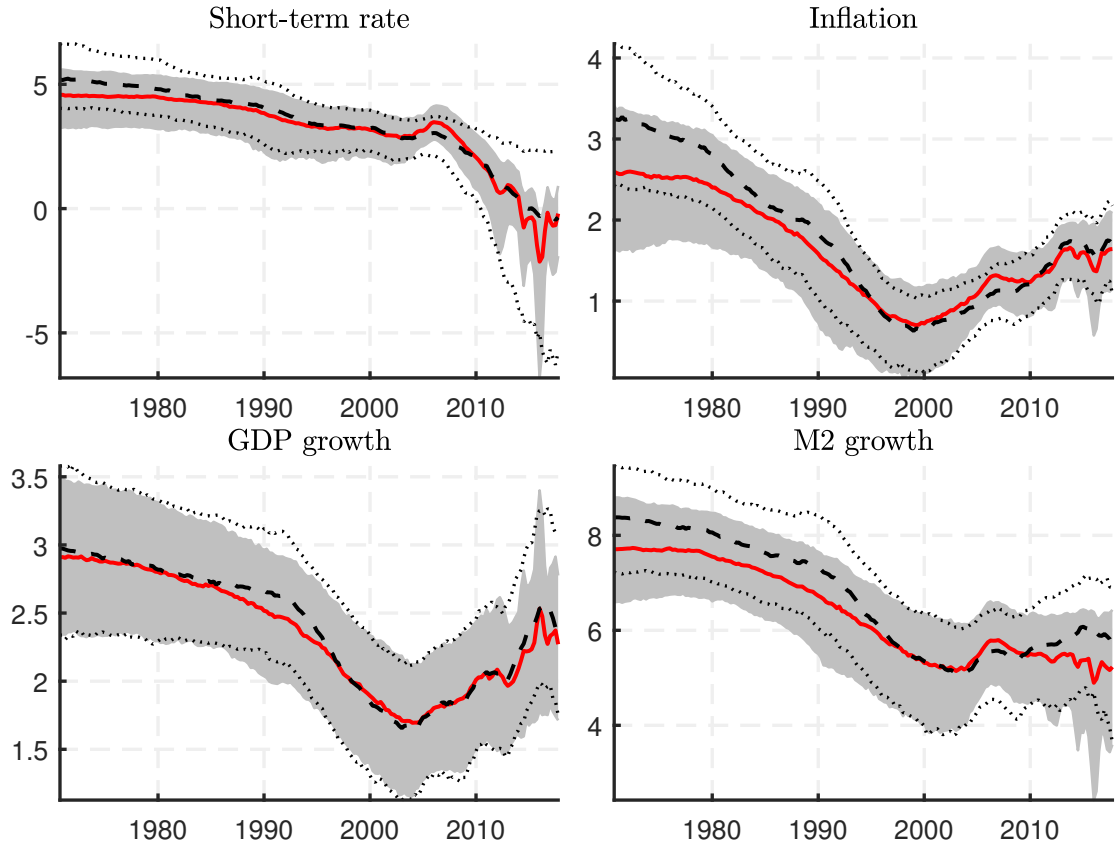
Trend inflation is at roughly three annualized percent in the seventies, constantly decreases to about 0.7 percent in 2000, and afterwards converges back to the ECB's inflation target of close below two percent. This implies that the median estimate for the trend real interest rate, which can be related to a measure of the natural interest rate, is below zero percent form 2014 onward—from 2016 onward even the 84th percentile is negative. The latter is in line with the findings from Brand, Bielecki, and Penalver (2018), using the Laubach and Williams (2003) approach, and Fries et al. (2018).[6]

Moreover, Figure 3.1 suggests that the decrease in trend inflation goes along with a decline in the trend growth rate of the money stock, thus, providing evidence in favor of the quantity theory (see Friedman, 1987). Around 2010, though, the link between inflation and money growth seems to weaken; while trend inflation continues to rise, trend money supply remains roughly on the same level until the sample end.

GDP trend growth falls from three annualized percent in the seventies to about 1.75 percent in 2005. Afterwards, it picks up again and approaches 2.5 percent at the end of the sample period. Hence, the anew rise of long-run output growth coincides with the implementation of the Hartz labor market reform in Germany. While the macroeconomic effects of these reforms are still controversial, several studies show that they indeed caused a more flexible labor market, and thus, an increase in employment, which leads to an increase in output (see, among others, Krause and Uhlig, 2012; Krebs and Scheffel, 2013; Hartung, Jung, and Kuhn, 2018).

[6]Fluctuations in the trend (real) interest rate can be due to shifts in the natural rate of interest or shifts in the inflation target. Since the TVP-SV-VAR cannot differentiate between both sources of variation, the results presented here should be taken with some caution.

Figure 3.1: Evolution of the time-varying trends



Notes: Posterior median of time-varying trends according to model A (solid line) and model B (dashed line) using the approximation (3.7). All figures are expressed in terms of annualized rates.

In addition, the Great Recession has only a minor impact on the trend estimates, suggesting that the models interpret it only as a temporary phenomenon that mainly affects the residuals' volatility. This result is consistent with the findings from Ball (2014), showing only a little impact of the Great Recession on German potential output estimates. One explanation for this result might be the so-called German labor market miracle (Burda and Hunt, 2011). The latter refers to the fact that while the drop of GDP in 2008/2009 was larger in Germany than, for instance, the United States, France, or the United Kingdom, unemployment increased by a lesser extend in Germany.[7] A likely rationale for these differences is the German short-time working scheme, which was gradually made more attractive for firms in the course of the Great Recession (see Brenke, Rinne, and Zimmermann, 2013, for a summary) and allowed firms to maintain their level

[7]Annual GDP growth in Germany was -5.6% in 2009. The figures for the US, France, and the UK are: -2.5%, -2.9%, and -4.2%. The unemployment rate increased in the same period by 0.3pp. in Germany, while it increased in the US, France, and the UK by 3.5pp., 1.6pp., and 1.9pp., respectively.

of employment during the crisis by reducing the hours worked per employee.[8] As pointed out by Rinne and Zimmermann (2012), particularly export-orientated firms from the manufacturing sector—suffering the most from the crisis—strongly benefited from this possibility. When global demand was recovering, these firms could quickly adapt and increase production. Hence, a hysteresis effect with regard to the unemployment rate could not build up and GDP trend growth remained largely unaffected.

3.4.2 Persistence

Subsequently, I analyze how the persistence of the series under investigation has changed over time. I follow Cogley (2005) and Cogley et al. (2010) by measuring persistence in terms of the predictability of the series.[9] Specifically, I approximate the time-varying multivariate R^2 statistics as the ratio between the series' conditional and unconditional variance:

$$R^2_{y_i,t,j} = 1 - \frac{s_{y_i}(\sum_{h=0}^{j-1} F_t^h V_{t+1} F'^h_t) s'_{y_i}}{s_{y_i}(\sum_{h=0}^{\infty} F_t^h V_{t+1} F'^h_t) s'_{y_i}}, \tag{3.8}$$

where s_{y_i} is a selection vector, picking the variable of interest. This measure is bounded between zero and one. Values close to zero imply that past shocks decay quickly, which makes the series less persistent and hence less predictable. Figure 3.2 plots $R^2_{y_i,t,j}$ for $j = 1$ and 4 quarters along with 68% probability bands. Obviously, there is considerable variation in the time-varying predictability of the series. The most persistent series is the short-term interest rate, plotted in the top panel. The R^2 is around 0.92 in the early 1970s, which implies that VAR pseudo-forecasts account for roughly 92 percent of the variation in the interest rate.[10] This figure steadily increases until the end of the sample with the R^2 statistics almost reaching one. However, the latter is obviously driven by the ELB, which prevents the interest rate from going further below zero. In contrast, the persistence of the shadow rate, estimated according to model B, also increases over time, but is far lower in the post-Great Recession period compared to EONIA. At the four-quarter ahead horizon (right column of Figure 3.2), both series are almost identically persistent.

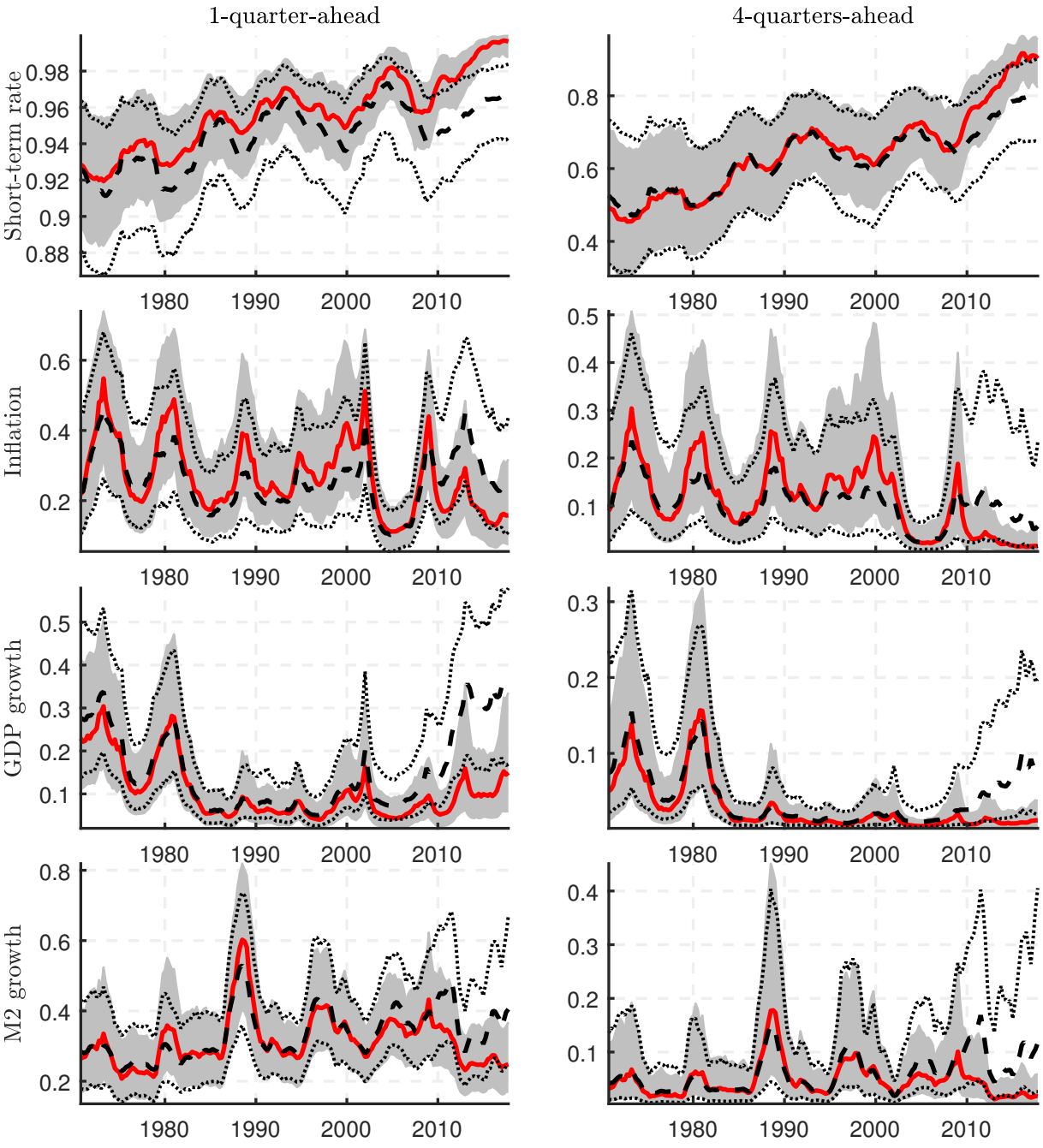
The persistence of inflation fluctuates around roughly 0.3 over time according to both models. When the ELB becomes binding, persistence according to model B is slightly higher than according to model A.

[8] Using structural VAR analysis, Balleer et al. (2016) find that the increase in the unemployment rate during the Great Recession was dampened by 1.29 percentage points due to short-time working, which amounts to roughly 466000 saved jobs.

[9] Alternative measures for persistence are the normalized spectrum of a variable at spectrum zero (Cogley and Sargent, 2005; Gambetti et al., 2008), or (in univariate models) the sum of the AR-coefficients in rolling regressions (Stock and Watson, 2005).

[10] It has to be noted that the forecasts used in this section are based on the full sample estimates, which is why I refer to them as pseudo-forecasts.

Figure 3.2: Evolving predictability



Notes: Posterior median of the estimates $R^2_{y_i,t,j}$ for model A and B; $j = 1$ (left column) and $j = 4$ (right column). Shaded areas and dotted lines refer to 68% equal-tailed point-wise posterior probability bands.

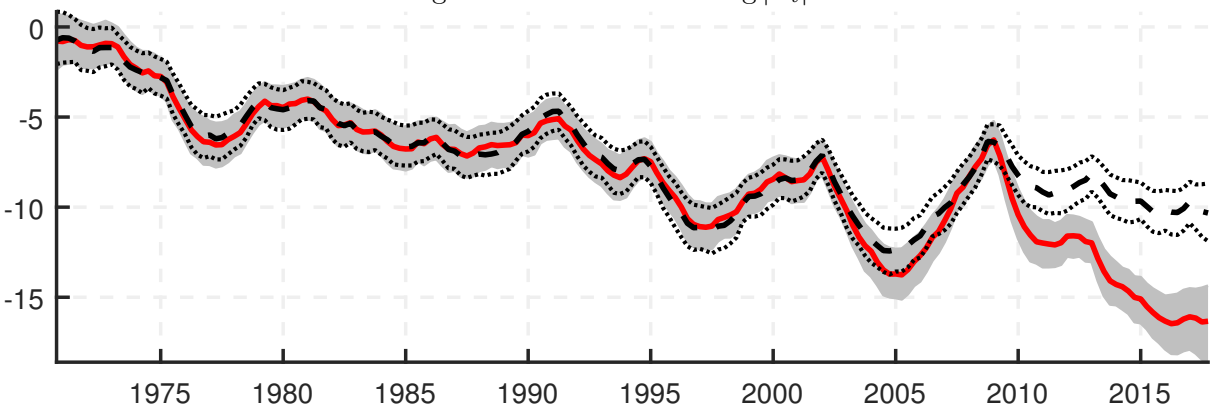
The graphs for GDP growth support the findings from Stock and Watson (2005), who provide evidence in favor of a declining persistence of German GDP growth until the early 1990s and a subsequent increase. Moreover, and in line with Pirschel and Wolters (2018), the results indicate

that German GDP growth is rather non-persistent. This is also reflected by the remarkable lower R^2 for GDP growth at 4-quarters-ahead. At this horizons, VAR pseudo-forecasts account for less than one percent of the variation. It is also much less persistent compared to inflation. At the beginning of the sample, VAR pseudo-forecasts account for roughly 25 percent of the variation of GDP growth rates. Until the early 2000's, this value increases to 30 percent, but drops to about 10 percent in 2008/2009. Following the Great Recession, persistence of GDP growth continues its upward trend.[11] As for inflation, model B yields a remarkably better predictability of GDP growth. Comparing again models A and B shows that output growth predictability is much higher when considering the shadow-rate, which accounts for unconventional monetary policy; at the end of the sample, persistence according to model B is almost twice as high compared to model A, indicating that the shadow rate contains useful information for the evolution of GDP growth. Finally, M2 growth predictability stays rather constant over time.

3.4.3 Volatility

The previous section has demonstrated considerable time-variation in the low-frequency properties of the series under investigation. In the following, I examine whether these changes are accompanied by fluctuations in the series' high-frequency properties. To this end, I investigate how business cycle volatility has evolved over time. Figure 3.3 plots the evolution of the log determinant of the VAR's residual coveriance matrix ($\log|\Omega_t|$) for model A (solid line) and model B (dashed line). Following Cogley and Sargent (2005), this measure is interpreted as the total size of shocks hitting the economy at each point in time.

Figure 3.3: Evolution of $\log|\Omega_t|$



Notes: Figure depicts $\log|\Omega|$ along with 68% equal-tailed point-wise posterior probability bands for model A (solid line) and model B (dashed line).

[11] A similar result is obtained by Benati (2008) for UK GDP growth, which is, however, roughly twice as persistent as German GDP growth.

Figure 3.3 comprises two implications. First, $\log|\Omega_t|$ steadily decreases over time, indicating a substantial decrease in short-run uncertainty of the system. Second, this decrease is far from monotonic. For instance, during the eighties $\log|\Omega_t|$ is almost constant, while the sharp drop during the first half of the nineties is almost totally compensated for by the increase in the second half of the nineties. In total, the figures for Germany until the early 2000s resemble the results from Benati (2008) for the UK. In contrast, Cogley and Sargent (2005) document for the US an increase in $\log|\Omega_t|$ until the early 1980s followed by a sharp decrease.

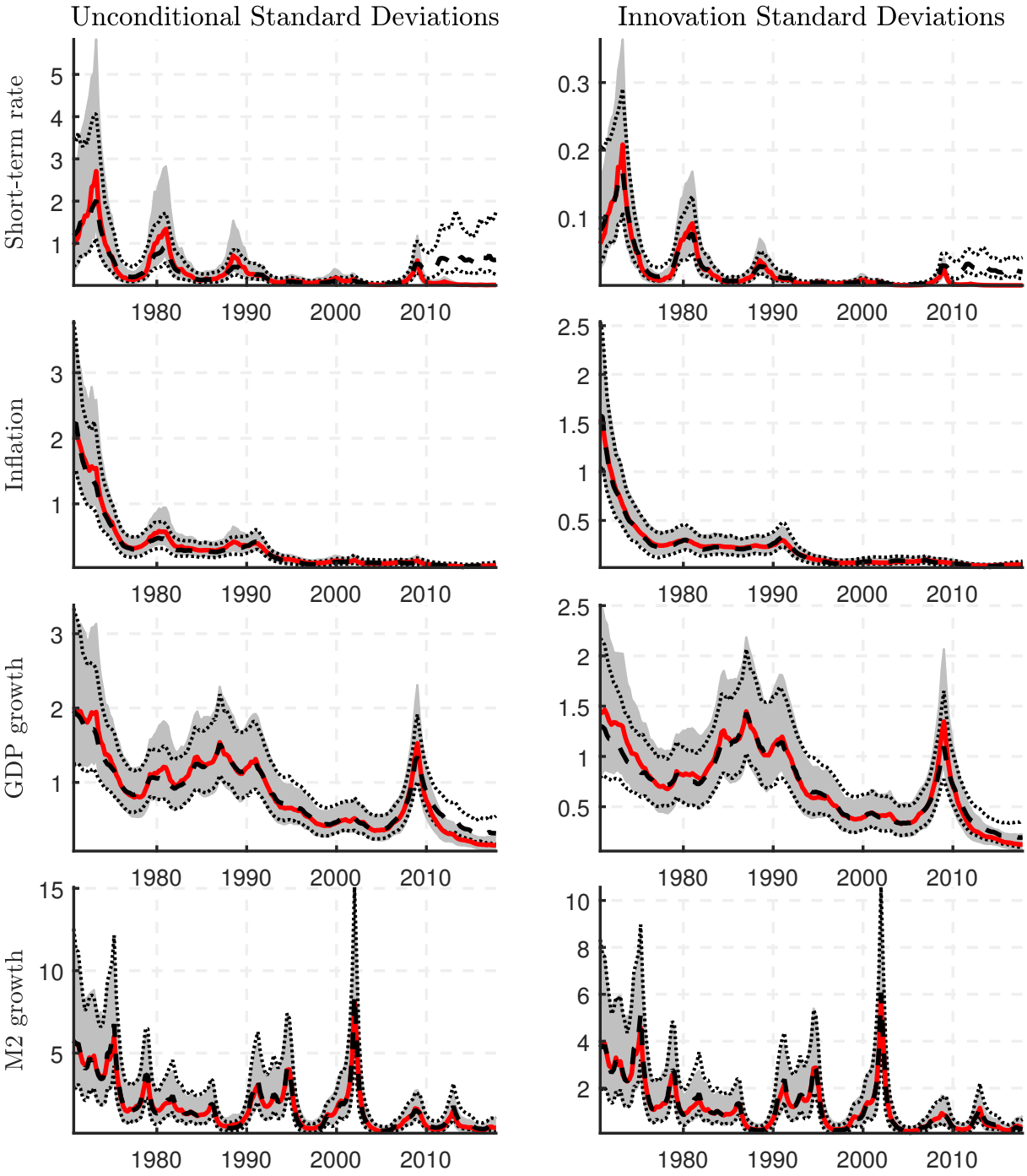
Since the last peak, which can be attributed to the Great Recession, $\log|\Omega_t|$ constantly declines, with the latest estimate of model A the lowest for the entire sample. The latter indicates that currently the German economy is remarkably less exposed to shocks. However, comparing the results from model A and B reveals that both estimates are virtually identical only until the Great Recession. Afterwards, according to model B, short-run uncertainty is higher. Although $\log|\Omega_t|$ from model B also declines since the peak during the Great Recession, it is on a higher level compared to the period from 1995 until 2008.

To gauge the reasons behind this evolution and to assess why the German economy behaves differently compared to the US, Figure 3.4 provides a closer look at both the unconditional standard deviation of each variable and the standard deviation of the reduced-form residuals, that is, the remaining elements of Ω_t. I approximate the unconditional standard deviations of the series by taking the limit of the conditional variance (the root of the denominator in (3.8)).

First, I consider the evolution of the unconditional standard deviations of the series (left column of Figure 3.4). Overall, each variable displays a strong decline in variability until the Great Recession, confirming the Great Moderation also in Germany. Moreover, the results resemble the ones of Stock and Watson (2005), showing a strong drop in volatility during the early 1970s and 1990s. GDP growth volatility decreases from close to two percentage points in the early 1970s to around one percentage point in the eighties. After a strong hike during the Great Recession, output growth volatility falls below pre-crisis levels. At the end of the sample period, it is on a historically low level of roughly 0.15pp. Hence, as pointed out for the US by, for example, Gadea Rivas, Gómez-Loscos, and Pérez-Quirós (2014), also in Germany the Great Recession seems to have only disrupted, but not ceased, the Great Moderation. GDP growth volatility was more than ten times higher in the early 1970s than today. Inflation exhibits a strong drop during the first half of the 1970s, from two percentage points to roughly 0.4pp. Afterwards it remains fairly stable until another drop in 1992.

After the Great Recession, the unconditional standard deviations from model A and B remain almost identical for inflation and M2 growth, implying that they are independent of the interest rate used, while they show differences for the short-term rate and output growth. M2 growth also stabilizes at a low level after the Great Recession; at the end of the sample, it is on pre-reunification levels.

Figure 3.4: Evolution of the covariance matrix



Notes: Posterior means of the unconditional standard deviations (left column) and the innovation standard deviations (right column) along with 68% equal-tailed point-wise posterior probability bands. All figures are reported in percentage points. The solid lines refer to model A; the dashed lines refer to model B.

Second, the right column of Figure 3.4 indicates that the decrease in the volatility of the series over time is caused by a strong reduction of the volatility of the reduced-form shock. However, it has to be noted that, in case of the interest rate, the innovation standard deviations are much smaller than the unconditional standard deviations, implying that the volatility of (reduced-form) shocks accounts only for a small fraction of fluctuations in the unconditional standard deviations. A likely explanation for this result is that the short-term interest rate—as the policy instrument—reacts to changes in output growth and inflation (according to a Taylor rule), while exogenous fluctuations in the instrument itself are avoided. With regard to the remaining variables, though, the reduction in the innovation standard deviations is of similar magnitude compared to the overall reduction of variability of the series (irrespective of the model used). For instance, according to both models, the variance of the shocks hitting GDP growth is today roughly seven times smaller than during the 1970s. Moreover, as already suggested by the trend estimates, the Great Recession mainly materializes as a strong increase in the volatility of the reduced-form residuals. Overall, these results indicate that much of the output growth stabilization is due to a reduction in the magnitude of the shocks.

Finally, according to model B, the innovation standard deviations of the interest rate are significantly higher compared to those of model A. The latter provides an explanation for the higher $\log|\Omega_t|$ of model B following the Great Recession. Thus, the results imply that focusing solely on actual interest rates underestimates the actual uncertainty of the system, since it ignores the impact of unconventional monetary policy, which is a consequence of the economic developments in the euro area following the Great Recession.

In total, the reduced-form analysis points at important changes in the German economy and a stabilization of the business cycle. Inflation dynamics are—expect for a decrease in trend inflation—rather unchanged since the early 1980s. In comparison, GDP growth exhibits a strong drop in both unconditional variability and the size of reduced-form shocks. This volatility reduction goes along with marked time-variation in trend output growth. At the end of sample, GDP trend growth is roughly on the same level as in 1990, while its variability is almost 90 percent lower.

3.5 Structural analysis

To get a deeper understanding of the drivers of the results presented in the previous section, this section provides a structural analysis based on impulse responses. Since impulse responses are only informative with regard to a one-time shock on the variables, but do not contain information on how important this shock has been on average or on how much of the historical variation in the variables can be explained by this shock, I also examine the forecast error variance decomposition (FEVD) and the historical decomposition of the identified shocks based on the TVP-SV-VAR.

3.5.1 Impulse response analysis

I aim at identifying three major macroeconomic shocks, namely a monetary policy shock, an aggregate demand shock, and an aggregate supply shock. To uniquely identify these shocks, I follow previous research (see, for instance, Gambetti et al., 2008; Benati, 2011; Belongia and Ireland, 2016) and postulate sign restrictions on the shocks' contemporaneous effects (see Table 3.1).

Table 3.1: Identification restrictions

Shocks/variables	Interest rate	Inflation rate	Output growth	M2 growth
Monetary policy	>0	<0	<0	<0
Aggregate demand	>0	>0	>0	>0
Aggregate supply		<0	>0	

Notes: Restrictions are imposed on impact. Blank entries remain unconstrained.

While the identifying assumptions summarized in Table 3.1 are commonly used in the literature, identifying monetary policy shocks during and after the Great Recession requires addressing two issues. First, since the short-term interest rate reached the ELB in the aftermath of the Great Recession, monetary policy decisions are probably better reflected in the central bank's assets (Gambacorta, Hofmann, and Peersman, 2014) or (indirectly) in the bond yield spread (Baumeister and Benati, 2013). Second, the transmission of monetary policy might has changed. Jannsen, Potjagailo, and Wolters (2018), for example, show that output and inflation are non-responsive to unexpected interventions of the monetary authority during the recovery phase of a financial crisis. Thus, to address these issues, I compute impulse responses for both model A and B. The latter uses the shadow rate as policy instrument and therefore should be more appropriate for the identification of monetary policy shocks after 2008/2009.

Implementation of the sign restrictions in the nonlinear model follows Baumeister and Peersman (2013). Specifically, as suggested by Koop, Pesaran, and Potter (1996), I compute generalized impulse responses (GIRFs) as the difference between the conditional expectation with and without a shock. To compute these conditional expectations, at each point in time, I use the laws of motion of the time-varying coefficients conditional on a randomly selected draw of the Gibbs sampler to project the model for 20 quarters into the future. The latter enables me to account for uncertainty stemming from variation of the time-varying coefficients. The time-dependent structural impact matrix is calculated using the efficient algorithm of Rubio-Ramirez, Waggoner, and Zha (2010). Further details are provided in the Appendix B.2.

Figure 3.5 provides a first impression of the time-varying structural dynamics by plotting the median responses of the four variables—according to model A—to the three shocks over all

periods (solid line), along with 68% posterior probability bands (dotted lines).[12] With regard to the interest rate, the figure shows that the shocks' transmission mechanism features noticeable differences. For instance, the response of the interest rate to a demand shock ranges between 0.02pp. and 0.15pp. five quarters after the shock occured. Concerning inflation and GDP growth, supply shocks exhibit time variation on impact, while the shock propagation is rather constant. Conversely, monetary policy- and demand shocks show less time variation on impact, while the shock propagation is more heterogeneous across periods. For example, the response of inflation to a demand shock varies between 0.01pp. and 0.09pp. five quarters after the shock occurred.[13] Regarding M2 growth, the responses show both time variation in the impact responses and the shock propagation.

Another way to look at the shock propagation is provided by Figure 3.6, which plots the median responses on impact (solid line) and one year after the shock has hit the economy (dashed line) for each point in time.[14] This facilitates detecting changes in both the shocks' magnitude and their persistence. Evidently, the impact responses show substantial time-variation. Most striking, the impact response of each variable to the shocks is decreasing over time. For instance, following an unexpected monetary policy tightening, inflation drops by about 0.4pp. in the early 1970s. In the 1980s, the impact response is merely 0.2pp. while it is below 0.1pp. after the Great Recession. An even stronger reduction of the impact response of inflation is obtained for supply shocks (-0.8pp. in 1970 vs. -0.1pp. in 2018). For output growth, a similar picture emerges, even though the decrease of the impact response is not as monotonic as for inflation. In fact, during the eighties and the Great Recession, the impact responses strongly increase in magnitude. However, the overall trend is unbroken.

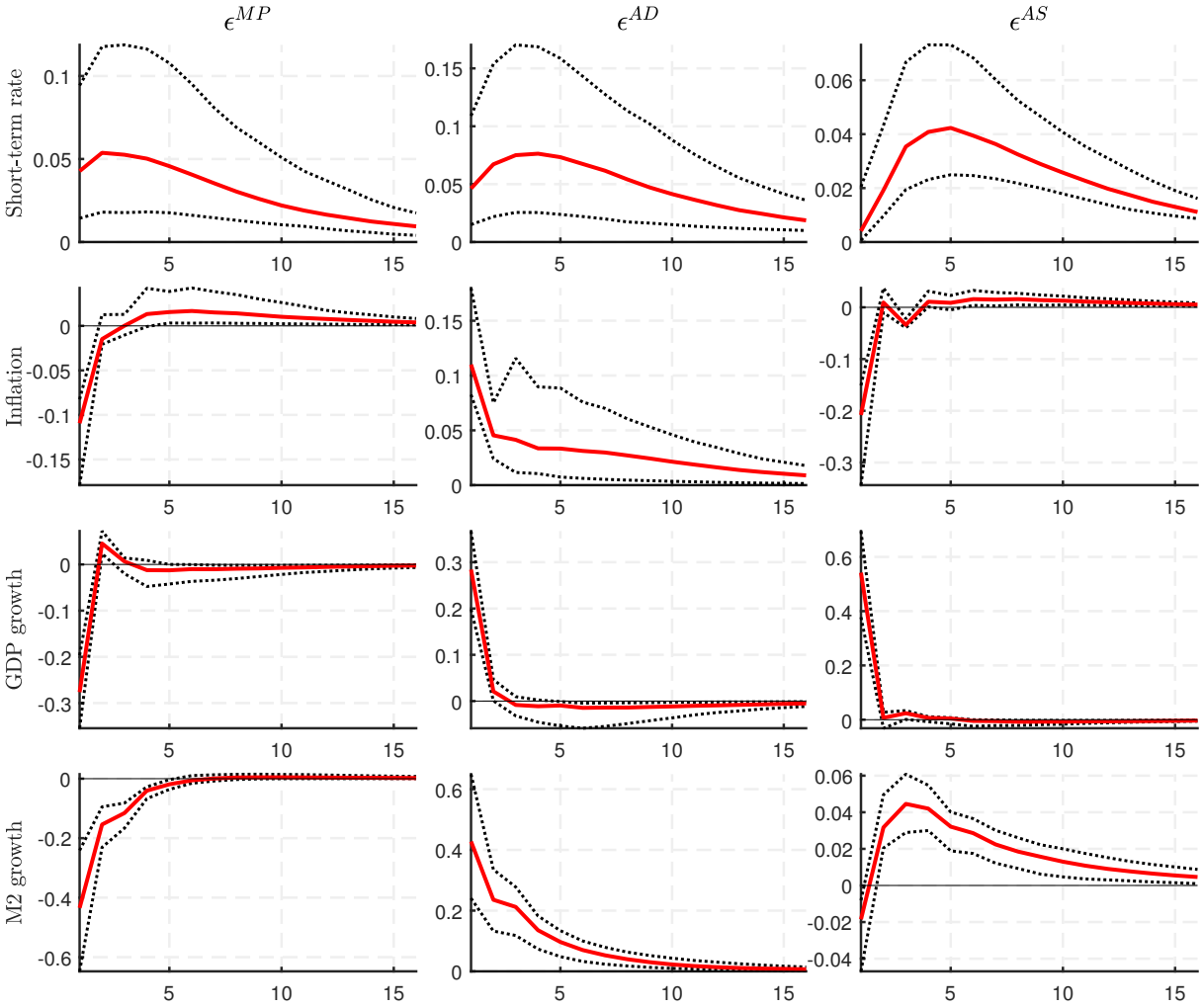
As a result of the smaller impact reactions, the responses after one year are also decreasing over time. Two features are worth discussing, though. First, there is evidence for a price puzzle in Germany until the end of the 1980s. Second, while the impact responses show a steady decrease over the entire sample, the responses after one year, exhibit—in most cases—only noticeable time-variation until the mid 1980s and remain almost constant (and close to zero for some variables) afterwards. Thus, the already low persistence of the shocks has further decreased. However, the major reduction has already taken place in the 1980s. Hence, the results suggest that the propagation of shocks has changed very little during the past 30 years, providing support for the good luck hypothesis.

3.5.2 Forecast error variance decomposition

Figure 3.7 presents the evolution of the posterior medians of the FEVD after 20 quarters for the four variables and the three shocks along with 68% posterior probability bands. The rows refer

[12] The median responses of model B are virtually identical, thus I do not report them.
[13] A similar result is found by Gambetti et al. (2008) for the US.
[14] The complete distributions of the GIRFs over time is provided in Appendix B.3.

Figure 3.5: Generalized impulse responses – median response over time

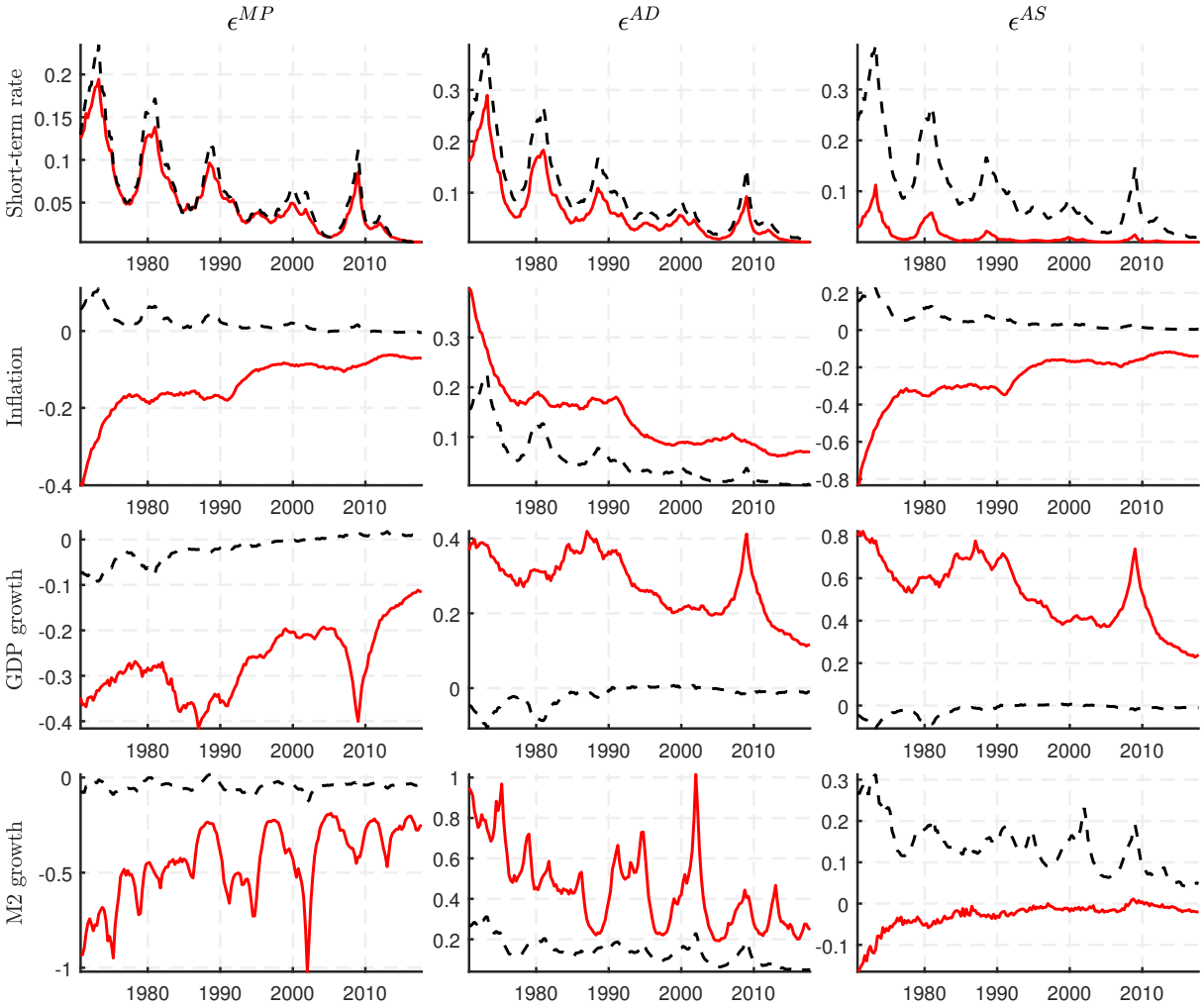


Notes: Median response of variables to identified shocks along with 68% equal-tailed point-wise posterior probability bands. ϵ^{MP}, ϵ^{AD}, and ϵ^{AS} refer to monetary policy, aggregate demand, and aggregate supply shocks, respectively.

to the variables, the columns to the shocks. All figures are expressed in terms of percentage contributions to the forecast error variance of the respective variable. While for output and inflation, the identified shocks constantly explain about 80 percent of the variation, for the interest rate, the contribution varies more strongly. Regarding the latter, the shocks identify up to 85 percent of the variation until 2005. Afterwards, the explanatory power of the shocks decreases, approaching about 65 percent in 2018. Regarding M2 growth, the shocks identify between 55 and 80 percent of the variation over time.

In the case of the short-term interest rate, monetary policy shocks account for roughly 20 percent of the variance throughout the entire sample. However, while the contribution is stable

Figure 3.6: Generalized impulse responses – responses over time

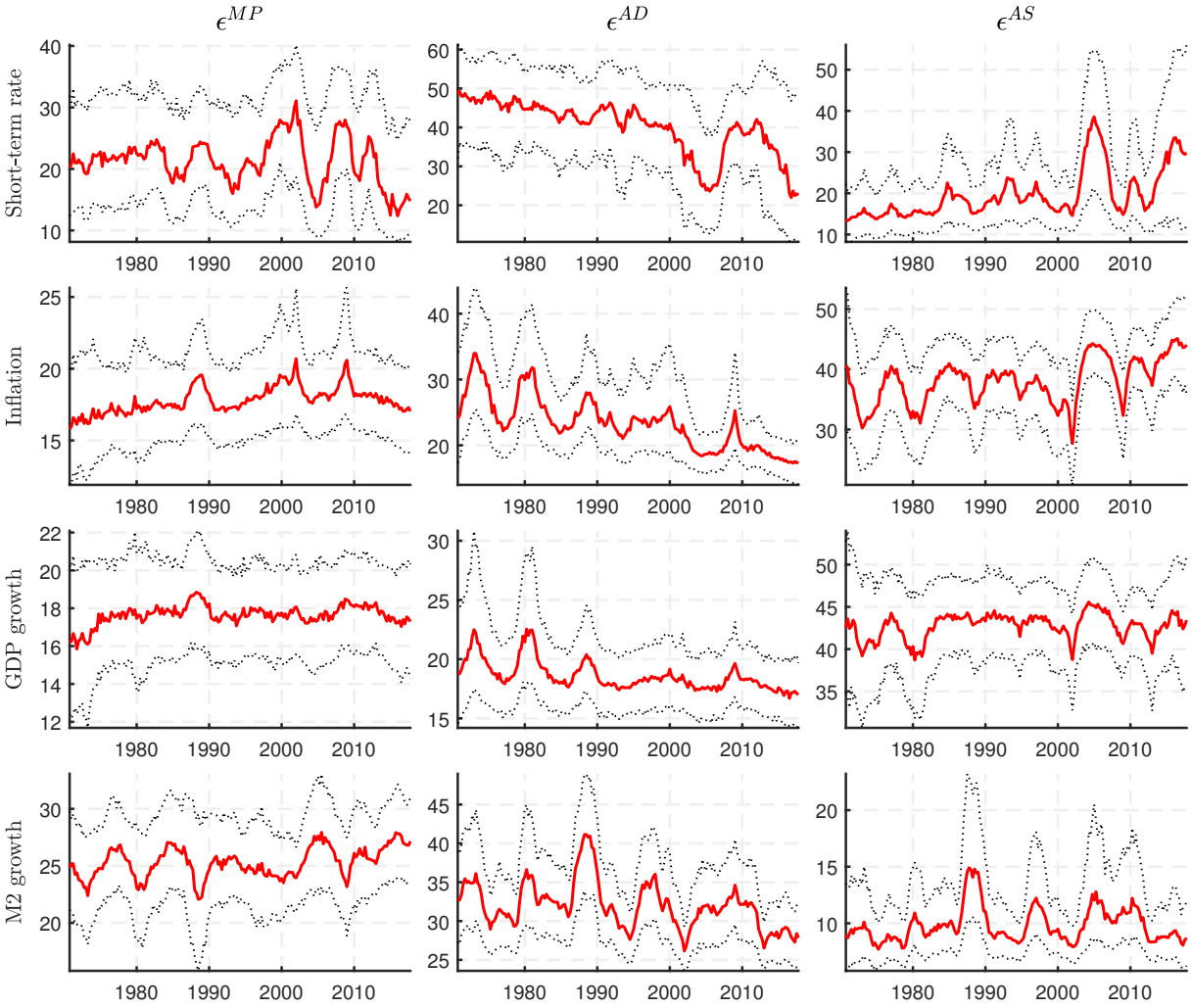


Notes: Responses to identified shocks on impact (solid line) and one year after the shock occurred (dashed line) over time. ϵ^{MP}, ϵ^{AD}, and ϵ^{AS} refer to monetary policy, aggregate demand, and aggregate supply shocks, respectively.

until the end of the nineties, it becomes volatile afterwards. The contribution of demand shocks exhibits a strong decrease over time; starting with a value of around 50 percent, the contribution falls to about 20 percent in 2018. In contrast, supply shocks show an upward trend and account for the largest part of the variance of the short-term rate at the end of the sample.

Regarding inflation, Figure 3.7 implies that monetary policy shocks account, on average, for 17 percent of the variation. In comparison, the contribution of supply shocks features an almost constant decline of in total 10pp., which is only temporarily interrupted by the burst of the dot-com bubble and the Great Recession. Supply shocks display an increasing contribution. At the

Figure 3.7: Forecast error variance decomposition



Notes: Median contribution of the shocks to the forecast error variance of all endogenous variables after 20 quarters along with 68% equal-tailed point-wise posterior probability bands. ϵ^{MP}, ϵ^{AD}, and ϵ^{AS} refer to monetary policy, aggregate demand, and aggregate supply shocks, respectively.

end of the sample, roughly 45 percent of the variation of inflation in Germany can be attributed to supply shocks.

The FEVD of GDP growth shows a slightly different pattern. In this case, the shocks' contribution is almost constant over time. The contribution of monetary policy shocks slightly increases during the 1970s from about 16 percent to roughly 18 percent and subsequently remains at this level. The contribution of demand shocks fluctuates around 18 percent throughout the sample. The largest fraction of the forecast error variance of German GDP growth is explained by supply shocks; from 1970 to 2018, it stays close to 45 percent.

The FEVD from models A and B (see Figure 3.12 in Appendix B.3) are almost identical until the Great Recession hits the German economy. Afterwards, according to model B, monetary policy shocks explain a larger fraction of the variables' variation—for the interest rate, the contribution is up to ten percentage points higher. Thus, the shadow rate captures, at least to some degree, EBC's unconventional monetary policy actions.

In summary, these results suggest that changes in the conduct of monetary policy play only a minor role in understanding the changing properties of the German business cycle, thus confirming the findings of Canova and Gambetti (2009) for the US and Benati (2008) for the UK, but casting doubt on the results of Fritsche and Kuzin (2005) and Buch et al. (2004) for Germany. Moreover, consistent with, for instance, Gambetti et al. (2008) or Gordon (2005) for the US, I find that changes in the magnitude of supply shocks and the transmission of demand shocks appear to be a far more important driver of the changing business cycle dynamics in Germany.

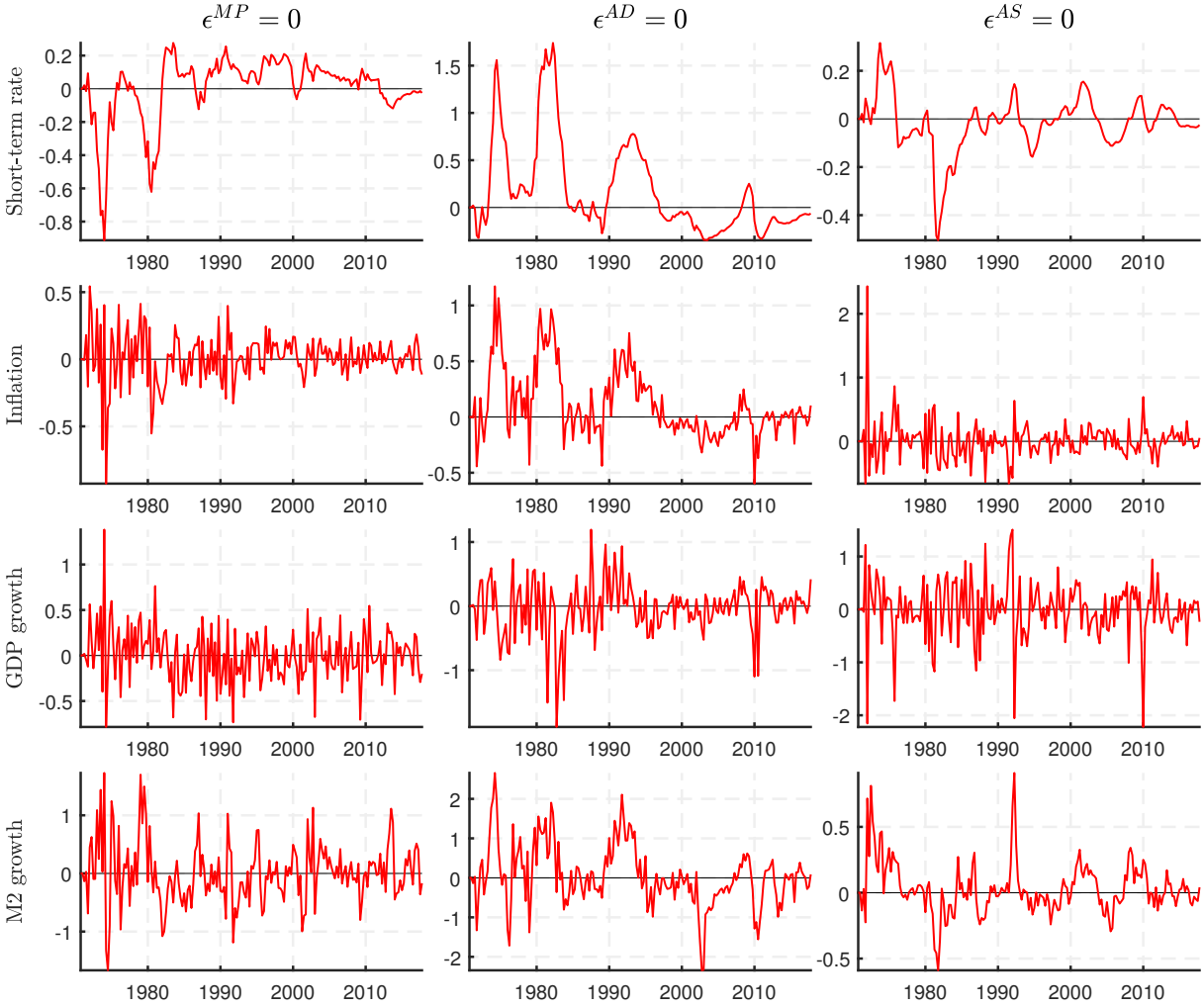
3.5.3 Counterfactual analysis

To assess the relative importance of each shock in generating the actual evolution of the variables, I conduct a counterfactual analysis.[15] Specifically, I follow Sims and Zha (2006b) and Benati (2008) by taking the output of the Gibbs sampler as given, set one shock at a time to zero, and then calculate how the variables would have evolved without these shocks taking place. Thus, large differences between the actual development and the ones of the respective counterfactual indicate that this shock was an important driver of the economy. Figure 3.8 presents the results from this exercise. The rows refer to the variables; the columns to the shocks being switched off. Values above (below) zero indicate that the counterfactual value of the series is smaller (larger) than the actual.

Figure 3.8 suggests both that monetary policy shocks exert only a little impact on the evolution of the German economy and that their impact is decreasing over time, especially regarding inflation. The largest difference between both paths is obtained for M2 growth, which would have been less volatile in the absence of monetary policy shocks during the 1970s. The results with regard to inflation are consistent with the findings from Sims and Zha (2006b); large parts of the fluctuations of inflation are associated to nonpolicy shocks. With regard to GDP growth, the impact of monetary policy shocks is rather constant over time. However, in many periods, the differences between the actual evolution and the counterfactual are negative. Thus, it is suggested that without unsystematic interventions of the monetary authority, GDP growth would have been slightly higher in many periods.

[15] Counterfactuals based on structural VARs are subject to the Lucas (1976) critique, which is why the results should be regarded with caution. However, if the differences between the counterfactual and the actual evolution of the economy are small, it is reasonable to assume that the public would not regard the counterfactual as the result of a new probability law. Thus, counterfactuals should not be ignored (see Sims, 1998; Sims and Zha, 2006a).

Figure 3.8: Historical decomposition – one shock equal to zero at a time



Notes: Median difference between actual and counterfactual evolution of the endogenous variables. Figures above (below) zero indicate that the counterfactual is smaller (larger) than the actual. ϵ^{MP}, ϵ^{AD}, and ϵ^{AS} refer to monetary policy, aggregate demand, and aggregate supply shocks, respectively.

As already suggested by Figure 3.7, demand shocks had a much larger impact on the economy. Without sudden hikes in aggregate demand, the short-term interest rate would have been up to 1.5pp. higher through the mid 1970s and 1980s. A similar picture is obtained for inflation; until the late 1990s it would have been lower in the absence of demand shocks in most periods. Afterwards, it is vice versa. The differences for output growth suggest that the impact of demand shocks has considerably decreased over time. Until 1985, the differences are mostly negative, indicating that GDP growth is higher in the counterfactual, on average, about one percentage point higher. From 1985 onward, the differences substantially decrease in magnitude

and fluctuate around zero. M2 growth is strongly affected by demand shocks during the whole sample period. However, while the differences are mostly positive until 1995, they are negative in almost each period since 1995.

Unexpected changes in aggregate supply exert the strongest effects on the evolution of output growth, with differences of up to two percentage points, for instance, during the Great Recession. The impact on the remaining variables, though, is considerably smaller. Moreover, and in contrast to the other shocks, the differences between the counterfactual excluding supply shocks and the actual evolution of the variables do not exhibit a strong decline in magnitude over time.

In sum, the findings from this exercise show that large parts of the reduction in business cycle volatility in Germany are due to a strong reduction in the response of the endogenous variables to demand and—more important—supply shocks. Thus, providing support for the good luck hypothesis.

3.6 Conclusion

The reduction of business cycle volatility has been found for several countries, including Germany. This chapter provides a more comprehensive view on this issue by means of a time-varying parameter VAR with stochastic volatility. I conducted both a reduced-form and a structural analysis. The former demonstrates that not only the volatility of output growth has substantially declined over time, but also the volatility of inflation, M2 growth, and the interest rate. These reductions were mainly driven by smaller variances of the reduced-form residuals. However, the series' persistence also shows slight variations over time, which provides evidence that good luck is not the only explanation for the Great Moderation in Germany. Using a structural identification based on sign restrictions, I examine how the responses of the variables to structural shocks have evolved over time. While monetary policy innovations account only for a minor part of the changing business cycle dynamics, the decreasing magnitude of supply shocks are a far more important contributor for business cycle stabilization in Germany. I also document slight changes in the response of the private sector to the identified shocks. However, in relation to the decline of the shocks' sizes, this effect plays only a minor role. In total, the results provide strong support in favor of the good luck hypothesis. However, I use only a small-scale model and focus on very broadly defined shocks. Future research should investigate the impact of a larger amount of information on the time-varying dynamics and try to extract other, more specific structural disturbances.

References

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2017). Tracking the Slowdown in Long-Run GDP Growth. *The Review of Economics and Statistics 99*(2), 343–356.

Aßmann, C., J. Hogrefe, and R. Liesenfeld (2009). The Decline in German Output Volatility: A Bayesian Analysis. *Empirical Economics 37*(3), 653–679.

Ball, L. (2014). Long-term Damage from the Great Recession in OECD Countries. *European Journal of Economics and Economic Policies: Intervention 11*(2), 149–160.

Balleer, A., B. Gehrke, W. Lechthaler, and C. Merkl (2016). Does Short-time Work Save Jobs? A Business Cycle Analysis. *European Economic Review 84*, 99–122.

Baumeister, C. and L. Benati (2013). Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound. *International Journal of Central Banking 9*(2), 165–212.

Baumeister, C. and G. Peersman (2013). Time-Varying Effects of Oil Supply Shocks on the US Economy. *American Economic Journal: Macroeconomics 5*(4), 1–28.

Belongia, M. T. and P. N. Ireland (2016). The Evolution of U.S. Monetary Policy: 2000–2007. *Journal of Economic Dynamics and Control 73*, 78–93.

Beltrao, K. I. and P. Bloomfield (1987). Determining the Bandwidth of a Kernel Spectrum Estimate. *Journal of Time Series Analysis 8*(1), 21–38.

Benati, L. (2008). The Great Moderation in the United Kingdom. *Journal of Money, Credit and Banking 40*(1), 121–147.

Benati, L. (2011). Would the Bundesbank Have Prevented the Great Inflation in the United States? *Journal of Economic Dynamics and Control 35*(7), 1106–1125.

Beveridge, S. and C. R. Nelson (1981). A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the "Business Cycle". *Journal of Monetary Economics 7*(2), 151–174.

Black, F. (1995). Interest Rates as Options. *The Journal of Finance 50*(5), 1371–1376.

Brand, C., M. Bielecki, and A. Penalver (2018). The Natural Rate of Interest: Estimates, Drivers, and Challenges to Monetary Policy. Occasional Paper Series 217, European Central Bank.

Brenke, K., U. Rinne, and K. F. Zimmermann (2013). Short-Time Work: The German Answer to the Great Recession. *International Labour Review 152*(2), 287–305.

Buch, C. M., J. Doepke, and C. Pierdzioch (2004). Business Cycle Volatility in Germany. *German Economic Review 5*(4), 451–479.

Burda, M. C. and J. Hunt (2011). What Explains the German Labor Market Miracle in the Great Recession? *Brookings Papers on Economic Activity 42*(1), 273–335.

Canova, F. and L. Gambetti (2009). Structural Changes in the US economy: Is There a Role for Monetary Policy? *Journal of Economic Dynamics and Control 33*(2), 477–490.

Carter, C. K. and R. Kohn (1994). On Gibbs Sampling for State Space Models. *Biometrika 81*(3), 541–553.

Cogley, T. (2005). How Fast Can the New Economy Grow? A Bayesian Analysis of the Evolution of Trend Growth. *Journal of Macroeconomics 27*(2), 179–207.

Cogley, T., G. E. Primiceri, and T. J. Sargent (2010). Inflation-Gap Persistence in the US. *American Economic Journal: Macroeconomics 2*(1), 43–69.

Cogley, T. and T. J. Sargent (2001). Evolving Post-World War II US inflation Dynamics. In B. S. Bernanke and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2001*, Volume 16, 331–373. National Bureau of Economic Research, Inc.

Cogley, T. and T. J. Sargent (2005). Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S. *Review of Economic Dynamics 8*(3), 262–302.

Friedman, M. (1987). Quantity Theory of Money. In J. Eatwell, M. Milgate, and P. Newman (Eds.), *The New Palgrave: A Dictionary of Economics*, Volume 4, 3–20. Palgrave Macmillan, London.

Fries, S., J.-S. Mésonnier, S. Mouabbi, and J.-P. Renne (2018). National Natural Rates of Interest and the Single Monetary Policy in the Euro Area. *Journal of Applied Econometrics 3*(6), 763–779.

Fritsche, U. and V. Kuzin (2005). Declining Output Volatility in Germany: Impulses, Propagation, and the Role of Monetary Policy. *Applied Economics 37*(21), 2445–2457.

Gadea Rivas, M., A. Gómez-Loscos, and G. Pérez-Quirós (2014). The Two Greatest. Great Recession vs. Great Moderation. Working Paper 1423, Banco de Espana.

Gambacorta, L., B. Hofmann, and G. Peersman (2014). The Effectiveness of Unconventional Monetary Policy at the Zero Lower Bound: A Cross-Country Analysis. *Journal of Money, Credit and Banking 46*(4), 615–642.

Gambetti, L., E. Pappa, and F. Canova (2008). The Structural Dynamics of U.S. Output and Inflation: What Explains the Changes? *Journal of Money, Credit and Banking 40*(2-3), 369–388.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics*, 169–193. Oxford University Press.

Gordon, R. J. (2005). What Caused the Decline in U.S. Business Cycle Volatility? NBER Working Paper 11777, National Bureau of Economic Research.

Hartung, B., P. Jung, and M. Kuhn (2018). What Hides Behind the German Labor Market Miracle? Unemployment Insurance Reforms and Labor Market Dynamics. CESifo Working Paper 7379, CESifo Group Munich.

Jacquier, E., N. G. Polson, and P. E. Rossi (1995). Models and Priors for Multivariate Stochastic Volatility. CIRANO Working Papers 95s-18, CIRANO.

Jannsen, N., G. Potjagailo, and M. H. Wolters (2018). Monetary Policy During Financial Crises: Is the Transmission Mechanism Impaired? *International Journal of Central Banking* (forthcoming).

Koop, G. M., M. H. Pesaran, and S. M. Potter (1996). Impulse Response Analysis in Nonlinear Multivariate Models. *Journal of Econometrics 74*(1), 119–147.

Krause, M. U. and H. Uhlig (2012). Transitions in the German Labor Market: Structure and Crisis. *Journal of Monetary Economics 59*(1), 64–79. Carnegie-NYU-Rochester Conference Series on Public Policy at New York University on April 15-16, 2011.

Krebs, T. and M. Scheffel (2013). Macroeconomic Evaluation of Labor Market Reform in Germany. *IMF Economic Review 61*(4), 664–701.

Laubach, T. and J. C. Williams (2003). Measuring the Natural Rate of Interest. *The Review of Economics and Statistics 85*(4), 1063–1070.

Lucas, R. J. (1976). Econometric Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy 1*(1), 19–46.

Mills, T. C. and P. Wang (2003). Have Output Growth Rates Stabilised? Evidence from the G-7 Economies. *Scottish Journal of Political Economy 50*(3), 232–246.

Pirschel, I. and M. H. Wolters (2018). Forecasting with Large Datasets: Compressing Information Before, During or After the Estimation? *Empirical Economics 55*(2), 573–596.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies 72*(3), 821–852.

Rinne, U. and K. F. Zimmermann (2012, Oct). Another Economic Miracle? The German Labor Market and the Great Recession. *IZA Journal of Labor Policy 1*(1), 3.

Rubio-Ramirez, J. F., D. F. Waggoner, and T. Zha (2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies 77*(2), 665–696.

Sims, C. A. (1998). The Role of Interest Rate Policy in the Generation and Propagation of Business Cycles: What Has Changed Since the '30s? Volume 42 of *Annual Research Conference*, 121–175. Federal Reserve Bank of Boston.

Sims, C. A. and T. Zha (2006a). Does Monetary Policy Generate Recessions? *Macroeconomic Dynamics 10*(2), 231–272.

Sims, C. A. and T. Zha (2006b). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review 96*(1), 54–81.

Stock, J. H. and M. W. Watson (2005). Understanding Changes in International Business Cycle Dynamics. *Journal of the European Economic Association 3*(5), 968–1006.

Summers, L. H. (2014). Reflections on the 'New Secular Stagnation Hypothesis'. In C. Teulings and R. Baldwin (Eds.), *Secular Stagnation: Facts, Causes and Cures*. CEPR Press.

Summers, P. M. (2005). What Caused the Great Moderation? Some Cross-Country Evidence. *Economic Review* (Q III), 5–32.

Wu, J. C. and F. D. Xia (2017). Time-Varying Lower Bound of Interest Rates in Europe. Research Paper 17-06, Chicago Booth.

B Appendix

B.1 Details on the model estimation

Priors

To estimate the model in (3.1), prior distributions for the AR-coefficients, the stochastic volatilities, and the contemporaneous relations of the volatilities have to be selected. Following Primiceri (2005), I specify these prior distributions using a training sample. In the following, variables denoted with OLS refer to OLS quantities based on this training sample. The training sample consists of the first 10 years of the entire sample, denoted by T_0.
I draw the VAR-coefficients subject to the following prior:

$$p(\beta_0) \sim N(\hat{\beta}_{OLS}, 4 \times V(\hat{\beta}_{OLS}). \tag{B.1}$$

The prior for the covariance of the VAR-coefficients, Q, follows an inverse-Wishart distribution:

$$p(Q) \sim IW(k_Q^2 \times T_0 \times V(\hat{\beta}_{OLS}), T_0). \tag{B.2}$$

The prior distribution for the stochastic volatilties and the contemporaneous relations follow normal distributions:

$$p(\log \sigma_0) \sim N(\log \hat{\sigma}_{OLS}, I_n), \tag{B.3}$$

$$p(A_0) \sim N(\hat{A}_{OLS}, 4 \times V(\hat{A}_{OLS})). \tag{B.4}$$

The priors for the covariances of $\log \sigma_0$ and A_0 are inverse-Wishart distributed:

$$p(\Psi) \sim IW(k_\Psi^2 \times (1+n) \times I_n, 4), \tag{B.5}$$

$$p(\Phi_i) \sim IW(k_\Phi^2 \times (i+1) \times V(\hat{A}_{i,OLS}), i+1), \qquad i = 1, \ldots, k-1. \tag{B.6}$$

where i denotes the respective VAR-equation that has non-zero and non-one elements in the lower-triangular matrix A_t, i.e. for $n = 4$ it is equation 2, 3, and 4. For the hyperparameters k_Q, k_Ψ, and k_Φ, I follow common practice by setting them to 3.5^{-4}, 0.001, and 0.001.

Specification of the Gibbs sampler

To simulate the posterior distribution of the coefficients, I apply the MCMC algorithm of Cogley et al. (2010), which combines features from the Primiceri (2005) and Cogley and Sargent (2005) algorithms. The algorithm consecutively draws from the conditional distributions. Denote any

vector of variables x over the sample T by $x^T = [x_1', \ldots, x_T']$, the Gibbs sampler takes the following form:

1. Initialize β_t, Σ^T, A^T, Q, Ψ, and Φ.
2. Draw β^T from $p(\beta^T|y^T, Q, \Sigma^T, A^T, \Psi, \Phi)$.
3. Draw Q from $p(Q|y^T, \beta^T, \Sigma^T, A^T, \Psi, \Phi)$.
4. Draw A^T from $p(A^T|y^T, \beta^T, Q, \Sigma^T, \Psi, \Phi)$.
5. Draw Φ from $p(\Phi|y^T, \beta^T, Q, \Sigma^T, A^T, \Psi)$.
6. Draw Ψ from $p(\Psi|y^T, \beta^T, Q, \Sigma^T, A^T, \Phi)$.
7. Draw Σ^T from $\tilde{p}(\Sigma^T|y^T, \beta^T, Q, A^T, s^T, \Psi, \Phi)$.

Step 2: Drawing the VAR-coefficient β^T

Draws for β_t are obtained by using the Carter and Kohn (1994) algorithm, i.e., I run the Kalman filter until T to obtain $\beta_{T|T}$ as well as $P_{T|T}$ and draw β_T from $N(\beta_{T|T}, P_{T|T})$. Subsequently, for $t = T-1, \ldots, 1$, I draw β_t from $N(\beta_{t|t}, P_{t|t})$ by recursively updating $\beta_{t|t}$ and $P_{t|t}$.

Step 3: Drawing the covariance of the VAR-coefficients Q

The posterior of the covariance of VAR-coefficients is inverse-Wishart distributed with scale matrix $\overline{Q} = Q_0 + e_t' e$, $e_t = \Delta\beta_t'$, and degrees of freedom $df_Q = T + T_0$, where Q_0 denote the prior scale for Q and prior degrees of freedom, respectively.

Step 4: Drawing the elements of A^T

To draw the elements of A_T, I follow Primiceri (2005) and rewrite the VAR in (3.1) as follows:

$$A_t(\tilde{y}_t - Z_t'\beta_t) = \tilde{y}_t^* = \Sigma_t u_t, \tag{B.7}$$

where, taking into account that β_T and $\tilde{y}_t$ are known, y_t^* is observable. Due to the lower-triangular structure of A_t^{-1}, this system can be written as a system of k equations:

$$\hat{\tilde{y}}_{1,t} = \sigma_{1,t} u_{1,t}, \tag{B.8}$$

$$\hat{\tilde{y}}_{i,t} = -\hat{y}_{[1,i-1]} a_{i,t} + \sigma_{i,t} u_{i,t}, \qquad i = 2, \ldots, k, \tag{B.9}$$

where $\hat{\tilde{y}}_{[1,i-1]} = [\hat{\tilde{y}}_{1,t}, \ldots, \hat{\tilde{y}}_{i-1,t}]$. $\sigma_{i,t}$ and $u_{i,t}$ refer to the i-th elements of σ_t and u_t. Thus, under the block diagonal assumption of Φ, the RHS of equation i does not include $\hat{y}_{i,t}$, implying that one can recursively obtain draws for $a_{i,t}$ by applying an otherwise ordinary Carter and Kohn (1994) algorithm equation-wise.

Step 5: Drawing the covariance Φ_i of the elements of A^T

Φ_i has an inverse-Wishart posterior with scale matrix $\overline{\Phi}_i = \Phi_{0,i} + \epsilon'_{i,t}\epsilon_{i,t}$, $\epsilon_{i,t} = \Delta a'_{i,t}$, and degrees of freedom $df_{\Phi_i} = T + df_{\Phi_{i,0}}$ for $i = 1, \ldots, k$. $\Phi_{0,i}$, and $df_{\Phi_{i,0}}$ denote prior scale and prior degree of freedoms, respectively.

Step 6: Drawing the covariance Ψ of log-volatilities

As in Step 6, Ψ has an inverse-Wishart distributed posterior with scale matrix $\overline{\Psi} = \Psi_0 + \varepsilon'_t\varepsilon_t$, $\varepsilon_t = \Delta \log \sigma_t'^2$, and degrees of freedom $df_\Psi = T + df_{\Psi_0}$, where Ψ_0 and df_{Ψ_0} denote the prior scale and the prior degree of freedoms, respectively.

Step 7: Drawing the volatilities

Following Cogley and Sargent (2005), I sample the stochastic volatilities one at a time using the Jacquier, Polson, and Rossi (1995) algorithm.

I employ 90000 burn-in iterations of the Gibbs sampler for each model and use every 10th draw of 10000 after burn-in draws for posterior inference. Convergence statistics are provided in the next section.

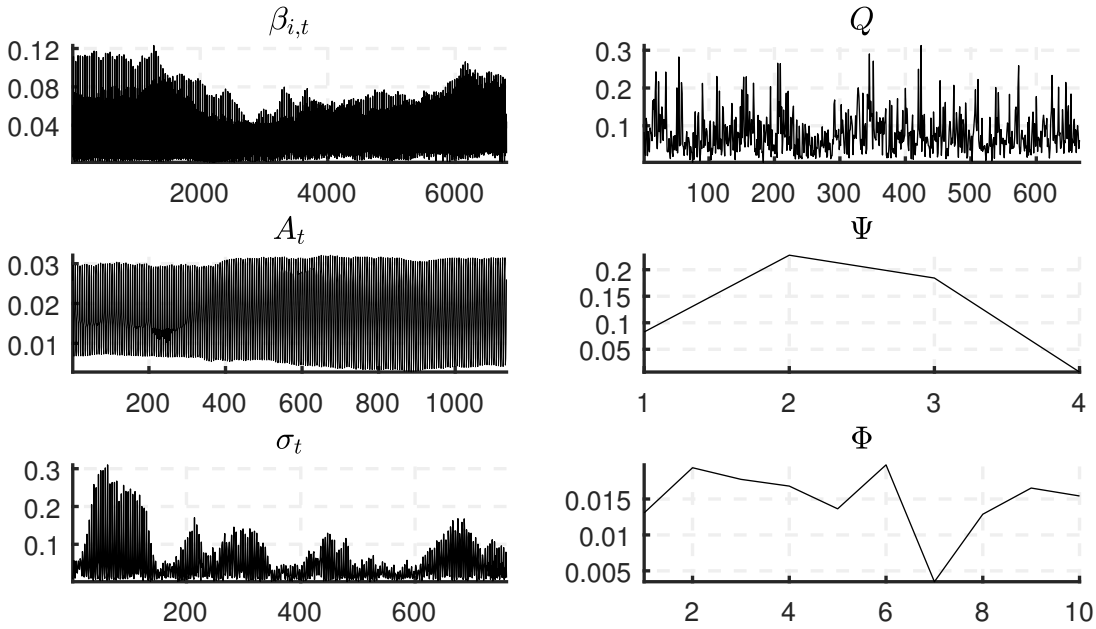
Convergence of the Gibbs sampler

Convergence of the Markov Chains is assessed by inspecting the draws' autocorrelation functions. To this end, I compute inefficiency factors (IFs) for the draws of the coefficients, which are defined as the inverse of the relative numerical efficiency measure introduced by Geweke (1992):

$$\text{RNE} = (2\pi)^{-1} \frac{1}{S(0)} \int_{-\pi}^{\pi} S(\omega) d\omega, \tag{B.10}$$

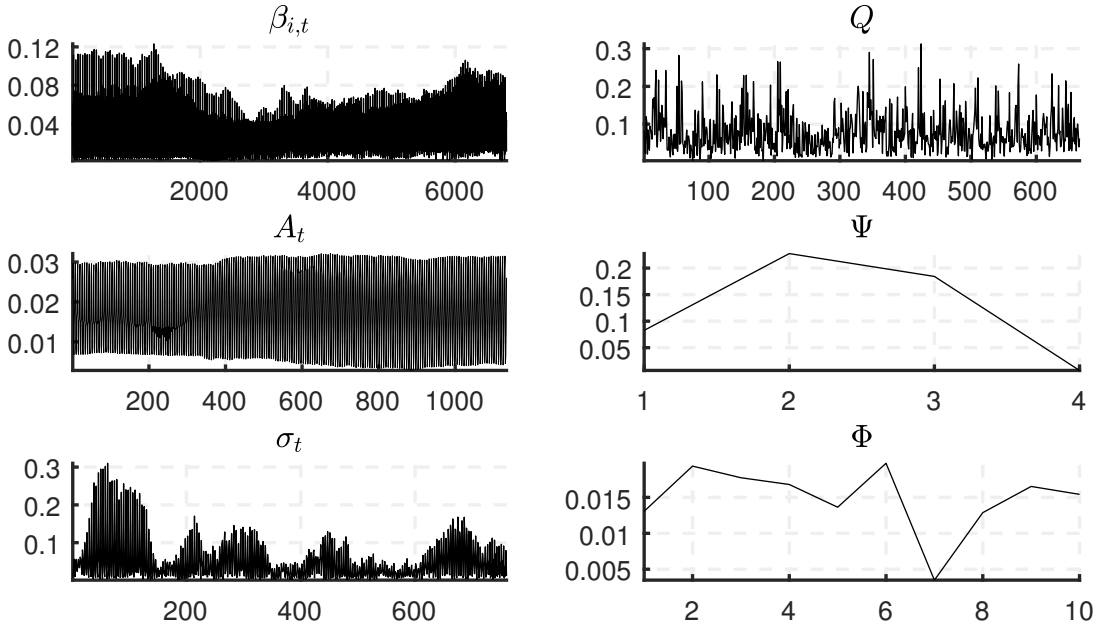
where $S(\omega)$ denotes the spectral density of the draws from the Gibbs sampler for the coefficient considered at frequency ω. I compute the latter quantity by smoothing the periodograms in the frequency domain by means of a Bartlett spectral window (Benati, 2008). The bandwidth parameter is automatically select via the method provided by Beltrao and Bloomfield (1987). As stressed by Primiceri (2005), IFs below 20 are regarded as efficient, implying that 20 times as many MCMC draws as from an uncorrelated sample have to be drawn. Figures 3.9 and 3.10 display the IFs for the coefficients of model A and B, respectively. For each coefficients the IFs are far below 20, suggesting that the draws come from the ergodic posterior distribution.

Figure 3.9: Inefficiency factors of model A



Notes: Inefficiency factors for the states (left column) and the hyperparameters (right column). Ordinate: Inefficiency factor, abscissa: parameter.

Figure 3.10: Inefficiency factors of model B



Notes: Inefficiency factors for the states (left column) and the hyperparameters (right column). Ordinate: Inefficiency factor, abscissa: parameter

B.2 Implementation of generalized impulse responses

Following Koop et al. (1996), GIRFs are calculated as the differences between two conditional expectations. Formally, the GIRF at horizon h of variables y to a shock of size ε and conditional on an initial condition I_{t-1} is defined as follows:

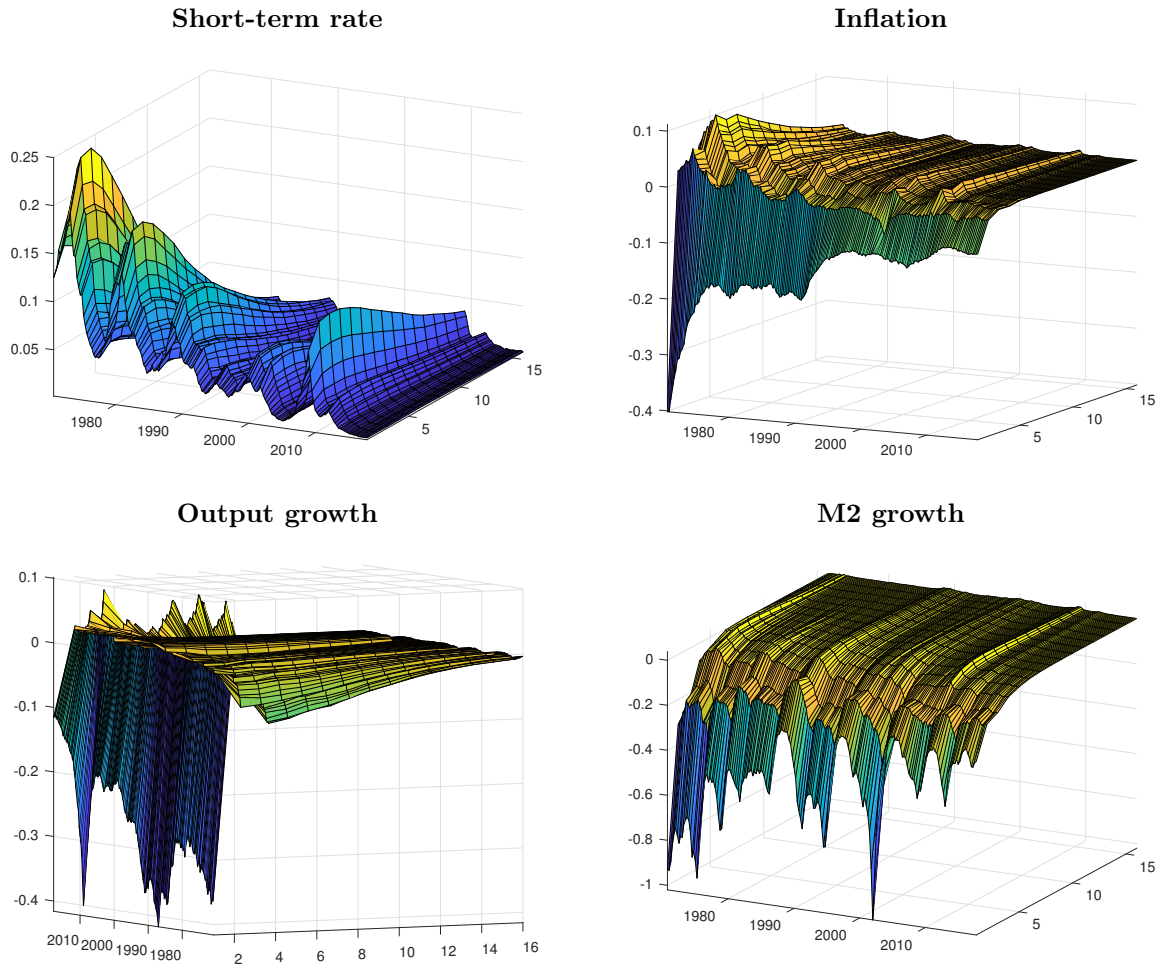
$$\text{GIRF}_y(h, \epsilon, I_{t-1}) = E[y_{t+h}|\epsilon, I_{t-1}] - E[y_{t+h}|I_{t-1}]. \tag{B.11}$$

To compute the right-hand side terms at each point in time, I use the laws of motions of the time-varying coefficients and a randomly selected draw from the Gibbs sampler to project the model h periods into the future. I employ for each initial condition 500 draws from the Gibbs sampler each with a shock hitting the system in the initial period and without this shock. I then average across the differences between both time paths to obtain the GIRF for the respective history.

The structural impact matrix, $B_{0,t}$, is obtained using the procedure of Rubio-Ramirez et al. (2010). Specifically, I decompose the time-varying covariance matrix of the VAR, Ω_t, according to $\Omega_t = P_t D_t P_t'$ and define $\tilde{B}_t = P_t D_t^{0.5}$. Moreover, I draw an $N \times N$ matrix, K, from a standard normal distribution and compute its QR decomposition, that is, I calculate Q and R (with all entries normalized to be positive) such that $K = QR$ holds. Finally, I obtain the structural impact matrix as $B_{0,t} = \tilde{B}_t Q'$. Using $u_{i,t} = B_{0,t}\varepsilon_{i,t}$, where $u_{i,t}$ and $\varepsilon_{i,t}$ denote the reduced-form and structural residuals, respectively, I impose a structural shock on variable i by setting $\varepsilon_{i,t} = \varepsilon_{i,t} + 1$. From the set of possible impulse responses I retain only those, which satisfy the imposed sign restrictions.

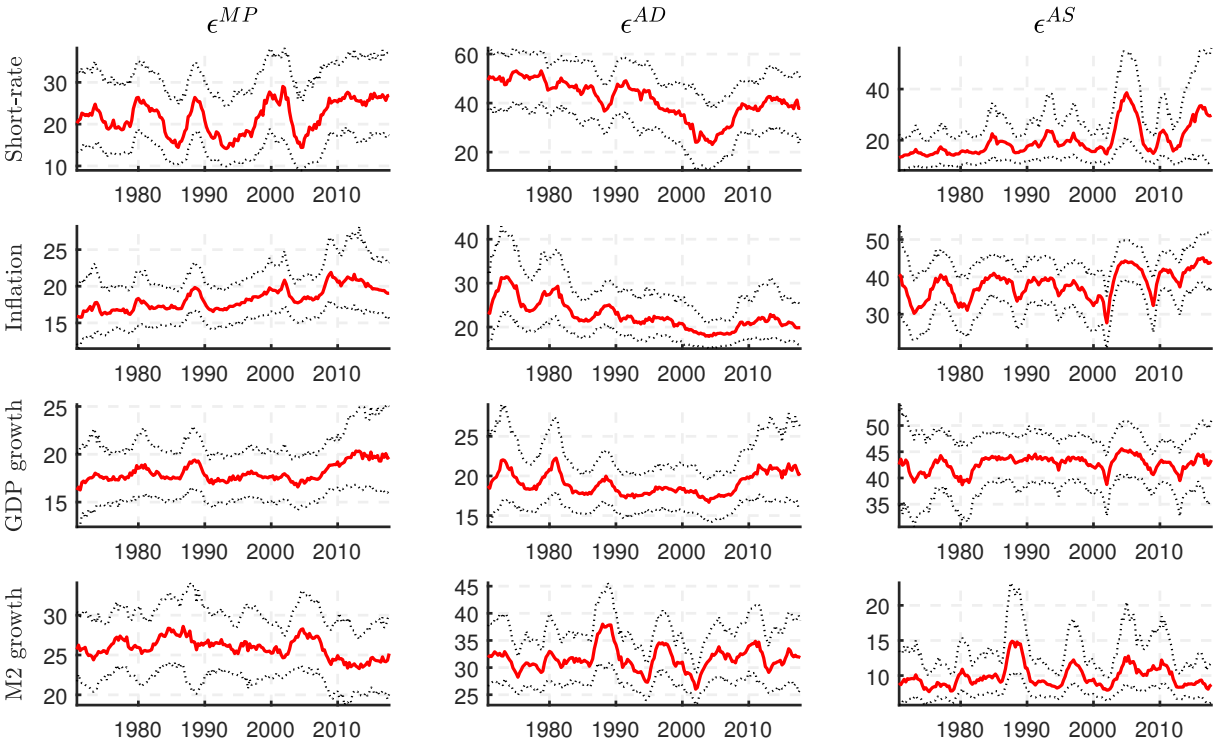
B.3 Additional figures

Figure 3.11: Median generalized impulse responses to a monetary policy shock over time



Notes: Median responses of the variables to a monetary policy shock over time according to model A.

Figure 3.12: Forecast error variance decomposition for model B



Notes: Median contribution of the shocks to the forecast error variance of all endogenous variables after 20 quarters along with 68% equal-tailed point-wise posterior probability bands. ϵ^{MP}, ϵ^{AD}, and ϵ^{AS} refer to monetary policy, aggregate demand, and aggregate supply shocks, respectively.

4 | Macroeconomic Uncertainty and Forecasting Macroeconomic Aggregates

Abstract Can information on macroeconomic uncertainty improve the forecast accuracy for key macroeconomic time series for the US? Since previous studies have demonstrated that the link between the real economy and uncertainty is subject to nonlinearities, I assess the predictive power of macroeconomic uncertainty in both linear and nonlinear Bayesian VARs. For the latter I use a threshold VAR that allows for regime-dependent dynamics conditional on the level of the uncertainty measure. I find that the predictive power of macroeconomic uncertainty in the linear VAR is negligible. In contrast, using information on macroeconomic uncertainty in a threshold VAR can significantly improve the accuracy of short-term point and density forecasts, especially in the presence of high uncertainty.

Keywords: Forecasting, BVAR, nonlinearity, threshold VAR, uncertainty
JEL-Codes: C11, C53, C55, E32

4.1 Introduction

Since the seminal contribution of Bloom (2009), the contractive effects of uncertainty shocks on the real economy are uncontroversial.[1] Moreover, recent studies show that uncertainty shocks have nonlinear effects. On the one hand, uncertainty shocks induce stronger effects during recessionary episodes or in times of financial distress (see, for instance, Caggiano, Castelnuovo, and Groshenny, 2014; Ferrara and Guérin, 2018; Alessandri and Mumtaz, 2019). On the other hand, the magnitude of the variables' response to the uncertainty shock depends on the shock's sign (Jones and Enders, 2016; Foerster, 2014). While a great deal of the literature focus on structural analysis of fluctuations in uncertainty, evidence regarding the impact of uncertainty on forecast performance is, however, rather sparse.

This chapter explores the link between economic uncertainty and forecast performance, making two contributions to the literature. First, I assess the predictive power of uncertainty in a linear model. I derive the baseline results using the large Bayesian VAR (BVAR) approach introduced by Bańbura, Giannone, and Reichlin (2010).[2] The impact of economic uncertainty on forecast performance is assessed by adding a recursively estimated version of the macroeconomic uncertainty index of Jurado, Ludvigson, and Ng (2015) to a medium-sized dataset of macroeconomic indicators for the US. Second, I investigate whether allowing for nonlinearity improves forecast accuracy relative to standard, linear models. To this end, I employ a threshold BVAR (T-VAR) that accounts for nonlinear relations between macroeconomic uncertainty and the real economy. This model allows to directly link the nonlineartiy to the threshold variable, which in my application is the uncertainty index mentioned above.[3] Moreover, the T-VAR facilitates the possibility of two distinct regimes, which can be interpreted as high and low uncertainty regimes. Since these regimes can differ in all of the model's parameters, the model allows for regime-dependent shock propagation processes and heteroscedasticity. As shown by several studies (for example, Barnett, Mumtaz, and Theodoridis, 2014; Clark and Ravazzolo, 2015; Alessandri and Mumtaz, 2017), although not in the context of uncertainty, both features can significantly increase forecast accuracy. To estimate the threshold VAR, I combine the Gibbs sampler provided by Chen and Lee (1995) with the large Bayesian VAR framework mentioned above and the hyperparameter estimation approach of Giannone, Lenza, and Primiceri (2015). The appealing property of this approach is that each of the model's parameters, including the tightness of the

[1] For the transmission of uncertainty shocks to the real economy, capital adjustment frictions (Bernanke, 1983; Caballero and Pindyck, 1996; Bachmann and Bayer, 2013) and financial frictions (Gilchrist, Sim, and Zakrajšek, 2014; Christiano, Motto, and Rostagno, 2014) have been found to be important.

[2] The large BVAR has been proven capable of processing a large number of economic indicators while generating precise forecasts (see Carriero, Kapetanios, and Marcellino, 2009; Koop, 2013, among others).

[3] As shown in Section 4.4 of this chapter, recessions and phases of high uncertainty do not inevitably coincide, which is why I do not condition the model on recessions.

prior on the model coefficients, the lag of the threshold variable, as well as the threshold level (and therefore the regimes) are estimated endogenously and are purely data driven.

First, I perform an in-sample analysis based on quarterly US data from 1960 to 2017 to demonstrate that the T-VAR yields reasonable full-sample estimates. I illustrate that the estimated high uncertainty regimes are similar, but do not fully coincide with the recession dates provided by the NBER business cycle dating committee. Using the threshold BVAR, I isolate state-dependent uncertainty shocks. To account for the model's nonlinearity, I compute generalized impulse responses à la Koop, Pesaran, and Potter (1996) with the modification of Kilian and Vigfusson (2011) that allows for nonlinear shock propagation. I show that the model is able to generate the effects of uncertainty shocks commonly found in the literature. I find that uncertainty shocks have both negative effects on the real economy and nonlinear effects, depending on the level of the uncertainty proxy. During episodes of high uncertainty, the effects of an uncertainty shock on labor market variables are much stronger. The peak response of the unemployment rate, for instance, is roughly twice the size in times of high uncertainty compared to normal times.

Second, I conduct a rigorous out-of-sample forecast exercise using a recursive estimation scheme that mimics the information set of the actual forecaster at each point in time. I evaluate the forecasts with respect to both point forecasts and predictive densities. The point forecasts are evaluated in terms of mean forecast errors and root mean squared forecast errors. The predictive densities are evaluated using log predictive scores and continuous ranked probability scores.

My main results are that information on economic uncertainty can improve forecast accuracy and that density forecasts benefit more from this information than point forecasts. Concerning the point forecasts, I find that adding the uncertainty proxy to the otherwise standard linear BVAR yields only marginal improvements. Although, in most cases, the T-VAR is outperformed by the linear specifications, interest and unemployment rate forecasts can be significantly improved. With regard to the predictive densities, the linear models are dominated by the T-VAR. Indeed, in most cases, each model overestimate the true uncertainty of the data, indicated by too wide predictive densities. Controlling for uncertainty regimes, though, reduces this bias and provides a better description of the data. This suggests that accounting for state-dependent disturbances is more important for forecasting purposes than state-dependent shock propagation. Finally, I document substantial variation of the model's predictive abilities over time and show that the gains in forecast accuracy are particularly high when uncertainty is high. Thus, the T-VAR can serve as a complement to existing approaches to get a better picture of the actual uncertainty surrounding the point estimate in times of high uncertainty.

This chapter adds to the literature investigating the predictive power of uncertainty indicators. Pierdzioch and Gupta (2017) and Balcilar et al. (2016) focus on forecasting recessions and show that information on uncertainty improves forecast accuracy. Segnon et al. (2018) and Bekiros, Gupta, and Paccagnini (2015) employ bivariate models including information on un-

certainty and suggest that uncertainty can be helpful in predicting GNP growth and oil prices already in small-scale models. None of these contributions considers a large set of indicators that an applied forecaster would use, or directly allows for nonlinearity with respect to the uncertainty measure.

The chapter is structured as follows. Section 4.2 describes the Bayesian VAR as well as the Bayesian threshold VAR and outlines the estimation methodology. Section 4.3 describes the dataset and the forecast methodology. Section 4.4 presents the in-sample results. Section 4.5 discusses the results from the forecast experiment. Section 4.6 concludes.

4.2 The models

In this section, I first describe a standard Bayesian VAR model, following which the Bayesian threshold VAR is outlined.

4.2.1 The Bayesian VAR

The VAR(p) is specified as follows:

$$y_t = c + \sum_{j=1}^{p} A_j y_{t-j} + \varepsilon_t \quad \text{with } \varepsilon_t \sim N(0, \Sigma), \tag{4.1}$$

where y_t and c are $n \times 1$ vectors of endogenous variables and intercept terms, respectively. ε_t denotes the vector of normally distributed residuals. A_j are $n \times n$ matrices of coefficients with $j = 1, \ldots, p$. I employ Bayesian estimation techniques to estimate the model. Specifically, I use the Minnesota prior developed by Litterman (1986), which assumes that every economic time series can be sufficiently described by a random walk with drift. Thus, the prior shrinks all coefficients on the main diagonal of A_1 towards one while the remaining coefficients are shrunk towards zero. Moreover, the classical Minnesota prior assumes a diagonal covariance matrix of the residuals. In the following, I use the generalized version of the classical Minnesota prior provided by Kadiyala and Karlsson (1997), which allows for a non-diagonal residual covariance matrix while retaining the idea of the Minnesota prior described above. As demonstrated by Bańbura et al. (2010), using a normal-inverse Wishart prior generates accurate forecasts despite the additional parameters to be estimated. In addition, I follow Doan, Litterman, and Sims (1984) as well as Sims (1993) by implementing the "sum-of-coefficents" and "co-peristence" prior. The former accounts for unit roots in the data; the latter introduces beliefs on cointegration relations among the series. Each prior is implemented using dummy observations. I estimate the tightness of the priors by applying the hierachical Bayesian procedure of Giannone et al. (2015). For details regarding the prior implementation and the estimation procedure, see Appendices C.1 and C.2.

4.2.2 The Bayesian threshold VAR

The threshold VAR is defined as follows:

$$y_t = \left(c_1 + \sum_{i=1}^{p} A_{1,i} y_{t-i} + \Omega_1^{0.5} \varepsilon_t \right) S_t + \left(c_2 + \sum_{i=1}^{p} A_{2,i} y_{t-i} + \Omega_2^{0.5} \varepsilon_t \right) (1 - S_t), \tag{4.2}$$

$$\text{with: } S_t = \begin{cases} 1, & \text{if } r_{t-d} \leq r^* \\ 0, & \text{if } r_{t-d} > r^* \end{cases} \tag{4.3}$$

where y_t is the vector of endogenous variables. Contrary to the linear VAR in (4.1), the intercept terms c_j and the matrices of coefficients A_j with $j \in \{1, 2\}$ are state dependent. The regime prevailing in period t depends on whether the level of the threshold variable, r, in period $t-d$ is below/above a latent threshold level, $\bar{r}$. This mechanism allows for different model dynamics depending on the respective regime. As in the previous section, I use natural conjugate priors for the VAR coefficients and implement the priors using dummy observations. Moreover, the elements of Λ are separately estimated for both regimes to obtain a sensible degree of shrinkage. I follow Chen and Lee (1995) for the threshold level and the delay coefficient:

$$p(d) = \frac{1}{d_{\max}} \quad \text{and } r^* \sim N(\bar{r}, v), \tag{4.4}$$

where $d_{\max} = 8$ denotes the maximal delay. $\bar{r}$ is sample average of r and $v = 10$. Since both the threshold value $\bar{r}$ and the delay coefficient d depend on the model parameters and Λ_j depends on $\bar{r}$ and d, the algorithm from the previous section is no longer appropriate. In fact, I combine the Metropolis Hastings step for estimating the amount of shrinkage (see Appendix C.2) with the Gibbs sampler introduced by Chen and Lee (1995) to simulate the posterior distribution of the model's parameters. In detail, the Gibbs sampler works as follows:

1. At iteration $k = 1$ set starting values for $d^k = d_0$, $r^{*k} = r_0$.
2. Initialize Λ_j at the posterior mode conditional on d^k and r^{*k}.
3. Draw Λ_j^k according to steps 2 and 3 from the algorithm in the previous section.
4. Draw $\Sigma_j^k | d^k, r^{*k}, \Lambda_j^k, y_j$, and $\beta_j^k | d, r^{*k}, \Lambda_j^k, \Sigma_j^k, y_j$ from their posteriors given by (C.7).
5. Draw a candidate value for r^{*k} by: $r^{**k} = r^{*k-1} + \Phi\epsilon$ with: $\epsilon \sim N(0, 1)$ and Φ is a scaling factor ensuring an acceptance rate of about 20%.
6. Accept the draw with probability

$$p^k = \min \left\{ 1, \frac{p(Y_t | r^*, \theta)}{p(Y_t | \bar{r}^{k-1}, \theta)} \right\} \tag{4.5}$$

where $p(\cdot)$ denotes the posterior density given all other parameters of the model.

7. Draw d from

$$p(d=i|Y_t,\theta) = \frac{p(Y_t|d,\theta)}{\sum_{d=1}^{d_0} p(Y_t|d,\theta)}, \quad \text{for: } i = 1,\dots,d_{\max}. \tag{4.6}$$

8. Generate $e_{j,T+1},\dots,e_{j,T+h}$ from $\epsilon_{j,t} \sim N(0,\Sigma_j^k)$ and compute h-step-ahead forecasts recursively by iterating (4.2) and (4.3) h periods into the future.

9. Redo until $k = D + R$.

I employ 25000 iterations of the Gibbs sampler and discard the first 20000 as burn-ins.

The key element of this model is the threshold variable r, which governs the regime dependency. Different specifications for r are proposed in the literature. Caggiano et al. (2014) and Caggiano, Castelnuovo, and Figueres (2017) argue that recessions are particularly informative regarding the identification of uncertainty shocks. These studies follow Auerbach and Gorodnichenko (2012) and use a moving average of GDP growth rates as threshold variable. Other studies emphasize the importance of the uncertainty proxy itself and condition on either the historic change (for example, Henzel and Rengel, 2017; Foerster, 2014) or the historic level of the uncertainty proxy (Jones and Enders, 2016; Berg, 2017a; Castelnuovo and Pellegrino, 2018, among others). Since this chapter aims at identifying uncertainty regimes, I follow the latter and specify r as the level of the uncertainty indicator.

However, nowadays there are various uncertainty proxies available, for example, stock market volatility (Bloom, 2009), newspaper-based indices (Baker, Bloom, and Davis, 2016), firm-level data-based indices (Bachmann, Elstner, and Sims, 2013), indices based on macroeconomic forecast errors (Rossi and Sekhposyan, 2015), and indices based on the residuals from factor augmented regressions (Jurado et al., 2015). I choose the macroeconomic uncertainty index provided by Jurado et al. (2015), a choice motivated by two factors. First, this proxy defines uncertainty in terms of the variation in the unforecastable component of macroeconomic variables and not in terms of the variables' raw volatility.[4] Second, and in contrast to other measures, it is based on a large number of economic indicators and, hence, should represent an aggregate uncertainty factor that affects many series, sectors, or markets (Jurado et al., 2015).[5]

I recursively construct the index to avoid that the index at a given point in time includes information that would not be available to the forecaster at this moment. As already pointed out by Jurado et al. (2015), the indices based on both in-sample forecasts and out-of-sample forecasts are highly correlated.

[4]The unforecastable component is defined as the expected squared forecast error of a series conditional on all available information.

[5]The macroeconomic uncertainty index is based on the FRED-MD database provided by McCracken and Ng (2016), which consist of 134 series representing broad classes of variables.

4.3 Data and forecast methodology

The dataset includes 11 quarterly US macroeconomic series from 1960Q3 through 2017Q4 covering a broad range of economic activity especially relevant for policymakers and central bankers.[6] The series are obtained via the Federal Reserve Economic Database (FRED). To study the impact of macroeconomic uncertainty on the forecast performance, I further augment the dataset with the economic uncertainty index developed by Jurado et al. (2015).

Most of the series enter the model in annualized log levels, that is, I take logarithms and multiply by 4, except for those series that are already expressed as annualized rates. For the stationary variables, I utilize a white noise prior ($\delta_i = 0$), whereas for integrated series a random walk prior ($\delta_i = 1$) is used. A detailed description of the data, their corresponding transformations and sources is provided in Table 4.1. For both models, I generate 1- up to 4-quarter-ahead forecasts by a recursive estimation scheme over an expanding window. The initial sample runs from 1960Q3 to 2004Q3. Thus, I generate forecasts for 2004Q4 until 2005Q3 in the first recursion. Subsequently, I iterated the procedure by updating the estimation sample with the observations from the next quarter until 2016Q4 is reached. This procedure generates a total of 50 forecasts for each horizon. Forecasts for horizons larger than one are obtained iteratively. The lag length in all VARs is set to four. While I estimate the model with both stationary and integrated variables, I report results solely in terms of annualized percentage growth rates. To this end, I transform the models' level forecasts for the integrated variables into growth rates based on these level forecasts.

Table 4.1: Dataset

Variable	Mnemonic	Source	Transformation
Real GDP	GDPC1	FRED	log ×400
CPI for All Urban Consumers: All Items	CPIAUCSL	FRED	log ×400
Industrial Production Index	INDPRO	FRED	log ×400
All Employees: Total Nonfarm	PAYEMS	FRED	log ×400
Civilian Unemployment Rate	UNRATE	FRED	–
Real Gross Private Domestic Investment	GPDIC1	FRED	log ×400
ISM Manufacturing: PMI Composite Index	NAPM	FRED	–
Personal Consumption Expenditures, Price Index	PCECTPI	FRED	log ×400
Capacity Utilization: Total Industry	TCU	FRED	–
Federal Funds Rate	FEDFUNDS	FRED	–
S&P 500 Composite - Price Index	S&PCOMP	FRED	log ×100
Macroeconomic Uncertainty Index	–	own calculations	–

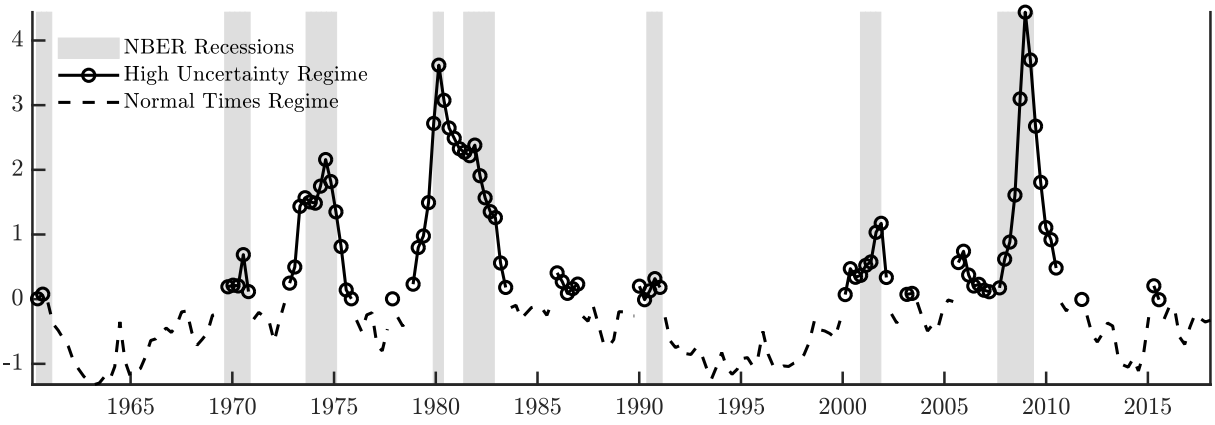
Notes: The macroeconomic uncertainty index is calculated using the codes provided by Jurado et al. (2015) modified to provide a recursively estimated index.

[6] Although a large Bayesian VAR is, in general, capable of processing a much higher number of economic indicators, even medium-sized BVARs produce accurate forecasts (see, for example, Bańbura et al., 2010; Koop, 2013; Berg, 2016).

4.4 In-sample analysis

Now that we have outlined the empirical setup, we turn to investigating the in-sample properties of the Bayesian threshold VAR, which are based on full-sample estimates. Figure 4.1 plots the macroeconomic uncertainty index along with NBER recessions. The solid-dotted line refers to the episodes of the endogenously identified high uncertainty regime, while the dashed line corresponds to the normal times regime. The figure reflects the common knowledge that macroeconomic uncertainty is countercyclical. Moreover, while the uncertainty regimes partly coincide with NBER recessions, they are more persistent and more frequently identified.[7] These discrepancies can be explained by differences in the concepts. NBER defines recessions as significant decline in economic activity, whereas the macroeconomic uncertainty index focuses on predictability. Obviously, the latter implies that booms and recoveries, which are characterized by high growth rates of macroeconomic aggregates, are excluded from the NBER recessions but can be part of the high uncertainty regime if the evolution of these aggregates is hard to predict during these episodes. Nevertheless, these results suggest that recessions are a useful proxy for uncertainty regimes. To directly identify regimes based on the prevailing level of uncertainty, however, might be more appropriate for capturing possible nonlinear dynamics.

Figure 4.1: Estimated uncertainty regimes



Notes: Shaded areas correspond NBER recessions. Solid-dotted line refers to the high uncertainty regime, i.e. the median estimate of $(1-S_t)$ from (4.2) and (4.3).

Having identified uncertainty regimes, we assess whether uncertainty has different effects on the economy depending on the prevailing regime. For this purpose, we perform a structural

[7]For example, according to the NBER Business Cycle Dating Committee, the recession induced by the burst of the dot-com bubble lasted for the entire year 2001, while the high uncertainty regime in turn starts in the first quarter of 2000 and lasts until the first quarter of 2002. The same holds for the Great Recession, which is dated from 2008Q1 until 2009Q2 according to the NBER. The high uncertainty regime begins already in 2007Q2 and then lasts until 2010Q2.

analysis based on impulse responses.[8] As the threshold VAR from Section 4.2.2 is nonlinear, standard impulse responses are not appropriate for capturing the effects of a shock. Thus, I follow Koop et al. (1996) and compute generalized impulse responses (GIRFs). Formally, the GIRF at horizon h of variable y to a shock of size ϵ and conditional on an initial condition I_{t-1} is defined as the difference between two conditional expectations:

$$\text{GIRF}_y(h, \epsilon, I_{t-1}) = E[y_{t+h}|\epsilon, I_{t-1}] - E[y_{t+h}|I_{t-1}], \tag{4.7}$$

where the terms on the right-hand side are approximated by a stochastic simulation of the model. I calculate for each initial condition 500 time paths of length h each with an uncertainty shock hitting the system in the initial period and without this shock. I then average across the differences between both time paths to obtain the GIRF for the respective history. To compute regime-dependent responses, I average over the GIRFs based on the histories of the normal times and high uncertainty regime, respectively. Moreover, I follow Kilian and Vigfusson (2011) and consider orthogonalized residuals to identify uncertainty shocks. The shocks are identified using a recursive estimation scheme based on a Cholesky decomposition with uncertainty ordered second and the S&P 500 ordered first. The latter allows real and nominal variables to react instantaneously to an uncertainty shock (see Bloom, 2009; Fernández-Villaverde et al., 2015; Baker et al., 2016, among others). Since the T-VAR captures regime-dependent shock sizes and shock propagation processes, I consider both a one standard deviation shock and a unit shock to assess whether differences in the responses are triggered by the size of the shock or by its propagation. Due to space constraints, I only present the results for GDP, GDP deflator, investment, consumption, the unemployment rate, and the federal funds rate.[9] The left column of Figure 4.2 plots the responses to a one standard deviation uncertainty shock that is different in magnitude across the regimes. The right column depicts the responses for the unit shock. The red line is the response in the high uncertainty regime; the dashed blue line corresponds to the normal times regime. Shaded areas and dotted lines refer to 68% error bands.

First, Figure 4.2 shows that independently of both the size of the shock and the regime, an increase in macroeconomic uncertainty operates as a negative demand shock. Private consumption drops persistently. A likely explanation for this is precautionary saving by households. The latter reduces the demand for investment goods and leads to a decline in investment, which is roughly twice as large as the drop in consumption. Moreover, the responses point at the existence of the real option effect. As a consequence of increased uncertainty, investors postpone investment decisions—if investment is (partially) irreversible—until business conditions become clearer (Bernanke, 1983). Finally, the unemployment rate persistently increases and follows a

[8] For generating the impulse responses, the variabels enter the model in logarithms multiplyed by 100 so that they can be interpreted as percentage deviations from the trend. Moreover, the macroeconomic uncertainty index is standardized to faciliate the interpretation of the shock sizes.

[9] The effects for the remaining variables are presented in Appendix C.3.

hump-shaped path with a peak effect occurring seven quarters subsequent to the impact period. These results are in line with previous studies (see, for instance, Caggiano et al., 2014; Caldara et al., 2016) and follow the predictions of theoretical models incorporating price rigidities (Basu and Bundick, 2017; Leduc and Liu, 2016).

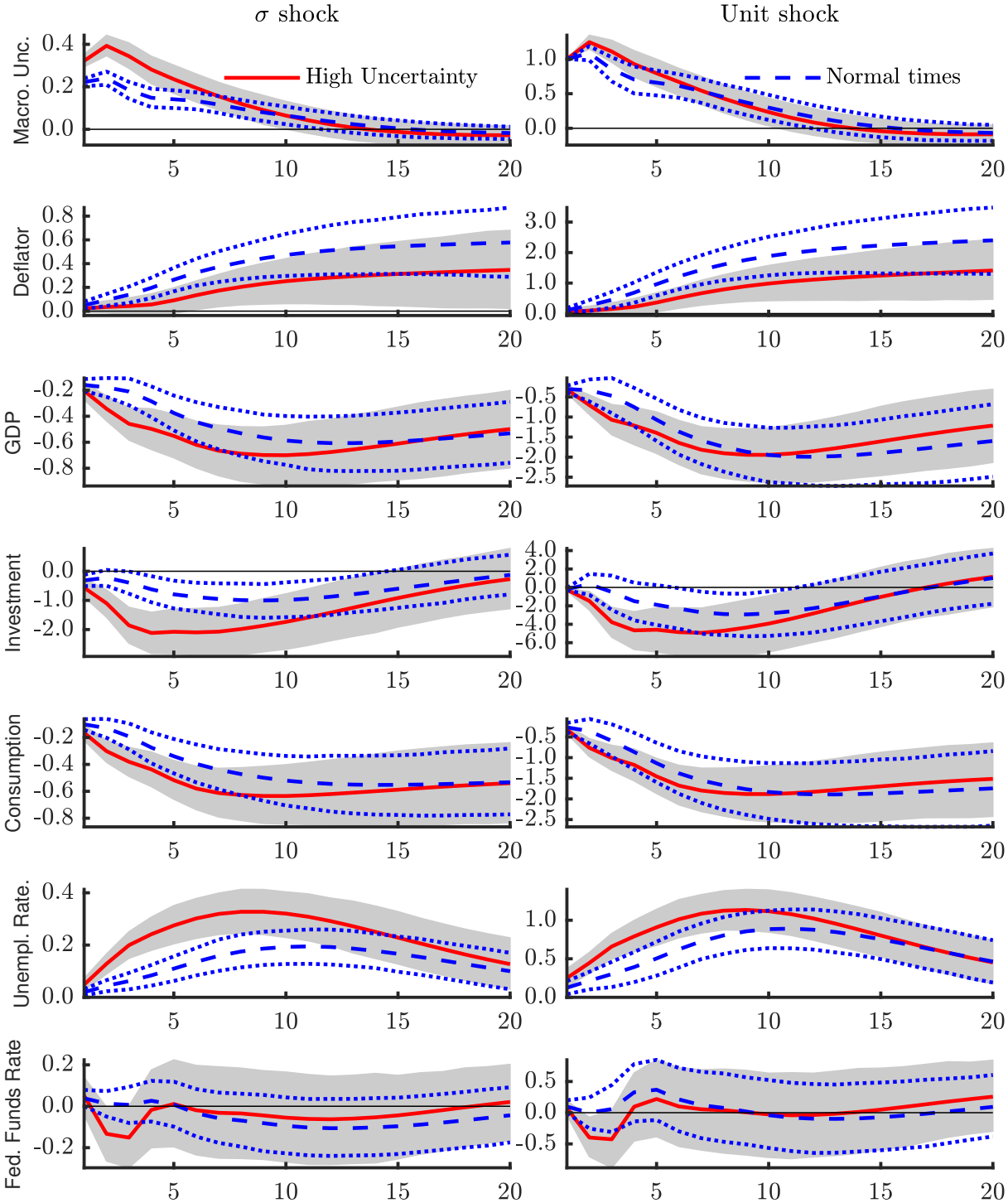
Evidence regarding the price response to an increase in uncertainty is mixed. Figure 4.2 depicts weak inflationary effects and supports the findings of Alessandri and Mumtaz (2019), Mumtaz and Theodoridis (2015, 2018), and Popescu and Smets (2010). Other studies stress the deflationary effects of uncertainty shocks (see, for instance, Christiano et al., 2014; Carriero et al., 2015). From a theoretical point of view, the responses provide evidence in favor of an "inverse Oi (1961)-Hartman (1972)-Abel (1973) effect". As pointed out by Born and Pfeifer (2014, 2017) and Fernández-Villaverde et al. (2015), given sticky prices, firms can set a price, which is either too low or too high. The former is obviously not optimal because the firm has to sell too many units at a too low price. However, in the latter case, the firm sells too few units but is compensated by a higher price per unit. Therefore, firms are prone to an upward bias in future prices, which can lead to inflationary effects of an uncertainty shock.

Second, the estimated size of the uncertainty shock is roughly 1.5 times larger in the high uncertainty regime than in the normal times regime (0.22 to 0.33). However, the persistence of the shock is significantly lower in the high uncertainty regime. Third, comparing the responses across regimes reveals statistically significant differences. The impact of the shock is much larger during times of high uncertainty. Investment, for instance, drops by roughly 0.5% in normal times compared with a decline by 2.0% in times of high uncertainty. The same pattern holds for the unemployment rate, which significantly increases to roughly twice as high in the high uncertainty regime (0.35% versus 0.17%). Thus, in line with previous studies, the contractionary effects of uncertainty shocks are especially large when uncertainty is already at a high level (Jones and Enders, 2016; Bijsterbosch and Guérin, 2013). These results suggest that using a linear model potentially underestimates the actual effect of a sudden hike in economic uncertainty. Finally, as in Caggiano et al. (2014) and Alessandri and Mumtaz (2019), monetary policy seems to react to uncertainty shocks only in crisis periods (either recessions or financial stress) by lowering the policy rate. However, the response is not distinguishable from zero.[10]

To arrive at a better impression of the relative importance of the shock's size and its propagation, Figure 4.3 depicts the differences in the responses between both regimes along with 68% error bands. Overall, the differences in the responses to the unit shock are larger, however, the corresponding error bands are wide and differences become insignificant after a few quarters for most variables. In contrast, the differences between the state-dependent responses of the one standard deviation shock are less pronounced but remain significantly different form zero longer. This suggests that the shock size is a very important factor for the state-dependency of

[10]The interest rate response documented by Caggiano et al. (2014) is somewhat larger. The latter analysis, however, does not allow for regime switches in the responses, which tends to increase the effect of a shock.

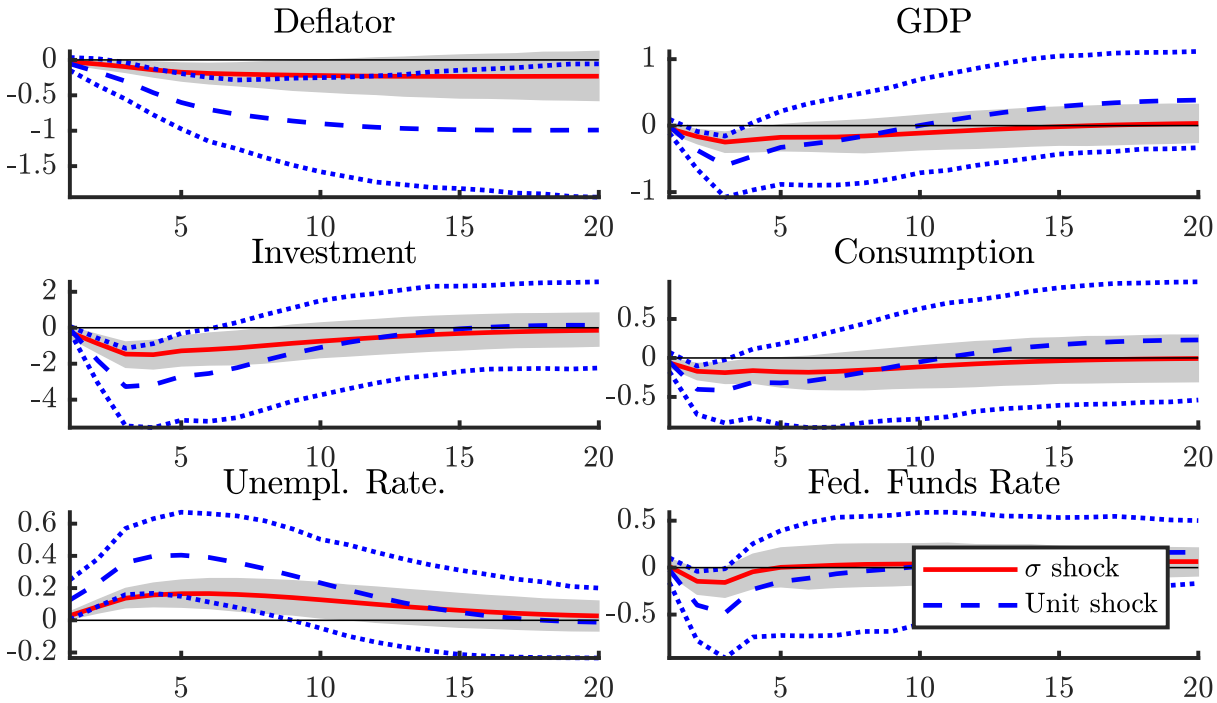
Figure 4.2: Generalized impulse responses to an uncertainty shock



Notes: The figure displays the impact of an uncertainty shock to selected variables in normal times and in times of high uncertainty. The left column refers to a one standard deviation innovation; the right column depicts a unit shock. The responses are generated using a recursive identification scheme with uncertainty ordered second. Gray shaded ares and dotted blue lines refer to 68% error bands. The macro uncertainty index enters the model standardized.

the responses. From a forecasting perspective, this might yield more accurate density forecasts, since the nonlinear model is potentially better at capturing the state-dependent disturbances.

Figure 4.3: Differences in generalized impulse responses between normal times and high uncertainty regime



Notes: The figure displays the differences in the responses between high uncertainty and normal times for a one standard deviation shock (red line) and a unit shock (blue line). Gray shaded ares and dotted blue lines refer to 68% error bands.

4.5 Forecast evaluation

In this section, the forecasts of the competing models are evaluated. I first discuss the measures used for the evaluation of both point forecasts and the predictive densities. Subsequently, the forecast performance is highlighted. In the following, j, i, and h denote the model, variable, and forecast horizon, respectively, for the forecast sample $t = 1, \ldots, N$.

4.5.1 Forecast metrics

A first impression of the models' forecast performance is provided by the mean forecast error (MFE), which indicates the average deviation of the forecast from the realization. Thus, positive

(negative) MFEs show that the model on average overestimates (underestimates) the true value. The MFE is defined as follows:

$$\text{MFE}_{i,j}^{h} = \frac{1}{N} \sum_{T=T_0+h} (\bar{y}_{i,T|T-h}^{j} - y_{i,T}), \tag{4.8}$$

where $\bar{y}_{i,T|T-h}^{j}$ and $y_{i,T}$ denote the mean of the model's predictive density and the corresponding realization. I further evaluate point forecasts in terms of the root mean squared forecast error (RMSFE):

$$\text{RMSFE}_{i,j}^{h} = \sqrt{\frac{1}{N} \sum (\bar{y}_{i,T|T-h} - y_{i,T})^2}. \tag{4.9}$$

While the MFE can be interpreted on its own, the RMSFE is only useful in assessing the accuracy of a model compared to that of other models. Therefore, I report the RMSFEs relative to a benchmark model ($\text{RMSFE}_{i,B}^{h}$):

$$\text{relative RMSFE}_{i,j}^{h} = \text{RMSFE}_{i,j}^{h} / \text{RMSFE}_{i,B}^{h}. \tag{4.10}$$

To test whether the forecasts are significantly different from each other, I apply the test provided by Diebold and Mariano (1995) adjusted for the small-sample correction of Harvey, Leybourne, and Newbold (1997).

To take into account the uncertainty around the point estimate, additionally I evaluate the predictive densities. Specifically, I apply the average log predictive score, which goes back to Good (1952) and has become a commonly accepted tool for comparing the forecast performance of different models (see Geweke and Amisano, 2010; Clark, 2012, among others). It is defined as the logarithm of the predictive density evaluated at the realized value:

$$\text{LS}_{i,j}^{h} = \frac{1}{N} \sum_{\log} p_t(y_{i,t+h}|j). \tag{4.11}$$

The predictive density, $p(y_{t+h}|j)$, is obtained by applying a kernel estimator on the forecast sample.[11] Hence, if the competing model has a lower log score than the benchmark, its forecasts are closer to the realizations with a higher probability. As for the RMSFE, the log scores are not informative on their own, which is why I report them relative to the benchmark model ($LS_{B,i}^{h}$):

$$\text{relative LS}_{i,j}^{h} = LS_{i,j}^{h} - LS_{i,B}^{h}. \tag{4.12}$$

[11] Since the predictive density is not necessarily Gaussian, I do not resort to the frequently used approximation of Adolfson, Lindé, and Villani (2007).

Furthermore, I evaluate the predictive densities in terms of the (average) continuous ranked probability score (CRPS) introduced by Matheson and Winkler (1976). As argued by, for instance, Gneiting and Raftery (2007), the CRPS has two advantages compared to the log scores. First, it can be reported in the same units as the respective variable and therefore facilitates a direct comparison of deterministic and probabilistic forecasts. Second, in contrast to log scores, CRPSs are both less sensitive to extreme outcomes and better able to assess forecasts close but not equal to the realization. I follow Gneiting and Ranjan (2011) and express the CRPS in terms of a score function:

$$S(p, y_{i,t+h}, \nu(\alpha)) = \int_{-0}^{1} \mathrm{QS}_{\alpha}(P(\alpha)^{-1}, y_{i,t+h})\nu(\alpha)d\alpha, \tag{4.13}$$

where $\mathrm{QS}_{\alpha}(P_t(\alpha)^{-1}, y_{i,t+h}) = 2(I\{y_{i,t+h} < P_t(\alpha)^{-1}\} - \alpha)(P_t(\alpha)^{-1} - y_{i,t+h})$ is the quantile score for forecast quantile $P_t(\alpha)^{-1}$ at level $0 < \alpha < 1$. $I\{y_{i,t+h} < P_t(\alpha)^{-1}\}$ is an indicator function taking the value one if $y_{i,t+h} < P_t(\alpha)^{-1}$ and zero otherwise. $\nu(\alpha)$ is a weighting function. Applying a uniform weighting scheme, yields the average CRPS:

$$\mathrm{CRPS}_{i,j}^{h} = S(p_t, y_{i,t+h}, 1). \tag{4.14}$$

To compute this expression, $P(\cdot)$ is approximated by the empirical distribution of forecasts and the integral is calculated numerically.[12] According to this definition, lower CRPSs imply that the predictive density is more closely distributed to the actual density. As for the log scores, I report the CRPS in terms of the average across all evaluation periods and relative to the benchmark model. To provide a rough gauge on whether these scores are significantly different from the benchmark, I follow D'Agostino, Gambetti, and Giannone (2013) by regressing the differences between the scores of each model and the benchmark on a constant. A t-test with Newey-West standard errors on the constant indicates whether these average differences are significantly different from zero.

Finally, I compute probability integral transforms (PITs) developed by Rosenblatt (1952) and popularized in economics by Diebold, Gunther, and Tay (1998). The PIT is defined as the CDF corresponding to the predictive density evaluated at the respective realizations:

$$z_{t+h}^{i} = \int_{-\infty}^{y_{t+h}^{i}} p_t(u)du \quad \text{for } t = 1, \ldots, N. \tag{4.15}$$

Thus, with regard to the respective predictive density, the PIT denotes the probability that a forecast is less or equal to the realization. For example, a realization that corresponds to the 10^{th}

[12] As shown by Smith and Vahey (2015), this procedure is more accurate than expressing the CRPS as the difference of two expectations and the approximation of these expecations using Monte Carlo draws (see Gneiting and Raftery, 2007; Panagiotelis and Smith, 2008).

percentile receives a PIT of 0.1. Hence, if the predictive densities match the true densities, the PITs are uniformly distributed over the unit interval. To assess the accuracy of the predictive density according to the PIT, it is convenient to divide the unit interval into k equally sized bins and count the number of PITs in each bin. If the predictive density equals the actual density, each bin contains N/k observations. In the following, I set $k = 10$, implying that each bin accounts for 10% of the probability mass. Moreover, I follow Rossi and Sekhposyan (2014) and compute 90% confidence bands by using a normal approximation to gauge significant deviation from uniformity.

4.5.2 Point forecasts

Table 4.2 summarizes the results of the forecast evaluation based on MFEs and RMSFEs. The dimension for measures is percentage points. While the models provide forecasts for each variable in the dataset, for the sake of brevity, I present results only for the variables depicted in Section 4.4, namely, inflation (measured in terms of the GDP deflator growth), GDP growth, consumption growth, investment growth, the unemployment rate, and the federal funds rate.[13] Let us begin by analyzing the results for MFE presented in the left panel of Table 4.2. The table shows that the benchmark VAR on average and in most cases overestimates the realization. Inflation for the next quarter, for instance, is overpredicted by 0.14 annualized percentage points. Adding the uncertainty index to the otherwise standard VAR (VAR^{U}) tends to increase this bias except for the unemployment rate and for investment growth. In the latter case, the MFE is on average over all horizons about one percentage point smaller. The MFEs of the threshold VAR (T-VAR) are distinct from the former ones. First, compared to the linear models, the MFEs from the T-VAR are in most cases larger. Only for certain variables and horizons (for example, output growth at $h=3$) reductions are detectable. Thus, identifying uncertainty regimes seems to be less fruitful for generating well-calibrated predictive means. Second, while the linear models consistently underpredict unemployment and overpredict the federal funds rate, the T-VAR overpredicts unemployment and underpredicts the federal funds rate. The latter result stems from the fact that the T-VAR predicts federal funds rate values below zero even though the federal funds rate is fixed at its lower bound.[14] Overall, the evaluation of the MFEs, thus, provides only little evidence in favor of both the VAR^{U} and the T-VAR. In fact, the benchmark model provides very competitive MFEs and in some cases outperforms the remaining models.

The right panel of Table 4.2 depicts the results for the RMSFE. With respect to the benchmark model (linear VAR), the RMSFEs are reported in absolute terms, while the remaining specifications are reported as ratios relative to the benchmark model, i.e. a figure below unity indicates that the model outperforms the benchmark specification. The differences between the

[13] Results for the remaining variables are available upon request.

[14] Berg (2017b) studies how this issue affects the forecast performance of linear VARs.

VAR and the VAR^U are again very small and in most cases insignificant, suggesting that the uncertainty index has on average only marginal impact on the forecast performance in a linear setting. Only for the federal funds rate, the VAR^U provides significantly smaller RMSFEs. The results for the threshold VAR are mixed. In most cases, the latter is outperformed by its linear counterparts, implying that identifying uncertainty regimes is not beneficial with regard to point forecasting. The worst relative performance is obtained for inflation forecasts. Moreover, neither for GDP growth, nor for investment or consumption growth, the T-VAR delivers a reduction in RMSFEs. While for the former indicators regime-dependency apparently does not pay off, unemployment and interest rate forecasts benefit significantly. Regarding the federal funds rate at the one and two-quarter ahead horizons, the T-VAR's forecasts are on average 14% and 8% more precise, respectively, while with regard to the unemployment rate forecast, accuracy increased by 6% and 7% for these horizons. These results are particularly appealing since labor market variables possess an especially strong regime dependency with regard to uncertainty (see Figures 4.2 and 4.3). In addition, these findings underpin the results of Barnett et al. (2014) and Alessandri and Mumtaz (2017). While the former demonstrates that regime-dependent VARs are inferior to linear VARs and VARs with time-varying parameters with regard to GDP growth and inflation, the latter provides evidence that financial variables particularly benefit from regime dependency. Thus, it is suggested that for activity variables there is, if any, only gradual structural change, which cannot be covered by a threshold VAR, while for labor market and financial variables the structural shift is more abrupt and thus can be captured by the T-VAR.

Figure 4.4 explores the models' forecast performance over time. To this end, I calculate four-quarter moving averages of the MFE (upper panel) and relative RMSFE (lower panel) for one-quarter-ahead forecasts of the unemployment rate (left column) and federal funds rate (right column). Evidently, the degree of predictability varies substantially over time. Regarding unemployment rate forecasts, the VAR^U and the T-VAR work particularly well during the Great Recession and the subsequent recovery when uncertainty was high. In the remaining periods, when uncertainty was rather low, the forecast performance is very similar (VAR^U) or even worse (T-VAR) compared to the linear VAR, suggesting that uncertainty is especially relevant when it is high. A similar pattern arises for the federal funds rate. The largest gains in forecast accuracy are obtained during 2008–2012 when uncertainty was high. However, in contrast to the unemployment rate, federal funds rate forecasts are also more precise from 2013–2016, while the short hike of the federal funds rate in 2012 is captured best by the linear VAR; both the VAR^U and the T-VAR strongly overestimate the actual increase. Overall, the results suggest that including information on economic uncertainty can improve point forecast accuracy for some variables and for short horizons, with the largest gains during episodes of high uncertainty.

Table 4.2: MFEs and RMSEs

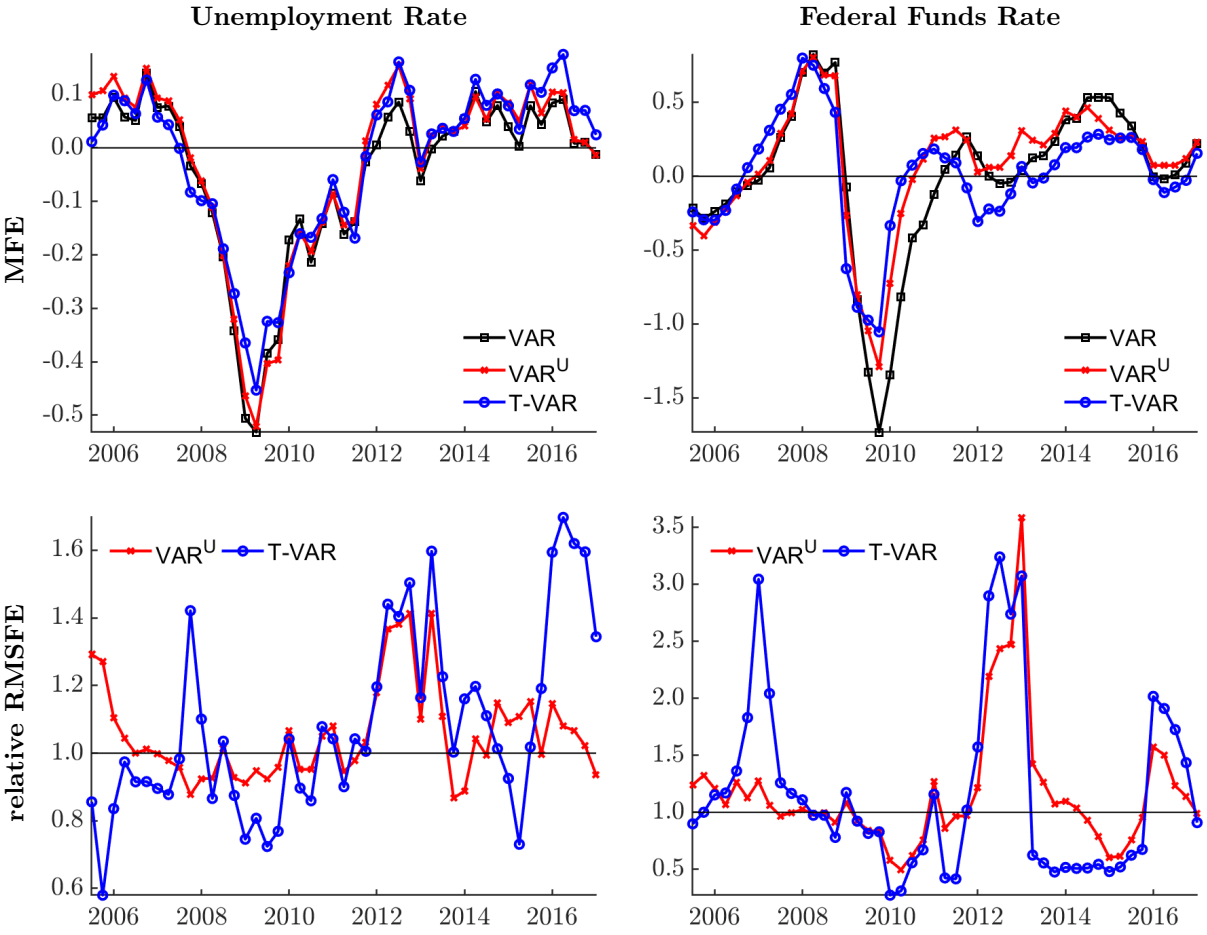
	MFE				**RMSFE**			
Specification	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
	Inflation							
VAR	0.14	0.17	0.18	0.18	0.95	1.02	1.09	1.15
VAR^{U}	0.16	0.21	0.25	0.27	1.10	1.06	1.02$\bullet$	1.04
T-VAR	0.35	-0.34	-0.84$\blacksquare$	1.20$\blacksquare$	1.19$\blacksquare$	1.08	1.28$\blacksquare$	1.45$\blacksquare$
	Output growth							
VAR	0.97$\blacksquare$	1.11$\blacksquare$	1.15$\blacksquare$	1.00$\blacksquare$	2.35	2.79	2.91	2.80
VAR^{U}	1.14$\blacksquare$	1.16$\blacksquare$	1.25$\blacksquare$	1.15$\blacksquare$	1.06$\bullet$	1.03	1.01	0.99
T-VAR	1.39$\bullet$	0.77	-0.05	0.79	1.17$\blacksquare$	1.04	0.92	0.97
	Investment growth							
VAR	4.30$\blacksquare$	5.44$\blacksquare$	5.65$\blacksquare$	4.48$\blacksquare$	10.27	15.01	15.52	13.76
VAR^{U}	3.36$\blacksquare$	4.00$\blacksquare$	4.64$\blacksquare$	3.83$\blacksquare$	0.96	0.99	1.03	0.97
T-VAR	4.88$\blacksquare$	0.55	-5.11$\blacksquare$	-4.42$\blacksquare$	1.18$\blacksquare$	0.92	1.00	1.06
	Consumption growth							
VAR	0.77$\blacksquare$	0.72$\blacksquare$	0.73$\blacksquare$	0.87$\blacksquare$	2.14	2.13	2.08	2.34
VAR^{U}	0.77$\blacksquare$	0.74$\blacksquare$	0.82$\blacksquare$	0.95$\blacksquare$	1.01	0.98	1.03	0.97
T-VAR	1.11$\blacksquare$	1.00$\blacksquare$	0.86	2.07$\blacksquare$	1.12$\blacksquare$	1.19$\blacksquare$	1.11	1.30$\blacksquare$
	Unemployment rate							
VAR	-0.04	-0.09	-0.15	-0.20	0.22	0.47	0.73	1.00
VAR^{U}	-0.03	-0.06	-0.11	-0.16	1.01	0.97	0.97	0.96
T-VAR	-0.03	0.08	0.29$\blacksquare$	0.50$\blacksquare$	0.94$\bullet$	0.93$\bullet$	1.06	1.13$\blacksquare$
	Federal funds rate							
VAR	0.01	0.09	0.17	0.27	0.65	1.16	1.42	1.58
VAR^{U}	0.08	0.19	0.29$\bullet$	0.40$\blacksquare$	0.90$\bullet$	0.98	1.06	1.14$\blacksquare$
T-VAR	0.04	-0.24	-0.51$\bullet$	-0.71$\blacksquare$	0.86$\bullet$	0.92$\bullet$	1.04	1.16$\blacksquare$

Notes: VAR and VAR^{U} denote the linear VAR both without macro uncertainty and including macro uncertainty, respectively. T-VAR refers to the threshold VARs. RMSFEs are reported in absolute terms for the benchmark model (linear VAR) and in ratios relative to the benchmark (VAR) for the remaining specifications. Ratios below unity indicate that the model outperforms the benchmark. $\blacksquare$, $\bullet$, and $\bullet$ denote that the errors are significantly different from zero (MFE) or the benchmark (RMSFE) on the 5%, 10%, and 15% level, respectively. Sample: 1960Q3–2017Q4.

4.5.3 Density forecasts

Subsequently, we evaluate the models' forecasts with respect to the predictive densities. Thus, apart from the predictive mean evaluated above, the variances have to be precisely estimated as

Figure 4.4: Forecast performance over time – point forecasts



Notes: The figure displays mean forecast errors (upper panel) and relative root mean squared forecast errors (bottom panel) computed as a four-quarter moving average over the forecast sample for unemployment and federal funds rate forecasts.

well to ensure an accurate predictive density. Table 4.3 sets out the results for the CRPS and the LS. First, we consider the results for the LS, which are reported in levels for the benchmark (linear VAR) and in differences for the remaining models. Positive differences indicate that the respective model outperforms the benchmark. With regard to the linear models, the LS provide a pattern similar to that in the previous section. Again, the differences between both models are rather small, indicating that the marginal impact of the uncertainty index in a linear setting is on average almost negligible. However, in some cases, already the linear VAR using additional information on economic uncertainty provides significantly better (lower) LS. Turning to the T-VAR reveals that for medium- to long-term forecasts, controlling for regime-dependency with respect to uncertainty leads to considerably less accurate predictive densities. Regarding short-term forecasts, though, the T-VAR provides, for most variables, remarkably better log scores,

with the largest improvements obtained for the activity variables. For instance, the LS for output growth at $h = 1$ is 19% lower than the benchmark's score. Hence, while the T-VAR is inferior in generating precise point forecasts for the activity variables, it is superior in computing the complete predictive distribution of these indicators and thus is better suited for describing the uncertainty around the point estimate.

In total, the CRPS underpin the findings of the LS. However, there are noteworthy differences in regard to the unemployment rate. While according to the LS the predictive distributions of the T-VAR are virtually identical to the ones of the benchmark, according the CRPS, the T-VAR provides significantly more accurate densities. For instance, the one-quarter-ahead CRPS for the unemployment rate is 16% lower than the benchmark's CRPS while the average log score is virtually identical. The latter suggests that the log scores regarding the unemployment forecasts are partly distorted by outliers. Overall, the evaluation of both the LS and CRPS underpins findings of previous studies demonstrating that nonlinearity is particularly useful in calibrating accurate predictive densities (see Chiu, Mumtaz, and Pintér, 2017; Huber, 2016; Groen, Paap, and Ravazzolo, 2013, among others). However, while the former studies mainly focus on forecasting output, inflation, and interest rates, this chapter shows that unemployment rate forecasts also benefit significantly. Figure 4.5 presents evidence on time-varying predictability. Similar to Figure 4.4, the T-VAR provides more accurate densities during the Great Recession and the subsequent recovery.

Between 2011 and the end of 2013, the T-VAR's entire forecast distribution is stretched by a few forecasts far away from the realizations, which leads to low log scores. Since the CRPS is better able to reward the observations close to the realization and is more robust to outliers, according to the CRPS, the T-VAR provides more precise densities even for this period and thus for almost the entire evaluation period. For the federal funds rate, the picture is more clear-cut. The LS indicate that the T-VAR is superior at the beginning of the Great Recession, but the CRPS display more accurate predictive densities for almost the entire sample. As for the unemployment rate, the T-VAR provides the best relative forecast performance during the Great Recession and the subsequent recovery when economic uncertainty was very high. In total, Figure 4.5 provides evidence for important changes in the predictive ability of the models.

Finally, I compute PITs to gauge the calibration of the predictive densities. Figure 4.6 facilitates a graphical inspection of the PITs and shows that the predictive densities look similar for the different models.[15] As I computed 50 forecasts for each horizon, each bin in Figure 4.6 should contain five observations (depicted by the solid black line) to ensure uniformity. Thus, the closer the histograms are to the solid black line, the more accurate are the predictive densities.

[15] Alternatively, one can also pursue more formal approaches to evaluate PITs; see, for instance, Rossi and Sekhposyan (2014). Since, the visual inspection offers straightforward conclusions, I do not resort to these methods.

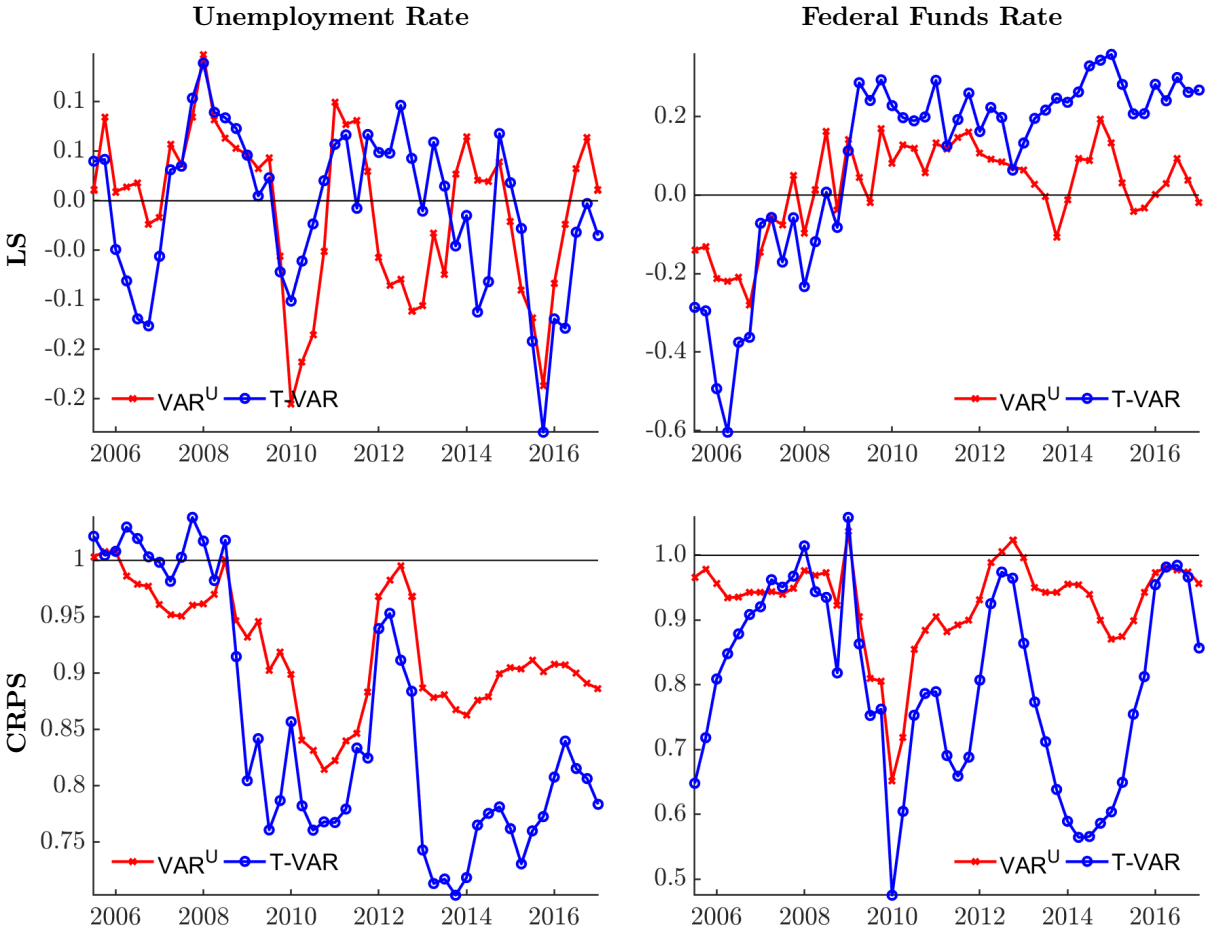
Table 4.3: CRPS and LS

	CRPS				LS			
Specification	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
	Inflation							
VAR	0.50	1.91	3.62	5.53	-2.17	-3.20	-3.75	-4.12
VAR^{U}	0.89	0.97	1.02	1.09●	0.04	0.10●	-0.04	-0.05
T-VAR	1.60●	3.20●	3.73●	4.24●	-0.36●	-1.23●	-1.29●	-1.31●
	Output growth							
VAR	3.55	10.40	5.73	5.94	-3.61	-4.72	-4.28	-4.30
VAR^{U}	1.00	0.99•	0.98•	0.99•	0.01	0.07●	0.01	-0.01
T-VAR	0.92●	1.59●	2.26●	3.03●	0.19●	-0.35●	-0.73●	-1.00●
	Investment growth							
VAR	16.64	54.25	32.74	34.62	-5.38	-6.53	-6.11	-6.07
VAR^{U}	0.97•	0.97•	0.98•	0.98•	-0.02	0.02	0.07	0.07
T-VAR	1.10●	1.79●	2.46●	3.12●	-0.04	-0.55●	-0.83●	-1.02●
	Consumption growth							
VAR	3.26	7.77	12.38	15.79	-3.56	-4.43	-4.86	-5.10
VAR^{U}	0.96●	0.95●	0.97●	0.96●	0.05●	0.04	0.03	0.01
T-VAR	0.87●	1.61●	2.14●	2.75●	0.23●	-0.41●	-0.67●	-0.96●
	Unemployment rate							
VAR	0.15	0.47	0.76	1.10	-1.99	-2.29	-2.49	-2.67
VAR^{U}	0.89●	0.92●	0.92●	0.91●	-0.01	0.05•	0.07•	0.07●
T-VAR	0.84●	1.00	0.94	0.89●	-0.00	-0.05	0.02	0.11●
	Federal funds rate							
VAR	0.36	1.09	1.77	2.42	-2.14	-2.73	-3.12	-3.39
VAR^{U}	0.91●	1.04	1.05	1.07	0.02	0.01	0.02	-0.03
T-VAR	0.79●	0.84●	0.93●	0.99	0.10●	0.16●	0.14●	0.13●

Notes: VAR and VAR^{U} denote the linear VAR without macro uncertainty and including macro uncertainty, respectively. T-VAR refers to the threshold VARs. The scores are reported in absolute terms for the benchmark model (linear VAR). For the remaining models LSs are expressed in differences to the benchmark and CRPSs in ratios to the benchmark model. A positive difference and a ratio below unity indicate the model outperforms the benchmark. ●, •, and • denote significance on the 5%, 10%, and 15% level, respectively, according to a t-test on the average difference in scores relative to the benchmark model with Newey-West standard errors. Sample: 1960Q3–2017Q4.

In case of inflation, output, investment, and consumption, the PITs appear hump-shaped, with significant departures from uniformity. In fact, the models assign too much probability to the center of the distribution with too many PIT-values around 0.5. The latter indicates that

Figure 4.5: Forecast performance over time – density forecasts



Notes: The figure displays log scores (upper panel) and continuous ranked probability scores (bottom panel) computed as a four-quarter moving average over the forecast sample for unemployment and federal funds rate forecasts.

the kurtosis of predictive densities at each horizon and recursion is higher than the kurtosis of true density, which implies that the models overestimate the actual uncertainty around the point estimate. This pattern is frequently found in the VAR forecasting literature—see, for example, Rossi and Sekhposyan (2014), Bekiros and Paccagnini (2015) or Gerard and Nimark (2008)—and can be caused by a too dense parametrization of the model.[16] With regard to one-quarter-ahead forecasts (blue bars), the T-VAR mitigates this issue by generating more forecasts that correspond to the lower percentiles of the actual distribution and thus provides a better description of the data. At higher horizons, however, the densities are again too wide. Regarding unemployment rate forecasts, the PITs of each model are closer to uniformity for $h=1$ and $h=2$;

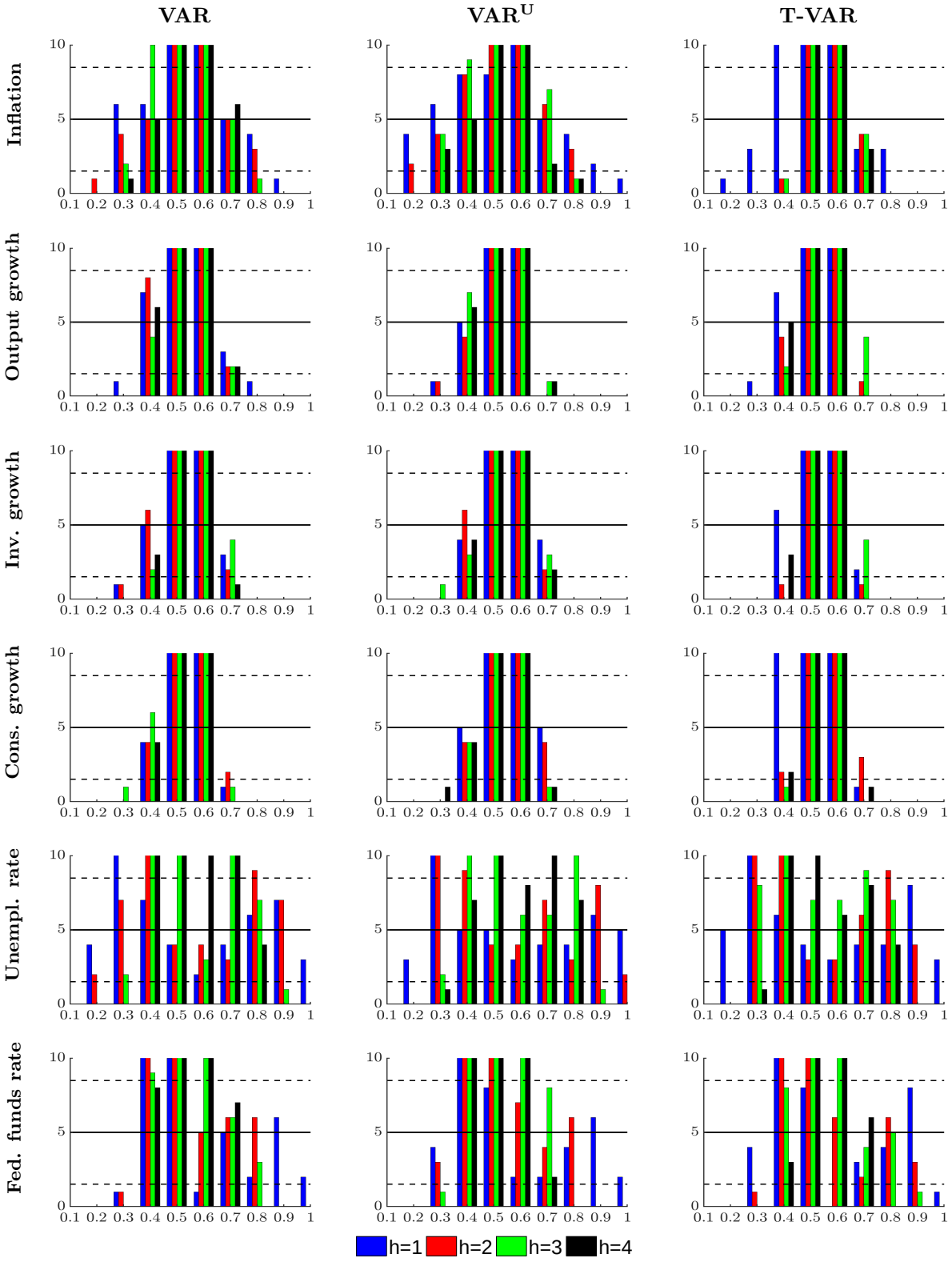
[16] Wolters (2015) demonstrates that this problem also applies to estimated DSGE models.

both the lower and the upper percentiles of the actual distribution are captured by the models. At the remaining horizons, the models again overestimate the actual uncertainty. The PITs for the interest rate forecasts appear to be right skewed, and thus missing the left tail of the actual distribution. The latter stems from the phase of extrodinary low interest rates at the end of the sample, which are barley captured by the models. Only the VAR^{U} is able to generate forecasts corresponding to the lower percentiles. Jointly with the results from Table 4.3, the evaluation of the PITs suggests that estimating regime-dependent covariance matrices with respect to the prevailing level of uncertainty helps calibrating accurate predictive densities.

4.6 Conclusion

Evidence from studies on the effects of uncertainty shocks suggests that uncertainty impacts real economy variables and that these impacts depend on the prevailing level of uncertainty. This chapter answers the questions of whether these insights can be used to achieve more accurate forecasts from VAR models and whether one has to account for nonlinearities to achieve this goal. I compared the forecast performance of different Bayesian VAR specifications. The analysis provides four main results. First, in a linear setting, point forecast accuracy cannot be significantly improved by considering information from the macroeconomic uncertainty index. Second, accounting for regime-specific model dynamics depending on the level of uncertainty improves the point forecast accuracy for unemployment rate and interest rate forecasts, while the accuracy for real activity variables deteriorates. Third, predictive densities benefit significantly from the macroeconomic uncertainty index both in a linear and nonlinear setting. However, the nonlinear model outperforms the linear models, especially at short horizons. The largest gains are obtained for unemployment rate forecasts. Moreover, and in contrast to the point forecasts, the threshold VAR also provides strong improvements for the predictive densities of the real activity variables. Finally, I document substantial variation in the models' predictive ability. In particular, during episodes of high uncertainty, the T-VAR provides strong gains in forecast accuracy with respect to the predictive densities. Thus, it can serve as a complement to existing approaches in arriving at a better picture of the actual uncertainty surrounding the point estimate in times of high uncertainty and especially for unemployment forecasts.

Figure 4.6: Probability integral transform (PITs)



Notes: The figure displays the cdf of the probability integral transforms (PITs). Solid and dashed black lines denote uniformity and 90% confidence bands, respectively.

References

Abel, A. B. (1973). Optimal Investment Under Uncertainty. *American Economic Review 73*(1), 228–233.

Adolfson, M., J. Lindé, and M. Villani (2007). Forecasting Performance of an Open Economy DSGE Model. *Econometrics Review 26*(2-4), 289–328.

Alessandri, P. and H. Mumtaz (2017). Financial Conditions and Density Forecasts for US Output and Inflation. *Review of Economic Dynamics 24*, 66–78.

Alessandri, P. and H. Mumtaz (2019). Financial Regimes and Uncertainty Shocks. *Journal of Monetary Economics 101*, 31–46.

Auerbach, A. J. and Y. Gorodnichenko (2012). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy 4*(2), 1–27.

Bachmann, R. and C. Bayer (2013). Wait-and-See Business Cycles? *Journal of Monetary Economics 60*(6), 704–719.

Bachmann, R., S. Elstner, and E. R. Sims (2013). Uncertainty and Economic Activity: Evidence from Business Survey Data. *American Economic Journal: Macroeconomics 5*(2), 217–249.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

Balcilar, M., R. Gupta, and M. Segnon (2016). The Role of Economic Policy Uncertainty in Predicting US Recessions: A Mixed-Frequency Markov-Switching Vector Autoregressive Approach. *Economics: The Open-Access, Open-Assessment E-Journal 10*(2016–27), 1–20.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian Vector Auto Regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Barnett, A., H. Mumtaz, and K. Theodoridis (2014). Forecasting UK GDP Growth and Inflation Under Structural Change. A Comparison of Models with Time-Varying Parameters. *International Journal of Forecasting 30*(1), 129–143.

Basu, S. and B. Bundick (2017). Uncertainty Shocks in a Model of Effective Demand. *Econometrica 85*(3), 937–958.

Bekiros, S., R. Gupta, and A. Paccagnini (2015). Oil Price Forecastability and Economic Uncertainty. *Economics Letters 132*, 125–128.

Bekiros, S. and A. Paccagnini (2015). Estimating Point and Density Forecasts for the US Economy with a Factor-Augmented Vector Autoregressive DSGE Model. *Studies in Nonlinear Dynamics & Econometrics 19*(2), 107–136.

Berg, T. O. (2016). Multivariate Forecasting with BVARs and DSGE Models. *Journal of Forecasting 35*, 718–740.

Berg, T. O. (2017a). Business Uncertainty and the Effectivness of Fiscal Policy in Germany. *Macroeconomic Dynamics* (forthcoming).

Berg, T. O. (2017b). Forecast Accuracy of a BVAR under Alternative Specifications of the Zero Lower Bound. *Studies in Nonlinear Dynamics & Econometrics 21*(2), 1081–1826.

Berg, T. O. and S. R. Henzel (2015). Point and Fensity Forecasts for the Euro Area Using Bayesian VARs. *International Journal of Forecasting 31*(4), 1067–1095.

Bernanke, B. S. (1983). Irreversibility, Uncertainty, and Cyclical Investment. *The Quarterly Journal of Economics 98*(1), 85–106.

Bijsterbosch, M. and P. Guérin (2013). Characterizing Very High Uncertainty Episodes. *Economics Letters 121*(2), 239–243.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica 77*(3), 623–685.

Born, B. and J. Pfeifer (2014). Policy Risk and the Business Cycle. *Journal of Monetary Economics 68*, 68–85.

Born, B. and J. Pfeifer (2017). Uncertainty-Driven Business Cycles: Assessing the Markup Channel. CESifo Working Paper 6303, Center for Economic Studies and Ifo Institute (CESifo), Munich.

Caballero, R. and R. S. Pindyck (1996). Uncertainty, Investment, and Industry Evolution. *International Economic Review 37*(3), 641–662.

Caggiano, G., E. Castelnuovo, and J. M. Figueres (2017). Economic Policy Uncertainty and Unemployment in the United States: A Nonlinear Approach. *Economics Letters 151*, 31–34.

Caggiano, G., E. Castelnuovo, and N. Groshenny (2014). Uncertainty Shocks and Unemployment Dynamics in U.S. Recessions. *Journal of Monetary Economics 67*, 78–92.

Caldara, D., C. Fuentes-Albero, S. Gilchrist, and E. Zakrajšek (2016). The Macroeconomic Impact of Financial and Uncertainty Shocks. *European Economic Review 88*, 185–207.

Carriero, A., T. E. Clark, and M. Marcellino (2013). Bayesian VARs: Specification Choices and Forecast Accuracy. *Journal of Applied Econometrics 30*(1), 46–73.

Carriero, A., G. Kapetanios, and M. Marcellino (2009). Forecasting Exchange Rates with a Large Bayesian VAR. *International Journal of Forecasting 25*(2), 400–417.

Carriero, A., H. Mumtaz, K. Theodoridis, and A. Theophilopoulou (2015). The Impact of Uncertainty Shocks under Measurement Error: A Proxy SVAR Approach. *Journal of Money, Credit and Banking 47*(6), 1223–1238.

Castelnuovo, E. and G. Pellegrino (2018). Uncertainty-Dependent Effects of Monetary Policy Shocks: A New-Keynesian Interpretation. *Journal of Economic Dynamics and Control 93*, 277–296.

Chen, C. W. S. and J. C. Lee (1995). Bayesian Inference of Threshold Autoregressive Models. *Journal of Time Series Analysis 16*(5), 483–492.

Chiu, C.-W. J., H. Mumtaz, and G. Pintér (2017). Forecasting with VAR models: Fat Tails and Stochastic Volatility. *International Journal of Forecasting 33*(4), 1124–1143.

Christiano, L., R. Motto, and M. Rostagno (2014). Risk Shocks. *American Economic Review 104*(1), 27–65.

Clark, T. E. (2012). Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility. *Journal of Business & Economic Statistics 29*(3), 327–341.

Clark, T. E. and F. Ravazzolo (2015). Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics 30*(4), 551–575.

Cogley, T. and T. J. Sargent (2001). Evolving Post-World War II US inflation Dynamics. In B. S. Bernanke and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2001*, Volume 16, 331–373. National Bureau of Economic Research, Inc.

D'Agostino, A., L. Gambetti, and D. Giannone (2013). Macroeconomic Forecasting and Structural Change. *Journal of Applied Econometrics 28*(1), 82–101.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review 39*(4), 863–883.

Diebold, F. X. and R. S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and Conditional Projection Using Realistic Prior Distributions. *Econometric Reviews 3*(1), 1–100.

Fernández-Villaverde, J., P. Guerrón-Quintana, K. Kuester, and J. Rubio-Ramírez (2015). Fiscal Volatility Shocks and Economic Activity. *American Economic Review 105*(11), 3352–3384.

Ferrara, L. and P. Guérin (2018). What are the Macroeconomic Effects of High-Frequency Uncertainty Shocks? *Journal of Applied Econometrics 33*(5), 662–679.

Foerster, A. T. (2014). The Asymmetric Effects of Uncertainty. *Economic Review-Federal Reserve Bank of Kansas City Q III*, 5–26.

Gerard, H. and K. Nimark (2008). Combining Multivariate Density Forecasts Using Predictive Criteria. Economics Working Papers 1117, Department of Economics and Business, Universitat Pompeu Fabra.

Geweke, J. and G. Amisano (2010). Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns. *International Journal of Forecasting 26*(2), 216–230.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior Selection for Vector Autoregressions. *Review of Economics and Statistics 97*(2), 436–451.

Gilchrist, S., J. W. Sim, and E. Zakrajšek (2014). Uncertainty, Financial Frictions, and Investment Dynamics. NBER Working Paper 20038, National Bureau of Economic Research, Inc.

Gneiting, T. and A. E. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Gneiting, T. and R. Ranjan (2011). Comparing Density Forecasts Using Threshold and Quantile Weighted Scoring Rules. *Journal of Business & Economic Statistics 29*(3), 411–422.

Good, I. J. (1952). Rational Decisions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 14*(1), 107–114.

Groen, J. J. J., R. Paap, and F. Ravazzolo (2013). Real-Time Inflation Forecasting in a Changing World. *Journal of Business & Economic Statistics 31*(1), 29–44.

Hartman, R. (1972). The Effects of Price and Cost Uncertainty on Investment. *Journal of Economic Theory 5*(2), 258–266.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting 13*(2), 281–291.

Henzel, S. R. and M. Rengel (2017). Dimensions of Macroeconomic Uncertainty: A Common Factor Analysis. *Economic Inquiry 55*(2), 843–877.

Huber, F. (2016). Density Forecasting Using Bayesian Global Vector Autoregressions with Stochastic Volatility. *International Journal of Forecasting 32*(3), 818–837.

Jones, P. M. and W. Enders (2016). The Asymmetric Effects of Uncertainty on Macroeconomic Activity. *Macroeconomic Dynamics 20*(5), 1219–1246.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring Uncertainty. *The American Economic Review 105*(3), 1177–1216.

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.

Kilian, L. and R. J. Vigfusson (2011). Are the Responses of the U.S. Economy Asymmetric in Energy Price Increases and Decreases? *Quantitative Economics 2*(3), 419–453.

Koop, G. M. (2013). Forecasting with Medium and Large Bayesian VARs. *Journal of Applied Econometrics 28*(2), 177–203.

Koop, G. M., M. H. Pesaran, and S. M. Potter (1996). Impulse Response Analysis in Nonlinear Multivariate Models. *Journal of Econometrics 74*(1), 119–147.

Leduc, S. and Z. Liu (2016). Uncertainty Shocks are Aggregate Demand Shocks. *Journal of Monetary Economics 82*, 20–35.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions – five years of experience. *Journal of Business & Economic Statistics 4*(1), 25–38.

Matheson, J. E. and R. L. Winkler (1976). Scoring Rules for Continuous Probability Distributions. *Management Science 22*(10), 1087–1096.

McCracken, M. W. and S. Ng (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Mumtaz, H. and K. Theodoridis (2015). The International Transmission of Volatility Shocks: An Empirical Analysis. *Journal of the European Economic Association 13*(3), 512–533.

Mumtaz, H. and K. Theodoridis (2018). The Changing Transmission of Uncertainty Shocks in the U.S. *Journal of Business & Economic Statistics 36*(2), 239–252.

Oi, W. Y. (1961). The Desirability of Price Instability Under Perfect Competition. *Econometrica 29*(1), 58–64.

Panagiotelis, A. and M. Smith (2008). Bayesian Density Forecasting of Intraday Electricity Prices Using Multivariate Skew t Distributions. *International Journal of Forecasting 24*(4), 710–727.

Pierdzioch, C. and R. Gupta (2017). Uncertainty and Forecasts of U.S. Recessions. Working Papers 201732, University of Pretoria, Department of Economics.

Popescu, A. and F. R. Smets (2010). Uncertainty, Risk-taking, and the Business Cycle in Germany. *Cesifo Economic Studies 56*(4/2010), 596–626.

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *The Annals of Mathematical Statistics 23*(3), 470–472.

Rossi, B. and T. Sekhposyan (2014). Evaluating Predictive Densities of US Output Growth and Inflation in a Large Macroeconomic Data Set. *International Journal of Forecasting 30*(3), 662–682.

Rossi, B. and T. Sekhposyan (2015). Macroeconomic Uncertainty Indices Based on Nowcast and Forecast Error Distributions. *The American Economic Review 105*(5), 650–655.

Segnon, M., R. Gupta, S. Bekiros, and M. E. Wohar (2018). Forecasting US GNP Growth: The Role of Uncertainty. *Journal of Forecasting 37*(5), 541–559.

Sims, C. A. (1993). A Nine-Variable Probabilistic Macroeconomic Forecasting Model. In J. H. Stock and M. W. Watson (Eds.), *Business Cycles, Indicators, and Forecasting*, NBER Book Series Studies in Business Cycles, 179–212. National Bureau of Economic Research, Inc.

Smith, M. S. and S. P. Vahey (2015). Asymmetric Forecast Densities for U.S. Macroeconomic Variables from a Gaussian Copula Model of Cross-Sectional and Serial Dependence. *Journal of Business & Economic Statistics 34*(3), 416–434.

Wolters, M. H. (2015). Evaluating Point and Density Forecasts of DSGE Models. *Journal of Applied Econometrics 30*(1), 74–96.

C Appendix

C.1 Prior implementation

For the prior implementation, I express the VAR(p) in (4.1) in companion form:

$$Y = XB + U, \tag{C.1}$$

with $Y = (y_1, ..., y_T)'$, $X = (X_1, ..., X_T)'$ with $X_t = (y'_{t-1}, ..., y'_{t-p}, 1)'$, $U = (\varepsilon_1, ...\varepsilon_t)'$ and $B = (A_1, ..., A_p, c)'$.

The normal-inverse Wishart prior takes the following form:

$$\Sigma \sim iW(\Psi, \underline{\alpha}) \text{ and } vec(B)|\Sigma \sim N(vec(\underline{B}), \Sigma \otimes \underline{\Omega}), \tag{C.2}$$

where $\underline{B}$, $\underline{\Omega}$, $\underline{\alpha}$, and Ψ are functions of hyperparameters. To implement these prior beliefs, I follow Bańbura et al. (2010) and augment the dataset with dummy observations:

$$Y^{D,1} = \begin{pmatrix} \text{diag}(\delta_1\sigma_1, \ldots, \delta_n\sigma_n)/\lambda_1 \\ 0_{n(p-1)\times n} \\ diag(\sigma_1, \ldots, \sigma_n) \\ 0_{1\times n} \end{pmatrix} X^{D,1} = \begin{pmatrix} J_p\text{diag}(\sigma_1, \ldots, \sigma_n)/\lambda_1 & 0_{np\times 1} \\ 0_{n\times np} & 0_{n\times 1} \\ 0_{1\times np} & \epsilon \end{pmatrix}. \tag{C.3}$$

δ_1 to δ_n denote the prior means of the coefficients on the first lag. δ_i is set to one, implying a random walk prior for non-stationary variables, and set to zero for stationary variables. σ_1 to σ_n are scaling factors, which are set to the standard deviations from univariate autoregressions of the endogenous variables using the same lag length as in the VAR. I impose a flat prior on the intercept terms by setting ε to 1/10000. The hyperparameter λ_1 controls the overall tightness of the prior. Hence, with increasing λ_1 the degree of shrinkage declines.

The "sum-of-coefficients" prior imposes the restriction that the sum of the coefficients of the lags of the dependent variables sum up to unity, whereas the lags of other variables sum up to zero. It is implemented by the following dummy observations:

$$Y^{D,2} = \text{diag}(\delta_1 y_1, \ldots, \delta_n y_n)/\lambda_2 \quad X^{D,2} = ((1_{1\times p})\text{diag}(\delta_1\mu_1, \ldots, \delta_n\mu_n)/\lambda_2 \;\; 0_{n\times 1}), \tag{C.4}$$

where μ_i denotes the sample average of variable i. The degree of shrinkage is determined by the hyperparameter λ_2. The prior becomes less informative for higher values of λ_2.

The "co-persistence" prior allows for possibility of stable cointegration relations among the variables. Sims (1993) proposes to add the following dummy observations to the sample to implement the prior:

$$Y^{D,3} = \text{diag}(\delta_1\mu_1, \ldots, \delta_n\mu_2)\lambda_3 \quad X^{D,3} = ((1_{1\times p})\text{diag}(\delta_1\mu_1, \ldots, \delta_n\mu_2))\lambda_3, \tag{C.5}$$

where λ_3 controls the degree of shrinkage of this prior. If λ_3 approaches zero, the prior becomes more tight. Defining $Y^* = [Y, Y^{D,1}, Y^{D,2}, Y^{D,3}]$, $X^* = [X, X^{D,1}, X^{D,2}, X^{D,3}]$, and $U^* = [U, U^{D,1}, U^{D,2}, U^{D,3}]$ yields the augmented dataset, which is used for inference via:

$$Y^* = X^*B + U^*. \tag{C.6}$$

The posterior expectations are determined by an OLS regression of Y^* on X^*. The posterior takes the form:

$$\Sigma|\lambda, y \sim IW(\tilde{\Sigma}, T + n + 2) \quad vec(B)|\Sigma, \lambda, y \ \sim N(vec(\hat{B}), \Sigma \otimes (X^{*'}X^*)^{-1}), \tag{C.7}$$

where $\hat{B}$ is the matrix of coefficients from the regression of Y^* on X^*, and $\hat{\Sigma}$ is the corresponding covariance matrix. In sampling B, I follow Cogley and Sargent (2001) and discard draws leading to an unstable VAR.

C.2 Determining the degree of shrinkage

The forecast performance of Bayesian VARs tends to be sensitive with respect to the choice of the hyperparameters, which in turn have to be choosen with care. The vector Λ collecting the hyperparameters consists of three elements: the overall tightness of the prior (λ_1), the extent to which the sum of coefficients on the lags of a variable are forced to unity (λ_2), and the extent to which co-persistence restrictions are imposed on the VAR coefficients (λ_3). Following the specifications (C.3), (C.4), and (C.5), the smaller λ_i, the more informative the prior. To get a reasonable degree of shrinkage, I apply the hierachical, fully Bayesian procedure of Giannone et al. (2015).[17] The posterior for such a hierarchical prior is obtained by applying Bayes' law

$$p(\Lambda|y) \propto p(y|\Lambda)p(\Lambda), \tag{C.8}$$

[17] Apart from this procedure, one can also determine the degree of shrinkage based on the in-sample fit compared to a parsimonious VAR (Bańbura et al., 2010), or by maximizing the marginal likelihood at each point in time (Carriero, Clark, and Marcellino, 2013). A comparison of these methods with respect to forecast accuracy is provided by Berg and Henzel (2015).

where $p(\lambda)$ is the prior density of the hyperparameters—the so-called hyperprior. The marginal likelihood of the model $p(y|\Lambda)$ is given by:

$$p(y|\lambda) = \int p(y|\theta, \Lambda)p(\theta|\Lambda)d\theta, \tag{C.9}$$

with θ denoting the vector of model parameters. As shown by Carriero et al. (2013) and Giannone et al. (2015), using conjugate priors results in a closed-form solution for the marginal likelihood:

$$p(Y|\Lambda) = k^{-1} \times |\Psi + (Y - X\underline{B})'(I + X\underline{\Omega}X')^{-1}(Y - X\underline{B})|^{-\frac{\underline{\alpha}+T}{2}}, \tag{C.10}$$

$$\text{with: } k = \pi^{\frac{Tn}{2}} \times |(I + X\underline{\Omega}X')^{-\frac{n}{2}} \times |\Psi|^{-\frac{\underline{\alpha}}{2}} \times \frac{\Gamma(\frac{\underline{\alpha}}{2})}{\Gamma\frac{T+\underline{\alpha}}{2}}, \tag{C.11}$$

where $\underline{B} = (X^{D\prime}X^D)^{-1}X^{D\prime}Y^D$, $\underline{\Omega} = (X^{D\prime}X^D)^{-1}$, and $\Psi = (Y^D - X^D\underline{B})^{-1}(Y^D - X^D\underline{B})$. $\Gamma(\cdot)$ denotes the n-variate gamma distribution and $\underline{\alpha} = n + 2$ in order to ensure existence of the prior mean of Σ (Kadiyala and Karlsson, 1997). The hyperpriors for λ_1, λ_2, λ_3, i.e., the priors for the hyperparameters, reflect the knowledge about the values of λ_i. I follow Giannone et al. (2015) by choosing uninformative priors using Gamma densities with modes equal to 0.2, 1, and 1 and standard deviations equal to 0.4, 1, and 1, respectively. I employ a Metropolis-Hastings step to simulate the distribution. Define D and R as the number of discarded and retained draws, respectively. The algorithm works as follows:

1. At iteration $k = 1$ initialize Λ at the posterior mode, which can be obtained by numerical optimization.[18]

2. Draw a candidate value Λ^* for the hyperparameters from a random walk proposal distribution $\Lambda^* \sim N(\delta^{k-1}, cH^{-1})$, where c is a scaling factor calibrated to ensure an acceptance rate of roughly 20% and H^{-1} is the inverse of the Hessian evaluated at the posterior mode.

3. Set $\Lambda^k = \Lambda^*$ with probability

$$\alpha^k = \min\left\{1, \frac{p(\Lambda^*|y)}{p(\Lambda^{k-1}|y)}\right\}. \tag{C.12}$$

   If $k < D$ redo, otherwise continue.

4. Conditional on Λ^k draw Σ^k and β^k from their posteriors given by (C.7).

[18] I use the Matlab globalsearch class based on the routine fmincon to obtain a global maximum.

5. Generate $\varepsilon^k_{T+1}, \ldots, \varepsilon^k_{T+h}$ from $\varepsilon_t \sim N(0, \Sigma^k)$ and calculate h-step-ahead forecasts recursively

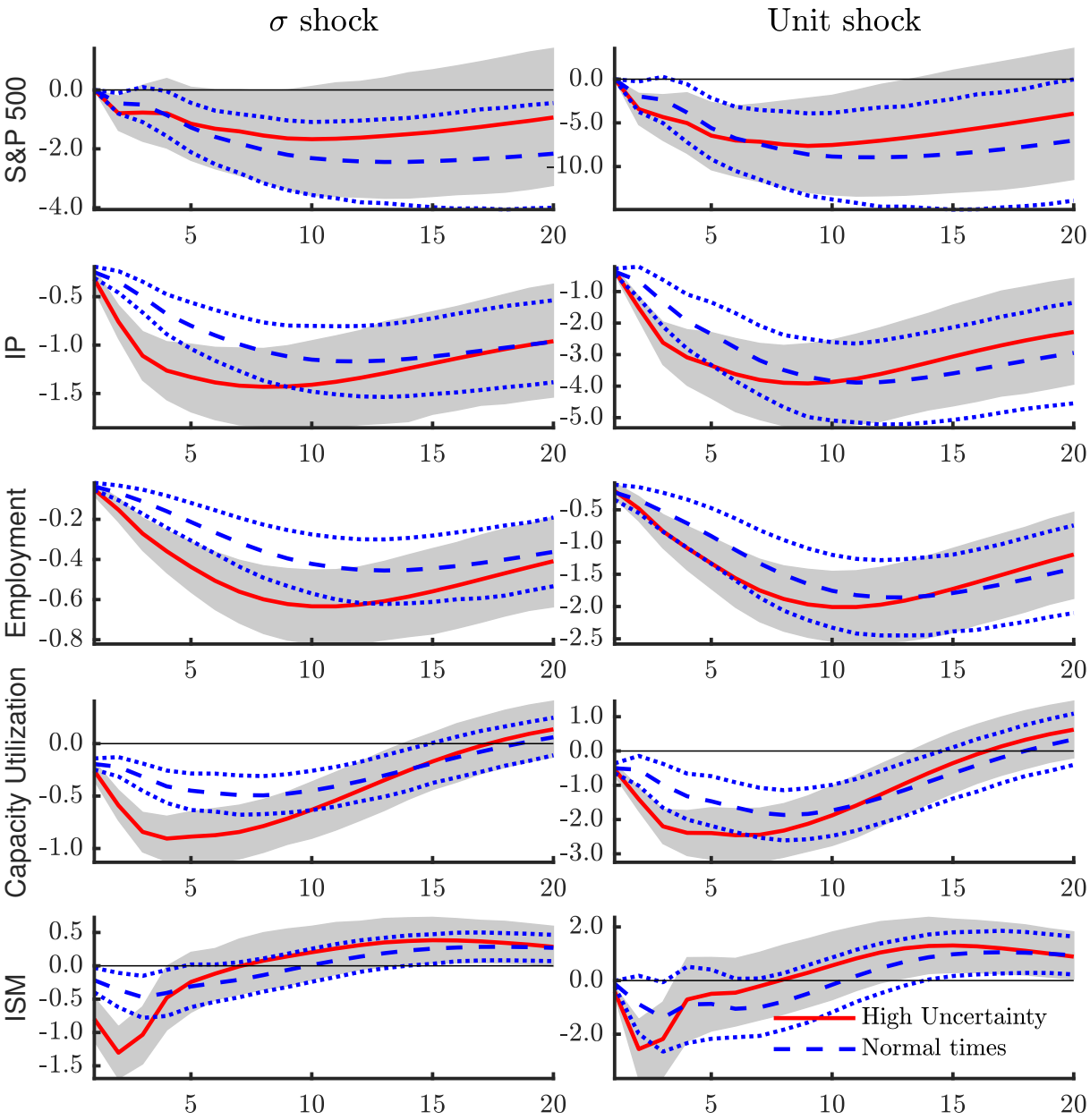
$$\hat{y}^k_{T+h} = c^k + \sum_{i=1}^{h-1} A^k_i \hat{y}^k_{T+h-i} + \sum_{i=h}^{p} A^k_i \hat{y}^k_{T+h-i} + u^k_{T+h}. \tag{C.13}$$

6. Iterate these steps until $j = D + R$.

Note that since Λ is independent of Σ and β, one can draw Λ until the sampler converges and subsequently draw successively Σ and β. Applying this algorithm yields R h-step-ahead forecasts from the joint posterior distribution. From 25000 draws, 5000 are used for inference.

C.3 Generalized impulse responses

Figure 4.7: Regime-dependent impact to an uncertainty shock



Notes: Impact of an uncertainty shock to selected variables in normal times and in times of high uncertainty. Left column refers to a one standard deviation innovation; the right column depicts a unit shock. Responses are generated using a recursive identification scheme with uncertainty ordered second. Gray shaded ares and dashed blue lines refer to 68% error bands. The macro uncertainty index enters the model standardized.

5 | Forecasting using Mixed-Frequency VARs with Time-Varying Parameters

(**with Markus Heinrich**)

Abstract We extend the economic forecasting literature by constructing a mixed-frequency time-varying parameter vector autoregression with stochastic volatility (MF-TVP-SV-VAR). The latter can take structural changes into account and can handle indicators sampled at different frequencies. We conduct a real-time forecast exercise to predict US key macroeconomic variables and compare the predictions of the MF-TVP-SV-VAR with several linear, nonlinear, mixed-frequency, and quarterly-frequency VARs. Our key finding is that the MF-TVP-SV-VAR delivers very accurate forecasts and, on average, outperforms its competitors. In particular, inflation forecasts benefit from this new forecasting approach. Finally, we assess the models' performance during the Great Recession and find that the combination of stochastic volatility, time-varying parameters, and mixed-frequencies generates very precise inflation forecasts.

Keywords: Time-varying parameters, Forecasting, Mixed-frequency models, Bayesian methods
JEL-Codes: C11, C53, C55, E32

5.1 Introduction

Macroeconomists and, in particular, macroeconomic forecasters face two major challenges. First, there are structural changes within an economy. Second, many economic time series are sampled at different frequencies and released with different publication lags. Several studies show that allowing for either structural changes or mixed-frequencies improves forecast performance considerably, and Carriero, Clark, and Marcellino (2013) assess a combination of both specifications in a univariate model. However, a multivariate assessment is still missing. Accordingly, the main contribution of this chapter is to fill this gap and examine the real-time forecast performance of a model incorporating drifting coefficients and indicators observed at different frequencies. Our main finding is that this forecasting approach delivers accurate forecasts for the variables considered and, in most cases, significantly improves upon existing approaches, especially for inflation forecasts.

Our work relates to two strands of the literature. The first strand concerns the importance of modeling structural change in forecasting. To account for both changes in the comovements of variables demonstrated by Cogley and Sargent (2001, 2005) and the decline of business cycle volatility highlighted by Kim and Nelson (1999) and McConnell and Pérez-Quirós (2000), time-varying parameter VARs with stochastic volatility (TVP-SV-VAR) are frequently used.[1]

The second strand deals with the fact that many key macroeconomic variables, for instance, GDP are unavailable at frequencies higher than quarterly, while most key indicators for these variables are published at a higher frequency. As an alternative to models that require all variables to be sampled at the same frequency, in the recent past, mixed-frequency models have attracted interest (for a survey, see Foroni and Marcellino, 2013). This class of models has two advantages. First, the researcher can refrain from any kind of time (dis)aggregation to use, for example, quarterly and monthly variables in one model. Second, by jointly modeling high and low frequency variables, the researcher is better able to track the economic development in real time and assess the usefulness and impact of higher-frequency information on the predictions.

This chapter combines these two strands of literature by using a mixed-frequency TVP-SV-VAR (MF-TVP-SV-VAR) based on Cimadomo and D'Agostino (2016) to forecast in real time four US macroeconomic variables: GDP growth, CPI inflation, the unemployment rate, and a short-term interest rate. As a combination of a MF-VAR and a TVP-SV-VAR, it can cope with indicators sampled at different frequencies and unbalanced datasets. To disentangle the relative impact on the forecast accuracy of the model's mixed-frequency part and the time variation in the model's coefficients, we compare the forecast performance of the MF-TVP-SV-VAR with several other specifications, including constant parameter VARs with and without mixed frequencies

[1] See Galí and Gambetti (2009), Baumeister and Benati (2013), or Koop and Korobilis (2014) for examples of structural analysis using TVP-SV-VARs.

and time-varying VARs without a mixed-frequency part. Furthermore, we evaluate the intra-quarterly inflow of information with regard to the current-quarter estimates (nowcasts).

Estimation of the mixed-frequency part is based on the idea that variables observed at a lower frequency can be expressed as higher-frequency variables with missing observations (Zadrozny, 1988).[2] Adopting this notion, Mariano and Murasawa (2010) derive a state-space representation for VARs with missing observations, called mixed-frequency VAR (MF-VAR). We follow Schorfheide and Song (2015) and apply the MF-VAR approach in a Bayesian framework.

Estimation of the TVP part basically follows Primiceri (2005). However, we treat those hyperparameters that determine the amount of time variation in the parameters as an additional layer and estimate them using Bayesian methods (see Amir-Ahmadi, Matthes, and Wang, 2018).[3] We generate forecasts up to one year ahead and evaluate these predictions in terms of both point and density forecasts. For the point forecast evaluation we resort to root mean squared forecast errors, while the predictive densities are evaluated using scoring rules.

Overall, our results provide evidence that the combination of mixed frequencies, stochastic volatility, and time-varying parameters provides very competitive point and density forecasts for each variable considered. We show that both nowcasts and forecasts benefit significantly from modeling intra-quarterly dynamics. In particular, the novel MF-TVP-SV-VAR generates, on average, more precise inflation forecasts than those generated by any other model considered. Using probability integral transforms, we compare the predictive densities of inflation forecasts generated by both the MF-TVP-SV-VAR and a quarterly TVP-SV-VAR and demonstrate that the former delivers an improved description of the data, especially in the short run. In fact, the MF-TVP-SV-VAR provides a better estimate of the actual uncertainty surrounding the point estimate and is able to produce more forecasts corresponding to the upper percentiles of the empirical distribution. Finally, we examine the mixed-frequency models' inflation forecasts during the Great Recession. We show that allowing for time variation in the VAR coefficients and stochastic volatility leads to more precise predictions for the steep downturn and the subsequent recovery than considering only one of these specifications. Regarding the remaining variables, the results are mixed; for unemployment rate forecasts, drifting coefficients are sufficient, for interest rate and GDP growth forecasts, stochastic volatility yields precise forecasts.

On the one hand, this chapter contributes to the ongoing discussion on how structural change affects forecast performance. D'Agostino, Gambetti, and Giannone (2013) forecast US inflation, unemployment, and short-term interest rates with TVP-SV-VARs and find that allowing for parameter instability significantly improves forecast accuracy. A detailed assessment of the forecast performance of models with time-varying coefficients relative to a variety of other nonlinear and

[2] An alternative approach, called mixed data sampling (MIDAS), is provided by Ghysels, Santa-Clara, and Valkanov (2004). For an assesment of this approach with regard to forecasting, see Clements and Galvão (2008).

[3] Amir-Ahmadi et al. (2018) show that the magnitude of the hyperparameters changes significantly when estimated on monthly data compared to quarterly data, which affects the time-variation in the model's coefficients.

linear time series approaches is provided by both Barnett, Mumtaz, and Theodoridis (2014) and Clark and Ravazzolo (2015). They underpin the findings of D'Agostino et al. (2013) and show that models with time-varying parameters improve forecast performance, especially in regard to inflation forecasts. Banbura and van Vlodrop (2018) illustrate that accounting for time-varying means in a Bayesian VAR substantially increases long-term forecast accuracy. Antolin-Diaz, Drechsel, and Petrella (2017) provide evidence in favor of decline in long-run US output growth and demonstrate that modeling this decline in a DFM increases nowcast accuracy.

On the other hand, this article extends the literature on forecasting with mixed-frequency models. Since the work of Giannone, Reichlin, and Small (2008), investigating the marginal impact of new information on nowcast accuracy, several studies have underpinned the benefits of modeling different frequencies with regard to forecasting.[4] The studies by Foroni, Guérin, and Marcellino (2015), Barsoum and Stankiewicz (2015), and Bessec and Bouabdallah (2015) extend this literature by considering mixed-frequency models with discrete regime switches.

The remainder of the chapter is as follows. Section 5.2 provides a description of the dataset and outlines the forecast setup. Section 5.3 depicts the competing models and explains the estimation methodology. Section 5.4 describes the measures used for the forecast comparison. Section 5.5 presents the results. Section 5.6 concludes.

5.2 Data and forecast setup

5.2.1 Dataset

We use an updated version of the dataset used by Clark and Ravazzolo (2015) consisting of four macroeconomic time series, three of which are sampled at monthly frequency and one is observed quarterly. The quarterly series is US real GDP; the monthly series are CPI, the unemployment rate, and the 3-month Treasury bill rate. GDP and CPI enter the models in log first differences to obtain real GDP growth rates and CPI inflation, respectively. The unemployment and interest rate remain untransformed. For the VARs estimated on quarterly frequency, the monthly indicators enter the models as quarterly averages; we do not apply any further transformation for the mixed-frequency models. We obtain real-time data on inflation, unemployment and GDP from the Archival FRED (ALFRED) database of the St. Louis Fed. Since the Treasury bill rate is not revised, we resort to the last available publication from the FRED database. The sample runs from January 1960 until September 2016. The first 10 years are used as a training sample to specify priors. Thus, the actual model estimation starts in January 1970.

Generally, macroeconomic variables are released with a publication lag, which implies that a certain vintage does not include the figures referring to the date of the vintage. The first release

[4]For example, Clements and Galvão (2008), Wohlrabe (2009), Kuzin, Marcellino, and Schumacher (2011), Foroni and Marcellino (2014), and Mikosch and Neuwirth (2015).

of quarterly GDP has a publication lag of roughly one month, thus—for example—the first figure for 2011Q4 is released at the end of 2012M1 and is then consecutively revised in the subsequent months 2012M2 and 2012M3. The value for the unemployment rate (CPI) is published in the first (second) week of the following month. Hence, following our previous example, at the end of 2012M1 the unemployment rate and CPI are available until 2011M12. Finally, the 3-month Treasury bill rate is available without any delay. Thus, we have so-called "ragged-edges" in our real-time dataset.

5.2.2 Forecast setup

In assessing the predictions we follow Schorfheide and Song (2015) and establish three different information sets. We assume that the forecasts are generated at the end of each month, when all current releases for the indicators are available. The first information set, called I1, relates to the first month of each quarter such that the forecaster has information up to the end of January, April, July, or October. In these months, the researcher has observations on inflation and unemployment until the end of the respective previous quarter and a first and preliminary estimate of GDP referring to the previous quarter. The second information set, called I2 (February, May, August, November), has one additional observation on inflation and unemployment referring to the current quarter and the first revision of GDP. The last set, I3 (March, June, September, December), includes one more observation on inflation and unemployment and the second GDP revision. Each information set is augmented with the observations of the T-Bill rate.

To assess the intra-quarterly inflow of information, we evaluate the nowcasts separately per information set. However, since the quarterly VARs, cannot cope with "ragged-edges" in the data, we estimate them in each recursion based on the balanced information set I1, which accounts for new information only in terms of data revisions.

We employ an expanding window to evaluate our forecasts for data vintages from January 1995 until September 2016, providing 261 estimation samples. The last one-year-ahead forecast refers to the third quarter 2017. The predictions are evaluated based on quarterly averages, implying that for the mixed-frequency approaches we time aggregate the predicted monthly time paths to quarterly frequency. To abstract from benchmark revisions, definitional changes, and other unforeseeable changes, we evaluate the GDP growth forecasts based on the second available estimate, that is the forecast for period $t + h$ is evaluated with the realization taken from the vintage published in $t+h+2$. Since the remaining variables are revised only rarely and slightly, we evaluate the forecast based on the latest vintage. The maximum forecast horizon h_{max} is set to 4 quarters, which implies that the mixed-frequency models generate forecasts for $h_m = 1, \ldots, 12$ months. Forecasts for horizons larger than one are obtained iteratively. We report results for 1, 2, 3, and 4 quarters ahead forecasts.

5.3 The models

Our baseline model is a standard VAR with all variables sampled at quarterly frequency. Based on this model, we evaluate the forecast performance of three extensions, namely, mixed-frequencies, stochastic volatilities, and time-varying parameters, as well as the forecast performance of combinations of these features. For the models exhibiting stochastic volatility, we use random walk stochastic volatility, which is a parsimonious and competitive specification (Clark and Ravazzolo, 2015). Throughout the chapter, we use n as the number of variables, which can be further split into n_q for quarterly and n_m for monthly variables, respectively, such that $n = n_q + n_m$. Finally, p denotes the lag order.

5.3.1 Quarterly VAR

Our baseline quarterly VAR (Q-VAR) reads:

$$y_t = B_0 + \sum_{i=1}^{p} B_i y_{t-i} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \Omega), \tag{5.1}$$

where y_t and B_0 denote $n \times 1$ vectors of variables and intercepts, respectively. B_i for $i, \ldots, p$ are $n \times n$ matrices of coefficients and Ω is the time-invariant $n \times n$ variance-covariance matrix.

5.3.2 Quarterly VAR with stochastic volatility

The quarterly VAR with stochastic volatility (Q-SV-VAR) does not assume constant residual variances and includes a law of motion for the (log) volatilities. Following Primiceri (2005), we decompose the time-varying covariance matrix of the reduced-form residuals into a lower-triangular matrix A_t and a diagonal matrix Σ_t according to:

$$A_t \Omega_t A_t' = \Sigma_t \Sigma_t', \tag{5.2}$$

where the diagonal elements of Σ_t are the stochastic volatilities and A_t has ones on the main diagonal and nonzero numbers for the remaining lower triangular elements, describing the contemporaneous relationships between the volatilities. This enables us to rewrite the VAR in (5.1) as:

$$y_t = B_0 + \sum_{i=1}^{p} B_i y_{t-i} + A_t^{-1} \Sigma_t u_t, \qquad u_t \sim (0, I_n). \tag{5.3}$$

The law of motions are modeled by defining σ_t as the vector of the diagonal elements of Σ_t and a_t as the vector of nonzero elements stacked by rows of A_t as follows:

$$\log \sigma_t = \log \sigma_{t-1} + e_t, \qquad e_t = (e_{1,t}, \ldots, e_{n,t})' \sim N(0, \Psi), \tag{5.4}$$

$$a_t = a_{t-1} + \upsilon_t, \qquad \upsilon_t = (\upsilon'_{1,t}, \ldots, \upsilon'_{n,t})' \sim N(0, \Phi), \tag{5.5}$$

with Ψ being diagonal and Φ being block diagonal where each block is related to each equation of the VAR in (5.3).

5.3.3 Quarterly VAR with time-varying parameters

The quarterly VAR with time-varying parameter is estimated in a homoscedastic specification (Q-TVP-VAR) and with stochastic volatility (Q-TVP-SV-VAR). The Q-TVP-VAR augments the baseline Q-VAR with random walk processes governing the evolution of the VAR coefficients:

$$y_t = Z'_t \beta_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \Omega), \tag{5.6}$$

$$\beta_t = \beta_{t-1} + \chi_t, \qquad \chi_t \sim N(0, Q), \tag{5.7}$$

where $Z_t = I_n \otimes [1, y'_{t-1}, \ldots, y'_{t-p}]$ contains all the right-hand side variables of the VAR and β_t is the vectorized matrix of the VAR coefficients. For the Q-TVP-SV-VAR, the stochastic volatility part from (5.4) and (5.5) is added to the model. Thus, the heteroscedastic model specification allows for changes in the magnitude of the shocks and for changes in the propagation of these shocks, whereas the homoscedastic version accounts for only the latter.

5.3.4 Mixed-frequency VAR

Estimation of the mixed-frequency VAR (MF-VAR) follows the Bayesian state-space approach of Schorfheide and Song (2015), which can be straightforwardly combined with the former VAR specifications. To this end, we partition our vector of variables $y_t = [y'_{q,t}, y'_{m,t}]'$, where $y_{m,t}$ collects the monthly variables, which potentially contain missing observations due to "ragged-edges" in the dataset. $y_{q,t}$ denotes the quarterly variables at monthly frequency. Since the quarterly variables are observed only in the last month of each quarter, $y_{q,t}$ contains missing observations for the first and second month of each quarter. To construct the measurement equation, we adopt the notion of Mariano and Murasawa (2003) and assume that quarterly GDP in log levels ($\log Y_{q,t}$) can be expressed as the geometric mean of an unobserved monthly GDP ($\log \tilde{Y}_{q,t}$):

$$\log Y_{q,t} = \frac{1}{3}(\log \tilde{Y}_{q,t} + \log \tilde{Y}_{q,t-1} + \log \tilde{Y}_{q,t-2}). \tag{5.8}$$

This expression implies that the quarterly series is a first-order approximation to an arithmetic mean of the unobserved monthly series (see Mitchell et al., 2005). Note that $\log Y_{q,t}$ is observed only every third month, whereas $\log \tilde{Y}_{q,t}$ is never observed. To arrive at an expression for quarterly GDP growth based on latent monthly GDP growth denoted by $\tilde{y}_{q,t}$, we subtract $\log Y_{q,t-3}$ from (5.8):

$$\Delta_3 \log Y_{q,t} = y_{q,t} = \frac{1}{3}\tilde{y}_{q,t} + \frac{2}{3}\tilde{y}_{q,t-1} + \tilde{y}_{q,t-2} + \frac{2}{3}\tilde{y}_{q,t-3} + \frac{1}{3}\tilde{y}_{q,t-4}, \tag{5.9}$$

where lower-case variables refer to logs. Combining the unobserved with the observed monthly variables in $\tilde{y}_t = [\tilde{y}'_{q,t}, y'_{m,t}]'$, we define the state vector as $z_t = [\tilde{y}'_t, \ldots, \tilde{y}'_{t-p+1}]$ and write the measurement equation as:

$$y_t = H_t z_t. \tag{5.10}$$

Assuming that GDP growth is ordered first in the model, H_t is given by:

$$H_t = \begin{bmatrix} H_{1,t} & H_{2,t} \end{bmatrix}', \tag{5.11}$$

$$H_{1,t} = \begin{bmatrix} 1/3 & 0_{1\times n-1} & 2/3 & 0_{1\times n-1} & 1 & 0_{1\times n-1} & 2/3 & 0_{1\times n-1} & 1/3 & 0_{1\times n-1} & 0_{1\times (p-4)n} \end{bmatrix}, \tag{5.12}$$

$$H_{2,t} = \begin{bmatrix} 0_{n-1\times 1} & I_{n-1} & 0_{n-1\times pn} \end{bmatrix}, \tag{5.13}$$

where $H_{1,t}$ translates the disaggregation constraint in (5.9) into the state-space framework. To replace the missing observations in z_t with estimated states, we follow Durbin and Koopman (2001) and employ a time-dependent vector of observables y_t and a time-varying matrix H_t. If an indicator exhibits a missing observation in period t, the corresponding entry in y_t and the corresponding row of H_t are deleted. Finally, the transition equation of the MF-VAR in state-space form is given by:

$$z_t = \mu + F z_{t-1} + \upsilon_t, \qquad \upsilon_t \sim N(0, S), \tag{5.14}$$

where μ and F contain the intercepts and AR-coefficients, respectively. S is a $pn \times pn$ variance-covariance matrix with the upper left $n \times n$ submatrix corresponding to Ω and all the remaining entries being zero.

To introduce stochastic volatility into the mixed-frequency framework, we postulate that the first $n \times n$ elements of S are equal to Ω_t with the same decomposition as in (5.2) and following the same law of motions as in (5.4) and (5.5). This yields the MF-SV-VAR. The MF-TVP-VAR is obtained by allowing F to change over time according to (5.7). Including both specifications leads to the MF-TVP-SV-VAR.

To summarize, we have a total of eight competing models for our forecast experiment:

1. MF-TVP-SV-VAR: Mixed-frequency VAR with time-varying parameters and stochastic volatility
2. MF-SV-VAR: Mixed-frequency VAR with stochastic volatility
3. MF-TVP-VAR: Mixed-frequency VAR with time-varying parameters
4. MF-VAR: Mixed-frequency VAR
5. Q-TVP-SV-VAR: Quarterly VAR with time-varying parameters and stochastic volatility
6. Q-SV-VAR: Quarterly VAR with stochastic volatility (benchmark)
7. Q-TVP-VAR: Quarterly VAR with time-varying parameters
8. Q-VAR: Quarterly linear VAR

5.3.5 Estimation procedure and prior specifications

All models are estimated with Bayesian estimation techniques, since most models depend on a large number of parameters and thus make estimation based on frequentist approaches infeasible. The mixed-frequency models are estimated with 4 lags; the quarterly models are estimated with 2 lags.[5] In the following, we provide a brief description of the estimation procedure and the prior specifications. A detailed description is provided in Appendices D.1 and D.2.

For the baseline Q-VAR we impose a Jeffrey's prior in order to abstract from shrinkage, since we use a small-scale VAR with only four variables. For the Q-SV-VAR we apply the algorithm of Carter and Kohn (1994) (hereafter CK) to draw the VAR coefficients and the mixture sampler of Kim, Shephard, and Chib (1998) (hereafter KSC) to draw the (log) volatilities.[6] We use normal priors for the diagonal elements of Σ_t and the lower-triangular elements of A_t. Inverse-Wishart priors are applied for the variance covariance matrix Ψ and Φ, respectively. We adopt the CK algorithm for the Q-TVP-VAR with a normal prior for β_t and inverse-Wishart prior for the variance-covariance matrix Q. The Q-TVP-SV-VAR combines both prior specifications of the Q-SV-VAR and Q-TVP-VAR and is estimated using the Gibbs sampler as in Del Negro and Primiceri (2015).

[5] We fix the lag order for the quarterly model at 2 to be consistent with the literature on US data (see, e.g., Primiceri, 2005; D'Agostino et al., 2013; Chan and Eisenstat, 2017). The monthly models have 4 lags to keep them computationally feasible since each additional lag increases the number of parameters by $n \times n \times T$. Furthermore, at least four lags are required to disaggregate quarterly GDP into monthly GDP (see (5.9)).

[6] Drawing the VAR coefficients using CK is equivalent to a GLS transformation of the model. Another possibility for drawing the volatilities is the Jacquier, Polson, and Rossi (1995) algorithm, which draws the volatilities one at a time. This single-move procedure, however, is computationally much less efficient and yields draws that are more autocorrelated (see Kim et al., 1998).

The amount of time variation in β_t, a_{it}, and $\log \sigma_{it}$ depends on the magnitude of the random walk variances Q, Ψ, and Φ, which are—in part—determined by the corresponding prior distributions. Hence, assigning sensible priors is crucial. The literature on TVP-SV-VAR commonly follows Primiceri (2005). However, these priors are calibrated for quarterly TVP-SV-VARs and it is not clear, whether they are useful in case of monthly data or other model specifications. Thus, we follow Amir-Ahmadi et al. (2018) and abstract from using partly exogenous values for the scale matrix of the inverse-Wishart prior by implementing another layer of priors for those hyperparameters.

The latent states in the mixed-frequency part of Models 1 to 4 are estimated using a CK algorithm with the Durbin and Koopman (2001) modification, which enables us to cope with "ragged-edges" in the dataset and yields draws for each missing indicator until the end of the sample. We initialize the latent states of the CK algorithm with a normal prior based on monthly constant GDP values throughout the quarter for the mean and an identity matrix for the variance-covariance matrix. After having drawn the latent states, implementation of the remaining specification is straightforward: instead of conditional on the observed data, the remaining coefficients are drawn conditional on the drawn states. Each model is estimated using 60000 draws. For posterior inference we use each 5th draw from the last 10000 draws.

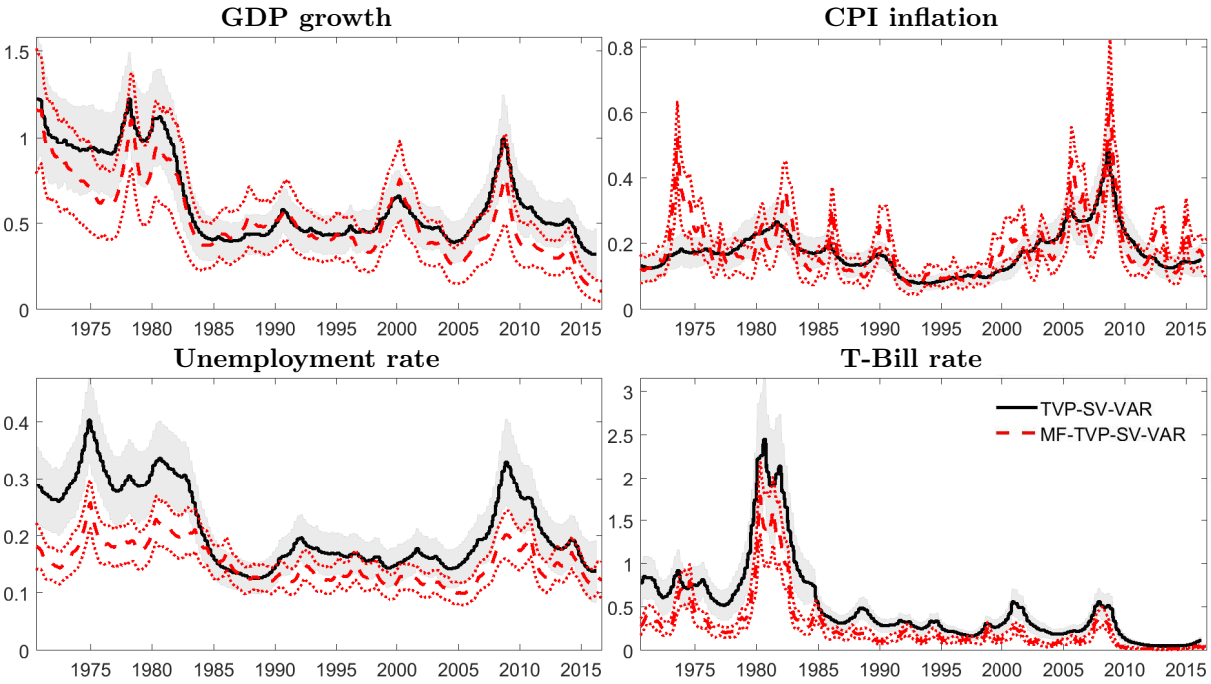
To illustrate the importance of modeling changes in volatility over time, Figure 5.1 plots the posterior means of standard deviations of reduced-form residuals from the MF-TVP-SV-VAR and Q-TVP-SV-VAR using the latest vintage of data.[7] We assume that the volatility estimates from the Q-TVP-SV-VAR are constant within a quarter to make them comparable across frequencies. The estimates of the Q-TVP-SV-VAR closely match the patterns of previous analysis and show significant time-variation (see, for instance, Primiceri, 2005; Clark and Ravazzolo, 2015). Until the mid 1980s, the estimated volatilities are quite high and then fall sharply, indicating the beginning of the Great Moderation. Except for the increase during the burst of the dot-com bubble in 2000 and the rise during the Great Recession, they remain roughly at the levels of the mid 1980s. At the end of the sample, however, there is again a decline in volatility, indicating a time during which the US was remarkably less exposed to absolute shocks hitting the economy. Thus, as suggested by Clark (2009) and Gadea Rivas, Gómez-Loscos, and Pérez-Quirós (2014), the Great Recession seems to have simply interrupted, but not ended, the Great Moderation. In fact, the latest volatility estimates for GDP growth and the unemployment rate are the lowest of the entire sample.

The estimates from the MF-TVP-SV-VAR closely track the evolution of the quarterly estimates. However, except for CPI inflation, they are somewhat smaller, indicating that using monthly information absorbs part of the fluctuations in the volatility. This finding confirms the

[7] We also examined the volatilites for different data vintages to investigate the impact of data revisions and different values for the hyperparameters. Analogous to Clark (2012), though, we obtain very similar estimates for the different vintages, and thus we report results only for the latest vintage.

results of Carriero, Clark, and Marcellino (2015), who employ a Bayesian mixed-frequency model without time variation in the AR-coefficients. The change in the VAR parameters over time is in turn far less pronounced (see Figure 5.4 in Appendix D.4). For both models, the largest variability is obtained for the coefficients of the interest rate equation. For instance, the impact of past interest rates on the current interest rate has increased, while the impact of inflation on interest rates has dropped to almost zero, which reflects the binding zero lower bound. Overall, the results for both models suggest that modeling variability in volatility is important for achieving precise forecasts.

Figure 5.1: Posterior means of standard deviations of reduced-form residuals



Notes: Figure depicts the posterior means of the residual standard deviations from the last vintage of data at monthly frequency. Quarterly estimates are assumed to be constant within a quarter. Shaded areas and dashed lines refer to 68% error bands.

5.3.6 Now- and forecasting

The quarterly models are estimated on balanced datasets containing all the available information from the previous quarter. To generate the now- and forecasts, we follow Cogley, Morozov, and Sargent (2005) and compute sequences of h_{max} normally distributed innovations with covariance Φ, Ψ, and Q to produce time paths for the elements of A_t, Σ_t, and β_t, respectively. Based on these trajectories, we simulate the vector of endogenous variable, y_t, h_{max} periods into the future. The first forecast is a nowcast, since it is generated in and refers to the respective current quarter.

Additional notation is helpful in describing how we obtain the predictive distributions for now- and forecasts from the mixed-frequency approaches. Let T_M denote the last month of the indicator that has the shortest publication lag and let $Z^{T_M} = [z_1, \ldots, z_{T_M}]$ denote the trajectory of simulated state vectors. Note that the CK algorithm provides draws for the latent states until T_M, which is why Z^{T_M} consists only of CK draws. To obtain $Z^{T_m+1:T_m+h_{max}}$, we again generate time paths for the elements of A_t, Σ_t, and β_t and simulate the state vector z_t forward using these time paths. Accordingly, if T_M belongs to I3, the CK algorithm provides draws for the entire last available quarter and by averaging over these draws we obtain the nowcasts. The forecasts are generated by averaging over the trajectories $Z^{T_m+1:T_m+h_{max}}$. However, if T_M belongs to I1 or I2, the CK algorithm does not provide draws of the latent states for the entire quarter since none of the indicators is available for the entire quarter. In this case, we average over the available CK draws and the simulated trajectories referring to this quarter to get the nowcast. The forecasts are calculated from the averages of the remaining trajectories.[8]

5.4 Forecast metrics

We evaluate our models' forecasts with respect to point and density forecasts. In the following, M, i, and h denote the model, variable, and forecast horizon, respectively, for the forecast sample $t = 1, \ldots, N$.[9] The point forecast accuracy is measured in terms of the root mean squared errors (RMSE):

$$\text{RMSE}_i^h = \sqrt{\frac{1}{N} \sum \left(\hat{y}_{t+h}^i - y_{t+h}^i\right)^2}. \tag{5.15}$$

However, the RMSE is a useful tool for assessing the accuracy of a model only when compared across different models, hence we report the RMSEs as ratios relative to a benchmark model:

$$\text{relative RMSE}_{M,i}^h = \frac{\text{RMSE}_{M,i}^h}{\text{RMSE}_{B,i}^h}, \tag{5.16}$$

where $\text{RMSE}_{B,i}^h$ refers to the RMSE of the benchmark Q-SV-VAR estimated with quarterly data.[10] To provide a formal test of whether the difference in forecast accuracy is significant, we apply Diebold and Mariano (1995) test.

[8] For instance, in February, the T-Bill rate is available until February (T_M), while inflation and the unemployment rate are available until January (T_{M-1}). Hence, the CK algorithm provides draws for each indicator until T_M. The figures for March (T_{M+1}) are generated using the time paths for A_t, Σ_t, and β_t. The forecast for the first quarter is the average over the figures of T_{M-1} to T_{M+1}.

[9] To facilitate readability, in the following we drop subscript M indicating the respective model in most cases.

[10] Since several studies demonstrate that VARs with stochastic volatility outperform constant volatility VARs (see, for instance, Clark, 2012; Clark and Ravazzolo, 2015; Chiu, Mumtaz, and Pintér, 2017), we abstract from using the Q-VAR as our benchmark.

In the past, evaluation of economic forecasts focused solely on point forecasts; however, more recently the uncertainty around the forecasts has become an important issue. To take this uncertainty into account, that is the remaining part of the predictive density, which is neglected by the measures outlined above, we further evaluate the predictive densities. Since the true density is not observed, however, evaluating the predictive density is less straightforward than evaluating point forecasts. The idea behind evaluation of density forecasts is to compare the distribution of observed data with the predictive density and assess whether the latter provides a realistic picture of reality. To this end, we apply both the log predictive scores and the continuous ranked probability score (CRPS). The log predictive score is computed by evaluating the predictive density at the realization.[11] In the following, we report average log scores:

$$\text{LS}_i^h = \frac{1}{N} \sum \log p_t(y_{t+h}^i), \tag{5.17}$$

where $p_t(\cdot)$ indicates the predictive density. According to (5.17), a higher average log score implies a more exact predictive density.[12] Again, we report results relative to the benchmark:

$$\text{relative LS}_{M,i}^h = LS_{M,i}^h - LS_{B,i}^h, \tag{5.18}$$

where $LS_{B,i}^h$ refers to the log score of the benchmark model. Furthermore, we evaluate the predictive densities in terms of the CRPS introduced by Matheson and Winkler (1976). As highlighted by, for example, Gneiting and Raftery (2007), CRPSs are both better able to evaluate forecasts close but not equal to the realization and less sensitive towards extreme outcomes. To compute the CRPS, we follow Gneiting and Ranjan (2011) and use the score function:

$$S(p_t, y_{t+h}^i, \nu(\alpha)) = \int_{-0}^{1} \text{QS}_\alpha(P_t(\alpha)^{-1}, y_{t+h}^i)\nu(\alpha)d\alpha, \tag{5.19}$$

where $\text{QS}_\alpha(P_t(\alpha)^{-1}, y_{t+h}^i) = 2(I\{y_{t+h}^i < P_t(\alpha)^{-1}\} - \alpha)(P_t(\alpha)^{-1} - y_{t+h}^i)$ is the quantile score for forecast quantile $P_t(\alpha)^{-1}$ at level $0 < \alpha < 1$. $I\{y_{t+h}^i < P_t(\alpha)^{-1}\}$ is an indicator function taking the value 1 when $y_{t+h}^i < P_t(\alpha)^{-1}$ and 0 otherwise. P_t^{-1} denotes the inverse of the cumulative predictive density function and $\nu(\alpha)$ is a weighting function. Applying a uniform weighting scheme, yields the average CRPS:

$$\text{CRPS}_i^h = \frac{1}{N} \sum S(p_t, y_{t+h}^i, 1). \tag{5.20}$$

[11] The log predictive score goes back to Good (1952) and has become a commonly accepted tool for comparing the forecast performance of different models (see Geweke and Amisano, 2010; Clark, 2012; Jore, Mitchell, and Vahey, 2010, among other).

[12] Since the predictive densities are not necessarily normal, the commonly used quadratic approximation of Adolfson, Lindé, and Villani (2007) may not be appropriate, which is why we smooth the empirical forecast distribution using a kernel estimator to obtain the predictive distribution.

Following (5.20), a lower value indicates a better score, which is evaluated as a ratio relative to our benchmark model:

$$\text{relative CRPS}^h_{M,i} = \frac{CRPS^h_{M,i}}{CRPS^h_{B,i}} \tag{5.21}$$

To obtain an approximate inference on whether the scores are significantly different from the benchmark, we follow D'Agostino et al. (2013) and regress the differences between the scores of each model and the benchmark on a constant. A t-test with Newey-West standard errors on the constant indicates whether these average differences are significantly different from zero.

Finally, we compute probability integral transforms (PITs) as in Diebold, Gunther, and Tay (1998). The PIT is defined as the CDF corresponding to the predictive density evaluated at the respective realizations:

$$z^i_{t+h} = \int_{-\infty}^{y^i_{t+h}} p_t(u)du \quad \text{for } t = 1, \ldots, N. \tag{5.22}$$

Thus, with regard to the respective predictive density, the PIT denotes the probability that a forecast is less than or equal to the realization. If the predictive densities equal the true densities, z^i_{t+h} is uniformly distributed over the 0-1 interval. To assess the accuracy of the predictive density according to the PIT, it is convenient to divide the unit interval into k equally sized bins and count the number of PITs in each bin. If the predictive density equals the actual density, each bin contains N/k observations.

5.5 Results

In this section, we discuss the results from our forecast experiment. Regarding the point forecasts, first we assess the nowcast accuracy of the models and resort to the information sets outlined in Section 5.2.2. Second, we evaluate the accuracy of the point forecasts and predictive densities with respect to horizons larger than 1, that is the respective subsequent quarters.[13] We provide results for the entire recursive sample (1995Q1–2016Q4) and for a shorter sample period of 2007Q1 until 2016Q4. The latter enables us to assess whether the rise in volatility during the Great Recession and the subsequent slowdown affect the forecast performance.

[13] We abstract from evaluating the nowcasts with respect to predictive densities. Depending on the information sets, the nowcasts of the mixed-frequency models consist of quarterly averages over draws from the CK algorithm and realizations. Therefore, the nowcast densities of the mixed-frequency models are very narrow compared to the quarterly models and thus hardly comparable.

5.5.1 Nowcast evaluation

Table 5.1 presents the results for the nowcasts taking into account the information sets I1 to I3. The results can be compared along five dimensions: quarterly- vs. mixed-frequencies, fixed-coefficients vs. time-varying coefficients, across information sets as well as variables, and across samples. First, we discuss the results with respect to the full sample as shown in the left panel of Table 5.1.

Comparing the mixed-frequency models with the quarterly models reveals that the MF-models generate more accurate nowcasts for each variable and each information set. On average, over all information sets and variables, the best nowcast performance is obtained by the MF-TVP-SV-VAR, which improves on the benchmark (Q-SV-VAR) by roughly 42%. The remaining mixed-frequency models provide, on average, gains ranging from 38% (MF-VAR) to 40% (MF-SV-VAR), indicating that apart from using monthly information, parameter variability is beneficial. Except for the Q-TVP-SV-VAR, which provides roughly the same performance as the benchmark, all quarterly models deliver inferior nowcast performance and thus provide—in line with the literature—strong support for mixed-frequency approaches.

Turning to the variables, we first look at GDP growth. In this case, the best performing model (MF-SV-VAR) provides up to 14% more accurate nowcasts compared to the benchmark. In contrast, the best performing quarterly model does not outperform the benchmark. A similar pattern emerges for inflation: the best performance is again provided by a mixed-frequency model (MF-TVP-SV-VAR), which improves on the benchmark by up to 60%, while the quarterly models are better than the benchmark by at most by 7%. The MF-TVP-SV-VAR provides, on average, much more precise forecasts than the Q-TVP-SV-VAR, which itself makes precise inflation forecasts (Faust and Wright, 2013). The largest difference between quarterly- and mixed-frequency models, is obtained for the unemployment rate. In this case, the MF-TVP-VAR improves on the benchmark by roughly 65%, while the quarterly models do no better than the benchmark

A comparison of the fixed-coefficients models with the time-varying coefficients models reveals that stochastic volatility seems to be a major determinant of precise nowcasts, which is consistent with, for instance, Carriero et al. (2015). In all but one case (unemployment rate at I1), the best performing model includes stochastic volatility. Allowing for time-varying parameters without stochastic volatility improves accuracy relative to the benchmark but is—in most cases—inferior to models with stochastic volatility. Inflation nowcasts especially benefit from combining both specifications. For instance, the relative RMSE of the MF-TVP-SV-VAR is about 5 percentage points lower than that of the MF-TVP.

Table 5.1: Real-time nowcast RMSEs

Model	1995-2016			2008-2016		
	I1	**I2**	**I3**	**I1**	**I2**	**I3**
	GDP growth					
MF-TVP-SV-VAR	0.91	0.90	0.91	0.86	0.84	0.87
MF-SV-VAR	**0.89**	**0.86**	**0.86**	**0.81**	**0.77**	**0.77**
MF-TVP-VAR	0.98	0.96	0.88	0.95	0.94	0.82
MF-VAR	0.99	0.93	0.91	0.98	0.88	0.90
Q-TVP-SV-VAR	1.06	1.09	1.07	1.04	1.08	1.08
Q-TVP-VAR	1.07	1.11	1.03	1.08	1.10	0.99
Q-VAR	1.12***	1.13***	1.13***	1.13	1.14***	1.14***
Q-SV-VAR	0.63	0.62	0.63	0.80	0.78	0.78
	Inflation					
MF-TVP-SV-VAR	**0.87***	**0.64****	**0.39****	**0.85**	**0.59**	0.31
MF-SV-VAR	0.99	0.68*	**0.39****	0.93	0.61	**0.30**
MF-TVP-VAR	0.92*	0.69**	0.40**	0.92	0.67	0.36
MF-VAR	0.99	0.69*	**0.39****	0.95	0.62	**0.30**
Q-TVP-SV-VAR	0.94	0.93	0.94	0.94	0.92	0.93
Q-TVP-VAR	0.93	0.95	0.98	0.98	0.96	0.99
Q-VAR	1.03	1.02	1.13	1.02	1.01	1.00
Q-SV-VAR	0.26	0.26	0.26	0.34	0.33	0.34
	Unemployment rate					
MF-TVP-SV-VAR	0.82***	0.61***	**0.34*****	**0.78***	**0.60****	0.34**
MF-SV-VAR	0.83***	0.61***	**0.34*****	0.85***	**0.60****	**0.32****
MF-TVP-VAR	**0.79*****	**0.60*****	**0.34*****	**0.78****	0.61*	0.34*
MF-VAR	0.86***	0.61***	**0.34****	0.84**	**0.60****	**0.32****
Q-TVP-SV-VAR	1.02	1.08	1.04	1.05	1.07	1.07
Q-TVP-VAR	1.07	1.04	1.03	1.09	1.10	1.04
Q-VAR	1.05***	1.07***	1.07***	1.06	1.07**	1.07**
Q-SV-VAR	0.27	0.26	0.26	0.36	0.35	0.35
	Interest rate					
MF-TVP-SV-VAR	**0.43*****	0.16***	–	**0.36***	**0.15***	–
MF-SV-VAR	0.44***	**0.15*****	–	**0.36***	**0.15***	–
MF-TVP-VAR	0.45***	0.17***	–	0.41*	0.16*	–
MF-VAR	0.59***	0.19***	–	0.57	0.18*	–
Q-TVP-SV-VAR	0.96	0.94	0.94	0.87**	0.86*	0.86*
Q-TVP-VAR	1.13**	1.15*	1.14*	1.09	1.07	1.10
Q-VAR	1.48***	1.46***	1.47*	1.56	1.52*	1.51*
Q-SV-VAR	0.36	0.36	0.36	0.37	0.38	0.39

Notes: The models are detailed in Section 5.3. RMSEs are reported in absolute terms for the benchmark model (the bottom row of each panel). For the remaining, the RMSEs are expressed as ratios relative to the benchmark model. A figure below unity indicates that the model outperforms the benchmark. Bold figures indicate the best performance for the variable and information set. $*$, $**$, and $***$ denote significance at the 15%, 10%, and 5% level, respectively, according to the Diebold-Mariano test with Newey-West standard errors. At I3 no interest rate forecast is computed by the mixed-frequency models, since the entire quarter is available.

Finally, we compare the RMSEs across information sets. In most cases, using more information—as expected—significantly reduces the RMSEs. In case of inflation forecasts, the improvements for the best performing models range from 13% at I1 to 61% at I3. With regard to the unemployment rate, the increases in forecast accuracy are even higher, with 18% at I1 and 66% at I3. When it comes to GDP growth, however, more information does not appear to increase forecast accuracy; the relative RMSEs for the best performing model (MF-TVP-SV-VAR) go from 0.89 at I1 to 0.86 at I3, providing some evidence that the variables used may not be the best predictors for GDP growth and that selecting the variables more carefully could improve GDP growth forecasts. Since the goal of this chapter is not to find the best GDP growth forecast, we leave this question for further research.

Overall, the results for the shorter sample are very similar to those for the full sample. The right panel of Table 5.1 shows that the relative nowcast performance of each model remains almost unchanged for unemployment and interest rate. However, with regard to GDP growth and inflation, the MF-models' relative performance improves in the shorter sample, which is characterized by a larger volatility of the series. The strongest gains are obtained for the MF-SV-VAR. Its GDP growth and inflation forecasts are roughly 8% more precise. Therefore, and in contrast to the entire sample, the best performance in the shorter sample, on average, is provided by the MF-SV-VAR, suggesting that stochastic volatility has become more important for precise nowcasts.

5.5.2 Forecast evaluation

We now investigate forecast performance. Since the marginal impact of an additional month of information becomes less important for forecasts at higher horizons, the RMSEs for higher horizons become more similar across the information sets.[14] Therefore, in the following we compute total RMSEs by averaging over the entire forecast sample.

The results in Table 5.2 indicate that mixed-frequency VARs provide competitive forecasts even for higher horizons and for both samples. Indeed, in the case of unemployment and interest rate forecasts, modeling within-quarter dynamics is particularly beneficial, since even the worst performing mixed-frequency VAR outperforms the best performing quarterly VAR on each horizon. In the following, we focus on the results for the full sample.

Overall, the most accurate forecasts for all indicators and on average are again provided by the MF-TVP-SV-VAR. It outperforms the baseline Q-VAR over all horizons and variables by roughly 12% on average. The best performance is obtained for the interest rate, with an average gain of about 30% relative to the benchmark. The MF-SV-VAR and the MF-TVP-VAR also provide very accurate forecasts, with average gains of around 10%. The MF-VAR yields roughly

[14] Figure 5.6 in Appendix D.4 plots the relative RMSE for each information set.

the same performance as the best performing quarterly model, namely, the Q-TVP-SV-VAR; both improve the forecast on average over all horizons and variables by about 2%.

Concerning the variables individually shows that the gains in forecast accuracy differ substantially across models. However, quarterly models outperform the mixed-frequency models only for GDP growth. Nevertheless, all of the RMSEs—except for the MF-VAR—are quite close to each other. For inflation, the best performance over all horizons is provided by the MF-TVP-SV-VAR. In particular, for one-quarter-ahead forecasts, it generates by far the most accurate predictions. On higher horizons, the Q-TVP-SV-VAR delivers virtually identical RMSEs, which indicates both that using monthly information becomes less important for higher horizons and that using time variation in all coefficients is crucial for inflation forecasts. The latter confirms the results of previous studies based on quarterly models (see D'Agostino et al., 2013; Barnett et al., 2014; Clark and Ravazzolo, 2015) by use of mixed-frequency models. As for the nowcasts, the most accurate unemployment rate forecasts are obtained by the MF-TVP-VAR. The differences from the MF-TVP-SV-VAR and MF-SV-VAR are, though, small. Using only quarterly data in turn provides significantly inferior RMSEs. The largest differences between the quarterly and the mixed-frequency models are obtained for the interest rate, where the RMSE of the best performing mixed-frequency model is roughly one-third the size of the RMSE of the best performing quarterly model (0.69 vs. 0.96).

Comparing the results across samples reveals that the relative RMSEs are very similar for each variable and model, suggesting that the sample has only minor influence on the results. However, while the models that incorporate stochastic volatility improve their relative forecast accuracy in the shorter sample, the performance of models without this feature tends to deteriorate, indicating that in this more volatile phase, stochastic volatility is more important for achieving precise forecasts. Moreover, and in contrast to the nowcast evaluation, the best performance on average is provided by the MF-TVP-SV-VAR, which improves on the benchmark by 17% and slightly outperforms the MF-SV-VAR.

Overall, the results are consistent with findings from previous studies, suggesting that the gains in accuracy due to variations in the VAR-coefficients are smaller than the gains induced by stochastic volatility. However, using models with both features provides more accurate forecasts for all variables in most cases.

Table 5.2: Real-time forecast RMSEs

Model	1995-2016			2008-2016		
	h = 2	**h = 3**	**h = 4**	**h = 2**	**h = 3**	**h = 4**
			GDP growth			
MF-TVP-SV-VAR	1.07	1.03	1.00	1.08	1.00	0.99
MF-SV-VAR	1.00	1.00	0.98	0.98	1.00	0.99
MF-TVP-VAR	**0.99**	0.99	1.01	**0.97**	**0.95**	1.00
MF-VAR	1.12***	1.14***	1.12***	1.11***	1.16***	1.15***
Q-TVP-SV-VAR	1.03	**0.98**	**0.96**	1.00	0.96	**0.94**
Q-TVP-VAR	1.17***	1.08**	1.07	1.17***	1.04	1.09
Q-VAR	1.17***	1.16***	1.15***	1.23***	1.19***	1.20***
Q-SV-VAR	0.63	0.64	0.66	0.82	0.85	0.85
			Inflation			
MF-TVP-SV-VAR	**0.87****	**0.91*****	**0.93*****	**0.85****	**0.92*****	**0.93*****
MF-SV-VAR	0.98	0.96	0.94	0.97	0.96	0.93
MF-TVP-VAR	0.93	0.95	0.94	0.93	0.97	0.96
MF-VAR	1.01	1.03	1.05	0.98	0.99	0.97
Q-TVP-SV-VAR	0.93**	0.93***	0.94***	0.92**	0.92***	0.94***
Q-TVP-VAR	0.93	0.95*	0.97	0.91	0.94	0.98
Q-VAR	1.00	1.07***	1.09***	0.99	1.06**	1.06
Q-SV-VAR	0.26	0.24	0.23	0.34	0.30	0.29
			Unemployment rate			
MF-TVP-SV-VAR	0.86*	0.93	0.98	0.86***	0.94***	0.98***
MF-SV-VAR	0.86**	0.91***	0.95	0.85***	0.91**	0.94
MF-TVP-VAR	**0.83****	**0.87****	**0.91**	**0.82****	**0.88****	**0.90***
MF-VAR	0.88**	0.94	0.94	0.85**	0.92	0.96
Q-TVP-SV-VAR	1.04	1.04	1.04	1.05	1.04	1.04
Q-TVP-VAR	1.02	1.01	0.98	1.12**	1.10***	1.05**
Q-VAR	1.07***	1.08***	1.08***	1.07**	1.07**	1.08**
Q-SV-VAR	0.48	0.73	0.97	0.68	1.04	1.39
			Interest rate			
MF-TVP-SV-VAR	0.59***	0.76***	0.84***	0.47***	0.59***	0.62***
MF-SV-VAR	**0.57*****	**0.71*****	**0.78*****	**0.44*****	**0.53*****	**0.57*****
MF-TVP-VAR	0.62***	0.81***	0.91***	0.56***	0.79***	0.89**
MF-VAR	0.72***	0.87***	0.94**	0.71***	0.90***	0.97**
Q-TVP-SV-VAR	0.95	0.96	0.97	0.81**	0.77***	0.81***
Q-TVP-VAR	1.14**	1.19***	1.17***	1.20**	1.33***	1.29***
Q-VAR	1.15***	1.16***	1.17***	1.34***	1.45***	1.48***
Q-SV-VAR	0.67	0.96	1.24	0.64	0.85	1.10

Notes: The models are detailed in Section 5.3. RMSEs are reported in absolute terms for the benchmark model (the bottom row of each panel). For the remaining, the RMSEs are expressed as ratios relative to the benchmark model. A figure below unity indicates that the model outperforms the benchmark. Bold figures indicate the best performance for the variable and horizon. *, **, and *** denote significance at the 15%, 10%, and 5% level, respectively, according to the Diebold-Mariano test with Newey-West standard errors.

5.5.3 Predictive density evaluation

The results for continuous rank probability scores (CRPS) are displayed in Table 5.3. The benchmark is reported in levels, while for the other models, the scores are reported as ratios relative to the benchmark. We focus on CRPS since it is more sensitive to distance and less sensitive to outliers than the log scores.[15]

We draw three main conclusions from the results. First, the sample period has only minor impact on the relative accuracy of the predictive densities. In fact, the CRPS are overall very similar, which is why we discuss results for both samples jointly.

Second, using within-quarter information significantly improves predictive densities; the mixed-frequency models provide better results on average over all variables and horizons. The best performance is again provided by the MF-TVP-SV-VAR, with an average improvement of roughly 13%. In contrast, the best performing quarterly model (Q-TVP-SV-VAR) improves on the benchmark on average only by 2%. With regard to the unemployment rate, the worst performing mixed-frequency model (MF-VAR) improves on the benchmark by 7% on average, and thus is better than each quarterly model, indicating that mixed-frequency is an important feature for unemployment forecasts. For the interest rate, we see a similar picture; apart from the Q-TVP-SV-VAR, none of the quarterly models outperform the mixed-frequency models. For GDP growth and inflation, the results are less obvious. The most accurate GDP growth forecasts over all horizons are provided by the Q-TVP-SV-VAR, though differences from its mixed-frequency counterpart are very small. Regarding inflation, only the mixed-frequency models with time-varying parameters outperform their quarterly counterparts. Investigating the forecast performance across the different horizons shows that the differences between the quarterly and mixed-frequency models become smaller with increasing horizons. Thus, within-quarter information is more valuable with respect to short-term forecasting.

Third, looking at the variables individually reveals that models using stochastic volatility and/or time-varying VAR-coefficients usually generate more accurate predictive densities than models without these features. As for the point forecast performance, the best performance for the interest rate forecasts is provided by the MF-SV-VAR, indicating that variation in the VAR-coefficients is only a minor issue in this case. The results for the unemployment rate and inflation are similar to the point forecast performance. In the case of the unemployment rate, the MF-SV-VAR and the MF-TVP-VAR have virtually identical performance, indicating that one can use either SV-models or TVP-models or both. For inflation, however, it is crucial to combine stochastic volatility and time-varying VAR-coefficients to obtain precise predictive densities.

[15] In general, the predictive distributions of the MF-models have a lower variance than those of the quarterly benchmark. Therefore, outlier realizations receive a very low log score in the case of MF-models, which distorts the overall results. However, as set out in Table 5.5 in Appendix D.3, both measures provide qualitatively similar results.

In summary, the results of the predictive density evaluation support the findings from the point forecast evaluation. Using mixed-frequency is beneficial irrespective of time-variation in parameters, stochastic volatility, variables, or forecast horizons. As shown by several studies using quarterly models, stochastic volatility significantly improves predictive densities (see, for example, Carriero et al., 2015; Carriero, Clark, and Marcellino, 2016; Chiu et al., 2017; Huber, 2016). We confirm this finding using mixed-frequency models. We add support to the results of D'Agostino et al. (2013) and Clark and Ravazzolo (2015) by demonstrating that, in general, combining stochastic volatility and time-varying parameters improves the accuracy of predictive densities for both quarterly and mixed-frequency models.

For a clearer picture of the predictive densities' calibration, we compute probability integral transforms (PITs). For the sake of brevity, Figure 5.2 only presents inflation predictions of the Q-TVP-SV-VAR (upper panel) and the MF-TVP-SV-VAR (bottom panel).[16] To ensure uniformity, each bin in Figure 5.2 should contain 20% of the forecasts. The most apparent difference between the models' histograms is that the MF-TVP-SV-VAR is much better at capturing the right tail of the distribution than is the Q-TVP-SV-VAR, especially at short horizons. Moreover, the histograms of the Q-TVP-SV-VAR are hump-shaped for $h=1$ and $h=2$, indicating that the predictive densities are too wide and the uncertainty around the point estimate is overestimated. This pattern is less pronounced for the MF-TVP-SV-VAR, which has histograms closer to uniformity. In summary, our results indicate that omitting within-quarter dynamics and computing quarterly averages from monthly variables ignores valuable information, which in most cases significantly improves forecast accuracy.

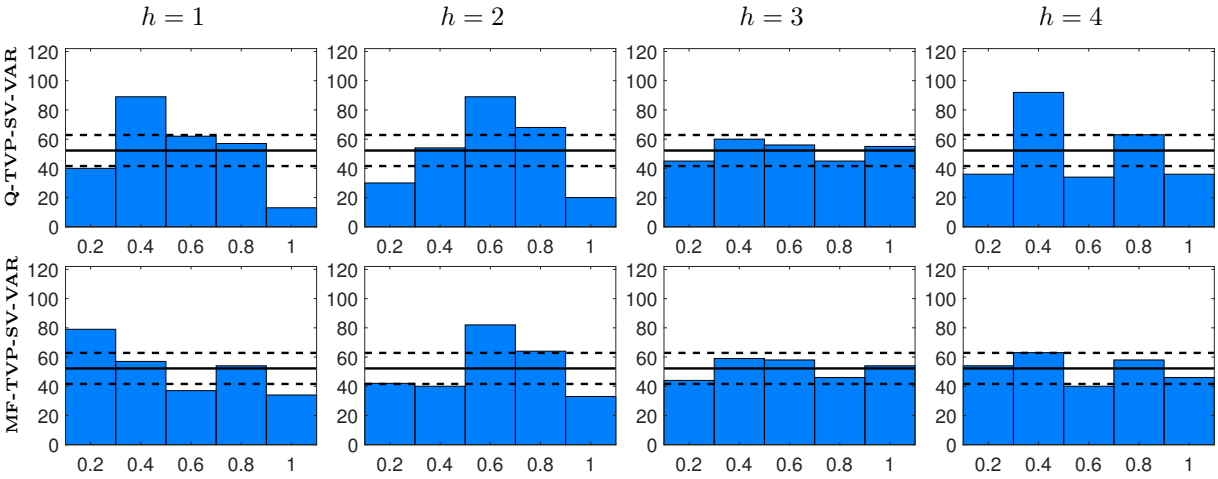
5.5.4 Forecasting during the Great Recession

The previous sections demonstrated that modeling intra-quarterly dynamics significantly improves forecast accuracy on average, in particular with regard to the novel MF-TVP-SV-VAR. We now take a closer look at the MF-models' absolute performance during the Great Recession, which is of great interest, because many structural and nonstructural models failed to provide accurate forecasts for the steep contraction and the following upswing in 2008/2009. Since the MF-models perform especially well for forecasting inflation, Figure 5.3 depicts real-time quarter-on-quarter CPI inflation growth (blue line) along with both the means (black line) and 60% as well as 90% error bands (shaded areas) from the predictive distributions, respectively.[17] The figure's columns refer to the data vintages of October 2008 until December 2008 and demonstrate how the arrival of new data points affects the forecasts. Consider the forecasts computed with the vintage of October 2008 (the first column). Note that in this month the models do not have any

[16] The PITs for the remaining variables, models, and horizons are presented in Appendix D.4, and paint similiar picture, further supporting the good performance of the MF-TVP-SV-VAR in terms of predictive density calibration.

[17] Figures for the remaining variables are provided in Appendix D.4.

Figure 5.2: Probability integral transforms for inflation forecasts

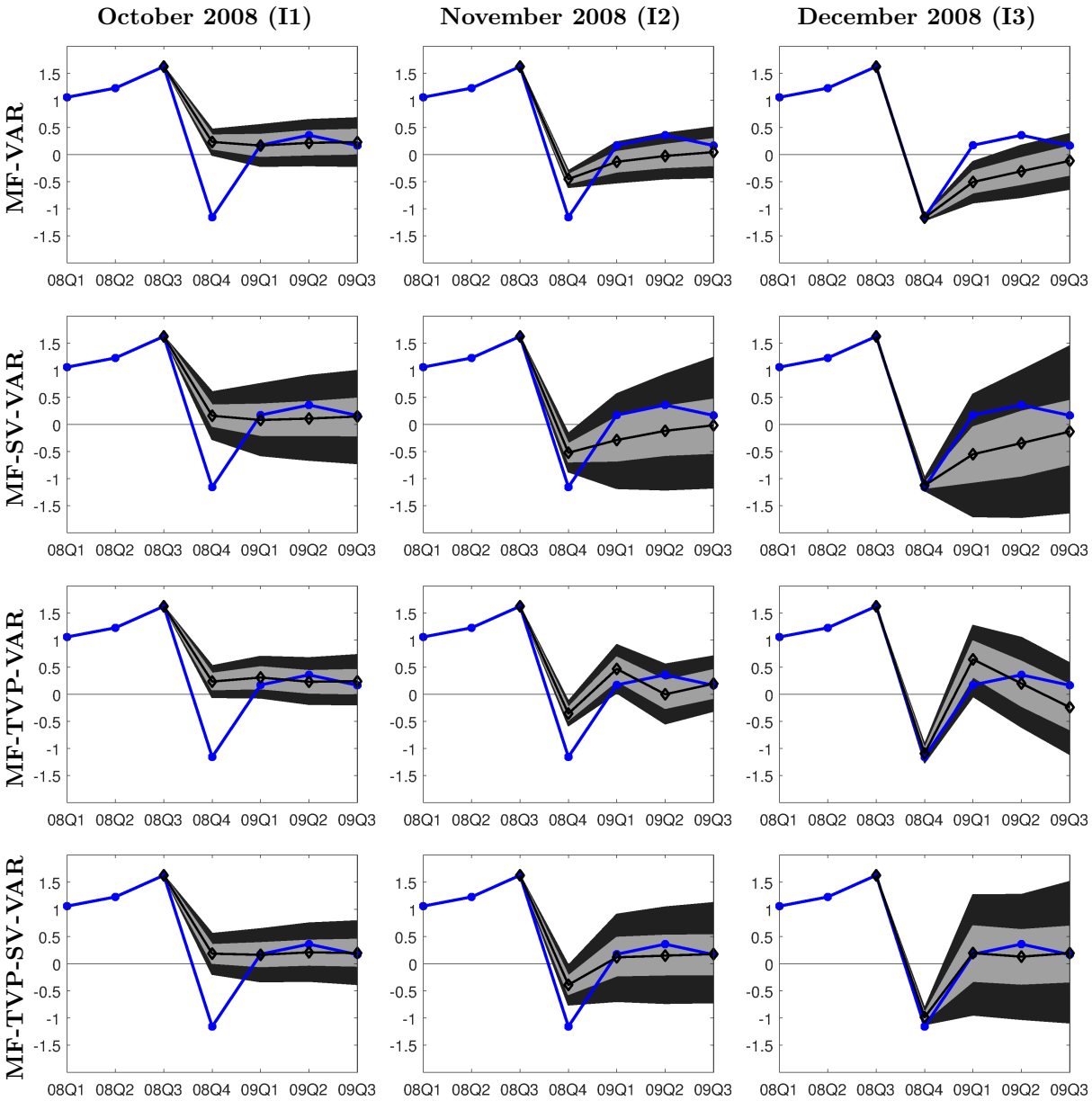


Notes: The rows refer to the PITs of the Q-TVP-SV-VAR and MF-TVP-SV-VAR, respectively. The columns refer to the forecast horizons. The solid line indicates uniformity and the dashed lines 90% confidence bands as in Rossi and Sekhposyan (2014). Sample: 1995-2016.

information on the current quarter except for the T-Bill rate of October. At this data vintage, the models' posterior means are rather close to each other for each horizon—for the current quarter all of them lie at roughly 0.25%, which is approximately 1 percentage point too high compared to the realization. In contrast, the forecast intervals show noteworthy differences. The MF-VAR and the MF-TVP-VAR deliver narrow intervals, which assign only a small fraction of probability mass to negative inflation rates—for the nowcast almost no probability mass. In contrast, the MF-SV-VAR and the MF-TVP-SV-VAR generate much wider intervals, clearly including negative growth rates. However, the realization is not included in any interval. In November 2008, the posterior means are still close to each other, but become much more pessimistic. Now each model correctly anticipates a negative growth rate for the nowcast—the nowcasts drop to about -0.5%. Thus, as indicated in Section 5.5.1, the forecast errors become remarkably smaller due to the additional monthly observations. Moreover, there are considerable differences in the posterior means for higher horizons. The models with fixed VAR-coefficients predict a very slow recovery with negative inflation rates until 2009Q3. The TVP-models predict—in line with the realizations—positive rates from 2009Q1 onward; the MF-TVP-SV-VAR almost exactly predicts the growth rate for 2009Q1. The same pattern holds for the forecasts from December 2008. Now each model produces a forecast error of almost zero for the nowcast with a very narrow forecast interval. The subsequent recovery, however, is much better predicted by the TVP-models. In summary, these results illustrate that the mixed-frequency models can translate intra-quarterly information into more precise point and density forecasts. Furthermore, this example supports the findings from Sections 5.5.1 and 5.5.2; it demonstrates the importance of stochastic volatility

for accurate nowcasts and the relevancy of time-varying parameters for precise forecasts. To improve forecast accuracy on average it is recommended to combine both specifications.

Figure 5.3: Inflation forecasts during the Great Recession



Notes: The rows refer to mixed-frequency models. The columns refer to the forecast origins, i.e., the information sets. The blue line indicates quarter-on-quarter real-time inflation growth; the black line is the mean of the predictive distribution. Shaded areas are 60% and 90% error bands from the predictive distributions.

5.6 Conclusion

Several studies show that modeling structural change improves forecast accuracy. We contribute to this discussion by investigating whether allowing for structural change in a mixed-frequency setup further improves performance. We use a Bayesian VAR that incorporates both time-varying parameters and stochastic volatility and can handle indicators sampled at different frequencies.

We conduct a rigorous real-time out-of-sample forecast experiment and generate predictions for GDP growth, CPI inflation, the unemployment rate, and the 3-month Treasury bill rate. Our findings show that modeling monthly dynamics results in substantially better forecast accuracy. Nowcasts and short-term forecasts especially benefit from within-quarter information, while for longer horizons, the advantages vanish. The MF-TVP-SV-VAR provides, on average, the best point and density forecast performance. In particular, inflation forecasts benefit considerably from modeling both monthly dynamics and structural change. For the remaining variables, the picture is more cloudy. The MF-SV-VAR delivers the best forecasts for the interest rate, while the MF-TVP-VAR provides superior forecasts for the unemployment rate. We obtain rather mixed results for the GDP growth rate forecasts; no model dominates over all horizons, though almost all nonlinear MF-models outperform their linear counterparts as well as the remaining quarterly models. Finally, we assess forecast performance during the Great Recession and demonstrate how the inflow of monthly information alters the inflation forecasts. We show that SV-models achieve the best performance for the downturn, while TVP-models are more precise in the subsequent recovery. Using the combined specification (MF-TVP-SV-VAR) is superior, on average.

Our models are small-scale VARs due to the large number of parameters that have to be estimated and our variables are rather standard in the literature. Thus, future research should focus on how to process a larger dataset in this model framework and on how to select the most informative indicators.

Table 5.3: Real-time forecast CRPS

Model	1995-2016			2008-2016		
	h = 2	h = 3	h = 4	h = 2	h = 3	h = 4
	GDP growth					
MF-TVP-SV-VAR	1.05	1.00	**0.97**	1.03	**0.93**	**0.93***
MF-SV-VAR	1.04	1.01	0.98	1.00	1.00	0.98
MF-TVP-VAR	**0.99**	1.05*	1.06***	**0.94***	0.96	1.03
MF-VAR	1.13***	1.17***	1.13***	1.15***	1.21***	1.19***
Q-TVP-SV-VAR	1.02	**0.98**	0.98	0.99	0.95	0.95
Q-TVP-VAR	1.15***	1.15***	1.14***	1.12***	1.08	1.11**
Q-VAR	1.19***	1.21***	1.18***	1.27***	1.24***	1.25***
Q-SV-VAR	0.35	0.34	0.36	0.44	0.46	0.46
	Inflation					
MF-TVP-SV-VAR	**0.90*****	**0.88*****	**0.88*****	**0.88****	**0.88*****	**0.89****
MF-SV-VAR	1.00	0.95*	0.95**	0.97	0.92*	0.9**
MF-TVP-VAR	0.97	0.92***	0.93***	0.98	0.93***	0.94
MF-VAR	1.06	1.02	1.05	1.00	0.93	0.91*
Q-TVP-SV-VAR	0.94***	0.91***	0.90***	0.93***	0.9***	0.91***
Q-TVP-VAR	0.96	0.94***	0.96	0.93	0.90	0.95
Q-VAR	1.03	1.06***	1.08***	1.01	1.02	1.00
Q-SV-VAR	0.12	0.12	0.12	0.15	0.15	0.14
	Unemployment rate					
MF-TVP-SV-VAR	0.84***	0.90***	0.94	0.89	0.95	0.97
MF-SV-VAR	0.83***	**0.87*****	**0.90*****	**0.82*****	**0.86*****	**0.88*****
MF-TVP-VAR	**0.83*****	0.87***	0.92**	0.85**	0.89*	0.91
MF-VAR	0.89***	0.93**	0.96	0.86***	0.90**	0.93*
Q-TVP-SV-VAR	1.02	1.03	1.02	1.05**	1.05**	1.04
Q-TVP-VAR	1.03	1.05	1.04	1.02	1.02	1.01
Q-VAR	1.10***	1.10***	1.11***	1.08***	1.10***	1.11***
Q-SV-VAR	0.23	0.35	0.47	0.34	0.53	0.72
	Interest rate					
MF-TVP-SV-VAR	0.53***	0.70***	0.78***	0.39***	0.52***	0.58***
MF-SV-VAR	**0.52*****	**0.68*****	**0.75*****	**0.37*****	**0.49*****	**0.55*****
MF-TVP-VAR	0.7***	0.91	1.00	0.72***	0.99	1.09
MF-VAR	0.90*	1.03	1.04	0.97	1.14	1.16*
Q-TVP-SV-VAR	0.97*	0.97	0.98	0.85***	0.84***	0.86***
Q-TVP-VAR	1.30***	1.31***	1.29***	1.46***	1.54***	1.53***
Q-VAR	1.43***	1.35***	1.30***	1.73***	1.75***	1.71***
Q-SV-VAR	0.34	0.50	0.66	0.29	0.41	0.55

Notes: The models are detailed in Section 5.3. The scores are reported in absolute terms for the benchmark model (the bottom row of each panel) and as ratios to the benchmark for the remaining models. A ratio below unity indicates that the model outperforms the benchmark. Bold figures indicate the best performance for the variable and horizon. $*$, $**$, and $***$ denote significance at the 15%, 10%, and 5% level, respectively, according to a t-test on the average difference in scores relative to the benchmark model with Newey-West standard errors.

References

Adolfson, M., J. Lindé, and M. Villani (2007). Forecasting Performance of an Open Economy DSGE Model. *Econometrics Review 26*(2-4), 289–328.

Amir-Ahmadi, P., C. Matthes, and M.-C. Wang (2018). Choosing Prior Hyperparameters: With Applications To Time-Varying Parameter Models. *Journal of Business & Economic Statistics* (forthcoming).

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2017). Tracking the Slowdown in Long-Run GDP Growth. *The Review of Economics and Statistics 99*(2), 343–356.

Banbura, M. and A. van Vlodrop (2018). Forecasting with Bayesian Vector Autoregressions with Time Variation in the Mean. Tinbergen Institute Discussion Paper 2018-025/IV, Tinbergen Institute.

Barnett, A., H. Mumtaz, and K. Theodoridis (2014). Forecasting UK GDP Growth and Inflation Under Structural Change. A Comparison of Models with Time-Varying Parameters. *International Journal of Forecasting 30*(1), 129–143.

Barsoum, F. and S. Stankiewicz (2015). Forecasting GDP Growth Using Mixed-Frequency Models with Switching Regimes. *International Journal of Forecasting 31*(1), 33–50.

Baumeister, C. and L. Benati (2013). Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound. *International Journal of Central Banking 9*(2), 165–212.

Bessec, M. and O. Bouabdallah (2015). Forecasting GDP over the Business Cycle in a Multi-Frequency and Data-Rich Environment. *Oxford Bulletin of Economics and Statistics 77*(3), 360–384.

Carriero, A., T. Clark, and M. Marcellino (2013). Real-Time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility. Working Paper 9312, Centre for Economic Policy Research, London.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Realtime Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 178*(4), 837–862.

Carriero, A., T. E. Clark, and M. Marcellino (2016). Common Drifting Volatility in Large Bayesian VARs. *Journal of Business & Economic Statistics 34*(3), 375–390.

Carter, C. K. and R. Kohn (1994). On Gibbs Sampling for State Space Models. *Biometrika 81*(3), 541–553.

Chan, J. C. C. and E. Eisenstat (2017). Bayesian Model Comparison for Time-Varying Parameter VARs with Stochastic Volatility. *Journal of Applied Econometrics 33*(4), 509–532.

Chiu, C.-W. J., H. Mumtaz, and G. Pintér (2017). Forecasting with VAR models: Fat Tails and Stochastic Volatility. *International Journal of Forecasting 33*(4), 1124–1143.

Cimadomo, J. and A. D'Agostino (2016). Combining Time Variation and Mixed Frequencies: An Analysis of Government Spending Multipliers in Italy. *Journal of Applied Econometrics 31*(7), 1276–1290.

Clark, T. E. (2009). Is the Great Moderation Over? An Empirical Analysis. *Economic Review Q IV*, 5–42.

Clark, T. E. (2012). Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility. *Journal of Business & Economic Statistics 29*(3), 327–341.

Clark, T. E. and F. Ravazzolo (2015). Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics 30*(4), 551–575.

Clements, M. P. and A. B. Galvão (2008). Macroeconomic Forecasting With Mixed-Frequency Data: Forecasting Output Growth in the United States. *Journal of Business & Economic Statistics 26*(4), 546–554.

Cogley, T., S. Morozov, and T. J. Sargent (2005). Bayesian Fan Charts for U.K. Inflation: Forecasting and Sources of Uncertainty in an Evolving Monetary System. *Journal of Economic Dynamics & Control 29*(11), 1893–1925.

Cogley, T. and T. J. Sargent (2001). Evolving Post-World War II US inflation Dynamics. In B. S. Bernanke and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2001*, Volume 16, 331–373. National Bureau of Economic Research, Inc.

Cogley, T. and T. J. Sargent (2005). Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S. *Review of Economic Dynamics 8*(3), 262–302.

D'Agostino, A., L. Gambetti, and D. Giannone (2013). Macroeconomic Forecasting and Structural Change. *Journal of Applied Econometrics 28*(1), 82–101.

Del Negro, M. and G. E. Primiceri (2015). Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum. *Review of Economic Studies 82*(4), 1342–1345.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review 39*(4), 863–883.

Diebold, F. X. and R. S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Number 9780198523543 in OUP Catalogue. Oxford University Press.

Faust, J. and J. H. Wright (2013). Forecasting Inflation. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Chapter 1, 2–56. Elsevier.

Foroni, C., P. Guérin, and M. Marcellino (2015). Markov-switching Mixed-Frequency VAR models. *International Journal of Forecasting 31*(3), 692–711.

Foroni, C. and M. Marcellino (2013). A Survey of Econometric Methods for Mixed-Frequency Data. Working Paper 2013/06, Norges Bank.

Foroni, C. and M. Marcellino (2014). A Comparison of Mixed Frequency Approaches for Nowcasting Euro Area Macroeconomic Aggregates. *International Journal of Forecasting 30*(3), 554–568.

Gadea Rivas, M., A. Gómez-Loscos, and G. Pérez-Quirós (2014). The Two Greatest. Great Recession vs. Great Moderation. Working Paper 1423, Banco de Espana.

Galí, J. and L. Gambetti (2009). On the Sources of the Great Moderation. *Amercian Economic Journal: Macroeconomics 1*(1), 26–57.

Geweke, J. and G. Amisano (2010). Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns. *International Journal of Forecasting 26*(2), 216–230.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The MIDAS Touch: Mixed Data Sampling Regression Models. CIRANO Working Papers 2004s-20, CIRANO.

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.

Gneiting, T. and A. E. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Gneiting, T. and R. Ranjan (2011). Comparing Density Forecasts Using Threshold and Quantile Weighted Scoring Rules. *Journal of Business & Economic Statistics 29*(3), 411–422.

Good, I. J. (1952). Rational Decisions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 14*(1), 107–114.

Huber, F. (2016). Density Forecasting Using Bayesian Global Vector Autoregressions with Stochastic Volatility. *International Journal of Forecasting 32*(3), 818–837.

Jacquier, E., N. G. Polson, and P. E. Rossi (1995). Models and Priors for Multivariate Stochastic Volatility. CIRANO Working Papers 95s-18, CIRANO.

Jore, A. S., J. Mitchell, and S. P. Vahey (2010). Combining Forecast Densities from VARs with Uncertain Instabilities. *Journal of Applied Econometrics 25*(4), 621–634.

Kim, C.-J. and C. R. Nelson (1999). Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov Switching Model of the Business Cycle. *The Review of Economics and Statistics 81*(4), 608–616.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies 65*(3), 361–393.

Koop, G. and D. Korobilis (2014). A New Index of Financial Conditions. *European Economic Review 71*, 101–116.

Kuzin, V., M. Marcellino, and C. Schumacher (2011). MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area. *International Journal of Forecasting 27*(2), 529–542.

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.

Mariano, R. S. and Y. Murasawa (2010). A Coincident Index, Common Factors, and Monthly Real GDP. *Oxford Bulletin of Economics and Statistics 72*(1), 27–46.

Matheson, J. E. and R. L. Winkler (1976). Scoring Rules for Continuous Probability Distributions. *Management Science 22*(10), 1087–1096.

McConnell, M. M. and G. Pérez-Quirós (2000). Output Fluctuations in the United States: What Has Changed since the Early 1980's? *American Economic Review 90*(5), 1464–1476.

Mikosch, H. and S. Neuwirth (2015). Real-Time Forecasting with a MIDAS VAR. Working Paper 15–377, KOF Swiss Economic Institute, ETH Zurich.

Mitchell, J., R. J. Smith, M. R. Weale, S. Wright, and E. L. Salaza (2005). An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth. *The Economic Journal 115*(501), 108–129.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies 72*(3), 821–852.

Rossi, B. and T. Sekhposyan (2014). Evaluating Predictive Densities of US Output Growth and Inflation in a Large Macroeconomic Data Set. *International Journal of Forecasting 30*(3), 662–682.

Schorfheide, F. and D. Song (2015). Real-Time Forecasting with a Mixed-Frequency VAR. *Journal of Business & Economic Statistics 33*(3), 366–380.

Wohlrabe, K. (2009). Forecasting with Mixed-Frequency Time Series Models. Munich Dissertations in Economics 9681, University of Munich, Department of Economics.

Zadrozny, P. A. (1988). Gaussian-Likelihood of Continuous-Time ARMAX Models When Data are Stocks and Flows at Different Frequencies. *Econometric Theory 4*(1), 108–124.

D Appendix

D.1 Priors

Apart from the VAR with constant volatilities, which uses a Jeffrey's prior, the priors for the remaining model specifications are based on a training sample, which consists of the first 10 years of the entire sample. In the following, variables denoted with OLS refer to OLS quantities based on this training sample. The length of the trainings sample is denoted by T_0.

VAR-coefficients

To keep the models comparable, we draw the VAR-coefficients for each nonlinear specification using the CK algorithm with the following prior:

$$p(\beta_0) \sim N(\hat{\beta}_{OLS}, 4 \times V(\hat{\beta}_{OLS})). \tag{D.1}$$

In the case of the VAR-SV, we use the first draw of the backward recursion of the CK algorithm, i.e., $\beta_{T|T}$, for each period. For the benchmark VAR we implement a diffuse Jeffrey's prior:

$$p(\beta, \Sigma) \propto |\Sigma|^{-(n+1)/2}. \tag{D.2}$$

The prior for the covariance of the VAR-coefficients Q follows an inverse-Wishart distribution:

$$p(Q) \sim IW(k_Q^2 \times T_0 \times V(\hat{\beta}_{OLS}), T_0). \tag{D.3}$$

Stochastic volatilities

The stochastic volatilities are drawn via the CK algorithm. Thus, additional priors for the diagonal elements of Σ_0 ($\log \sigma_0$), and the lower-triangular elements of A_0 ($a_{i,0}$), are required. We follow Primiceri (2005) in defining these priors distribution as:

$$p(\log \sigma_0) \sim N(\log \hat{\sigma}_{OLS}, I_n), \tag{D.4}$$

$$p(A_0) \sim N(\hat{A}_{OLS}, 4 \times V(\hat{A}_{OLS})). \tag{D.5}$$

The priors for the covariance of $\log \sigma_0$ and A_0 are inverse-Wishart distributed:

$$p(\Psi) \sim IW(k_\Psi^2 \times (1+n) \times I_n, 4), \tag{D.6}$$

$$p(\Phi_i) \sim IW(k_\Phi^2 \times (i+1) \times V(\hat{A}_{i,OLS}), i+1), \qquad i = 1, \ldots, k-1, \tag{D.7}$$

where i denotes the respective VAR-equation that has non-zero and non-one elements in the lower-triangular matrix A_t, i.e., for $n=4$ it is equation 2, 3, and 4.

Latent observations

The missing values of the quarterly series expressed at monthly frequency are replaced with an estimated latent state by applying a time-dependent CK algorithm. We initialize the unobserved state variable z_t with z_0 as actual observations from the monthly variables and constant values for the quarterly variables in levels from the last observations of our training sample:

$$p(z_0) \sim N(z_L, I_{np}). \tag{D.8}$$

Hence, $z_L = [\tilde{y}_0', \ldots, \tilde{y}_{0-p+1}']$ where $\tilde{y}_i$ contains actual values, if observed, and constant values in levels, thus zero growth rates, for missing observations.

Hyperparameters

The variability of β_t, a_t, and $\log \sigma_t$ depends on Q, Ψ, and Φ, respectively, and thus on the hyperparameters k_Q, k_Ψ, and k_Φ. Therefore, we follow Amir-Ahmadi et al. (2018) and use priors for those hyperparameters. Since we do not have any a priori knowledge about the true values of any of our models, we use uniform priors:

$$p(k_i) \sim U(1e^{-10}, 1), \qquad i = Q, \Phi, \Psi. \tag{D.9}$$

The lower and upper bound of the distribution are chosen to cover a broad range of possible values, including the default values used by Primiceri (2005).[18]

D.2 Specification of the Gibbs sampler

To estimate the models we employ a Gibbs sampler that consecutively draws from the conditional distribution. In the following, the general form of the MCMC algorithm according to Primiceri (2005) with the Del Negro and Primiceri (2015) correction is outlined. To include the estimation of the hyperparameters, an additional Metropolis Hastings step is added to the Gibbs sampler. Denoting any vector of variables x over the sample T by $x^T = [x_1', \ldots, x_T']'$, the Gibbs sampler takes the following form:

1. Initialize β_t, Σ^T, A^T, s^T, Q, Ψ, Φ, k_Q, k_Φ, and k_Ψ.
2. Draw $\tilde{y}^T$ from $p(\tilde{y}^T | y^T, \beta^T, Q, \Sigma^T, A^T, \Psi, \Phi)$.
3. Draw β^T from $p(\beta^T | \tilde{y}^T, Q, \Sigma^T, A^T, \Psi, \Phi)$.

[18] To ensure convergence of the MH-algorithm in the case of the MF-TVP-SV-VAR, the lower bound for k_Q is choosen to be higher than the value in Primiceri (2005).

4. Draw Q from $p(Q|\tilde{y}^T, \beta^T, \Sigma^T, A^T, \Psi, \Phi)$.
5. Draw A^T from $p(A^T|\tilde{y}^T, \beta^T, Q, \Sigma^T, \Psi, \Phi)$.
6. Draw Φ from $p(\Phi|\tilde{y}^T, \beta^T, Q, \Sigma^T, A^T, \Psi)$.
7. Draw Ψ from $p(\Psi|\tilde{y}^T, \beta^T, Q, \Sigma^T, A^T, \Phi)$.
8. Draw s^T from $\tilde{p}(s^T|\tilde{y}^T, \beta^T, Q, \Sigma^T, A^T, \Psi, \Phi)$.
9. Draw Σ^T from $\tilde{p}(\Sigma^T|\tilde{y}^T, \beta^T, Q, A^T, s^T, \Psi, \Phi)$.
10. Draw k_Q from $p(k_Q|Q) = p(Q|k_Q)p(k_Q)$.
    Draw k_Ψ from $p(k_\Psi|\Psi) = p(\Psi|k_\Psi)p(k_\Psi)$.
    Draw k_ϕ from $\prod_{i=1}^{k-1} p(k_\Phi|\Phi_i) = p(\Phi_i|k_\Phi)p(k_\Phi)$.

The second step of this Gibbs sampler refers to drawing the latent observations. Since there are no latent observations in the quarterly models, the Gibbs sampler omits Step 2 for these models. Steps 3 to 8 belong to the block of drawing the joint posterior of $\tilde{p}(\theta, s^T|\tilde{y}^T, \Sigma^T)$ by drawing θ from $p(\theta|\tilde{y}^T, \Sigma^T)$ where $\theta = [\beta^T, A^T, Q, \Phi, \Psi]$. Subsequently, we draw s^T from $\tilde{p}(s^T|\tilde{Y^T}, \Sigma^T, \theta)$, and then Σ_t from $\tilde{p}(\Sigma_t|s^T, \theta)$. $\tilde{p}$ denotes the draws based on the approximate likelihood due to the KSC step, while p refers to draws based on the true likelihood (for further detail, see Del Negro and Primiceri, 2015). In Step 10, we include the Metropolis-Hastings within the Gibbs sampler to draw our hyperparameters.

For ease of exposition, in the following we use $\tilde{y}^T$ to indicate the data used in each step of the algorithm. If one considers quarterly models, however, $\tilde{y}^T$ has to be replaced by y^T. We employ 50000 burn-in iterations of the Gibbs sampler for each model and use every 5th draw of 10000 after burn-in draws for posterior inference.

Step 2: Drawing latent states z_t

Let $z_T = [z_1, \ldots, z_T]$ denote the sequence of state vectors consisting of the unobserved monthly states. Draws for z_t are obtained by using the CK algorithm, i.e., we run the Kalman filter until T to obtain $z_{T|T}$ as well as $P_{T|T}$ and draw z_T from $N(z_{T|T}, P_{T|T})$. Subsequently, for $t = T-1, \ldots, 1$, we draw z_t from $N(z_{t|t}, P_{t|t})$ by recursively updating $z_{t|t}$ and $P_{t|t}$.

Step 3: Drawing the VAR-coefficient β^T

Conditional on the drawn states or the actual data, sampling the AR-coefficients proceeds as in Step 2 using the CK algorithm.

Step 4: Drawing the covariance of the VAR-coefficients Q

The posterior of the covariance of VAR-coefficients is inverse-Wishart distributed with scale

matrix $\overline{Q} = Q_0 + e_t'e$, $e_t = \Delta\beta_t'$, and degrees of freedom $df_Q = T + T_0$, where Q_0 and T_0 denote the prior scale for Q and prior degrees of freedom, respectively.

Step 5: Drawing the elements of A^T

To draw the elements of A_T we follow Primiceri (2005) and rewrite the VAR in (5.6) as follows:

$$A_t(\tilde{y}_t - Z_t'\beta_t) = \tilde{y}_t^* = \Sigma_t u_t, \tag{D.10}$$

where, taking into account that β_T and $\tilde{y}_t$ are known, y_t^* is observable. Due to the lower-triangular structure of A_t^{-1}, this system can be written as a system of k equations:

$$\hat{\tilde{y}}_{1,t} = \sigma_{1,t}u_{1,t}, \tag{D.11}$$

$$\hat{\tilde{y}}_{i,t} = -\hat{y}_{[1,i-1]}a_{i,t} + \sigma_{i,t}u_{i,t}, \qquad i = 2, \ldots, k, \tag{D.12}$$

where $\hat{\tilde{y}}_{[1,i-1]} = [\hat{\tilde{y}}_{1,t}, \ldots, \hat{\tilde{y}}_{i-1,t}]$. $\sigma_{i,t}$ and $u_{i,t}$ refer to the i-th elements of σ_t and u_t. Thus, under the block diagonal assumption of Φ, the RHS of equation i does not include $\hat{y}_{i,t}$, implying that one can recursively obtain draws for $a_{i,t}$ by applying an otherwise ordinary CK algorithm equation-wise.

Step 6: Drawing the covariance Φ_i of the elements of A^T

Φ_i has an inverse-ishart posterior with scale matrix $\overline{\Phi}_i = \Phi_{0,i} + \epsilon_{i,t}'\epsilon_{i,t}$, $\epsilon_{i,t} = \Delta a_{i,t}'$, and degrees of freedom $df_{\Phi_i} = T + df_{\Phi_{i,0}}$ for $i = 1, \ldots, k$. $\Phi_{0,i}$, and $df_{\Phi_{i,0}}$ denote prior scale and prior degree of freedoms, respectively.

Step 7: Drawing the covariance Ψ of log-volatilities

As in Step 6, Ψ has an inverse-Wishart distributed posterior with scale matrix $\overline{\Psi} = \Psi_0 + \varepsilon_t'\varepsilon_t$, $\varepsilon_t = \Delta \log \sigma_t'^2$, and degrees of freedom $df_\Psi = T + df_{\Psi_0}$, where Ψ_0 and df_{Ψ_0} denote the prior scale and the prior degree of freedoms, respectively.

Step 8: Drawing the states of the mixture distribution s^T

Conditional on the volatilities, we independently draw a new value for the indicator matrix s^T from (see Kim et al., 1998):

$$PR(s_{i,t} = j|\tilde{y}^{**}, h_{i,t}) \propto q_j f_N(\tilde{y}^{**}|2h_{i,t} + m_j - 1.2704, \nu_j^2). \tag{D.13}$$

Step 9: Drawing the volatilities

The elements of Σ_t are drawn using the KSC algorithm. To this end, we employ the VAR rewritten as in (D.10). Taking squares and logarithms, we get

$$\tilde{y}_t^{**} = 2r_t + \nu_t, \tag{D.14}$$

and for the volatility process:

$$h_t = h_{t-1} + \varepsilon_t, \tag{D.15}$$

where $\tilde{y}_{i,t}^{**} = \log((\tilde{y}_{i,t}^{*})^2 + c)$, $\nu_{i,t} = \log u_{i,t}^2$, $h_{i,t} = \log \sigma_{i,t}$, and c is set to a small but positive number to increase the robustness of the estimation process. To transform this non-Gaussian system (ν_t is distributed according to a χ^2-distribution with one degree of freedom) into a Gaussian system, we resort to Kim et al. (1998) and consider a mixture of seven normal densities with component probabilities q_j, means $m_j - 1.2704$, and variances ν_j^2. The values for $\{q_j, m_j, \nu_j^2\}$ are chosen to match the moments of the log $\chi^2(1)$ distribution are given by Table 5.4.

Table 5.4: Gaussian mixtures for approximating the log-$\chi^2(1)$ distribution

ω	q_j	m_j	ν_j^2
1	0.0073	-10.1300	5.7960
2	0.1056	-3.9728	2.6137
3	0.0000	-8.5669	5.1795
4	0.0440	2.7779	0.1674
5	0.3400	0.6194	0.6401
6	0.2457	1.7952	0.3402
7	0.2575	-1.0882	1.2626

Kim et al. (1998).

Conditional on s^T—the indicator matrix, governing composition of the mixture distribution for every ν_t, $t = 1, \ldots, T$—the CK algorithm enables us to recursively get draws for:

$$h_{t|t+1} = E(h_t|h_{t+1}, \tilde{y}^t, A^T, B^T, Q, s^T, \Psi, \Phi), \tag{D.16}$$

$$H_{t|t+1} = VAR(h_t|h_{t+1}, \tilde{y}^t, A^T, B^T, Q, s^T, \Psi, \Phi). \tag{D.17}$$

Step 10: Drawing the hyperparameters k_Q, k_Ψ, and k_Φ

The prior hyperparameters of the scale matrix of the variance covariance matrix Q, Ψ, and

Φ are drawn with a Metropolis within Gibbs step. Amir-Ahmadi et al. (2018) show that the acceptance probability for each draw i can be simplified to:

$$\alpha_{k_X}^i = min\left(\frac{p(X|k_X^*)p(k_X^*)q(k_X^*|k_X^{i-1})}{p(X|k_X^{i-1})p(k_X^{i-1})q(k_X^{i-1}|k_X^*)}, 1\right), \tag{D.18}$$

where $X = \{Q, \Psi, \Phi\}$. $p(X|k_X^*)$ denotes prior distribution of X, while $p(k_X^*)$ indicates the prior for the hyperparameter. $q(k_X^*|k_X^{i-1})$ labels the proposal distribution. We apply the random walk chain algorithm:

$$k_X^* = k_X^{i-1} + \xi_t, \qquad \xi_t \sim N(0, \sigma_{k_X}^2), \tag{D.19}$$

and the standard deviation σ_{k_X} is adjusted in every 500th step of the burn-in period by:

$$\sigma_{k_X}^{adj} = \sigma_{k_X} \frac{\overline{\alpha}_{k_X}}{\alpha^*}, \tag{D.20}$$

where $\overline{\alpha}_{k_X}$ is the average acceptance rate over the 500 draws and $\alpha^* = 0.4$ is the target acceptance rate. We initialize k_X with the values used by Primiceri (2005), $k_Q = 0.01$, $k_\Psi = 0.1$, and $k_\Phi = 0.01$, and the standard deviation by $\sigma_{k_X} = 0.01$.

D.3 Log scores

Table 5.5: Real-time forecast log scores

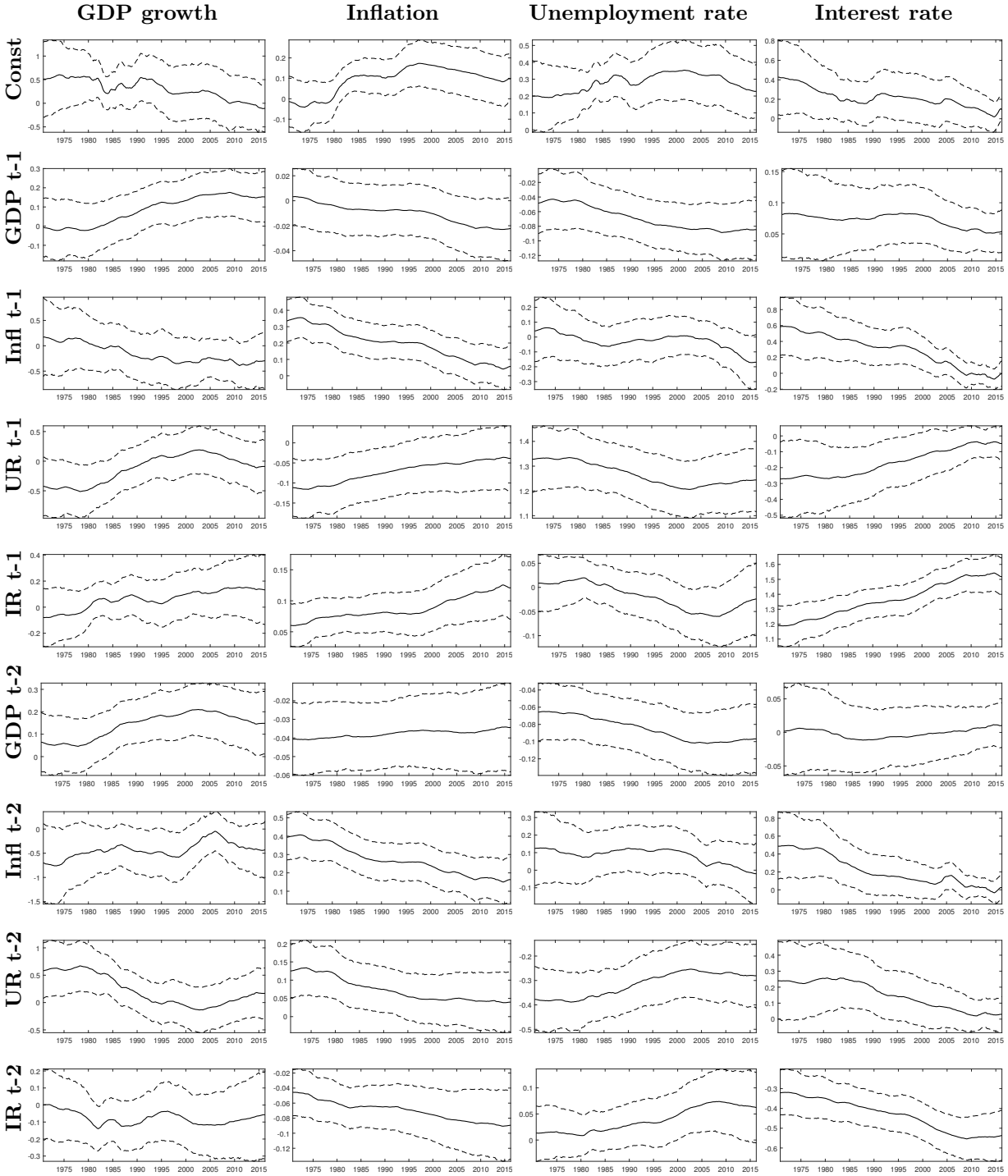
Model	1995-2016			2008-2016		
	h = 2	**h = 3**	**h = 4**	**h = 2**	**h = 3**	**h = 4**
	GDP growth					
MF-TVP-SV-VAR	-0.33***	**0.04**	**0.08***	-0.47	**0.18*****	**0.17*****
MF-SV-VAR	-1.61***	-0.16	0.01	-3.03**	-0.37	0
MF-TVP-VAR	-0.04	-0.15***	-0.15***	0.02	-0.03	-0.11***
MF-VAR	-0.2***	-0.21***	-0.17***	-0.35***	-0.28***	-0.26***
Q-TVP-SV-VAR	**0**	0.01	0	**0.04**	0.08	0.05
Q-TVP-VAR	-0.16***	-0.24***	-0.24***	-0.1*	-0.12	-0.17*
Q-VAR	-0.21***	-0.26***	-0.24***	-0.28***	-0.27***	-0.3***
Q-SV-VAR	-0.94	-0.91	-0.97	-1.19	-1.24	-1.25
	Inflation					
MF-TVP-SV-VAR	**0.07**	**0.12*****	**0.16*****	**0.11**	**0.09**	**0.15*****
MF-SV-VAR	-0.03	0.03	0.05	0.04	0.09	0.12**
MF-TVP-VAR	-0.81	-0.65	-0.17	-2.01	-1.71	-0.5
MF-VAR	-2.23*	-0.62	-0.52	-5.35	-1.38	-1.05
Q-TVP-SV-VAR	0.07***	0.07***	0.1***	0.11***	0.03	0.04
Q-TVP-VAR	-0.53	-0.3	-0.33	-1.24	-0.71	-0.76
Q-VAR	-1.56	-0.51	-0.63	-3.75	-1.1	-1.34
Q-SV-VAR	0.18	0.09	0.11	-0.06	-0.1	-0.05
	Unemployment rate					
MF-TVP-SV-VAR	-0.02	0.38**	**0.34***	0.06	0.66*	0.63
MF-SV-VAR	**0.21*****	0.34**	0.14	**0.3*****	0.66	0.27
MF-TVP-VAR	0.21***	**0.35****	0.3	0.2**	0.75***	0.71
MF-VAR	-1.07	-0.67	-0.82	-2.79	-1.68	-2.02
Q-TVP-SV-VAR	0.04	0.24	0.26	0.04	0.56	0.61
Q-TVP-VAR	0.05	0.26	0.31	0.23	**0.81**	**0.93**
Q-VAR	-0.63***	-0.23	-0.31**	-1.37**	-0.4	-0.63*
Q-SV-VAR	-0.55	-1.18	-1.52	-1.02	-2.09	-2.55
	Interest rate					
MF-TVP-SV-VAR	0.77***	0.43***	0.3***	0.9***	0.59***	0.46***
MF-SV-VAR	**0.8*****	**0.47*****	**0.35*****	**1.03*****	**0.7*****	**0.58*****
MF-TVP-VAR	0.15	-0.12	-0.18**	-0.18*	-0.44***	-0.49***
MF-VAR	-0.14	-0.27***	-0.23***	-0.5***	-0.59***	-0.53***
Q-TVP-SV-VAR	0.01	0	-0.04	0.02	0.02	0.02
Q-TVP-VAR	-0.46***	-0.48***	-0.45***	-0.83***	-0.83***	-0.82***
Q-VAR	-0.6***	-0.52***	-0.42***	-1.01***	-0.93***	-0.82***
Q-SV-VAR	-0.8	-1.18	-1.49	-0.37	-0.78	-1.12

Notes: The models are detailed in Section 5.3. The scores are reported in absolute terms for the benchmark (the bottom row of each panel) and in differences to the benchmark for the remaining models. A positive difference indicates that the model outperforms the benchmark. Bold figures indicate the best performance for the variable and horizon. $*$, $**$, and $***$ denote significance at the 15%, 10%, and 5% level, respectively, according to a t-test on the average difference in scores relative to the benchmark model with Newey-West standard errors.

D.4 Additional figures

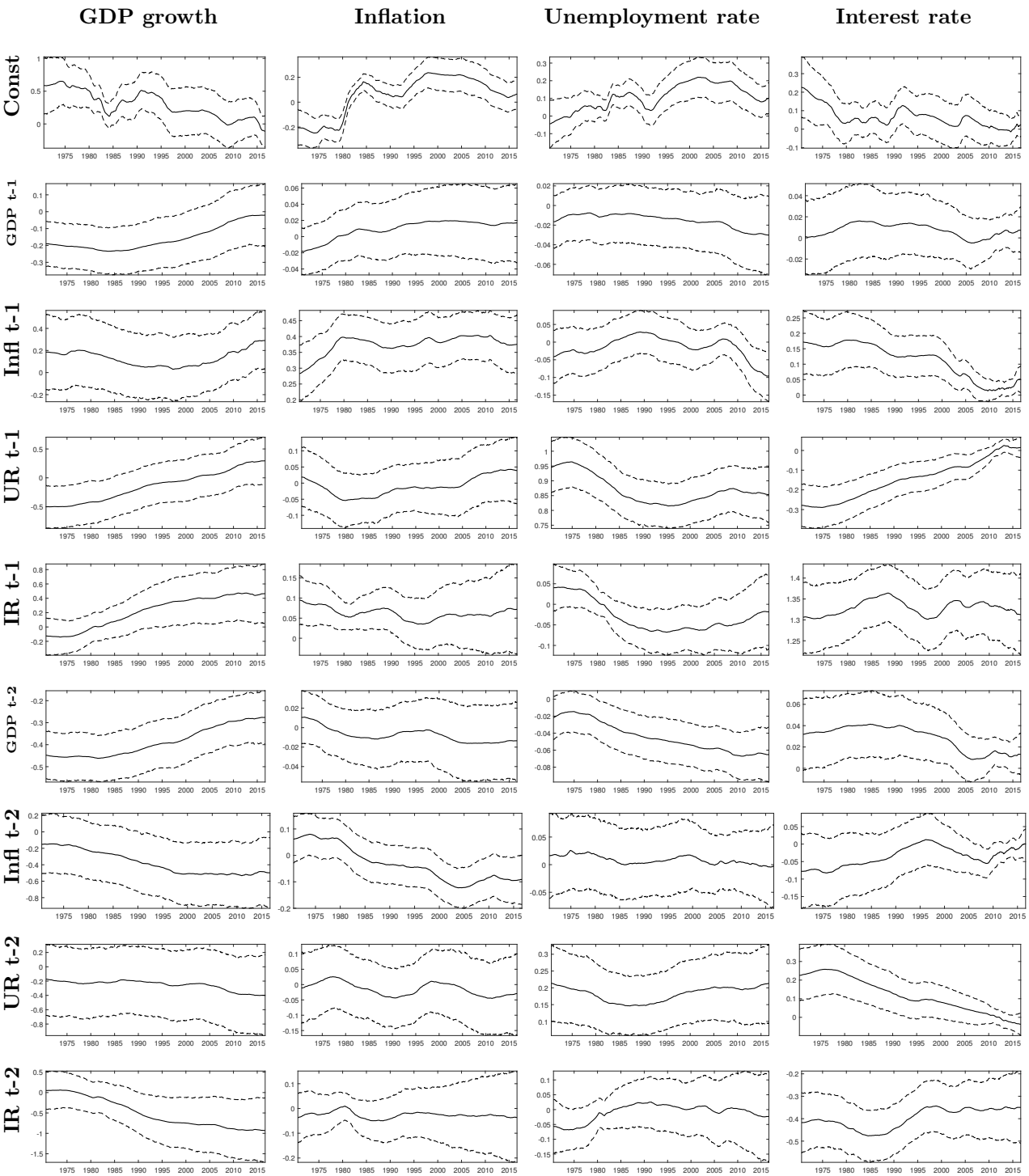
Time-Varying Parameters

Figure 5.4: Time-varying parameters of the Q-TVP-SV-VAR

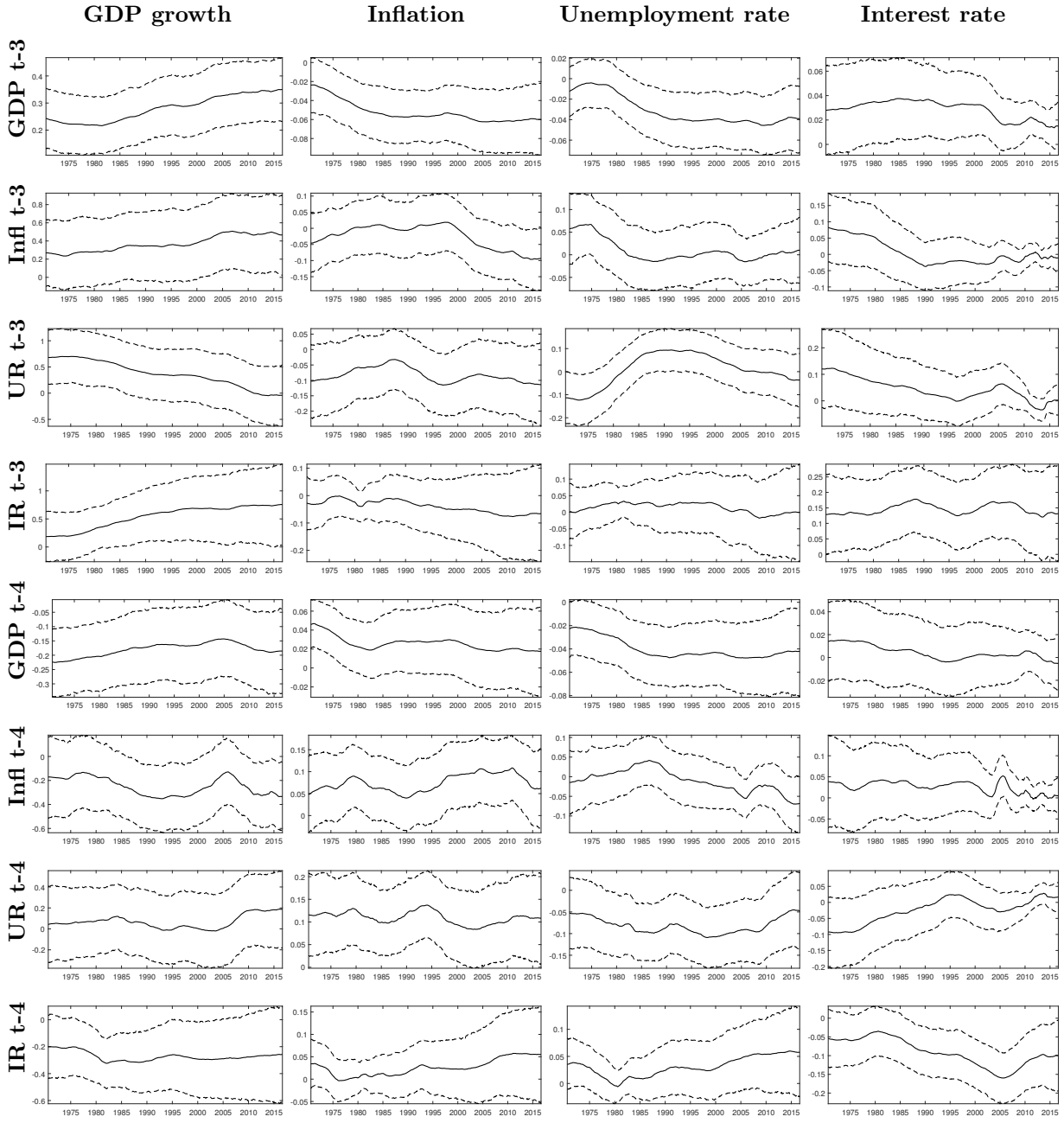


Notes: Figure depicts the time-varying parameters from the Q-TVP-SV-VAR. Columns refer to the variable and rows to the constant/lagged variable on which the variable is regressed. The dashed lines indicate 68% error bands. Results are based on the last data vintage

Figure 5.5: Time-varying parameters of the MF-TVP-SV-VAR



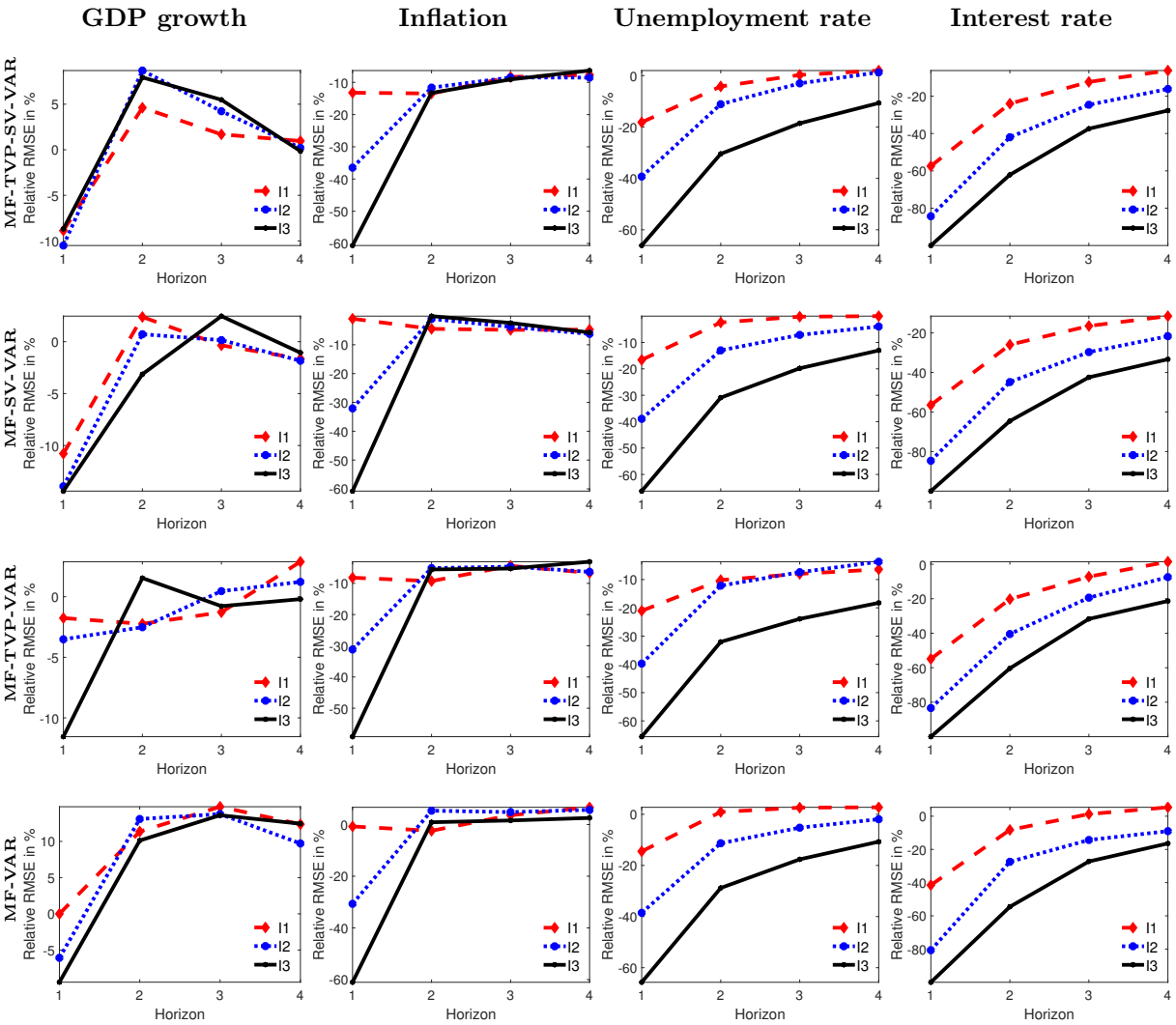
Time-varying parameters of the MF-TVP-SV-VAR (continued)



Notes: Figure depicts the time-varying parameters from the MF-TVP-SV-VAR. Columns refer to the variable and rows to the constant/lagged variable on which the variable is regressed. The dashed lines indicate 68% error bands. Results are based on the last data vintage.

Relative RMSEs

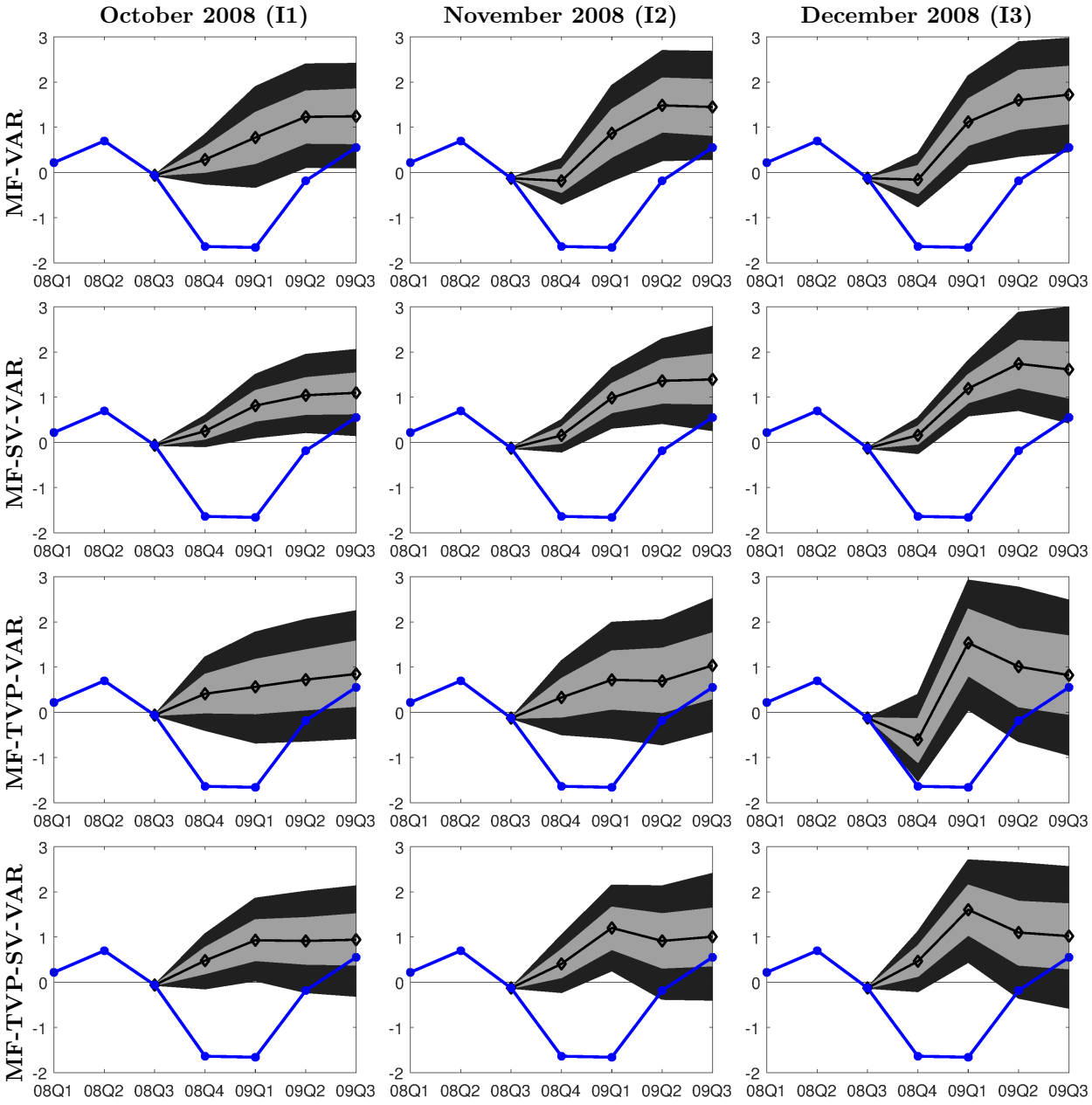
Figure 5.6: Relative RMSEs



Notes: Figure depicts the relative RMSEs in terms of percentage gains compared to the benchmark model. Red, blue, and black lines refer to the information sets I1, I2, and I3 as outlined in Section 5.2.2, respectively.

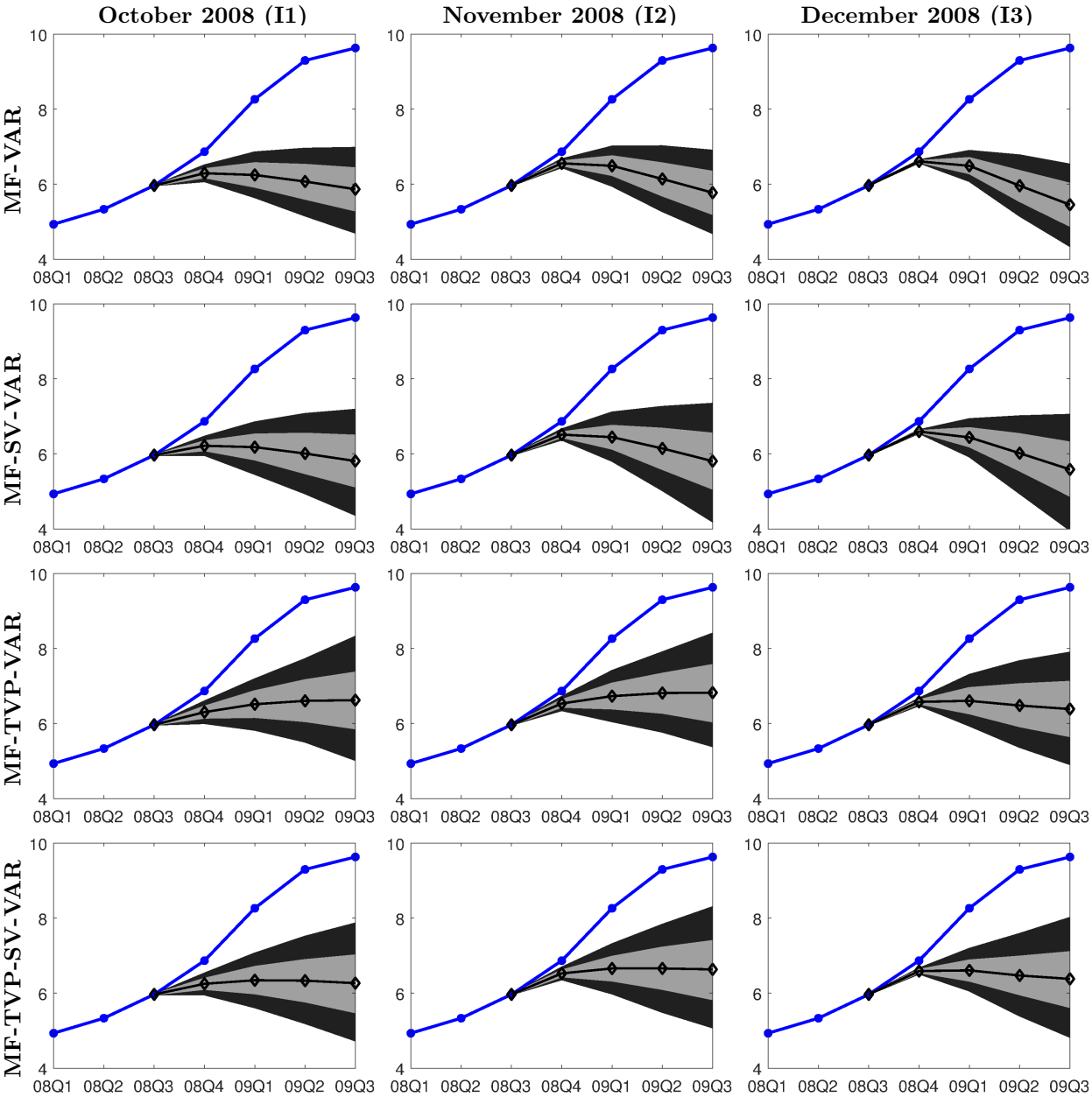
Forecasting during the Great Recession - GDP growth, unemployment rate, and interest rate

Figure 5.7: GDP growth forecasts during the Great Recession



Notes: The rows refer to mixed-frequency models. The columns refer to the forecast origins, i.e., the information sets. The blue line indicates quarter-on-quarter real-time GDP growth; the black line is the mean of the predictive distribution. Shaded areas are 60% and 90% error bands from the predictive distributions.

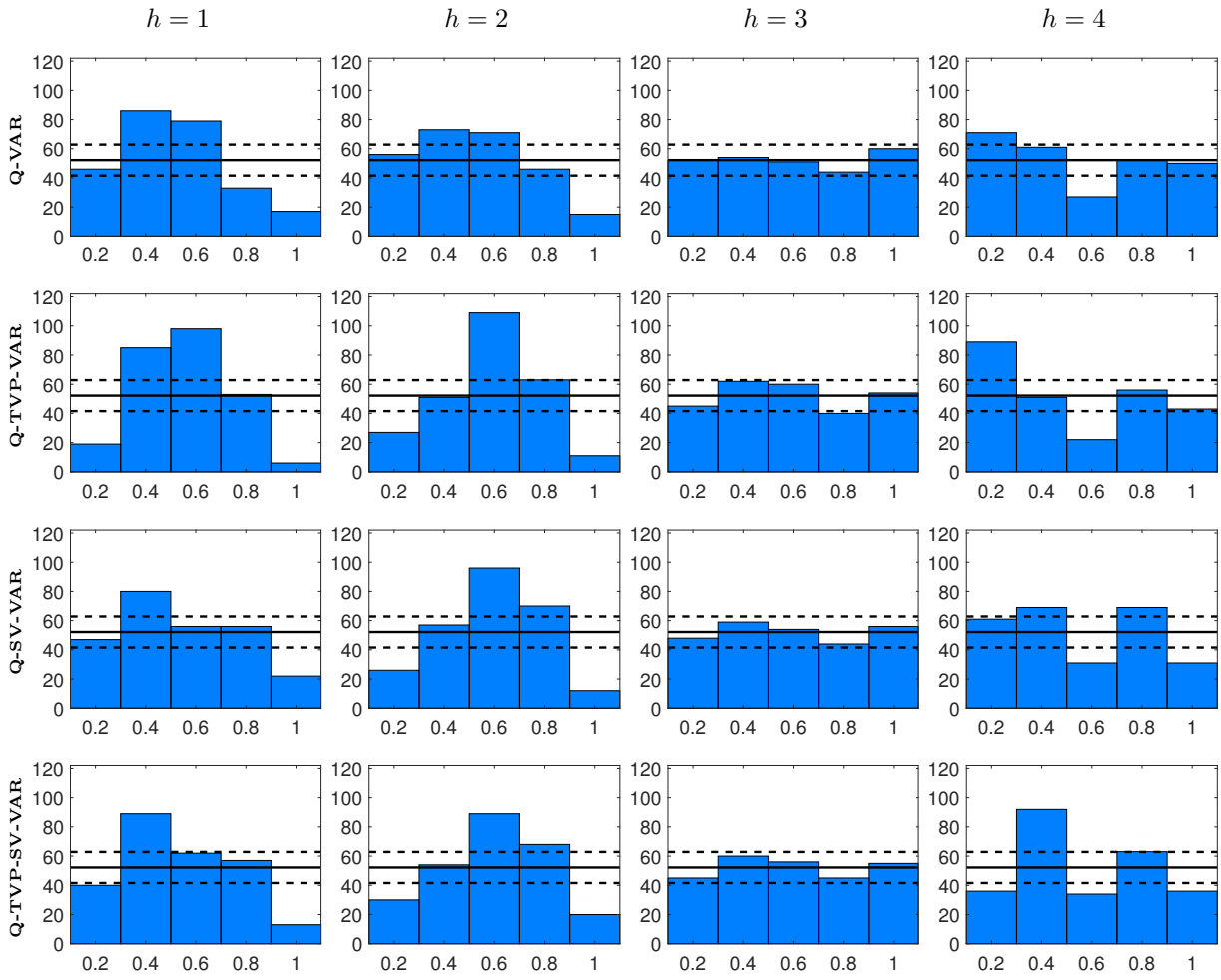
Figure 5.8: Unemployment rate forecasts during the Great Recession



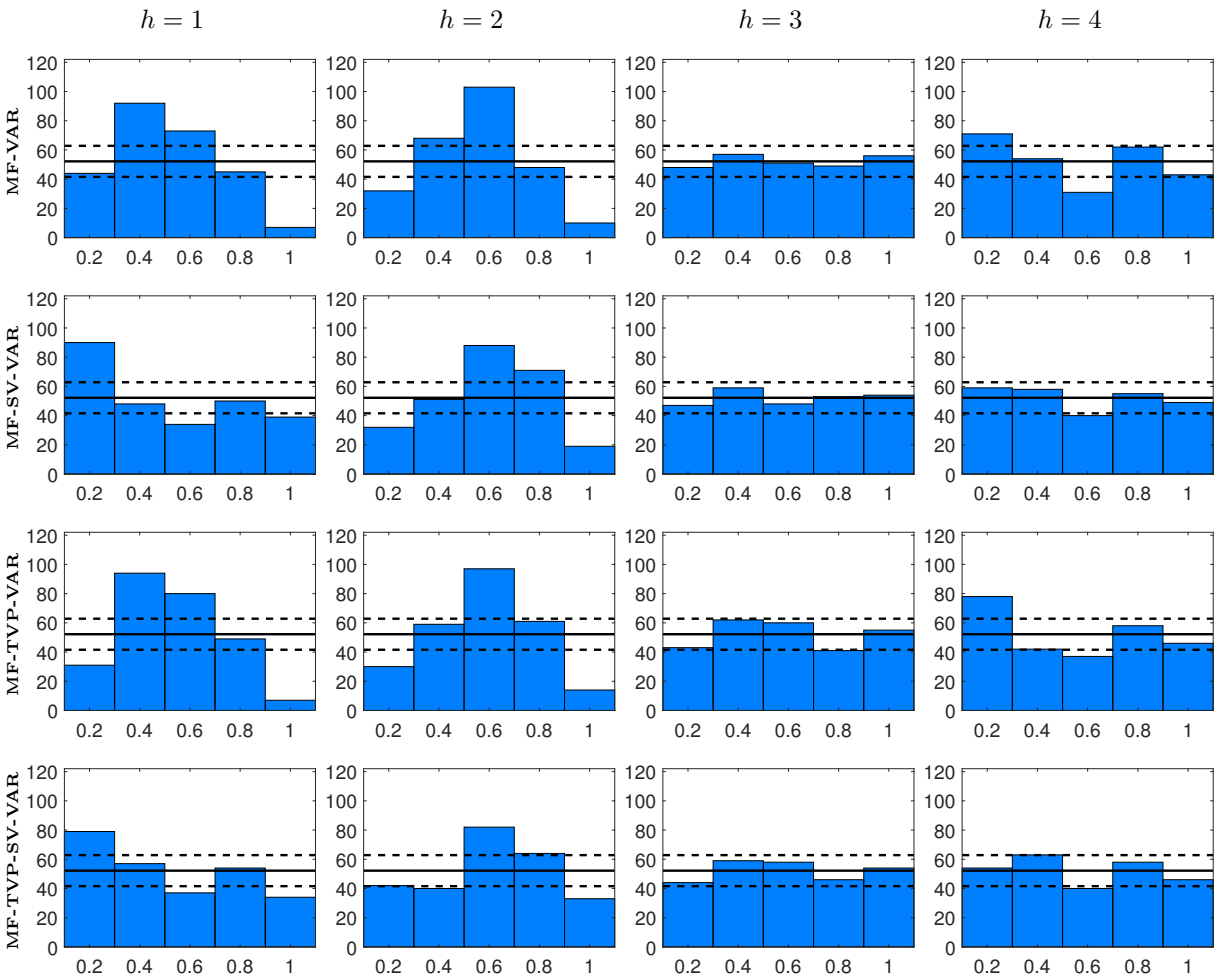
Notes: The rows refer to mixed-frequency models. The columns refer to the forecast origins, i.e., the information sets. The blue line indicates quarter-on-quarter real-time unemployment rate; the black line is the mean of the predictive distribution.Shaded areas are 60% and 90% error bands from the predictive distributions.

Probability integral transforms

Figure 5.9: Probability integral transforms for inflation forecasts

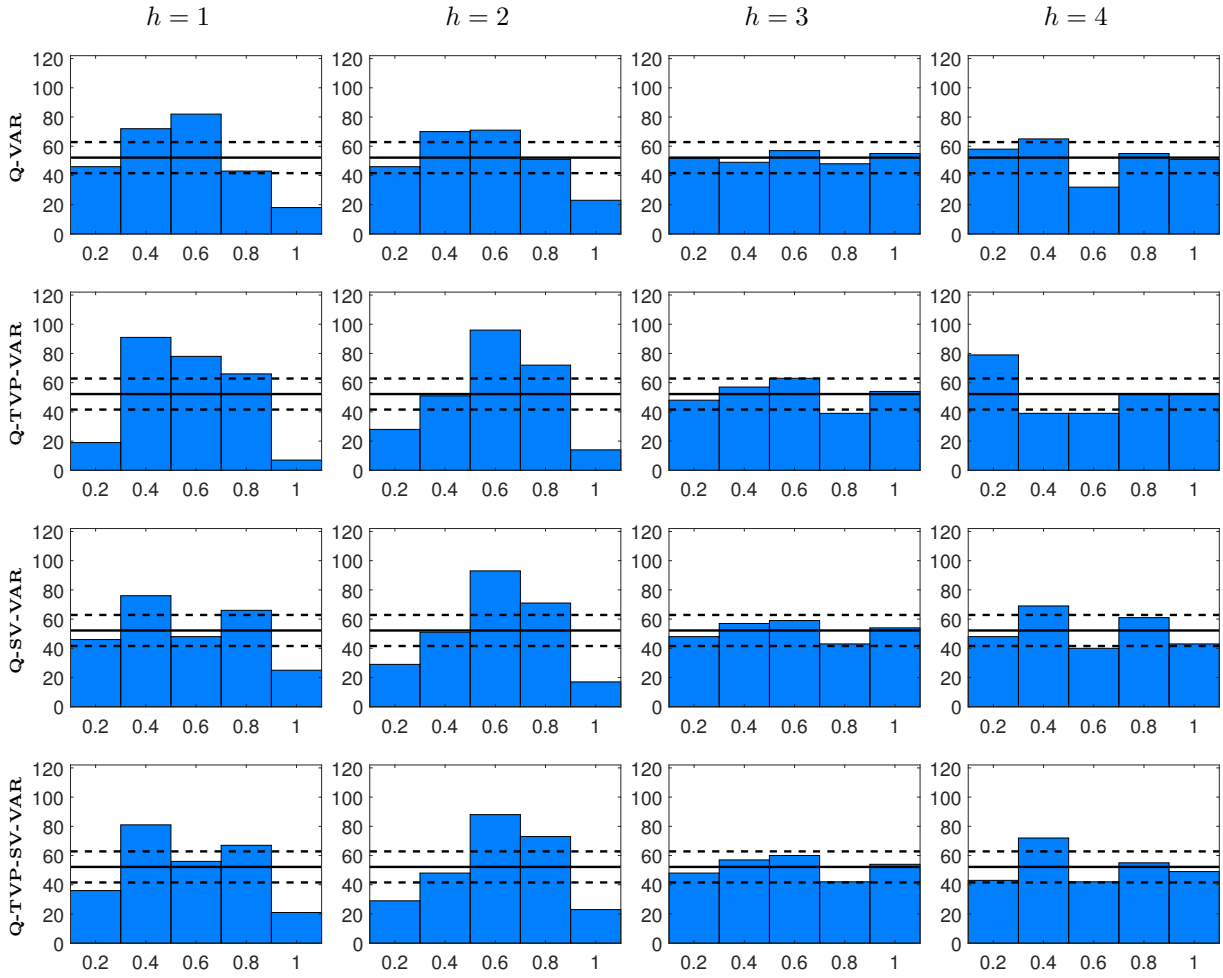


Probability integral transforms for inflation Forecasts (continued)

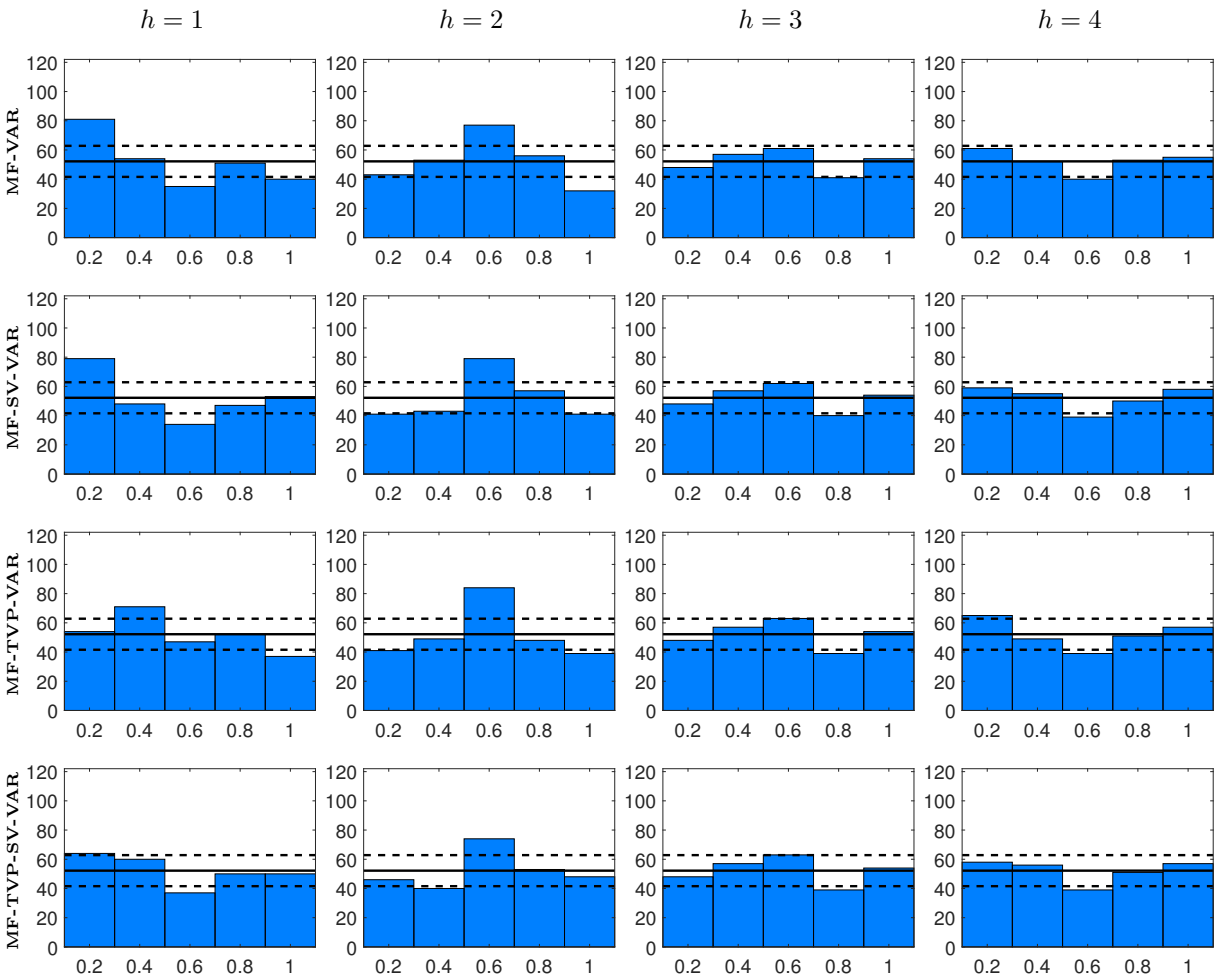


Notes: The rows refer to the PITs of each model. The columns refer to the forecast horizons. The solid line indicates uniformity and the dashed lines 90% confidence bands as in Rossi and Sekhposyan (2014).

Figure 5.10: Probability integral transforms for GDP growth forecasts

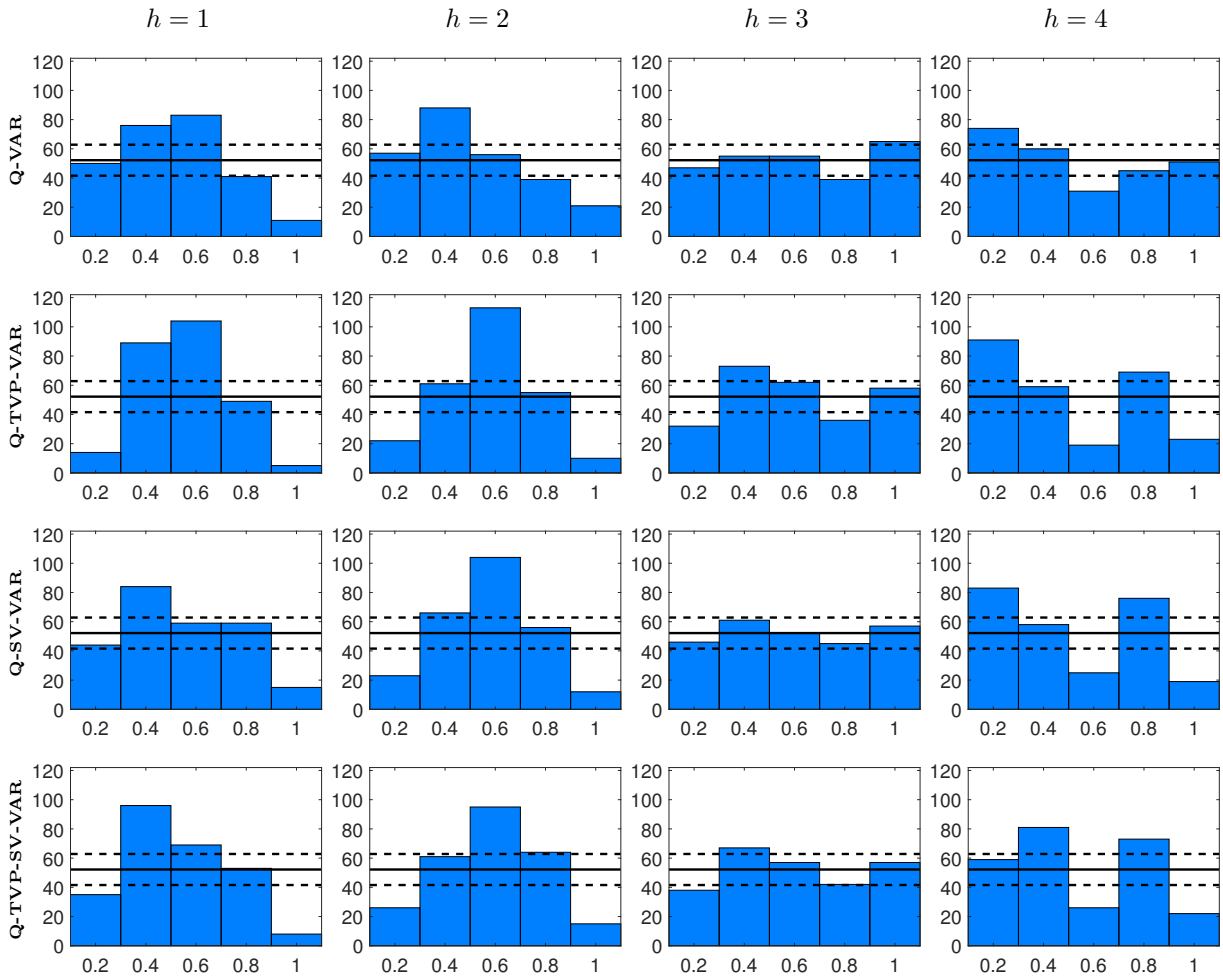


Probability integral transforms for GDP growth forecasts (continued)

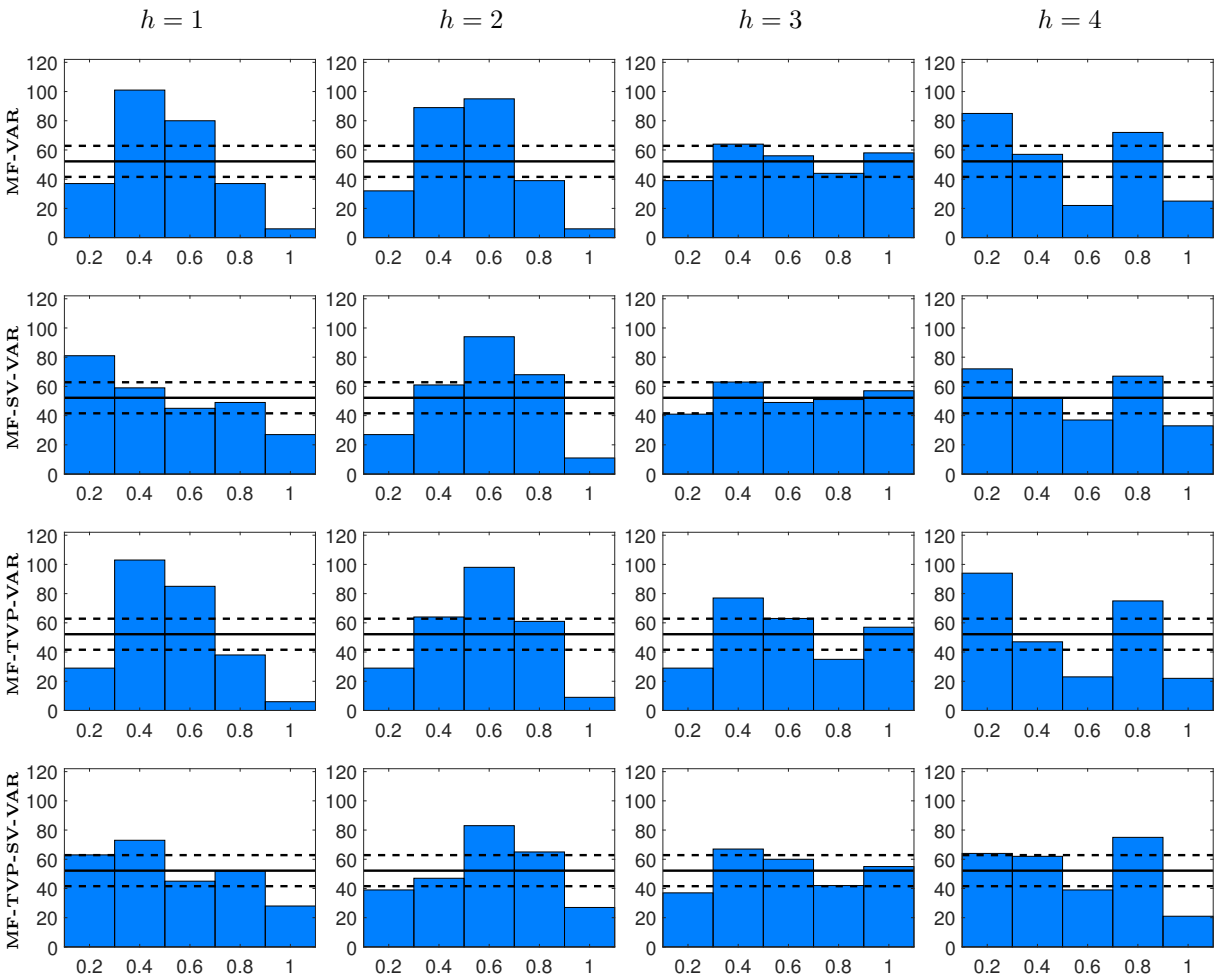


Notes: The rows refer to the PITs of each model. The columns refer to the forecast horizons. The solid line indicates uniformity and the dashed lines 90% confidence bands as in Rossi and Sekhposyan (2014).

Figure 5.11: Probability integral transforms for unemployment rate forecasts

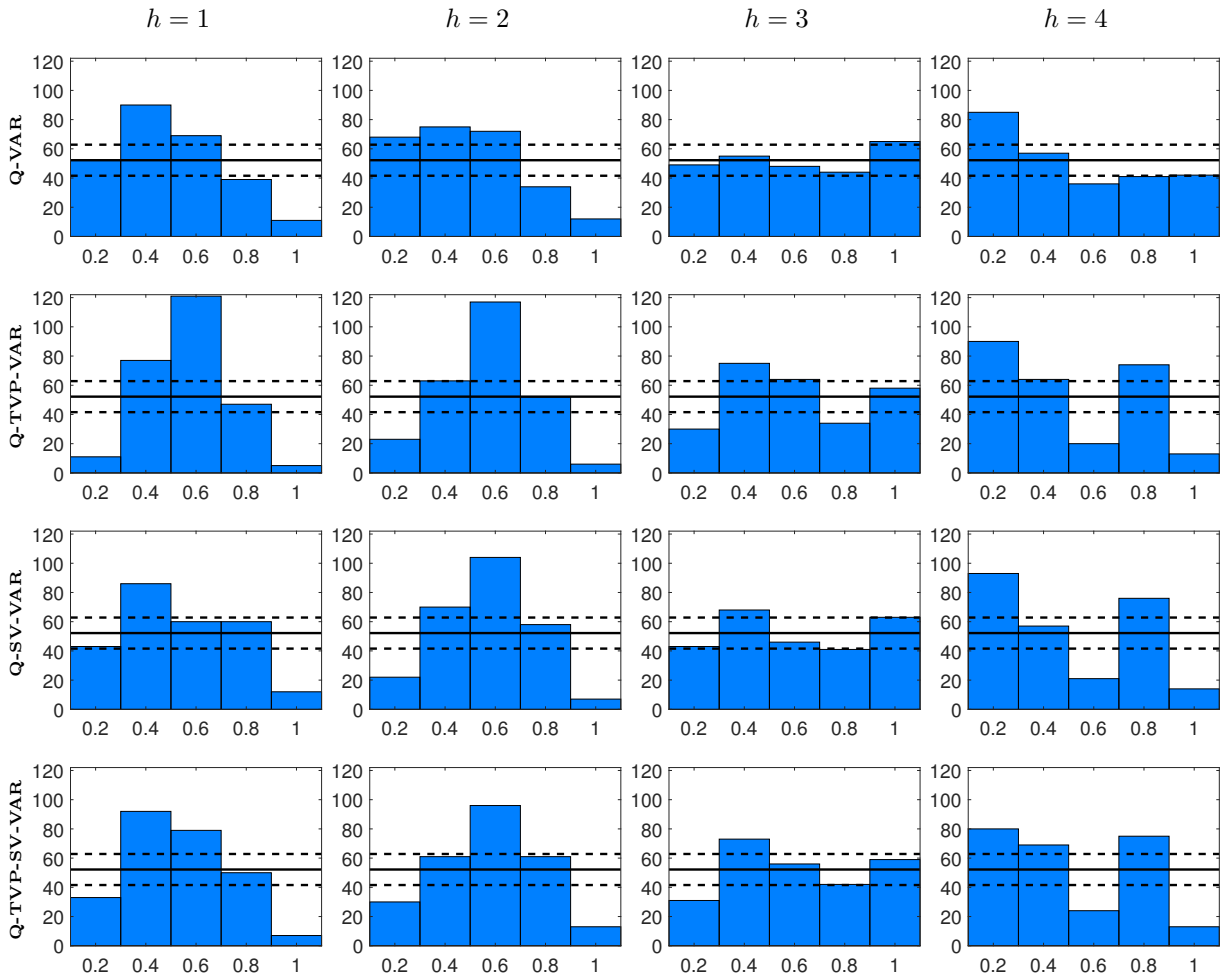


Probability integral transforms for unemployment rate forecasts (continued)

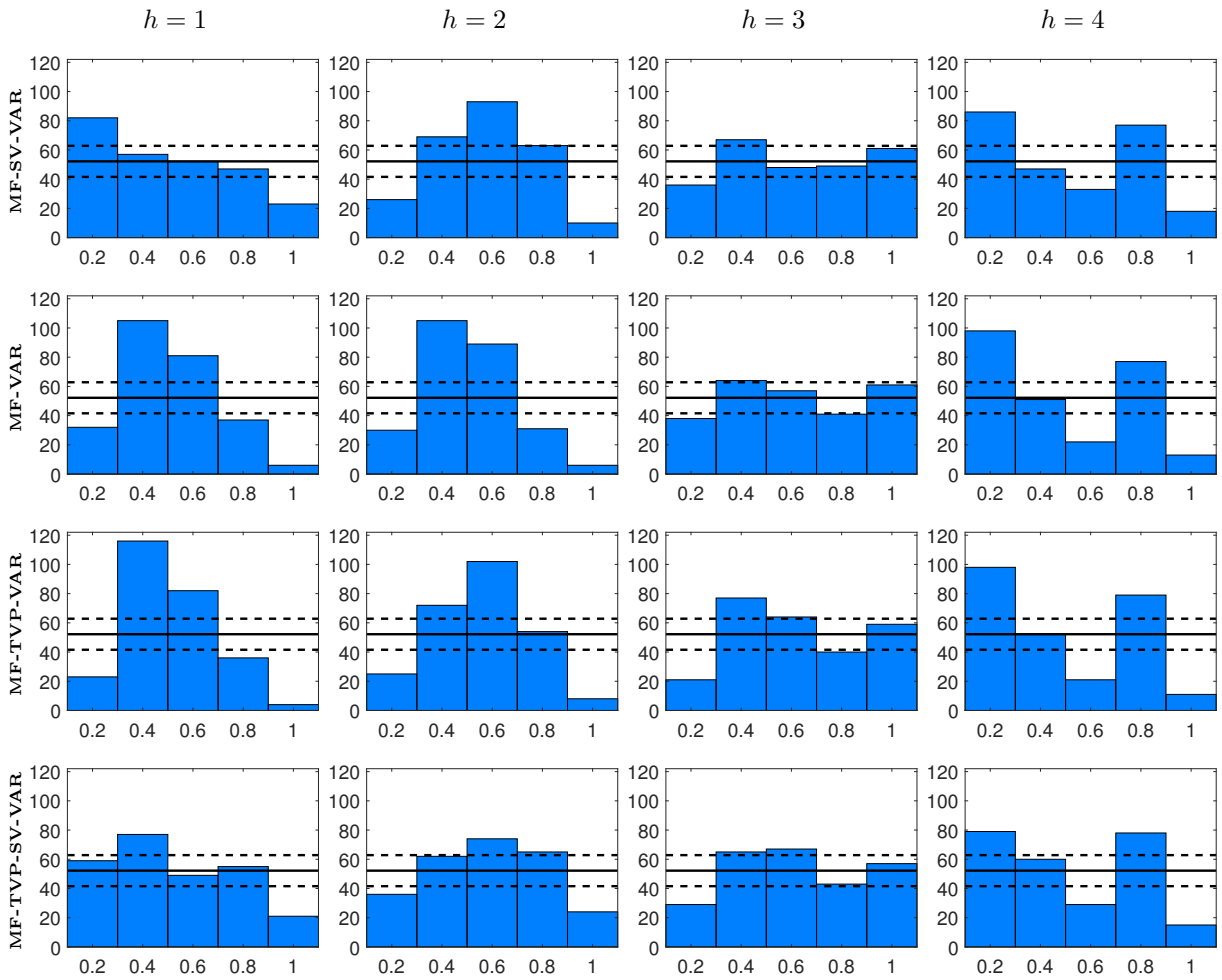


Notes: The rows refer to the PITs of each model. The columns refer to the forecast horizons. The solid line indicates uniformity and the dashed lines 90% confidence bands as in Rossi and Sekhposyan (2014).

Figure 5.12: Probability integral transforms for interest rate forecasts



Probability integral transforms for interest rate forecasts (continued)



Notes: The rows refer to the PITs of each model. The columns refer to the forecast horizons. The solid line indicates uniformity and the dashed lines 90% confidence bands as in Rossi and Sekhposyan (2014).

Erklärung zum selbständigen Verfassen der Arbeit

Ich erkläre hiermit, dass ich meine Doktorarbeit *"Macroeconomics, Nonlinearities, and the Business Cycle"* selbstständig und ohne fremde Hilfe angefertigt habe und dass ich als Koautor maßgeblich zu den weiteren Fachartikeln beigetragen habe. Alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehnenden Ausführungen der aufgeführten Beiträge wurden besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert.

Ort, Datum

Unterschrift