Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

ECONOMICS

Ciclo 31°

Settore Concorsuale: 13/A1 – ECONOMIA POLITICA

Settore Scientifico Disciplinare: SECS-P/01 – ECONOMIA POLITICA

Essays in Applied Macroeconometrics

Presentata da: Graziano Moramarco

Coordinatore Dottorato

Prof. Marco Casari

Supervisore

Prof. Roberto Golinelli

Co-supervisore

Prof. Paolo Manasse

Esame finale anno 2019

To my father

Contents

Acknowledgments

I would like to thank my supervisors, Roberto Golinelli and Paolo Manasse, for their invaluable support, their advice and their time. I am truly grateful. I also wish to extend my thanks to the external evaluators of this thesis, Carlo Favero and Mario Forni, for their reviews. Special thanks to Barbara Rossi for giving me the opportunity to spend a research period at Universitat Pompeu Fabra in Barcelona.

I am grateful to all faculty members of the Department of Economics at the University of Bologna, as well as researchers from other institutions, who have offered me feedback and encouragement at various stages of the PhD program. I would like to acknowledge helpful comments and suggestions by Piergiorgio Alessandri, Gianni Amisano, Luca Fanelli, Gergely Gánics, Nicola Gennaioli, Domenico Giannone, Barbara Rossi, and by participants at seminars and conferences where I presented my work. The usual disclaimer applies.

In developing my research agenda, I have greatly benefited from my previous work experience at Prometeia. I would like to express my deepest thanks to Flavio Cocco, to whom I am indebted for both my professional and personal development. I also owe special thanks to Giuseppe Lusignani and Paolo Onofri.

Chapter 1

Measuring Global Macroeconomic Uncertainty

Abstract

This paper provides new measures of global and country-specific macroeconomic uncertainty using a global vector autoregressive (GVAR) model. Uncertainty is measured as the time-varying dispersion of forecasts that results from the distribution of model parameters estimated by bootstrap methods on recursive windows. The proposed approach takes into account the international propagation of uncertainty, captures global uncertainty shocks and allows to quantify cross-country spillovers. Country-specific indices are computed by averaging standardized uncertainty across real and financial variables and global indices are constructed as GDP-weighted averages of country-specific measures. As a result of global economic linkages, uncertainty is highly correlated across countries. Moreover, the paper exploits the error correction representation of the GVAR to distinguish between short-run and long-run uncertainty measures, and shows that such distinction may help reconcile popular indicators of uncertainty, such as the VIX and the index of economic policy uncertainty (EPU) by Baker, Bloom and Davis (2016).

1.1 Introduction

Economic uncertainty has been a major concern at a global level during the last decade. It is often mentioned among the factors that negatively affect economic activity (e.g. ECB 2009; Stock and Watson 2012; Bloom, Floetotto, Jaimovich, Sapora-Eksten and Terry 2012) and its international transmission plays a key role in shaping macroeconomic outlooks (e.g. IMF 2012). As uncertainty is not directly observable, a growing literature has proposed a variety of methods to measure it and capture its fluctuations (Bloom 2014). Within this strand of research, however, some issues need to be further investigated. In particular, how best to measure global uncertainty, how to characterize the international propagation of uncertainty and how to quantify cross-country spillovers remain to some extent open questions. Several indicators of uncertainty have been developed for a number of countries separately (e.g. Baker, Bloom and Davis 2016; Scotti 2016; Ozturk and Sheng 2018), while global approaches that have been recently proposed (Berger, Grabert and Kempa 2017; Mumtaz and Theodoridis 2017) leave the cross-country transmission of uncertainty largely unexplored and are subject to additional limitations, such as limited scope in terms of variables or countries covered.

Furthermore, existing indicators of uncertainty show at times remarkable differences. In particular, the divergence between measures of economic policy uncertainty and stock market volatility has recently attracted market participants' and policymakers' attention (ECB 2017). More generally, uncertainty is a multifaceted concept and different indicators are likely to capture different aspects of it: it would be desirable to have a framework that reconciles the existing measures by providing an explanation of their differences.

This paper deals with such issues. It proposes new measures of global and country-specific macroeconomic uncertainty using a global vector autoregressive (GVAR) model. First introduced by Pesaran, Schuermann and Weiner (2004), the GVAR is a high-dimensional, flexible multi-country model, which allows to explicitly account for the international propagation of uncertainty, to capture common (global) uncertainty shocks and to measure cross-country uncertainty spillovers. At the same time, the proposed approach distinguishes between short-run and long-run uncertainty measures, and shows that this may help reconcile existing indicators of uncertainty.

Uncertainty is measured as the conditional standard deviation of forecasts resulting from the dispersion of global parameter estimates, which is tracked over time by iterating a bootstrap procedure over recursive sample windows. In each window, (i) the distribution of parameters of the GVAR is estimated by bootstrap methods, (ii) parameter uncertainty translates into distributions of (pseudo-)out-of-sample forecasts and (iii) forecast uncertainty is measured for all variables in the global model (real GDP levels, inflation rates, short-term interest rates, exchange rates and stock market prices). Uncertainty therefore rises whenever point forecasts become less reliable estimates of the expected future values of variables.[1] Since cross-country economic linkages are explicitly modeled in the GVAR, each

[1] In this respect, the advantage of using a high-dimensional model with a limited amount of restrictions, such as the GVAR, is that this reduces the dependence of the uncertainty measures on arbitrary modeling

variable-specific measure of uncertainty is consistent not only with those of other domestic variables, but also with time profiles of uncertainty in the rest of the world. To provide comprehensive measures of macroeconomic uncertainty, variable-specific uncertainties are aggregated into country-specific indices, which turn out to be highly correlated with each other as a result of global interdependencies. Global indices are then computed as weighted averages of the country-specific indices, using GDP levels at purchasing power parity (PPP) as weights. Finally, this framework allows to measure uncertainty spillovers between countries by bootstrapping individual country models one at a time and computing the effects on other countries' forecast uncertainty. Results are reported for U.S. uncertainty spillovers.

The paper develops measures of short-run and long-run uncertainty. The distinction is based on whether uncertainty concerns only short-run economic relationships or also the long-run behavior of the economy,[2] and is addressed using the error correction representation of the GVAR model. More specifically, short-run uncertainty is measured as the dispersion of forecasts that is obtained when the short-run parameters are treated as uncertain (i.e. bootstrapped), while the long-run (cointegrating) parameters are assumed to be known (i.e. their estimates are taken as the true parameter values and are fixed across bootstrap iterations). On the other hand, long-run uncertainty is measured as the dispersion of forecasts obtained when all parameters are treated as uncertain. Importantly, both types of measures turn out to have remarkable similarities with popular indicators of uncertainty, suggesting that the distinction between short-run and long-run uncertainty may be helpful to interpret the differences between existing indicators (cf. Barrero et al. 2016). In particular, the short-run uncertainty index is shown to be broadly consistent with the VIX index of expected stock market volatility and with the macro uncertainty index developed by Jurado, Ludvigson and Ng (2015), while the long-run uncertainty index is more similar to the economic policy uncertainty (EPU) index by Baker, Bloom and Davis (2016).[3]

---

choices.

[2]Cf. Barrero, Bloom and Wright (2016), where the distinction relates instead to the forecast horizon.

[3]The economic literature suggests at least three reasons why the distinction between short-run and long-run uncertainty may be relevant. First, the relative importance of short-run as opposed to long-run unpredictability arguably varies across economic agents. For instance, "financial risk management has generally focused on short-term risks rather than long-term risks" (Engle 2011). On the other hand, long-run uncertainty is central to a number of policy issues. For instance, uncertainty about potential (long-run) output affects the reliability of the output gap estimates, which are key inputs for both fiscal policy and monetary policy (Orphanides and van Norden 2002). Also, reducing uncertainty about long-run inflation and interest rates is often seen as critical for the effectiveness of monetary policy (Bernanke 2007; Gürkaynak, Sack and Swanson 2005; Orphanides and Williams 2002). Second, forecasters may be not equally exposed to these

By focusing on parameter uncertainty as a source of forecast uncertainty, the paper aims to provide measures that are conceptually closer to Knightian (or radical) uncertainty than to risk, as parameter uncertainty undermines individuals' confidence in the estimated probability distributions of economic outcomes. Measuring uncertainty with the estimated volatility of shocks, as is often done in the literature, implicitly relies on the assumption that probability distributions can be treated as known, which pertains more to the concept of risk.[4]

The remainder of the paper is organized as follows. Section 1.2 reviews the related literature, highlighting the contributions of this paper. Section 1.3 presents the methodology used to measure uncertainty. Section 1.4 introduces the empirical implementation. Section 1.5 presents the results. Section 1.6 concludes.

1.2 Related literature

A thriving literature has investigated fluctuations in uncertainty and developed methods to measure it. Among the proposed proxies, some are based on observable variables, while others are model-based. Measures of the first type include indices of option-implied stock market volatility, such as the VIX index in the United States, measures of disagreement among professional forecasters, the dispersion of survey forecast errors and newspaper word

two types of uncertainty, depending on the specification of their forecasting models. As documented by the literature on cointegration, modeling long-run economic relationships is not invariably beneficial for forecast performance (see, for example, Hoffman and Rasche 1996; Lin and Tsay 1996), which means that it may be reasonably omitted under certain circumstances. Third, the effects of short-run and long-run uncertainty on economic activity may differ. For instance, investment may be more responsive to long-run than to short-run uncertainty (Barrero, Bloom and Wright 2016).

[4] Moreover, the literature has emphasized the importance of parameter uncertainty in several contexts. In a seminal paper, Brainard (1967) showed that parameter uncertainty affects policymakers' optimal choices, whereas uncertainty about additive error terms can be ignored when setting the policy variables, at least under standard quadratic objective functions. A subsequent literature has expanded on the implications of parameter uncertainty for monetary policy (Wieland 2000; Söderström 2002). Theoretical models with parameter uncertainty and learning have been proposed to improve on rational expectations models by accounting for key macro puzzles, such as the equity premium puzzle (Hansen 2007; Collin-Dufresne, Johannes and Lochstoer 2016; Weitzman 2007). In finance, uncertainty about the degree of predictability of returns affects the relationship between the investment horizon and the optimal portfolio allocation (Xia 2001). This paper aims to provide comprehensive measures of uncertainty, covering a wide range of such specific types of uncertainty. Another paper that stresses the importance of measuring parameter uncertainty is Orlik and Veldkamp (2014).

counts. Bloom (2009) pioneered the use of stock market volatility indices. Bachmann, Elstner and Sims (2013) use forecast disagreement and ex-post forecast errors from German and U.S. business survey data. Rossi and Sekhposyan (2015) propose an index based on the unconditional probabilities of observing the realized forecast errors, using the U.S. Survey of Professional Forecasters. Rossi and Sekhposyan (2017) adopt the same methodology to construct uncertainty indices for the Euro Area and its member countries and investigate spillover effects. Baker, Bloom and Davis (2016) measure economic policy uncertainty in the United States and in other 11 major economies, using the frequency of newspaper articles containing words related to uncertainty, the economy and policy.

Observable indicators are unlikely to be perfectly correlated with underlying uncertainty. Bekaert, Hoerova and Lo Duca (2013) underscore that movements in the VIX index reflect changes in risk aversion as well as changes in uncertainty. Measures of disagreement and survey forecast error distributions may be affected by the omission of relevant information in surveys (empirical evidence on information rigidities in survey forecasts can be found in Coibion and Gorodnichenko 2012). Lahiri and Sheng (2010) argue that disagreement is a weak proxy for forecast uncertainty when the volatility of aggregate shocks changes over time and differences in point forecasts across forecasters depend solely on idiosyncratic errors that are uncorrelated with observed common shocks.

To address some of the limitations of observable measures, model-based measures of uncertainty have recently been proposed in the literature. Jurado, Ludvigson and Ng (2015) measure uncertainty in the United States through a factor-augmented vector autoregression (FAVAR) with stochastic volatility, using a large dataset of monthly macro and financial indicators. Orlik and Veldkamp (2014) estimate a hidden state model for the U.S. GDP growth and measure uncertainty as the conditional standard deviation of forecasts, taking into account parameter uncertainty. Berger, Grabert and Kempa (2016) use a dynamic factor model with stochastic volatility to measure output growth uncertainty in a multi-country setting. Output growth in each country is decomposed into a common (or global) and country-specific latent factors, plus an effect of domestic inflation, assumed as exogenous. Berger, Grabert and Kempa (2017) extend the approach by also considering inflation uncertainty. Mumtaz and Theodoridis (2017) use a factor model with stochastic volatility to decompose uncertainty in eleven OECD countries into country-specific and common components. Using U.S. data, Carriero, Clark and Marcellino (2017) jointly estimate uncertainty and its impact on the economy through a large VAR in which stochastic volatility is driven by common factors.

Hybrid approaches combining direct measurement and model-based methods have also been developed. Scotti (2016) measures uncertainty in a group of advanced economies by averaging the observed squared forecast errors associated with Bloomberg median expectations across several real-activity indicators. A dynamic factor model is used to assign weights to the indicators in the construction of the uncertainty measures. Similarly to Lahiri and Sheng (2010), Ozturk and Sheng (2018) decompose the uncertainty of a typical forecaster into common uncertainty, i.e. the variance of the consensus forecast error, and idiosyncratic uncertainty, i.e. disagreement. Using survey forecast data, they estimate the common component through a stochastic volatility model and measure disagreement as the interquartile range of forecasts. They construct time series of the two components for a large set of countries and derive a measure of global uncertainty as the PPP-weighted average of country-specific uncertainties.

Differences between short- and long-run components of uncertainty, in terms of drivers and effects on investment and employment, have been investigated by Barrero, Bloom and Wright (2016), using macro- and firm-level data on option-implied volatility at different horizons.

This paper contributes to research on macroeconomic uncertainty along several dimensions. First, it provides comprehensive measures of global uncertainty, country-specific uncertainty and international spillovers of uncertainty, using a GVAR model to account for economic linkages across a large set of countries and covering both real and financial variables. Among the other global approaches to uncertainty, Mumtaz and Theodoridis (2017) investigate common and country-specific uncertainty in OECD countries only. Berger et al. (2017) also adopt a global perspective, but their measures of uncertainty are based on GDP growth and inflation only, thus ignoring financial variables. Also, they impose strong restrictions on the international interactions between variables, in order to achieve identification of common components of uncertainty. Ozturk and Sheng (2018) develop an index of global uncertainty covering 45 countries and 8 variables for each country, but do not model global economic interrelations nor analyze the transmission of uncertainty. Spillovers of uncertainty have been investigated by Rossi and Sekhposyan (2017) in the context of the Euro Area only. Second, the paper develops distinct measures of uncertainty about short-run and long-run macroeconomic relationships, and shows that they match important features of widely-used indices of uncertainty. The proposed interpretation of the results is to some extent in keeping with Barrero et al. (2016), who distinguish short- and long-run uncertainty based on the forecast horizon. Third, the paper measures forecast uncertainty that incorporates parameter uncertainty, whereas the other model-based approaches do not focus on parameter uncertainty, except for Orlik and Veldkamp (2014) who measure parameter uncertainty in a univariate model. The paper also contributes to the GVAR literature. Cesa-Bianchi, Pesaran and Rebucci (2014) use a GVAR model to study the relationship between asset price volatility and economic activity. Unlike in this paper, they do not construct GVAR-based measures of uncertainty, but use observed proxies for uncertainty as an input to the model.

Finally, a literature review on both macro and micro uncertainty is provided by Bloom (2014). ECB (2016) also surveys the literature on macroeconomic uncertainty.

1.3 Methodology

This section illustrates the econometric framework used to measure global macroeconomic uncertainty and to distinguish short- and long-run uncertainty.

1.3.1 The GVAR model

The GVAR model (Pesaran et al. 2004) results from the aggregation of country-specific VARX* models, in which domestic macroeconomic variables are related to their foreign counterparts. To reduce the dimensionality of the parameter space, the foreign variables are built as cross-country weighted averages, using weights based on international trade flows. The foreign aggregates are treated as weakly exogenous in each VARX*, which implies that the estimation is performed at the country level. Here, the GVAR model is estimated on quarterly data and all VARX* models include two lags for both the domestic and the foreign variables.

Consider a generic country i, with $i = 1, ..., N$, where N is the total number of countries. Denote with $\mathbf{x}_{it}$ the $k_i \times 1$ vector of domestic macroeconomic variables of country i at time t and with $\mathbf{x}^*_{it}$ the $k^*_i \times 1$ vector of foreign variables. The VARX* model for country i can be written as

$$\mathbf{x}_{it} = \mathbf{a_0}_i + \mathbf{a_1}_i t + \sum_{j=1}^{2} \mathbf{\Phi}_{ji}\mathbf{x}_{i,t-j} + \sum_{l=0}^{2} \mathbf{\Lambda}_{li}\mathbf{x}^*_{i,t-l} + \boldsymbol{\nu}_{it} \tag{1.1}$$

where $\mathbf{a_0}_i$ and $\mathbf{a_1}_i$ are $k_i \times 1$ vectors of constants and trend coefficients, respectively, $\mathbf{\Phi}_{ji}$, for $j = 1, 2$, and $\mathbf{\Lambda}_{li}$, for $l = 0, 1, 2$, are $k_i \times k_i$ and $k_i \times k^*_i$ matrices of parameters, respectively, and $\boldsymbol{\nu}_{it} \sim iid(\mathbf{0}, \mathbf{\Sigma}_i)$ is the vector of errors.

Denote with k the total number of endogenous variables in the global economy, i.e. $k = \sum_i^N k_i$. Domestic and foreign variables can be expressed in terms of the $k \times 1$ stacked vector of global endogenous variables $\mathbf{x}_t$:

$$\begin{pmatrix} \mathbf{x}_{it} \\ \mathbf{x}^*_{it} \end{pmatrix} = \mathbf{W}_i \begin{bmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \\ \vdots \\ \mathbf{x}_{Nt} \end{bmatrix} = \mathbf{W}_i \mathbf{x}_t$$

where $\mathbf{W}_i$ is the $(k_i + k_i^*) \times k$ matrix of country-specific trade-based weights.

The error correction reparametrization of the country-specific model, or VECX*, distinguishes long-run (cointegrating) relationships between variables and short-run dynamics. Defining $\mathbf{z}_{it} = (\mathbf{x}'_{it}, \mathbf{x}^{*\prime}_{it})'$, it can be written as

$$\Delta \mathbf{x}_{it} = \bar{\mathbf{a}}_{\mathbf{0}i} - \mathbf{\Pi}_i \left[\mathbf{z}_{i,t-1} - \boldsymbol{\gamma}_i (t-1)\right] - \mathbf{\Phi}_{2i} \Delta \mathbf{x}_{i,t-1} + \mathbf{\Lambda}_{0i} \Delta \mathbf{x}^*_{it} - \mathbf{\Lambda}_{2i} \Delta \mathbf{x}^*_{i,t-1} + \boldsymbol{\nu}_{it} \qquad (1.2)$$

where $\mathbf{\Pi}_i$ is a $k_i \times (k_i + k_i^*)$ matrix of parameters, $\bar{\mathbf{a}}_{\mathbf{0}i}$ is a $k_i \times 1$ vector of constants and $\boldsymbol{\gamma}_i$ is a $(k_i + k_i^*) \times 1$ vector of trend coefficients.[5] Given (1.1), $\mathbf{a}_{\mathbf{0}i} = \bar{\mathbf{a}}_{\mathbf{0}i} - \mathbf{\Pi}_i \boldsymbol{\gamma}_i$ and $\mathbf{a}_{\mathbf{1}i} = \mathbf{\Pi}_i \boldsymbol{\gamma}_i$. The rank r_i of matrix $\mathbf{\Pi}_i$ represents the number of long-run relationships between the variables in $\mathbf{z}_{it}$. In particular, $\mathbf{\Pi}_i = \boldsymbol{\alpha}_i \boldsymbol{\beta}'_i$, where $\boldsymbol{\alpha}_i$ is the $k_i \times r_i$ matrix of loadings and $\boldsymbol{\beta}_i$ is the $(k_i + k_i^*) \times r_i$ matrix of cointegrating vectors (see Johansen 1995). Also, given (1.1) it is readily seen that $\mathbf{\Pi}_i = \left(\mathbf{I}_{k_i} - \sum_{j=1}^2 \mathbf{\Phi}_{ji}, -\sum_{l=0}^2 \mathbf{\Lambda}_{li}\right)$, where $\mathbf{I}_{k_i}$ is the $k_i \times k_i$ identity matrix.

Stacking all country-specific VARX* models provides a vector autoregressive representation of the global economy:

$$\mathbf{G}\mathbf{x}_t = \mathbf{a}_\mathbf{0} + \mathbf{a}_\mathbf{1} t + \mathbf{H}_1 \mathbf{x}_{t-1} + \mathbf{H}_2 \mathbf{x}_{t-2} + \boldsymbol{\nu}_t$$

where $\mathbf{a}_\mathbf{0} = (\mathbf{a_0}'_1, \mathbf{a_0}'_2, \cdots, \mathbf{a_0}'_N)'$ and $\mathbf{a}_\mathbf{1} = (\mathbf{a_1}'_1, \mathbf{a_1}'_2, \cdots, \mathbf{a_1}'_N)'$ are the $k \times 1$ stacked vectors of global constants and trends, respectively, $\boldsymbol{\nu}_t = (\boldsymbol{\nu}'_{1t}, \boldsymbol{\nu}'_{2t}, \cdots, \boldsymbol{\nu}'_{Nt})'$ is the $k \times 1$ stacked vector of errors and

$$\mathbf{G} = \begin{pmatrix} (\mathbf{I}_{k_1}, -\mathbf{\Lambda}_{01})\,\mathbf{W}_1 \\ (\mathbf{I}_{k_2}, -\mathbf{\Lambda}_{02})\,\mathbf{W}_2 \\ \vdots \\ (\mathbf{I}_{k_N}, -\mathbf{\Lambda}_{0N})\,\mathbf{W}_N \end{pmatrix}, \quad \mathbf{H}_1 = \begin{pmatrix} (\mathbf{\Phi}_{11}, \mathbf{\Lambda}_{11})\,\mathbf{W}_1 \\ (\mathbf{\Phi}_{12}, \mathbf{\Lambda}_{12})\,\mathbf{W}_2 \\ \vdots \\ (\mathbf{\Phi}_{1N}, \mathbf{\Lambda}_{1N})\,\mathbf{W}_N \end{pmatrix}, \quad \mathbf{H}_2 = \begin{pmatrix} (\mathbf{\Phi}_{21}, \mathbf{\Lambda}_{21})\,\mathbf{W}_1 \\ (\mathbf{\Phi}_{22}, \mathbf{\Lambda}_{22})\,\mathbf{W}_2 \\ \vdots \\ (\mathbf{\Phi}_{2N}, \mathbf{\Lambda}_{2N})\,\mathbf{W}_N \end{pmatrix}$$

[5] The VECX with restricted trends corresponds to case IV in Pesaran et al. (2000).

The reduced form of the global model can be written as

$$\mathbf{x}_t = \mathbf{c_0} + \mathbf{c_1}t + \mathbf{F}_1\mathbf{x}_{t-1} + \mathbf{F}_2\mathbf{x}_{t-2} + \boldsymbol{\varepsilon}_t \tag{1.3}$$

where

$$\begin{aligned} \mathbf{c_0} &= \mathbf{G}^{-1}\mathbf{a_0}, \qquad & \mathbf{c_1} &= \mathbf{G}^{-1}\mathbf{a_1}, \\ \mathbf{F}_1 &= \mathbf{G}^{-1}\mathbf{H}_1, \qquad & \mathbf{F}_2 &= \mathbf{G}^{-1}\mathbf{H}_2 \end{aligned}$$

and $\boldsymbol{\varepsilon}_t = \mathbf{G}^{-1}\boldsymbol{\nu}_t$ is the vector of reduced-form global residuals. Assuming that $\boldsymbol{\nu}_t$ is normally distributed, $\boldsymbol{\varepsilon}_t \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \mathbf{G}^{-1}\boldsymbol{\Sigma}(\mathbf{G}^{-1})'$ and $\boldsymbol{\Sigma}$ is the $k \times k$ covariance matrix of $\boldsymbol{\nu}_t$. In particular, $\boldsymbol{\Sigma}$ is a block matrix whose (i, j) block is the $k_i \times k_j$ matrix $\boldsymbol{\Sigma}_{ij}$ of shock covariances between country i and country j, with $\boldsymbol{\Sigma}_{ii} = \boldsymbol{\Sigma}_i$ from model (1.1).

Finally, it is useful to express the model in companion form:

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0 \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{0} \end{bmatrix} t + \begin{bmatrix} \mathbf{F}_1 & \mathbf{F}_2 \\ \mathbf{I}_k & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \end{bmatrix}$$

In what follows, the companion form will be denoted as

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{c}}_\mathbf{0} + \tilde{\mathbf{c}}_\mathbf{1}t + \tilde{\mathbf{F}}\tilde{\mathbf{x}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t \tag{1.4}$$

1.3.2 Time-varying uncertainty

In order to derive time profiles of short- and long-run uncertainty from the GVAR model, a bootstrap procedure is implemented over recursive windows.[6] The measurement methodology can be described as follows.

1. The GVAR is estimated over time using recursive windows. The shortest window spans the period 1979Q4-2000Q1, then the sample is extended by one-quarter increments, up to 1979Q4-2016Q1. The country-specific VECX* models are estimated on each window, i.e. they are all re-estimated every time an additional quarter is included in the sample. To this aim, window-specific foreign variables are constructed, based on trade patterns that were observed at the end of the window under consideration (Section 1.4 provides

[6]Bootstrap techniques were first applied to the GVAR model by Dees et al. (2007a, 2007b).

details). Let us define the *baseline* GVAR in a given quarter as the GVAR estimated on actual data over the window ending in that quarter. Consider a generic window w ending in period T_w. The baseline GVAR for quarter T_w can be written as

$$\mathbf{x}_t = \widehat{\mathbf{c}}_{\mathbf{0}}^{(w)} + \widehat{\mathbf{c}}_{\mathbf{1}}^{(w)} t + \widehat{\mathbf{F}}_1^{(w)} \mathbf{x}_{t-1} + \widehat{\mathbf{F}}_2^{(w)} \mathbf{x}_{t-2} + \widehat{\boldsymbol{\varepsilon}}_t^{(w)} \tag{1.5}$$

where $\widehat{}$ denotes estimates and $t = 1, 2, \ldots, T_w$.

2. In each sample window, a non-parametric bootstrap of the estimates is performed. First, I simulate alternative historical paths that the variables in the global model might have followed within the sample window, given the dynamics and the empirical distribution of errors of model (1.5). Then, I re-estimate all the VECX* models and, accordingly, the global model on the simulated time series.

   In window w,

   (a) the window-specific baseline GVAR model (1.5) produces a $k \times T_w$ matrix of global residuals $\widehat{\boldsymbol{E}}^{(w)} = \left(\widehat{\boldsymbol{\varepsilon}}_1^{(w)}, \widehat{\boldsymbol{\varepsilon}}_2^{(w)}, \ldots, \widehat{\boldsymbol{\varepsilon}}_{T_w-1}^{(w)}, \widehat{\boldsymbol{\varepsilon}}_{T_w}^{(w)} \right)$;

   (b) in the generic b-th bootstrap iteration ($b = 1, \ldots, B$), the T_w columns of matrix $\widehat{\boldsymbol{E}}^{(w)}$ are resampled. Then, artificial time series for all the variables are generated through an in-sample simulation of the baseline GVAR (1.5) using the resampled residuals as shocks. Denoting iteration b in window w with the superscript (w, b), let us call $\boldsymbol{\varepsilon}_t^{(w,b)}$ the bootstrap shocks, generated by randomly drawing columns from $\widehat{\boldsymbol{E}}^{(w)}$ with replacement. The simulated time series are given by

$$\mathbf{x}_t^{(w,b)} = \widehat{\mathbf{c}}_{\mathbf{0}}^{(w)} + \widehat{\mathbf{c}}_{\mathbf{1}}^{(w)} t + \widehat{\mathbf{F}}_1^{(w)} \mathbf{x}_{t-1}^{(w,b)} + \widehat{\mathbf{F}}_2^{(w)} \mathbf{x}_{t-2}^{(w,b)} + \boldsymbol{\varepsilon}_t^{(w,b)}$$

   with $\mathbf{x}_0^{(w,b)} = \mathbf{x}_0$ and $\mathbf{x}_{-1}^{(w,b)} = \mathbf{x}_{-1}$.
   Iteration-specific foreign variables $\mathbf{x}_{it}^{*(w,b)}$ are then constructed using the window-specific trade weight matrix $\mathbf{W}_i^{(w)}$ for every i;

   (c) in each bootstrap iteration, all the VECX* models are re-estimated on the simulated data. Two alternative cases are considered:

      i. uncertainty about both short-run and long-run parameters is considered. Accordingly, all parameters are re-estimated in each iteration, also allowing for iteration-specific cointegration ranks $r_i^{(w,b)}$ (details on rank selection are provided below):

$$\Delta\mathbf{x}_{it}^{(w,b)} = \widehat{\widehat{\mathbf{a}}}_{\mathbf{0}i}^{(w,b)} - \widehat{\boldsymbol{\alpha}}_i^{(w,b)}\widehat{\boldsymbol{\beta}}_i^{(w,b)\prime}\left[\mathbf{z}_{i,t-1}^{(w,b)} - \widehat{\boldsymbol{\gamma}}_i^{(w,b)}(t-1)\right] + \\ - \widehat{\boldsymbol{\Phi}}_{2i}^{(w,b)}\Delta\mathbf{x}_{i,t-1}^{(w,b)} + \widehat{\boldsymbol{\Lambda}}_{0i}^{(w,b)}\Delta\mathbf{x}_{it}^{*(w,b)} - \widehat{\boldsymbol{\Lambda}}_{2i}^{(w,b)}\Delta\mathbf{x}_{i,t-1}^{*(w,b)} + \widehat{\boldsymbol{\nu}}_{it}^{(w,b)} \tag{1.6}$$

ii. only short-run parameter uncertainty is considered. The long-run vectors $\boldsymbol{\beta}_i$ and $\boldsymbol{\gamma}_i$ are fixed at their baseline estimates $\widehat{\boldsymbol{\beta}}_i^{(w)}$ and $\widehat{\boldsymbol{\gamma}}_i^{(w)}$ across all iterations (the cointegration rank is fixed at $\widehat{r}_i^{(w)}$):

$$\Delta\mathbf{x}_{it}^{(w,b)} = \widehat{\widehat{\mathbf{a}}}_{\mathbf{0}i}^{(w,b)} - \widehat{\boldsymbol{\alpha}}_i^{(w,b)}\widehat{\boldsymbol{\beta}}_i^{(w)\prime}\left[\mathbf{z}_{i,t-1}^{(w,b)} - \widehat{\boldsymbol{\gamma}}_i^{(w)}(t-1)\right] + \\ - \widehat{\boldsymbol{\Phi}}_{2i}^{(w,b)}\Delta\mathbf{x}_{i,t-1}^{(w,b)} + \widehat{\boldsymbol{\Lambda}}_{0i}^{(w,b)}\Delta\mathbf{x}_{it}^{*(w,b)} - \widehat{\boldsymbol{\Lambda}}_{2i}^{(w,b)}\Delta\mathbf{x}_{i,t-1}^{*(w,b)} + \widehat{\boldsymbol{\nu}}_{it}^{(w,b)} \tag{1.7}$$

As a result, in either case B estimates of the GVAR model are obtained for each quarter from 2000Q1 to 2016Q1:

$$\mathbf{x}_t^{(w,b)} = \widehat{\mathbf{c}}_{\mathbf{0}}^{(w,b)} + \widehat{\mathbf{c}}_{\mathbf{1}}^{(w,b)}t + \widehat{\mathbf{F}}_1^{(w,b)}\mathbf{x}_{t-1}^{(w,b)} + \widehat{\mathbf{F}}_2^{(w,b)}\mathbf{x}_{t-2}^{(w,b)} + \widehat{\boldsymbol{\varepsilon}}_t^{(w,b)} \tag{1.8}$$

3. Each of the B window-specific GVAR versions produces forecasts out of the estimation window. More specifically, dynamic pseudo-out-of-sample forecasts are jointly calculated for all the variables in the global economy, taking as initial values for each variable the last two actual values within the estimation window. Recalling the companion form (1.4) of the GVAR model and denoting forecasts with the superscript (f), the h-step-ahead forecasts of the model estimated on window w in iteration b can be expressed as

$$\mathbf{x}_{T_w+h}^{(f)(b)} = \mathbf{S}\left(\widehat{\tilde{\mathbf{F}}}^{(w,b)}\right)^h \tilde{\mathbf{x}}_{T_w} + \mathbf{S}\sum_{\tau=0}^{h-1}\left(\widehat{\tilde{\mathbf{F}}}^{(w,b)}\right)^\tau \left[\widehat{\tilde{\mathbf{c}}}_{\mathbf{0}}^{(w,b)} + \widehat{\tilde{\mathbf{c}}}_{\mathbf{1}}^{(w,b)}\left(T_w + h - \tau\right)\right]$$

where $\mathbf{S} = (\mathbf{I}_k, \mathbf{0}_{k\times k})$ is a selection matrix. This is repeated for all windows.

The outcome of the procedure consists of multivariate distributions of forecasts in all quarters from 2000Q1 to 2016Q1. In each quarter, variable-specific uncertainty is measured as the standard deviation of 4-quarter-ahead forecasts. Denoting with $x_{v,t}$ the generic v-th variable in the global vector $\mathbf{x}_t$ and with $u_{v,t}$ the corresponding uncertainty measure,

$$u_{v,t} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(x_{v,t+4}^{(f)(b)} - \frac{1}{B}\sum_{b=1}^{B} x_{v,t+4}^{(f)(b)}\right)^2}$$

The procedure delivers time series of uncertainty for all variables in the global model. Each time series is then standardized by subtracting the mean and dividing by the standard deviation. Aggregate indices of uncertainty are computed for each country by averaging across the respective domestic variables. Like Jurado et al. (2015), I assign equal weights to the variables (using principal components yields almost identical results). Finally, an index of global uncertainty is calculated as a weighted average of country-specific uncertainties. The weights are given by the annual GDP levels in PPP terms (in each quarter, the previous year's GDP data are used).

This approach provides country-level uncertainty indicators that incorporate substantial information about international macroeconomic dynamics. Relying on a GVAR ensures that (i) all countries are jointly simulated in the sample, which is reflected in the bootstrap distributions of estimates, and (ii) all variables in the global economy are jointly forecast out of the sample and depend on each other.

The distinction between long- and short-run uncertainty is summarized by (1.6) and (1.7). Long-run uncertainty is the standard deviation of forecasts that is obtained when all parameters, including the cointegrating vectors, are treated as unknown. Accordingly, it is measured by considering (1.6) in the bootstrap procedure. Short-run uncertainty is the standard deviation of forecasts obtained when only the short-run parameters are treated as unknown. It is measured by considering equation (1.7). In the case of long-run uncertainty, the number of long-run relationships (i.e. the number of common stochastic trends) is also treated as unknown. Accordingly, the cointegration ranks are re-estimated in each bootstrap iteration. Each iteration-specific rank $\widehat{r}_i^{(w,b)}$ (as well as each baseline rank $\widehat{r}_i^{(w)}$) is determined by the Johansen trace test[7], provided that this ensures stability, as explained below. In the case of short-run uncertainty, the cointegration ranks are allowed to vary across windows but not across iterations.

The measures of uncertainty may be unduly inflated by explosive roots in (1.8). For this reason, in each iteration I check whether the estimated models are dynamically stable, i.e. whether all the eigenvalues of the companion matrices are less than or equal to 1 in modulus.

[7]The 5% critical values are considered.

I perform the stability analysis both on the country-specific models and on the resulting global model. Uncertainty is measured using stable models only. At the country level, the cointegration rank $r_i^{(w,b)}$ is allowed to deviate from the indications of the Johansen test if the estimated rank results in an unstable country model. In that case, I select the highest rank that makes the model stable. Since this does not ensure the stability of the global model, I also check the eigenvalues of the global companion matrix $\hat{\tilde{\mathbf{F}}}_{w,b}$. If the global model is unstable, the bootstrap iteration is repeated until a stable model is found.

To further mitigate the impact of extreme forecasts on the uncertainty measures, I remove those iteration-specific forecasts that are outliers with respect to a particularly relevant variable, namely U.S. GDP (of course, the interconnectedness of variables in the GVAR implies that such extreme values are transmitted across variables throughout the global economy). The iteration-specific forecasts are omitted whenever the 4-quarter-ahead forecast of the U.S. GDP lies more than 3 standard deviations away from the average forecast across iterations.

1.4 Empirical implementation

1.4.1 Countries and variables

The proposed approach is implemented for the 33 countries considered in Cesa-Bianchi, Pesaran and Rebucci (2014). Sixteen countries are aggregated into three areas, therefore 20 economies are included in the GVAR: Australia, Brazil, Canada, China, the Euro area, India, Japan, a Latin American area, Mexico, New Zealand, Norway, Saudi Arabia, South Africa, South-East Asia, South Korea, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The composition of the three areas is the following: the Euro area includes Austria, Belgium, Finland, France, Germany, Italy, the Netherlands and Spain; the Latin American area comprises Argentina, Chile and Peru; South-East Asia is composed by Indonesia, Malaysia, Philippines, Thailand and Singapore.

As mentioned above, the variables included in the GVAR model are real GDP levels, CPI quarterly inflation rates, short-term interest rates, exchange rates with respect to the U.S. dollar and equity price indices. Exchange rates and equity indices are expressed in real terms by deflating the nominal values using the consumer price index (this definition of "real exchange rate" is often used in the GVAR literature, see Pesaran et al. 2004 and Dees et al. 2007a). Logarithms are used for the real GDP, exchange rates and equity indices. Each interest rate is transformed to $0.25\,[1 + \ln(R_t/100)]$, where R_t is the rate expressed in percentage values on an annual basis. Domestic and foreign GDP, inflation and exchange rates are included in all the VARX* models (except for the domestic exchange rate in the U.S. model, as the U.S. dollar is the numeraire currency). Domestic short-term interest rates are included as endogenous in all VARX* models except for Saudi Arabia and for countries that experienced skyrocketing interest rates (higher than 100% on an annual basis) during major crises in the 80s and 90s (Brazil, Mexico, the Latin American area and Turkey), while foreign interest rates are included in all country models except for the United States. Stock market indices are included for the major financial economies, namely the United States, the Euro area, the United Kingdom and Japan. Foreign equity indices are considered in the models for the Euro area, the U.K. and Japan. As usual in the GVAR literature, the U.S. model has fewer weakly exogenous variables, given the special status of the United States in the global economy: I include foreign GDP, inflation and exchange rate.

For any pair of countries i and j, the weight assigned to j in the construction of i's foreign variables is based on the average of i's exports to j and i's imports from j. In particular, to calculate window-specific foreign variables I use the average trade weights observed in the 3 years prior to the window end year. The weights used to aggregate countries into areas are based on annual GDP levels in PPP. In each quarter, the aggregation weights are computed as the GDP shares in the previous year.

1.4.2 Data

The quarterly dataset used in this paper extends up to 2016Q1 the data used in Cesa-Bianchi, Pesaran and Rebucci (2014), which span the period 1979Q1-2013Q1. The data sources used for the period 2013Q2-2016Q1 are reported in Appendix 1.A.

Unlike financial data, GDP and inflation data are typically revised, which raises the question of whether the accuracy of uncertainty measures can be improved by using real-time vintage data (see e.g. Clements 2017). On the other hand, Jurado et al. (2015) point out that the use of real-time data may actually lead to biased estimates of uncertainty, since a substantial amount of information on macro variables becomes available to economic agents and forecasters well before official data releases. The present study does not use real-time vintages, which may be considered for future extensions. However, I did not include revisions to the original data used in Cesa-Bianchi et al. (2014).

1.5 Results

This section presents the uncertainty indices constructed using the proposed approach.[8] All the results are obtained with 1000 bootstrap iterations and all the indices are expressed in standardized units.[9]

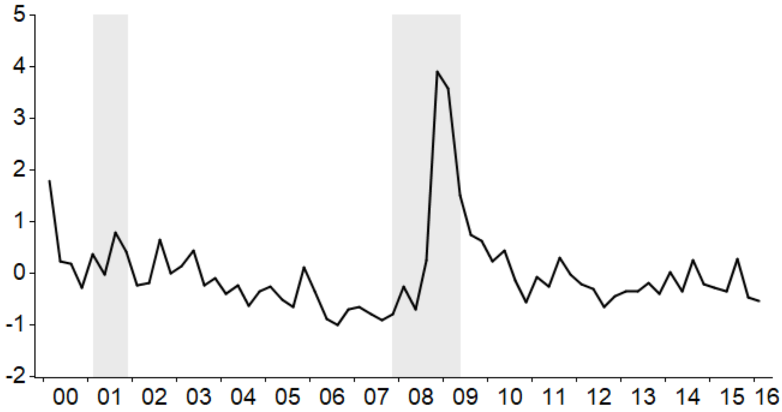
Figure 1.1 shows the index of global short-run uncertainty over the period 2000Q1-2016Q1. The index peaks around the Lehman Brothers collapse in 2008Q4, when variable-specific uncertainties rise on average 4 standard deviations above their means. It then drops during 2009 and 2010, and exhibits only minor peaks afterwards. Figure 1.2 plots the index of global long-run uncertainty. The index surges during the Great Recession of 2008-2009, decreases in 2010, rises again in 2011 and gradually subsides afterwards. The two plots show some similarities, stemming from the fact that the two indices are affected by the same shocks (the residuals used to generate the bootstrap samples coincide in the two cases, since the baseline model is the same). In particular, at the height of the global financial crisis, large shocks lead to increases in both short-run and long-run uncertainty. On the other hand, the two indices also exhibit remarkable differences, reflecting their different scope. The short-run uncertainty index concerns short-run fluctuations and the adjustment towards the long-run equilibrium. It is a constrained measure of uncertainty, since the long-run parameters are fixed in the bootstrap procedure, and it is estimated on stationary time series, namely the first differences (or growth rates) of trending variables and the cointegration residuals. During 2009 and 2010, when first differences typically return to normal after experiencing large deviations from their means, short-run uncertainty quickly reverts to lower values. On the contrary, the long-run uncertainty index is an unconstrained measure of uncertainty: in this case, all parameters are allowed to vary across bootstrap iterations, including the equilibrium relationships between variables in levels. Therefore, long-run uncertainty is effectively measured on non-stationary time series. As the shocks realized during and after the global crisis had permanent or highly persistent effects on the levels of the variables, increases in long-run uncertainty persist after short-run uncertainty abates. As the figures show, short-run uncertainty rises more sharply than long-run uncertainty, in standardized terms, during the global crisis.[10]

[8] The code, written for EViews and R, makes use of software developed by Baier, T. and E. Neuwirth (2007), "Excel :: COM :: R." *Computational Statistics*, 22 (1), 91–108.

[9] The complete set of country-specific indices is reported in Appendix 1.B, Figures 1.9 and 1.10.

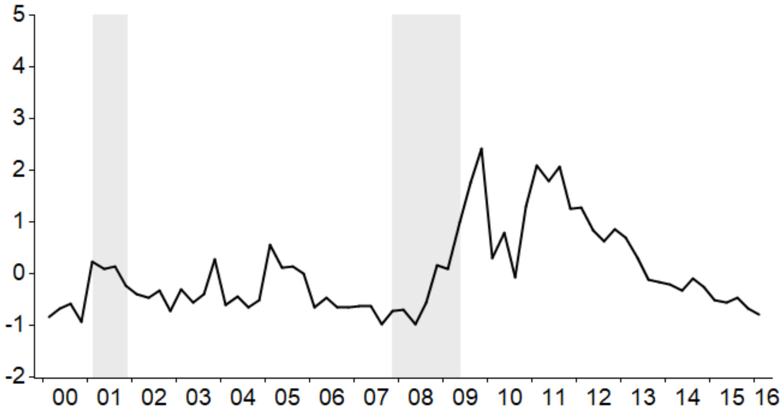
[10] In absolute terms, long-run uncertainty is systematically higher due to the additional variability of cointegrating vectors. Figure 1.11 in Appendix 1.B shows the different magnitudes of the pre-standardization

Figure 1.1: Global short-run uncertainty index



Notes: The index is calculated as the PPP GDP-weighted average of the country-specific short-run uncertainty indices and is expressed in standardized units. Each country-specific index is calculated as the average uncertainty across the domestic variables included in the GVAR model. The data are quarterly and span the period 2000Q1-2016Q1. Shaded areas are NBER recession periods.

Figure 1.2: Global long-run uncertainty index



Notes: The index is calculated as the PPP GDP-weighted average of the country-specific long-run uncertainty indices and is expressed in standardized units. Each country-specific index is calculated as the average uncertainty across the domestic variables included in the GVAR model. The data are quarterly and span the period 2000Q1-2016Q1. Shaded areas are NBER recession periods.

Commonalities in uncertainty across countries clearly emerge from Figures 1.3 and 1.4, which plot short- and long-run uncertainty, respectively, for the U.S., the Euro area, the U.K., China and India. The strong co-movement between country-specific indicators reflects both common (global) uncertainty shocks, as captured by cross-country contemporaneous correlations of residuals in the GVAR, and the dynamic propagation of uncertainty from any country to the others. Global patterns of uncertainty appear therefore to be dominant.[11]

Let us now examine how the short- and long-run uncertainty indices relate to different popular measures of uncertainty. Figure 1.5 compares the short-run uncertainty index (SRU) for the United States with the VIX, i.e. the index of option-implied volatility in the S&P500, and with the U.S. macro uncertainty index developed by Jurado et al. (2015) (JLN henceforth). All three measures peak in 2008Q4 and the magnitudes of their increases during the financial crisis are highly comparable, as well as the subsequent declines in the period 2009-2010. The correlation between SRU and VIX is 0.74, while the correlation between SRU and JLN is 0.59. Thus, the SRU index appears broadly in line with financial markets' volatility expectations over short horizons (the VIX measures 30-day-ahead expected volatility). Also, the JLN index is constructed using stationary time series, which seems consistent with a short-run perspective focusing on cyclical fluctuations rather than on trends.

Barrero, Bloom and Wright (2016) find that economic policy uncertainty is more tightly linked to long-run than to short-run components of uncertainty. Figure 1.6 contrasts the GVAR-based long-run uncertainty index (LRU) for the United States with the EPU index developed by Baker, Bloom and Davis (2016).[12] Unlike the indicators in Figure 1.5, both the LRU index and the EPU index exhibit relatively high values in the period 2010-2013, compared to the respective pre-crisis averages. The correlation between LRU and EPU is 0.67. While the results do not imply a systematic relationship between the two measures, it is arguably the case that uncertainty about the long run has been relevant for a variety of primary policy issues debated in the aftermath of the crisis, such as those relating to financial regulation and public debt sustainability.

The VIX and EPU indices emerge therefore as primary benchmarks for the different

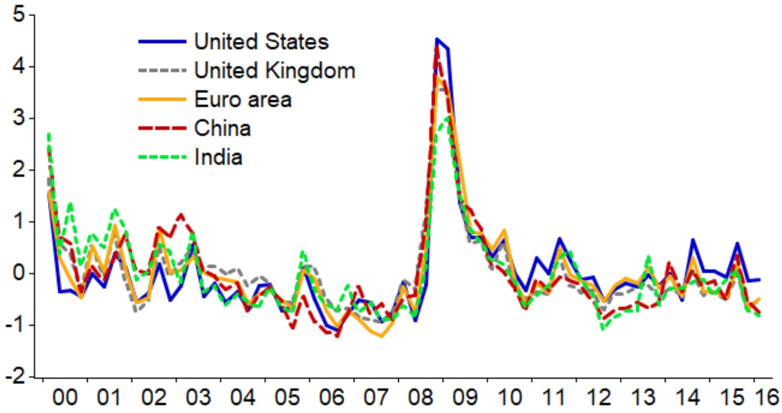
---

measures for the United States. Figure 1.12 compares global short-run uncertainty with the measure of uncertainty calculated using a GVAR in first differences, where $\mathbf{\Pi}_i = \mathbf{0}$ in (1.2) for every i.

[11] All cross-country correlations are reported in Appendix 1.B, Tables 1.4 and 1.5.

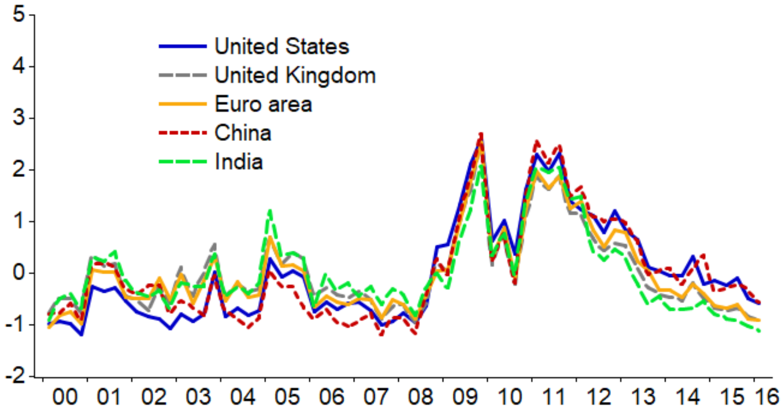
[12] The index considered here is the overall EPU index, which combines the news-based EPU index with three additional measures of policy uncertainty: an index of tax expirations, a measure of forecast disagreement over consumer prices and a measure of forecast disagreement over federal/state/local government purchases. The overall index is available at www.policyuncertainty.com.

Figure 1.3: Short-run uncertainty indices for U.S., Euro area, U.K., China and India



Notes: Each index is calculated as the average short-run uncertainty across the domestic variables included in the GVAR model and is expressed in standardized units. The data are quarterly and span the period 2000Q1-2016Q1.

Figure 1.4: Long-run uncertainty indices for U.S., Euro area, U.K., China and India



Notes: Each index is calculated as the average long-run uncertainty across the domestic variables included in the GVAR model and is expressed in standardized units. The data are quarterly and span the period 2000Q1-2016Q1.

uncertainty measures developed in this paper. While many events that trigger increases in economic policy uncertainty also have repercussions on stock market volatility,[13] the indices show remarkable differences, which have already been ascribed to several factors by Baker et al. (2016).[14] It is worth adding here that there exist widespread concerns about stock market-listed firms focusing too much on short-term outcomes (a phenomenon known as short-termism, see Davies, Haldane, Nielsen and Pezzini 2014; Asker, Farre-Mensa and Ljungqvist 2015), which contrasts with the long-term focus of many policy issues (Barrero et al. 2016). The results presented in this paper may be supportive of the idea that the VIX reflects a stronger focus on short-run issues, while economic policy uncertainty is more related to long-run issues, at least in the period under consideration (cf. Barrero et al. 2016).

Table 1.1 shows the correlations between the different uncertainty measures considered in this section.

1.5.1 Spillovers of uncertainty

The proposed GVAR-based methodology can be used to quantify the international spillovers of uncertainty. This section illustrates such feature by presenting results on the spillovers of U.S. uncertainty. These are quantified by measuring the component of each country's forecast uncertainty that only depends on parameter uncertainty in the U.S. model. In practice, in step (2c) of the algorithm presented in Section 1.3.2, only the U.S. VARX* is re-estimated on simulated data.[15] As a result, in each iteration the global model is built by combining the bootstrapped U.S. model with the baseline models for the other countries, and forecasts are produced accordingly. For each variable in the global model, the resulting forecast uncertainty is expressed as a fraction of total variable-specific uncertainty, i.e. the uncertainty that is obtained when all countries are bootstrapped. Finally, the ratios thus derived are averaged across domestic variables for each country. The average ratio is taken as a measure of the strength of the spillovers, both direct and indirect (i.e. transmitted through third-party countries).

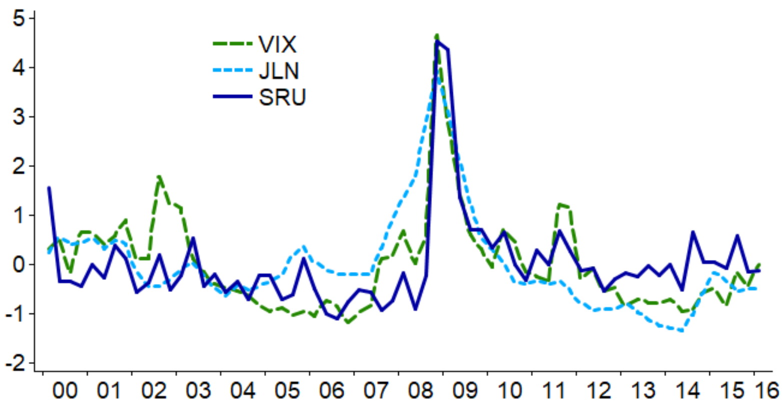
Figure 1.7 plots the spillovers of U.S. short-run uncertainty on the Euro area, the U.K.,

[13]This has been the case, for instance, with the Lehman Brothers default and the Troubled Asset Relief Program (TARP) in 2008, as well as with the Euro crisis of 2011.

[14]In particular, as argued by Baker et al. (2016): (i) the VIX has a short horizon, while the news-based component of EPU has no specific horizon; (ii) policy issues do not necessarily relate to equity returns; (iii) the VIX covers publicly traded firms only.

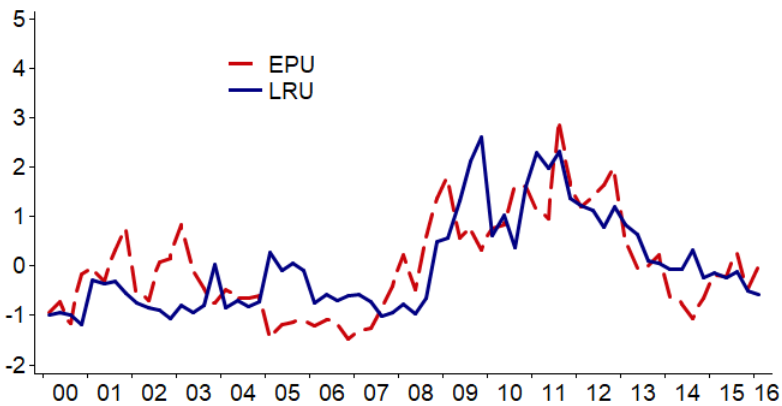
[15]Note however that the data are jointly simulated for all countries in step (2b), so that the U.S. foreign variables are also simulated.

Figure 1.5: U.S. short-run uncertainty: a comparison



Notes: In this figure, SRU is the U.S. short-run uncertainty index from Figure 1.3, VIX is the volatility index by the Chicago Board Options Exchange and JLN is the updated version (as of March 2017, source: www.sydneyludvigson.com) of the 12-month-ahead macro uncertainty index originally proposed in Jurado et al. (2015). Data are quarterly. For VIX and JLN, quarterly data are obtained by averaging daily and monthly data, respectively. VIX and JLN are standardized by subtracting the means and dividing by the standard deviations over the 2000Q1-2016Q1 interval.

Figure 1.6: U.S. long-run uncertainty compared to EPU



Notes: In this figure, LRU is the U.S. long-run uncertainty index from Figure 1.4, EPU is the overall index of U.S. economic policy uncertainty by Baker, Bloom and Davis (2016), which combines the news-based index with other three measures of uncertainty (an index of tax expirations, a measure of forecast disagreement over consumer prices and a measure of forecast disagreement over federal/state/local government purchases). The source for EPU is www.policyuncertainty.com and quarterly data are obtained by averaging monthly data. EPU is standardized by subtracting the mean and dividing by the standard deviation over the 2000Q1-2016Q1 interval.

Table 1.1: Correlations of uncertainty measures

	VIX	JLN	EPU	EPU_n	SRU	LRU
VIX	1.00					
JLN	0.70	1.00				
EPU	0.54	0.15	1.00			
EPU_n	0.61	0.16	0.89	1.00		
SRU	0.74	0.59	0.44	0.41	1.00	
LRU	0.12	-0.04	0.67	0.35	0.33	1.00

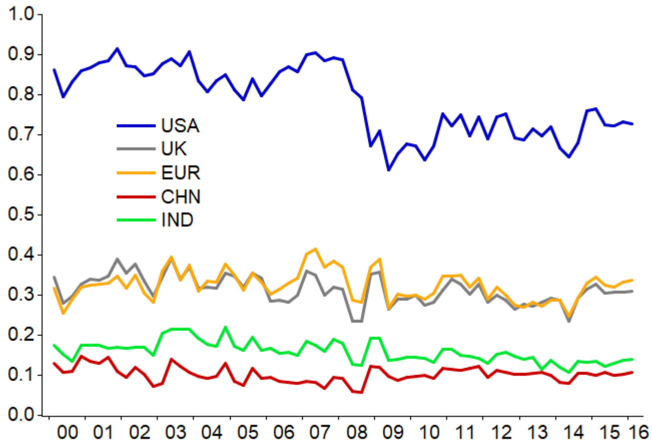
Notes: In this table, SRU is the U.S. short-run uncertainty index, LRU is the U.S. long-run uncertainty index, VIX is the volatility index by the Chicago Board Options Exchange, JLN is the U.S. macro uncertainty index by Jurado et al. (2015), EPU is the overall index of U.S. economic policy uncertainty by Baker, Bloom and Davis (2016) and EPU_n is the news-based component of EPU. All correlations are computed over the period 2000Q1-2016Q1.

China and India, as well as the ratio for the United States itself. Table 1.2 reports the average spillovers over the period 2000Q1-2016Q1 for all countries. Over the measurement period, the direct and indirect effects of domestic parameter uncertainty amount on average to almost 80% of total short-run uncertainty in the U.S. The ratio falls in 2008, signaling larger effects of international uncertainty starting in the crisis period. For both the Euro area and the U.K., the total spillover effects of U.S. uncertainty exceed 30% of total short-run uncertainty on average, while for China and India the spillovers represent lower fractions of total uncertainty (10% and 16% on average, respectively). Also, the spillovers do not exhibit large fluctuations over time. Figure 1.8 and Table 1.3 report the results on spillovers of U.S. long-run uncertainty. The ranking of countries is broadly similar, while the ratios are lower than their short-run counterparts, which means that U.S. uncertainty weighs relatively more on short-run relationships than on long-run ones.

1.6 Concluding remarks

This paper has developed global and country-specific indices of macroeconomic uncertainty. The proposed methodology is based on a GVAR model and provides comprehensive indicators by measuring uncertainty about real and financial variables in a large number of countries simultaneously. Since countries are interconnected in the GVAR, this approach takes into

Figure 1.7: Spillovers of U.S. short-run uncertainty



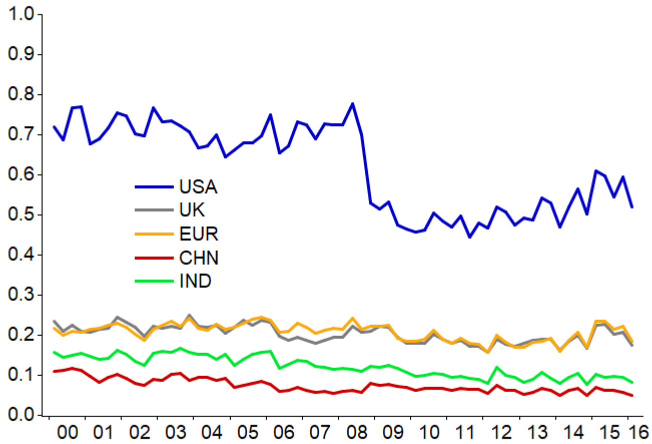
Notes: For each country (U.S., Euro area, U.K., China and India), the figure plots the component of forecast uncertainty that only depends on U.S. parameter uncertainty, expressed as a fraction of total forecast uncertainty. Quarterly data for the period 2000Q1-2016Q1.

Table 1.2: Average spillovers of U.S. short-run uncertainty (2000Q1-2016Q1)

USA	0.784	KOR	0.168
CAN	0.427	NOR	0.168
EUR	0.323	SEA	0.167
UK	0.313	NZL	0.163
CHE	0.257	IND	0.158
BRA	0.231	LAM	0.154
MEX	0.224	SAU	0.138
JAP	0.218	AUS	0.131
SWE	0.211	TUR	0.126
ZAF	0.178	CHN	0.101

Notes: For each country, the table reports the component of forecast uncertainty that only depends on U.S. parameter uncertainty, expressed as a fraction of total forecast uncertainty and averaged over the period 2000Q1-2016Q1. Countries are identified by ISO codes. For the three areas, the following codes are used: EUR = Euro area, LAM = Latin American area and SEA = South-East Asia.

Figure 1.8: Spillovers of U.S. long-run uncertainty



Notes: For each country (U.S., Euro area, U.K., China and India), the figure plots the component of forecast uncertainty that only depends on U.S. parameter uncertainty, expressed as a fraction of total forecast uncertainty. Quarterly data for the period 2000Q1-2016Q1.

Table 1.3: Average spillovers of U.S. long-run uncertainty (2000Q1-2016Q1)

USA	0.616	KOR	0.127
CAN	0.309	SEA	0.122
MEX	0.216	IND	0.121
EUR	0.206	LAM	0.120
UK	0.202	AUS	0.118
CHE	0.191	ZAF	0.114
BRA	0.172	NOR	0.113
JAP	0.161	SAU	0.106
SWE	0.137	TUR	0.097
NZL	0.132	CHN	0.073

Notes: For each country, the table reports the component of forecast uncertainty that only depends on U.S. parameter uncertainty, expressed as a fraction of total forecast uncertainty and averaged over the period 2000Q1-2016Q1. Countries are identified by ISO codes. For the three areas, the following codes are used: EUR = Euro area, LAM = Latin American area and SEA = South-East Asia.

account the international propagation of uncertainty as well as common uncertainty shocks, so that each country-specific index incorporates uncertainty about the global economy. The country-specific indices turn out to be highly correlated, which makes global measures particularly informative. The approach is also used to quantify the international spillovers of uncertainty.

Moreover, the paper proposes the distinction between short-run and long-run uncertainty as a possible way to reconcile existing indicators of uncertainty. In particular, it develops indices of short-run and long-run uncertainty based on the error-correction properties of the GVAR model and shows that such measures encompass key features of different indicators of uncertainty. The short-run uncertainty index is highly correlated with the VIX index of stock market volatility, while the long-run uncertainty index is closer to the EPU index by Baker, Bloom and Davis (2016). The results therefore suggest a possible interpretation of the differences between existing indicators, as reflecting different weights assigned to short-run and long-run issues. Further research may explore this interpretation by investigating how commonly used measures of uncertainty relate to the short-run or long-run concerns of different economic agents.

Bibliography

[1] Asker, J., J. Farre-Mensa and A. Ljungqvist (2015), "Corporate Investment and Stock Market Listing: A Puzzle?" *The Review of Financial Studies*, 28 (2), 342-390.

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

[3] Baker, S. R., N. Bloom and S. Davis (2016), "Measuring Economic Policy Uncertainty." *The Quarterly Journal of Economics*, 131 (4), 1593-1636.

[4] Barrero, J. M., N. Bloom and I. Wright (2016). "Short and Long Run Uncertainty." SIEPR Discussion Paper 16-030.

[5] Bekaert, G., M. Hoerova and M. Lo Duca (2013), "Risk, Uncertainty and Monetary Policy." *Journal of Monetary Economics*, 60 (7), 771-88.

[6] Berger, T., S. Grabert and B. Kempa (2016), "Global and Country-Specific Output Growth Uncertainty and Macroeconomic Performance." *Oxford Bulletin of Economics and Statistics*, 78 (5), 694-716.

[7] Berger, T., S. Grabert and B. Kempa (2017), "Global macroeconomic uncertainty." *Journal of Macroeconomics*, 53, 42-56.

[8] Bernanke, B. (2007), "Federal Reserve Communications." Speech at the Cato Institute 25th Annual Monetary Conference, Washington, November 14.

[9] Bloom, N. (2009), "The Impact of Uncertainty Shocks." *Econometrica* 77 (3), 623-685.

[10] Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten and S. Terry (2012), "Really Uncertain Business Cycles." NBER Working Paper no. 18245.

[11] Brainard, W. (1967), "Uncertainty and the Effectiveness of Policy." *American Economic Review*, 57 (2), 411-425.

[12] Carriero, A., T. E. Clark and M. Marcellino (2017), "Measuring Uncertainty and Its Impact on the Economy." Federal Reserve Bank of Cleveland Working Paper no. 16-22R.

[13] Cesa-Bianchi, A., M. H. Pesaran and A. Rebucci (2014), "Uncertainty and Economic Activity: A Global Perspective." CESifo Working Paper no. 4736.

[14] Clements, M. P. (2017), "Assessing Macro Uncertainty In Real-Time When Data Are Subject To Revision." *Journal of Business & Economic Statistics.*, 35 (3), 420-433.

[15] Collin-Dufresne, P., M. Johannes and L. A. Lochstoer (2016), "Parameter Learning in General Equilibrium: The Asset Pricing Implications." *American Economic Review*, 106 (3), 664–698.

[16] Coibion, O. and Y. Gorodnichenko (2012), "What Can Survey Forecasts Tell Us About Informational Rigidities?" *Journal of Political Economy*, 120 (1), 116-159.

[17] Davies, R., A. G. Haldane, M. Nielsen and S. Pezzini (2014), "Measuring the Costs of Short-Termism." *Journal of Financial Stability*, 12, 16-25.

[18] Dees, S., F. di Mauro, M. H. Pesaran and L. V. Smith (2007a), "Exploring the International Linkages of the Euro Area: a Global VAR Analysis." *Journal of Applied Econometrics*, 22 (1), 1-38.

[19] Dees, S., S. Holly, M. H. Pesaran and L. V. Smith (2007b), "Long Run Macroeconomic Relations in the Global Economy." *Economics - The Open-Access, Open-Assessment E-Journal*, 3, 1-20.

[20] Deutsche Bundesbank (2016), "Measuring Macroeconomic Uncertainty and its Impact on Investment in the Euro Area." *Monthly Report*, 68 (1), January, 42-45.

[21] ECB (2009), "Uncertainty and the Economic Prospects for the Euro Area." *Monthly Bulletin*, August, 58-61.

[22] ECB (2016), "The Impact of Uncertainty on Activity in the Euro Area." *Economic Bulletin*, 8, 55-74.

[23] ECB (2017), "Assessing the Decoupling of Economic Policy Uncertainty and Financial Conditions." *Financial Stability Review*, May, 135-143.

[24] Engle, R. F. (2011), “Long-Term Skewness and Systemic Risk.” *Journal of Financial Econometrics*, 9 (3), 437-468.

[25] Forbes, K. (2016), “Uncertainty about uncertainty.” Speech at J.P. Morgan Cazenove “Best of British” Conference, London, 23 November.

[26] Gürkaynak, R., B. Sack and E. Swanson (2005), “The Sensitivity of Long-term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models.” *American Economic Review*, 95 (1), 425-36.

[27] Hansen, L. P. (2007), “Beliefs, Doubts and Learning: Valuing Economic Risk.” *American Economic Review: Papers & Proceedings*, 97 (2), 1–30.

[28] Hoffman, D. L. and R. H. Rasche (1996), “Assessing Forecast Performance in a Cointegrated System.” *Journal of Applied Econometrics*, 11 (5), 495–517.

[29] IMF (2012), “World Economic Outlook: Coping with High Debt and Sluggish Growth.” October.

[30] Johansen, S. (1995), “Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.” Oxford University Press, Oxford.

[31] Jurado, K., S. C. Ludvigson, and S. Ng (2014), “Measuring Uncertainty: Supplementary Material.” www.aeaweb.org/aer/app/10503/20131193_app.pdf

[32] Jurado, K., S. C. Ludvigson, and S. Ng (2015), “Measuring Uncertainty.” *The American Economic Review*, 105 (3), 1177-1216.

[33] Lahiri, K., and X. S. Sheng (2010), “Measuring Forecast Uncertainty by Disagreement: The Missing Link.” *Journal of Applied Econometrics*, 25 (4), 514-38.

[34] Lin, J. L. and R. S. Tsay (1996), “Co-integration Constraint and Forecasting: an Empirical Examination.” *Journal of Applied Econometrics*, 11 (5), 519–538.

[35] Mumtaz, H. and K. Theodoridis (2017), “Common and Country Specific Economic Uncertainty.” *Journal of International Economics*, 105, March 2017, 205-216.

[36] Pesaran, M. H., T. Schuermann and S. M. Weiner (2004), “Modeling Regional Interdependencies Using a Global Error Correcting Macroeconometric Model.” *Journal of Business and Economic Statistics*, 22 (2), 129-62.

[37] Pesaran, M. H., Y. Shin and R. J. Smith (2000), "Structural Analysis of Vector Error Correction Models with Exogenous I(1) Variables." *Journal of Econometrics*, 97 (2), 293-343.

[38] Orlik, A. and L. Veldkamp (2014), "Understanding Uncertainty Shocks and the Role of Black Swans." NBER Working Paper no. 20445.

[39] Orphanides, A. and S. van Norden (2002), "The Unreliability of Output-Gap Estimates in Real Time." *Review of Economics and Statistics*, 84 (4), 569-583.

[40] Orphanides, A. and J. C. Williams (2005), "Inflation Scares and Forecast-Based Monetary Policy." *Review of Economic Dynamics*, 8 (2), 498-527.

[41] Ozturk, E. O. and X. S. Sheng (2018), "Measuring Global and Country-Specific Uncertainty." *Journal of International Money and Finance*, 88, 276–295.

[42] Rossi, B. and T. Sekhposyan (2015), "Macroeconomic Uncertainty Indices Based on Nowcast and Forecast Error Distributions." *American Economic Review: Papers & Proceedings*, 105 (5), 650-655.

[43] Rossi, B. and T. Sekhposyan (2017), "Macroeconomic Uncertainty Indices for the Euro Area and its Individual Member Countries." *Empirical Economics*, 53 (1), 41-62.

[44] Scotti, C. (2016), "Surprise and Uncertainty Indexes: Real-Time Aggregation of Real-Activity Macro Surprises." *Journal of Monetary Economics*, 82, 1-19.

[45] Söderström, U. (2002), "Monetary Policy with Uncertain Parameters." *The Scandinavian Journal of Economics*, 104 (1), 125–145.

[46] Stock, J. and M. Watson (2012), "Disentangling the Channels of the 2007-2009 Recession." *Brookings Papers on Economic Activity*, Spring, 81-135.

[47] Weitzman, M. L. (2007), "Subjective Expectations and Asset-Return Puzzles." *American Economic Review*, 97 (4), 1102–30.

[48] Wieland, V. (2000). "Monetary Policy, Parameter Uncertainty and Optimal Learning." *Journal of Monetary Economics*, 46 (1), 199-228.

[49] Xia, Y. (2001), "Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation." *The Journal of Finance*, 56 (1), 205-246.

Appendix 1.A Data sources

The dataset used in this paper was obtained by extending up to 2016Q1 the data used by Cesa-Bianchi et al. (2014). The sources for the period 2013Q2-2016Q1 are the following.

For real GDP, the year-on-year percent changes (not seasonally adjusted) from the IMF's International Financial Statistics (IFS) were used to extend the GDP level series for all countries except China and Singapore. For China, the year-on-year GDP growth data (seasonally adjusted) released by the National Bureau of Statistics of China/Thomson Reuters were used. For Singapore, I used the year-on-year percent changes in GDP at constant prices (in 2010 Singapore Dollars, seasonally adjusted) released by the Department of Statistics of Singapore.

For CPI, the year-on-year percent changes (not seasonally adjusted) from the IFS (item: "Consumer Prices, All items") were used for all countries to extend the index levels. For Argentina, due to missing values in the series, the IFS data were used only up to 2013Q4; from 2014Q1 onwards, I used the year-on-year percent changes (seasonally adjusted) provided by the OECD.

For all the short-term interest rates, I used quarterly changes to extend forward the levels in the original dataset. Following Cesa-Bianchi et al. (2014), I used the IFS data for China (item: "Deposit Rate"); New Zealand (item: "Discount Rate"); Canada, Malaysia, Philippines, South Africa, Sweden, U.K., U.S. (item: "Treasury Bill Rate"); Australia, Finland, Indonesia, Japan, South Korea, Singapore, Spain, Switzerland and Thailand (item: "Money Market Rate"). For Austria, Belgium, France, Germany, Italy and the Netherlands I used the 3-month Euribor rate (from Datastream). For India, I used the 3-month Treasury bill yield provided by the Reserve Bank of India. For Norway, the IFS item "3 Months Forward Rate (Dollar)" was used.

For exchange rates, all the data were collected from Datastream (source: Global Treasury Information Services). Quarter-on-quarter percentage changes were used to extend the original data.

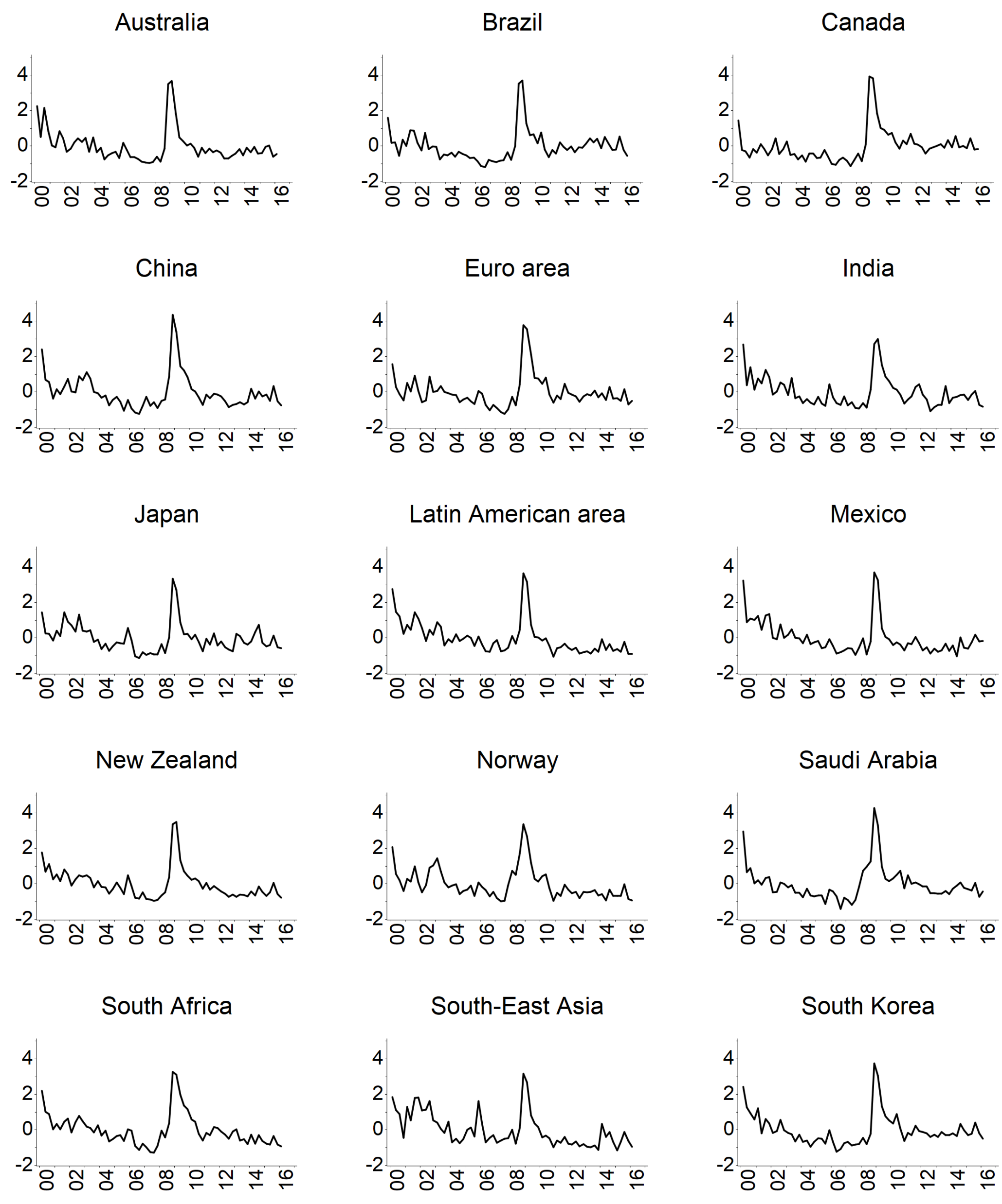
The time series for the equity indices were provided by Datastream. Quarter-on-quarter percentage changes were used to extend the original data.

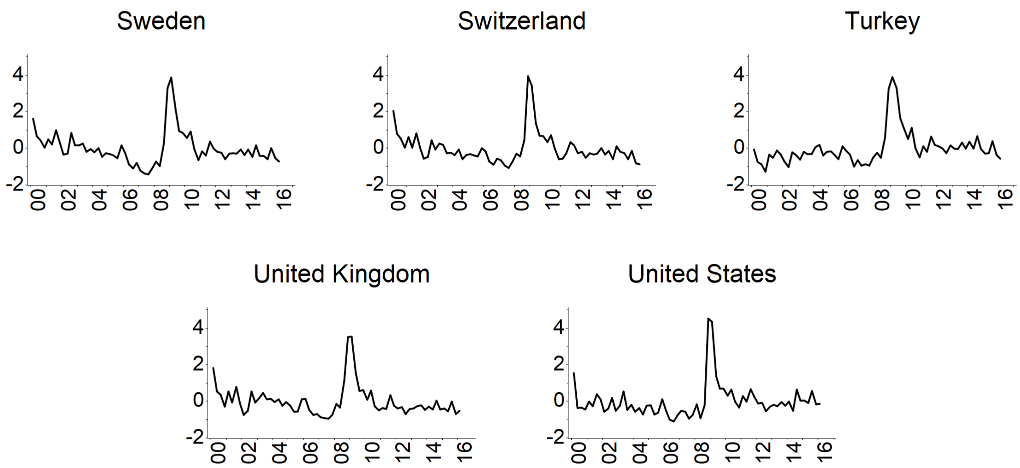
The 2013-2015 data on imports (c.i.f.) and exports are from the IMF's Direction of Trade Statistics.

The data on annual GDP levels in purchasing power parity terms (current international $) are from the World Bank's World Development Indicators.

Appendix 1.B Figures and tables

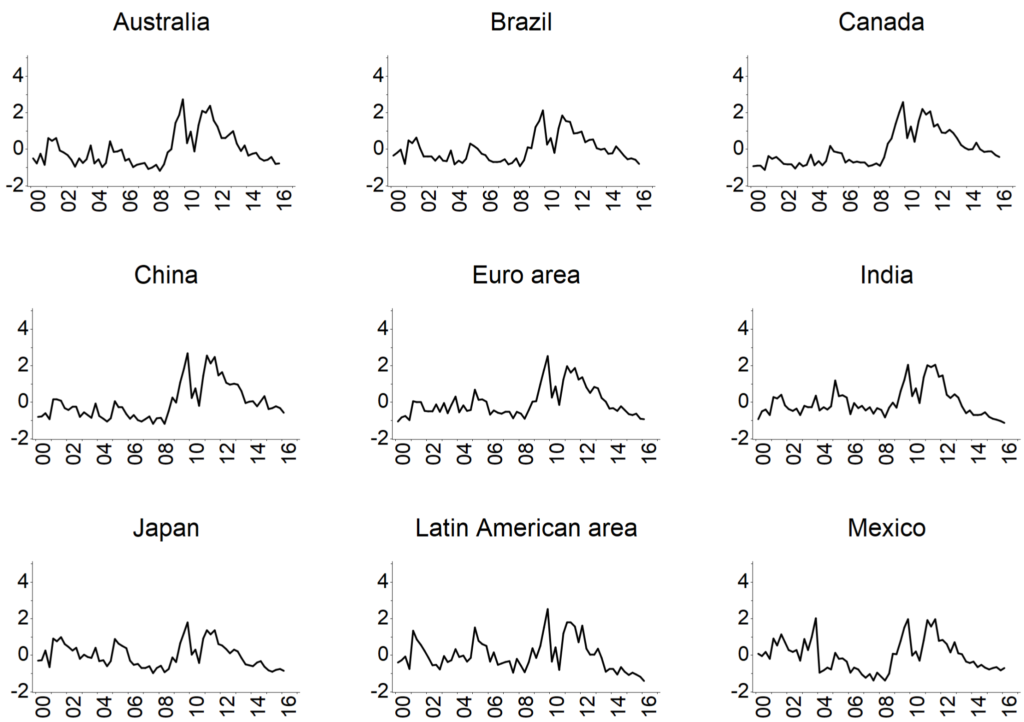
Figure 1.9: Country-specific short-run uncertainty indices

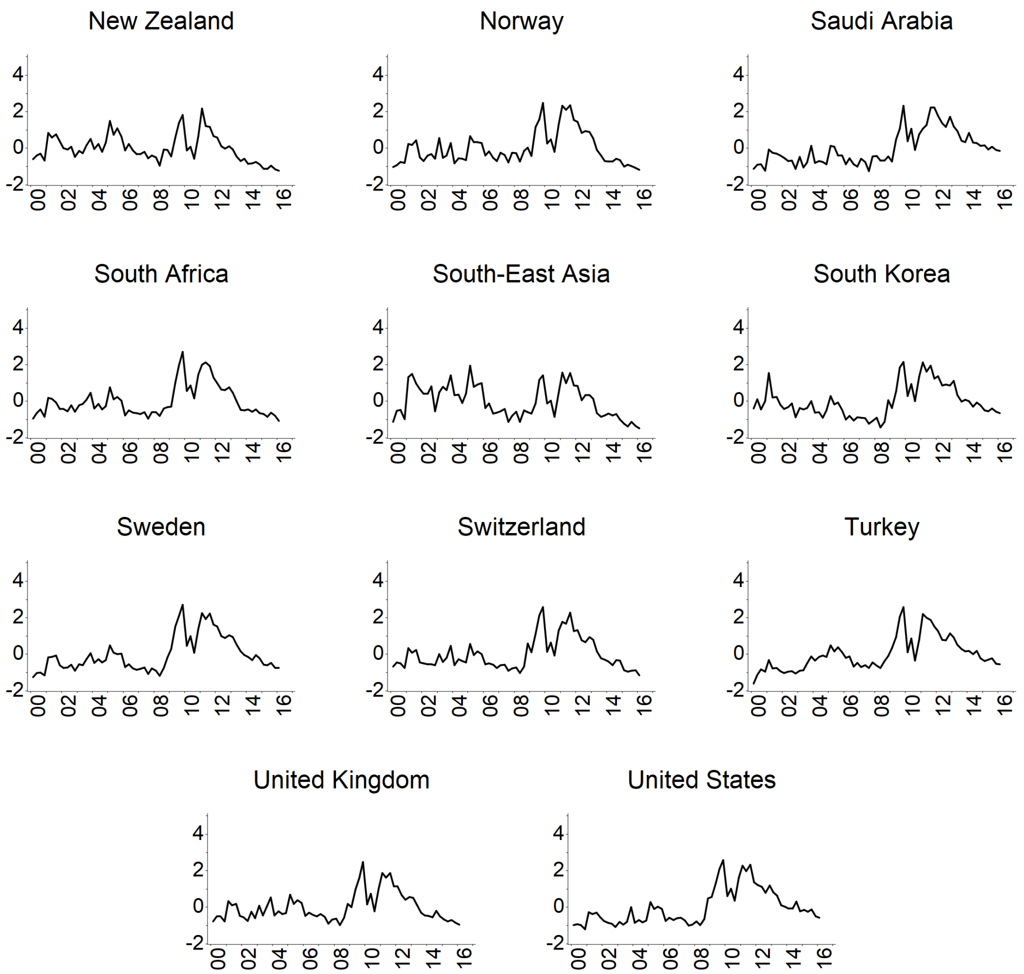




Notes: Each index is calculated as the average short-run uncertainty across the domestic variables included in the GVAR model and is expressed in standardized units. The data are quarterly and span the period 2000Q1-2016Q1.

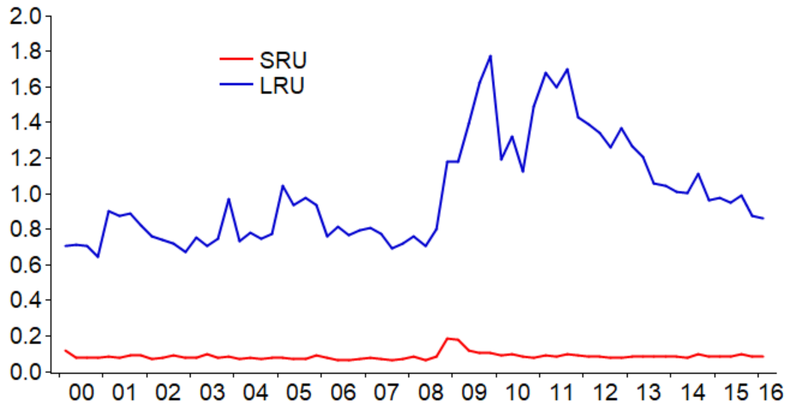
Figure 1.10: Country-specific long-run uncertainty indices





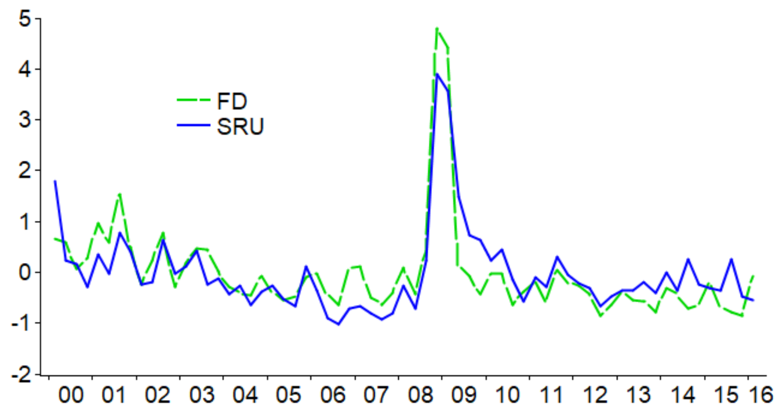
Notes: Each index is calculated as the average long-run uncertainty across the domestic variables included in the GVAR model and is expressed in standardized units. The data are quarterly and span the period 2000Q1-2016Q1.

Figure 1.11: Magnitudes of U.S. short-run and long-run uncertainty



Notes: This figure compares the U.S. short-run uncertainty index (SRU) and the U.S. long-run uncertainty index (LRU) without applying the standardization described in the paper. Here, for each U.S. variable, the variable-specific absolute measures of short-run and long-run uncertainty are both divided by the average of the long-run measure. SRU and LRU are then calculated by averaging the variable-specific uncertainty measures. The data are quarterly and span the period 2000Q1-2016Q1.

Figure 1.12: First-difference GVAR and short-run uncertainty



Notes: In this figure, SRU is the global short-run uncertainty index and FD is the global uncertainty index obtained by applying the same measurement procedure to a GVAR model in first differences (i.e. without cointegration). Both indices are expressed in standardized units. The data are quarterly and span the period 2000Q1-2016Q1.

Table 1.4: Cross-country correlations of short-run uncertainty

	AUS	BRA	CAN	CHE	CHN	EUR	UK	IND	JPN	KOR	LAM	MEX	NOR	NZL	SAU	SEA	SWE	TUR	USA	ZAF
AUS	1.00																			
BRA	0.85	1.00																		
CAN	0.81	0.93	1.00																	
CHE	0.91	0.92	0.90	1.00																
CHN	0.89	0.89	0.86	0.92	1.00															
EUR	0.88	0.93	0.94	0.97	0.90	1.00														
UK	0.90	0.87	0.87	0.97	0.91	0.97	1.00													
IND	0.93	0.84	0.79	0.90	0.88	0.87	0.87	1.00												
JPN	0.85	0.89	0.82	0.87	0.87	0.89	0.85	0.86	1.00											
KOR	0.91	0.92	0.87	0.94	0.89	0.90	0.88	0.90	0.87	1.00										
LAM	0.88	0.78	0.67	0.88	0.89	0.80	0.86	0.87	0.84	0.84	1.00									
MEX	0.90	0.81	0.71	0.87	0.86	0.81	0.85	0.89	0.85	0.89	0.94	1.00								
NOR	0.79	0.75	0.71	0.87	0.88	0.85	0.90	0.79	0.79	0.77	0.87	0.79	1.00							
NZL	0.95	0.85	0.82	0.94	0.93	0.91	0.93	0.92	0.89	0.91	0.91	0.90	0.87	1.00						
SAU	0.84	0.83	0.83	0.90	0.87	0.85	0.88	0.82	0.78	0.88	0.82	0.82	0.85	0.86	1.00					
SEA	0.77	0.69	0.57	0.74	0.77	0.71	0.73	0.81	0.86	0.75	0.88	0.83	0.76	0.85	0.66	1.00				
SWE	0.91	0.92	0.90	0.96	0.90	0.98	0.95	0.90	0.89	0.92	0.83	0.84	0.83	0.94	0.83	0.74	1.00			
TUR	0.66	0.80	0.89	0.77	0.69	0.86	0.79	0.62	0.64	0.68	0.49	0.48	0.61	0.67	0.67	0.40	0.82	1.00		
USA	0.83	0.93	0.97	0.91	0.87	0.93	0.89	0.81	0.83	0.88	0.72	0.78	0.72	0.85	0.84	0.61	0.90	0.85	1.00	
ZAF	0.90	0.85	0.83	0.94	0.91	0.92	0.92	0.90	0.86	0.90	0.86	0.84	0.86	0.94	0.85	0.79	0.95	0.72	0.82	1.00

Notes: Countries are identified by ISO codes. For the three areas, the following codes are used: EUR = Euro area, LAM = Latin American area and SEA = South-East Asia.

Table 1.5: Cross-country correlations of long-run uncertainty

	AUS	BRA	CAN	CHE	CHN	EUR	UK	IND	JPN	KOR	LAM	MEX	NOR	NZL	SAU	SEA	SWE	TUR	USA	ZAF
AUS	1.00																			
BRA	0.97	1.00																		
CAN	0.92	0.91	1.00																	
CHE	0.95	0.94	0.90	1.00																
CHN	0.96	0.95	0.95	0.92	1.00															
EUR	0.95	0.93	0.92	0.98	0.93	1.00														
UK	0.94	0.93	0.88	0.98	0.90	0.98	1.00													
IND	0.90	0.88	0.81	0.94	0.84	0.94	0.96	1.00												
JPN	0.85	0.85	0.64	0.84	0.75	0.82	0.86	0.86	1.00											
KOR	0.94	0.93	0.87	0.91	0.94	0.90	0.89	0.84	0.81	1.00										
LAM	0.82	0.82	0.66	0.87	0.73	0.85	0.90	0.93	0.90	0.78	1.00									
MEX	0.82	0.80	0.63	0.82	0.76	0.78	0.81	0.76	0.87	0.82	0.75	1.00								
NOR	0.91	0.90	0.84	0.95	0.88	0.96	0.96	0.96	0.83	0.84	0.88	0.77	1.00							
NZL	0.73	0.73	0.55	0.79	0.62	0.78	0.84	0.88	0.89	0.67	0.93	0.72	0.83	1.00						
SAU	0.85	0.81	0.90	0.80	0.90	0.84	0.78	0.71	0.57	0.83	0.56	0.58	0.77	0.42	1.00					
SEA	0.63	0.59	0.39	0.68	0.51	0.67	0.72	0.77	0.87	0.60	0.83	0.73	0.71	0.90	0.37	1.00				
SWE	0.95	0.93	0.96	0.96	0.95	0.98	0.95	0.89	0.75	0.90	0.77	0.73	0.92	0.70	0.88	0.58	1.00			
TUR	0.87	0.85	0.94	0.88	0.89	0.92	0.89	0.82	0.61	0.81	0.70	0.57	0.85	0.62	0.87	0.46	0.95	1.00		
USA	0.93	0.92	0.99	0.92	0.96	0.94	0.91	0.84	0.68	0.88	0.70	0.67	0.87	0.60	0.89	0.44	0.97	0.94	1.00	
ZAF	0.94	0.91	0.88	0.96	0.90	0.98	0.98	0.95	0.85	0.89	0.87	0.80	0.95	0.82	0.78	0.72	0.95	0.88	0.90	1.00

Notes: Countries are identified by ISO codes. For the three areas, the following codes are used: EUR = Euro area, LAM = Latin American area and SEA = South-East Asia.

Chapter 2

Financial Cycles and GDP Predictions in the United States

Abstract

This paper documents the predictive potential of financial-cycle indicators for real economic activity by means of an extensive comparative evaluation conducted on a large dataset for the United States. The analysis combines standard methods for assessing in-sample and out-of-sample predictive performance with a variety of techniques for data-rich environments. The results indicate that two housing and balance-sheet indicators, namely a cyclically-adjusted house price-rent ratio and the liabilities-income ratio of the non-corporate business sector, have been the most powerful predictors of U.S. GDP over a range of horizons between 1 and 7 years during the last decades. Their predictive ability appears comparatively robust to the choice of evaluation method and stable over time. Overall, financial-cycle variables provide valuable predictive content, both when considered individually and when pooled. Large data-intensive models and forecast combinations that use all available predictors generally fail to produce forecast gains over the best single-predictor models. As an additional indication of the practical relevance of the findings, the paper shows that small models based on the best financial-cycle indicators beat the IMF forecasts over all horizons.

2.1 Introduction

Investigating financial cycles and macro-financial linkages is essential to understand business cycles. Following the global financial crisis and the Great Recession, the interactions between asset price dynamics, balance sheets and real economic activity have become increasingly important ingredients of macroeconomic analysis (see, e.g., Taylor and Uhlig [eds.] 2016; Gertler and Gilchrist 2018; Mian and Sufi 2014b). In particular, house prices and credit have received a great deal of attention and have been argued to provide "the most parsimonious description of the financial cycle" (Borio 2012).

From an empirical perspective, recent research has delivered substantial results on the predictive potential of financial-cycle variables for business cycles (e.g., Mian and Sufi 2014b; Jordà, Schularick and Taylor 2014). However, comprehensive comparative evaluations of alternative predictors are still needed to answer questions such as: are housing and credit-cycle indicators generally more powerful predictors of real activity than other macroeconomic and financial variables? How best to exploit information on financial cycles to predict output? Do any specific indicators, among those related to credit, balance sheets and housing, stand out as effective predictors? Since the wild movements in output of the last decades have been largely blamed on financial excesses, there are reasons to assume that variables containing information on both financial conditions and observed economic fundamentals, such as debt-to-income ratios and asset price-to-earnings ratios, may convey superior predictive content about future economic activity.

This paper offers a novel assessment of the predictive power of macroeconomic and financial indicators for U.S. real GDP in the last four decades, with a special focus on housing and credit-cycle indicators. The evaluation is extensive both in terms of data and in terms of methodologies. Regarding the data, the analysis is conducted on a dataset of 272 quarterly indicators, which contains the FRED-QD macroeconomic database developed at the St. Louis Fed (McCracken and Ng 2015), but also includes additional variables. Concerning the methodologies, the paper explores a variety of approaches for performance measurement, variable selection and shrinkage, in order to exploit the possibilities offered by a data-rich environment. The evaluation takes into account both in-sample and out-of-sample measures of predictive importance over several horizons. In-sample performance is assessed through measures that range from the regression R^2 of single-predictor autoregressive distributed lag (ARDL) models to statistical significance in post-LASSO regressions and best subset selection. Out-of-sample evaluation is conducted using both direct forecasts from ARDL models and iterated forecasts from bivariate and multivariate VAR models, including data-intensive variants such as LASSO VARs, Large Bayesian VARs and factor models using principal components. Forecast combinations are also examined as an alternative way to pool information.

The main findings can be summarized as follows. First, the paper documents that the log house price-to-rent ratio, computed over the aggregate stock of owner-occupied housing and corrected for business-cycle fluctuations (Davis, Lehnert and Martin 2008; Contessi and

Kerdnunvong 2015), has been the most powerful predictor of U.S. GDP over medium/long horizons (i.e. longer than 3 years), the result being robust across the different evaluation methods considered. The ratio also features among the top predictors for horizons of 1-3 years, while it appears relatively less valuable for 1-quarter-ahead forecasts. The unadjusted log price-rent ratio also ranks among the best predictors, although it generally falls short of its adjusted counterpart. Second, the aggregate balance sheet of the noncorporate business sector also provides outstanding predictive information. Specifically, the ratio of nonfinancial noncorporate business sector liabilities to disposable business income exhibits the highest out-of-sample performance for horizons of 1-3 years and in general ranks among the most powerful predictors based on both in-sample and out-of-sample criteria.[1]. The paper also shows that other housing and credit indicators rank prominently overall, although their importance is more sensitive to the choice of the evaluation criteria and more unstable over time. This is the case for the credit-GDP ratio and residential fixed investment. Other variables that rank highly based on several evaluation metrics include the OECD composite leading indicator, the unfilled orders for durable goods and the ratio of household net worth to disposable income. Large forecasting models, which pool information across time series at the estimation stage, do not improve over the best single-predictor models. Also, standard forecast combinations using equal weights or Bayesian model averaging do not generate gains in forecast accuracy. At the same time, pooling financial-cycle variables provides better results than pooling all available information, at least for multi-year horizons.

The main contribution of the paper is a set of stylized facts on macro-financial interactions. Overall, the empirical analysis corroborates the importance of financial-cycle indicators for predictions of business cycles and offers fresh evidence for research on financial vulnerabilities and risk-taking. As far as the housing sector is concerned, on the one hand the results confirm the leading role of house prices in capturing real-financial interactions, on the other hand they indicate that using an appropriate valuation ratio, as opposed to simple price movements, considerably improves forecast performance. In particular, asset price-to-earnings ratios, which are commonly used to detect market overheating and to test for bubbles (e.g. Phillips et al. 2015), seem particularly suitable for investigating the effects of asset booms and the associated financial vulnerabilities on economic fluctuations. The cyclically-adjusted price-to-rent ratio (CAPR) (see Contessi and Kerdnunvong 2015), which stands out in the evaluation, is the housing-market counterpart of the popular cyclically-adjusted price-to-earnings ratio (CAPE) proposed by Campbell and Shiller (1998) for the

[1] The result is all the more interesting given that the nonfinancial noncorporate business sector includes the activities associated with tenant-occupied housing.

equity market, using a simple methodology that consists in dividing an asset price index in real terms by the ten-year moving average of real earnings on the index.

The remainder of the paper is structured as follows. Section 2.2 reviews the related literature. Section 2.3 introduces the dataset. Section 2.4 presents the in-sample evaluation. Section 2.5 presents the (pseudo) out-of-sample evaluation. This includes results from bivariate models (ARDL and VAR), small multivariate models as well as high-dimensional models (Large Bayesian VAR, LASSO VAR and factor models) and forecast combinations. Section 2.6 concludes.

2.2 Related literature

Evaluating predictors of economic activity is one of the traditional tasks of empirical macroeconomics. In this field, a number of papers have dealt with broad-based evaluations using large datasets of predictors. In particular, Stock and Watson (2003) assess forecasts of GDP growth in several countries including the United States, with a special focus on the role of asset prices. Marcellino, Stock and Watson (2003) and Banerjee, Marcellino and Masten (2005) evaluate leading indicators for the Euro Area. Rossi and Sekhposyan (2014) analyze density forecasts of U.S. output growth using the Stock and Watson (2003) updated dataset.

Based on their own evaluation and on previous literature, Stock and Watson (2003) conclude that the evidence on the predictive usefulness of asset prices (including interest rates) for GDP growth is generally mixed. When taken together, asset prices have clear predictive content and tend to perform better than non-financial variables. However, no single variable stands out as a reliable predictor. In particular, they find strong empirical instability of predictive relations and show that in-sample evaluations typically provide poor guidance for forecasting. Interestingly, financial variables that should have substantial predictive content based on economic theory, such as stock prices and log dividend yields, often perform poorly in practice. Others, most notably the term spread of interest rates, have higher predictive content only in some periods. Moreover, Stock and Watson (2003) find that simple combinations of forecasts lead to improvements in forecast accuracy compared to benchmark AR forecasts and forecasts from individual predictors.

More recently, Claessens and Kose (2017) have surveyed the literature on the interactions between asset prices and the real economy. A general conclusion is that equity and house prices have some predictive content for economic activity. Empirical studies provide evidence that equity prices tend to affect investment in particular, while house prices have a larger impact on consumption. Standard channels through which asset prices affect economic outcomes, according to the economic theory, include wealth and redistribution effects, substitution effects, collateral and credit effects. Also, asset prices are forward-looking and therefore tend to lead output growth. Among others, Bluedorn et al. (2016) find that they help predict recessions in advanced economies (see also IMF 2000). However, research also suggests that there are significant limits to the forecasting power of asset prices (Claessens and Kose 2017).

In the wake of the crisis, the notion of financial cycles, as distinct from business cycles, has been increasingly investigated. At a basic level, financial cycles can be defined as cycles in credit, housing and equity markets, which are typically longer than business cycles (Claessens, Kose and Terrones 2011; Drehmann, Borio and Tsatsaronis 2012; Borio 2012). A richer definition has been proposed, for instance, by Borio (2012), who characterizes the financial cycle as the "interactions between perceptions of value and risk, attitudes towards risk and financing constraints, which translate into booms followed by busts". The concept of financial cycle conveys the idea that credit, asset prices and financial leverage are mutually reinforcing, or at least tend to co-move.[2] In this respect, while credit and housing are generally recognized as key components of the financial cycle, the role of equity prices is more questionable. Drehmann et al. (2012) show that credit and property prices co-vary closely, whereas equity price variability concentrates at relatively higher frequencies. Claessens, Kose and Terrones (2011) also point out that cycles in credit and house prices appear to be the most highly synchronized within countries, while credit and equity cycles exhibit the highest synchronization across countries. Borio and Drehmann (2009) show that a leading indicator based on equity prices alone failed to give early warnings of the recent banking system distress, and that incorporating information on property prices has proved effective in several countries, including the United States. Interactions between credit and house prices can be especially strong as a result of collateral constraints and the use of housing wealth as collateral (Claessens, Kose and Terrones 2011, Borio 2012, Piazzesi and Schneider 2016).

Jordà, Schularick and Taylor (2013, 2014, 2016) provide stylized facts on the relationship between long trends in credit and business cycles in advanced economies. They find that the aftermath of lending booms is typically characterized by deeper recessions and slower recoveries than normal, as a result of debt overhang and deleveraging (see also Reinhart and

[2]While the study of macro-financial linkages has gained new popularity in recent years, the idea that interactions between credit, asset prices and leverage represent the key mechanisms at the root of financial crises is a long-standing one in economics and was central in the work of authors such as Charles Kindleberger and Hyman Minsky (Kindleberger 1978, Minsky 1982).

Rogoff 2009). Moreover, they show that the unprecedented increase of the credit-GDP ratio in advanced economies starting in the mid-20th century was mainly driven by a boom in mortgage credit, especially to households, and find that real estate credit has become a more powerful predictor of financial turmoil during the same period. When the 2007-2008 financial crisis broke out, real estate credit accounted for around two thirds of total bank credit in the United States.

In a series of influential contributions, Mian and Sufi (2013, 2014a, 2014b, 2018) provide extensive evidence on the prominent role of housing wealth and household debt in the recent financial crisis, the Great Recession and the subsequent slow recovery. In particular, they document that the deterioration in the housing net worth of households is a key explanatory factor of the drop in U.S. employment during the 2007-2009 crisis.[3] Leamer (2007, 2015) argues that housing is the economic sector with the largest contribution to recessions in the United States. Buiter (2010) investigates the effects of housing wealth on consumption within a theoretical framework. He shows that in an overlapping generations model increases in house prices that are justified by fundamentals boost consumption only if the birth rate is positive, whereas in a representative agent model (with zero birth rate) there is a wealth effect of housing on consumption if changes in house prices reflect speculative bubbles. A recent survey on housing is provided by Piazzesi and Schneider (2016). As they point out, “the first half of the 2000s saw not only the largest housing boom in postwar U.S. history, but also new research that introduced an explicit role for housing in macroeconomics”.

Relatedly, the financial crisis and its aftermath have spurred a renewed interest in asset price bubbles. Building on a literature from the 1980’s and 1990’s (e.g. Diba and Grossman 1988, Evans 1991), new bubble detection tests have been proposed, such as the test by Phillips et al. (2015). Brunnermeier and Oehmke (2013) provide an overview of the literature on bubbles, financial crises and systemic risk. Broadly in line with the theoretical framework adopted by Kindleberger (1974) and Minsky (1982), they distinguish two phases of financial crises. In the run-up phase, bubbles and imbalances gradually build up, setting the stage for financial vulnerability. The second phase is the crisis phase, in which risks materialize and a

[3] One of the key questions Mian and Sufi address is why the housing market appears to have had much more dramatic effects on economic activity than the stock market in recent decades. In fact, while the bursting of the stock market bubble of the late 1990’s only resulted in a mild recession, the housing market crash of 2006-2007 was associated with the most severe recession after the Great Depression. According to Mian and Sufi (2014a), the main reason “is that the marginal propensity to consume out of a housing-wealth shock is much higher — housing wealth is a levered asset held by lower net-worth households. The rich are the primary owners of tech stocks, and they respond much less to a decline in wealth. The larger MPC of indebted home owners is crucial for understanding why the housing crash was so much worse than the tech crash.”

financial meltdown occurs. The interactions of asset prices, leverage and investment amplify the effects of financial crashes on the real economy.[4]

Finally, this paper relates to the literature on valuation ratios, such as price-earnings, price-dividend and price-rent ratios, and their predictive power for output. Campbell (1999) provides theoretical justification for using the stock market log price-dividend ratio to forecast output growth. Empirically, he finds that the variable shows no clear predictive ability. Sommer, Sullivan and Verbrugge (2011) calculate that approximately half of the run-up in the U.S. house price-rent ratio between 1995 and 2005 can be explained as the equilibrium response of the housing market to changes in fundamentals, such as interest rates, the required down payment and income. The cyclically-adjusted price-to-earnings ratio (CAPE) has been introduced by Campbell and Shiller (1998) for the equity market (see also Shiller 2015) as a way to smooth spikes due to expansions and recessions. Contessi and Kerdnunvong (2015) use the cyclically-adjusted price-rent (CAPR) ratio to test for bubbles using the Phillips et al. (2015) test.[5] Like in this paper, they use data on the aggregate stock of owner-occupied housing collected by Davis et al. (2008) (see the next section). While there has long been an interest in valuation ratios as potential predictors of economic activity (Campbell 1999, Stock and Watson 2003), this paper is the first to show the outstanding predictive power of the CAPR in recent decades.

2.3 Variables and data

Most of the analysis in the paper is conducted using a large quarterly dataset (272 variables) that combines the FRED-QD dataset[6] (Tables 2.35 to 2.48) with the variables contained in Table 2.1, which I will refer to as the "reduced dataset". To stress the macroeconomic relevance of the analysis (and to circumvent some pitfalls that result from high dimensionality), additional results are presented using the reduced dataset only, which contains a selection of particularly popular macroeconomic variables along with a number of meaningful financial-cycle indicators. Importantly, some of these financial-cycle indicators are not present in

[4]The literature on financial imbalances, bubbles and crises is too broad to be covered exhaustively here. Two popular books that are related to the topics of this paper are Reinhart and Rogoff (2009) and Shiller (2015).

[5]They find three bubbly periods between 1960Q1 and 2014Q1: 1965Q3-1968Q4, 1977Q4-1978Q1 and 2000Q2-2006Q1.

[6]Available at https://research.stlouisfed.org/econ/mccracken/fred-databases/. This paper uses the 2018-06 vintage.

FRED-QD.[7] Throughout the analysis, lowercase labels identify variables from the reduced dataset and uppercase labels identify variables from FRED-QD.

The analysis is conducted using data over the period 1960Q1-2017Q4. The dataset is unbalanced: Table 2.2 summarizes the structure of the full dataset by reporting the number of available and unavailable series (after transformation) at several key dates in which data availability increases. In particular, 85% of the transformed series are available in 1960Q2, approximately 90% are available in 1967Q1 and approximately 95% are available in 1974Q1.[8]

The FRED-QD dataset is made up of 248 variables, which cover in detail a wide spectrum of macro areas. McCracken and Ng (2015) classify the variables into 14 groups: National Income and Product Accounts (NIPA); Industrial Production; Employment and Unemployment; Housing; Inventories, Orders, and Sales; Prices; Earnings and Productivity; Interest Rates; Money and Credit; Household Balance Sheets; Exchange Rates; Stock Markets; Non-Household Balance Sheets; Other.

At a different level of detail, the reduced dataset is meant to combine a broad picture of the macroeconomy with a special focus on the financial cycle. Key macro variables include CPI inflation, interest and exchange rates, money growth, unemployment, oil price inflation, the OECD composite leading indicator and business confidence index, and fiscal policy as captured by public debt growth in real terms and the public debt/GDP ratio. Financial-cycle variables include: total credit to the non-financial sector, the credit-GDP ratio, households' mortgage debt and mortgage-income ratio, house prices and price-rent ratios, residential fixed investment, housing starts, stock market prices and the cyclically-adjusted price-earnings ratio, and households' interest payments as a fraction of disposable income. The dataset also includes the comprehensive National Financial Conditions Index (NFCI) by the Chicago Fed to capture a variety of other financial factors. In more detail, the cyclically-adjusted house price-rent ratio (CAPR) is constructed by dividing house prices in real terms by the 10-year

[7] The data transformations suggested by McCracken and Ng (2015) are applied to the FRED-QD raw data, the only adaptation being the use of year-on-year changes/growth rates instead of quarter-on-quarter changes/growth rates. While several variables in Table 2.1 are also included in FRED-QD, the reduced dataset generally adopts different transformations compared to FRED-QD, so that in practice the only duplicate is the BAA corporate spread. When the analysis is conducted on the full dataset, both types of transformations are included and, overall, the transformations in the reduced dataset achieve better results in terms of predictive power.

[8] All time series in FRED-QD end in 2017Q4. In the reduced dataset, all time series end in 2017Q3 or 2017Q4 except those for the CAPR and the simple price-rent ratio, which end in 2016Q1 (based on original data available at http://www.lincolninst.edu/resources/) and have not been extrapolated forward.

average real rents:

$$CAPR_t = \frac{HPI_t}{\frac{1}{40}\sum_{i=1}^{40} R_{t-1}}$$

where HPI_t denotes the house price index in real terms at time t (in quarters) and R_t is the imputed real rent at time t.[9] Both prices and rents are computed using data on the aggregate stock of owner-occupied houses.[10]

Although the focus of the paper is on empirical predictive ability, the inclusion of price-to-rent and price-to-earnings ratios in (log) levels also relates to theoretical considerations that are critical for investigating booms and busts in asset prices. In particular, based on standard asset pricing models such as the dividend discount model, in the absence of bubbles an asset price should exhibit the same order of integration as the dividends or earnings paid by the asset. Such logic provides the foundations for several bubble detection tests proposed in the literature, such as the Diba and Grossman (1988) test. However, as shown by Evans (1991), the stationarity of a price-to-earnings ratio, as determined by unit root and cointegration tests, can be equally consistent with periodically collapsing bubbles, a point that has been recently developed by Phillips et al. (2015) to propose a new test.[11] Irrespective of considerations on bubbles, the levels of price-to-earnings ratios are generally monitored by investors and researchers to make judgments on asset market valuations (e.g., Shiller 2015) and appear to be more suitable metrics then the price levels or price growth rates to evaluate asset market expensiveness and to investigate its effects on the economy.

Figures 2.1 and 2.2 plot the two variables that will stand out as the top performers throughout the analysis: the log CAPR ratio and the nonfinancial noncorporate business sector liabilities-to-income ratio (NNBLI).

---

[9]To construct the CAPR from 1960Q1, the time series of rents is extrapolated backward to 1950Q1 using the quarterly growth rate of the rent of primary residence, as released by the U.S. Bureau of Labor Statistics and provided by the FRED website (https://fred.stlouisfed.org/).

[10]See Davis et al. (2008) for methodological details. The raw data on house prices and rents are provided by the Lincoln Institute for Land Policy at http://www.lincolninst.edu/resources/. The homeownership rate in the United States in the period 1965-2017 ranged between 63% and 69% (see https://www.census.gov/housing/hvs/index.html).

[11]The augmented Dickey-Fuller test for unit roots indicates that the CAPR ratio is (trend) stationary over the period 1960Q1-2016Q1. At the same time, as mentioned earlier, the test by Phillips et al. (2015) signals multiple bubbly periods.

2.4 In-sample evaluation

The in-sample evaluation of predictive power is conducted using autoregressive distributed lag (ARDL) models for multi-period-ahead cumulative GDP growth, estimated over the timespan 1974Q1-2017Q4 for the full dataset and 1971Q1-2017Q4 for the reduced dataset. The starting date for the full dataset is chosen so as to ensure that approximately 95% of the time series do not have missing data in the sample (see Table 2.2). These series are included in the analysis, while shorter series are excluded here but will be used in the out-of-sample evaluation of section 2.5. In the reduced dataset, 1971Q1 is chosen because it is the starting period of the shortest series (*nfci*).

From a forecast-oriented perspective, the in-sample evaluation can be seen as a preliminary analysis that gives a sense of which variables better capture the variability in GDP growth and are potentially more useful for out-of-sample predictions. In addition, consistent results across the in-sample and the out-of-sample analyses would provide a basic indication of the robustness and stability of the predictive relationships.

Let h denote the prediction horizon. Throughout the analysis, five values of h are considered: 1, 4, 12, 20 and 28 quarters.[12]

2.4.1 Autoregressive distributed lag (ARDL) models with one predictor at a time

The first in-sample approach consists in predicting the GDP level h steps ahead using a bivariate model that includes one predictor at a time plus lags of GDP growth. The economic significance of alternative indicators as predictors is then evaluated using the R^2 of the regressions. Let us therefore consider ARDL regressions of the following type:

$$y_t^h = \beta_0 + \beta_1(L)y_{t-h} + \beta_{2,i}(L)x_{i,t-h} + u_t \tag{2.1}$$

where y_t^h is the log approximation of the cumulative GDP growth rate over a period of length h, i.e. $y_t^h = \ln(GDP_t) - \ln(GDP_{t-h})$, y_t is the log approximation of the year-on-year GDP

[12]For $h \geq 12$ the number of observations used for estimation is the same for all predictors in the dataset. For $h = 4$, the estimation sample ends in 2017Q4 for all variables except *capr* and *pr*, for which the sample ends in 2017Q1 (as data are available until 2016Q1. See note 8). Finally, for $h = 1$ the regressions using *capr* and *pr* have 5 missing observations and those using *cred_gdp* and *cred* have 1 missing observation at the end of the sample.

growth rate at time t, i.e. $y_t = \ln(GDP_t) - \ln(GDP_{t-4})$, $x_{i,t}$ is the i-th candidate predictor, u_t is the error term, β_0 is a constant, $\beta_1(L)$ and $\beta_{2,i}(L)$ are lag polynomials, such that $\beta_1(L)y_{t-h} = \sum_{j=1}^{p} \beta_{1,j} y_{t-h+1-j}$ and $\beta_{2,i}(L)x_{i,t-h} = \sum_{j=1}^{q} \beta_{2,i,j} x_{i,t-h+1-j}$. To ensure perfect comparability, the lag length is fixed across models. In particular, both p and q are set to 5, which appears adequate to account for serial correlation given the quarterly frequency of the data.

Table 2.3 reports the regression R^2 of the models. Variables are listed in descending order of R^2 and only the best 30 predictors are shown for each horizon, out of the 272 predictors in the full dataset. The adjusted log price-rent ratio (label: *capr*) dominates over longer horizons. It is the best predictor for $h = 12, 20, 28$ and the second best for $h = 4$. The OECD composite leading indicator (*cli*) is a particularly strong predictor over short horizons. It ranks first for $h = 1, 4$, then it gradually falls down the ranking as the horizon increases (it is 5th for $h = 12$, 9th for $h = 20$ and 37th for $h = 28$). Most of the top positions are occupied by financial-cycle indicators. Private residential fixed investment, both as a share of GDP and in growth rates (labels: *prfi_gdp* and *PRFIx* or *prfi* respectively), performs well over short horizons, while the mortgage/income ratio (*mortg_inc*) and the credit/GDP ratio (*cred_gdp*) are effective predictors over longer horizons. The noncorporate liabilities-income ratio or NNBLI (FRED-QD label: *NNBTILQ027SBDIx*) also ranks highly for $h \geq 4$.

The usefulness of the log CAPR appears even more remarkable if the lag orders p and q are selected by the Bayes information criterion (BIC) (estimating p first and then q conditional on p, given the maximum lag length of 5). In this case, it is the best predictor for all horizons except $h = 1$.

The comparatively good performance of financial-cycle ratios such as *capr*, *mortg_inc* and *cred_gdp* over long horizons does not appear to be determined by the omission of a time trend in GDP growth. When the latter is included in the models, *capr* is still the best predictor for $h = 12$ and the third best predictor for $h = 20, 28$, while *cred_gdp* and *mortg_inc* rank first and second, respectively, for $h = 20, 28$.

2.4.2 Post-LASSO regressions

While in section 2.4.1 predictors are considered one at a time, the second approach presented here selects predictors after pooling all available data. In this context, where the number of predictors exceeds the number of observations, the LASSO regression is a viable method to deal with such a task. After a LASSO-based variable-selection step, the statistical significance of the shortlisted predictors can be tested using OLS regressions to provide a measure of predictive power. When interpreting the results, it should be kept in mind that, since predictors outnumber observations and therefore regressors are perfectly collinear, it is in general not possible to pin down which variables are truly explanatory of GDP, and the LASSO can only select one subset of predictors that performs particularly well. Still, the two-step procedure can offer valuable evidence on which types of variables prove useful for predictions in a multivariate setting. Consider the following large ARDL regression:

$$y_t^h = \beta_0 + \gamma_0 t + \sum_{j=1}^{p} \left(\beta_{1,j} y_{t-h+1-j} + \bar{\beta}'_{2,j} \bar{x}_{t-h+1-j}\right) + u_t \tag{2.2}$$

where $\bar{x}_t$ is the vector containing all available predictors, i.e. $\bar{x}_t = (x_{1,t}, \ldots, x_{i,t}, \ldots, x_{n,t})'$, n is the total number of predictors and $\bar{\beta}_{2,j} = (\beta_{2,i,j}, \ldots, \beta_{2,n,j})'$. A deterministic trend with coefficient γ_0 is also taken into consideration. p is the maximum number of lags.

In the LASSO regression, a penalty on the ℓ_1 norm of the parameter vector shrinks some coefficients estimates to zero. To estimate model (2.2), the LASSO solves the following minimization problem:

$$\min_{\beta} \quad \|\bar{y}^h - \iota\beta_0 - \bar{X}\beta\|_2^2 + \lambda\|\beta\|_1$$

where $\bar{y}^h = \left(y_1^h, \ldots, y_T^h\right)'$, $\bar{X} = (\bar{X}_1, \ldots, \bar{X}_T)'$, $\bar{X}_t = (y_{t-h}, \ldots, y_{t-h+1-p}, \bar{x}'_{t-h}, \ldots, \bar{x}'_{t-h+1-p}, t)'$, ι is a $T \times 1$ vector of ones, T is the length of the sample, $\beta = (\beta'_1, \beta'_2, \gamma_0)'$, $\beta_1 = (\beta_{1,1}, \ldots, \beta_{1,p})'$ $\beta_2 = (\bar{\beta}'_{2,1}, \ldots, \bar{\beta}'_{2,p})'$ and λ denotes the penalty parameter. The absence of β_0 from the penalty term implies that the constant is included with certainty in the shrunk model. The selected predictors are those that minimize the mean square error (MSE) given the shrinkage determined by λ. The penalty parameter is tuned using a 10-fold cross-validation procedure. Two values of λ are selected to report the results: the value that minimizes the cross-validated MSE and the largest value at which the MSE lies within one standard deviation of the minimal MSE. As the partition of the sample into folds is random [13], the procedure is repeated 1000 times and the estimated λ's are averaged across iterations.

Once the LASSO has been used to perform variable selection, the significance of the selected regressors is tested by running a post-LASSO step, which estimates the shrunk model by OLS using heteroskedasticity- and autocorrelation-consistent (HAC) standard errors. To

[13] For the procedure to be applicable to time series data, the sample is partitioned by selecting rows of the matrix $(\bar{y}^h, \bar{X})$, which ensures that the lag ordering is maintained.

facilitate the analysis, Tables 2.4 to 2.7 report the outcome of the post-LASSO step for the case $p = 1$. For each horizon, the tables present the shortlist of predictors selected by the LASSO, with an indication of the significance of the respective coefficients. Variables from the reduced dataset are included first in the tables. Variables in Tables 2.4 and 2.5 are selected using the value of λ that minimizes the cross-validated MSE, whereas in Tables 2.6 and 2.7 the one-standard-error rule applies, which results in more parsimonious models (note however that the BIC favors the models obtained using the MSE-minimizing λ).

As can be seen in Tables 2.4 and 2.5, in the case of minimum cross-validated MSE, six variables are significant at the 5% level for at least three horizons out of five considered. The log CAPR (*capr*) and the NNBLI ratio (*NNBTILQ027SBDIx*) are significant for every $h \geq 12$. The term spread of interest rates (*GS10TB3Mx*) is significant for horizons up to 12 quarters. The other variables are real M1 money creation (*M1REALx*), the change in private inventories as a share of GDP (*A014RE1Q156NBEA*) and an employment indicator (*USMINE*). When the one-standard-error rule is considered, the CAPR ratio is still highly significant at $h = 12$ and $h = 20$. Three out of the six variables mentioned above are still significant at the 5% level for at least three horizons (NNBLI, the inventories/GDP ratio and the employment indicator). In particular, the NNBLI ratio is now significant at $h = 1$, $h = 4$ and $h = 28$. In addition, an index of job seeking (*HWIx*) is significant at the 5% level for $h = 4$, $h = 12$ and $h = 20$. Another money variable (*MZMREALx*) is significant for $h \geq 12$. The term spread of interest rates is still significant for $h = 4, 12$ and real M1 for $h = 12, 28$.

For both cross-validation methods, the set of indicators that are significant for at least two horizons includes credit or balance-sheet variables, labor market indicators, the OECD composite leading indicator and price indices.

2.4.3 Best subset selection for different model sizes

In the previous sections, the analysis was carried out using either one-predictor models or large models. This section proposes an additional step of in-sample evaluation by investigating which predictors are the most useful for a range of different model sizes. This is implemented on the reduced dataset (Table 2.1), for reasons of economic relevance - given the popularity of macro variables contained therein - as well as for econometric reasons (as recalled in section 2.4.2, special caution is required in the interpretation of results when predictors outnumber observations) and for computational convenience.

Let us consider the following multivariate ARDL model:

$$y_t^h = \beta_0 + \gamma_0 t + \sum_{j=1}^{5} \phi_j' \bar{z}_{t-h+1-j} + u_t \tag{2.3}$$

where $\bar{z}_t$ is the vector of variables other than GDP that are included in Table 2.1, while ϕ_j is a vector of coefficients. Five lags are included for all variables. Given model (2.3), best subset selection consists in finding, for $k = 1, 2, \ldots, K$, the k-sized subset of predictors that minimizes the in-sample MSE of the model, i.e.

$$\min_{\phi} \quad \|\bar{y}^h - \iota\beta_0 - Z\phi'\|_2^2 \quad \text{subject to} \quad \|\phi\|_0 \leq k$$

where $\bar{y}^h = \left(y_1^h, \ldots, y_T^h\right)'$, $\bar{Z} = (\bar{Z}_1, \ldots, \bar{Z}_T)'$, $\bar{Z}_t = (\bar{z}_{t-h}', \bar{z}_{t-h-1}', \ldots, \bar{z}_{t-h-4}', t)'$, $\phi = (\phi_1, \ldots, \phi_5, \gamma_0)'$. As before, the constant is always included.

Equation (2.3) has to be estimated using all possible combinations of k regressors, with $k = 1, 2, \ldots, K$, where "regressor" here means any lagged term for one of the 26 predictors in Table 2.1. While exact computations are likely to become infeasible even for moderate values of K, as they require exploring a number of models equal to $\sum_{k=1}^{K} \binom{k}{K}$, recently developed algorithms use mixed integer optimization to deliver accurate yet fast approximations.[14] As the in-sample MSE invariably decreases when additional regressors are included, the BIC is used to detect the best k.

In this setup, two criteria are used to evaluate the usefulness of variables as predictors: the order of appearance (i.e. the smallest model size at which a given predictor is selected) and the inclusion in the best model as selected by BIC.[15]

Tables 2.8 to 2.12 display the results given $K = 30$. For each model size k, a colored cell indicates that the variable in row is included in the (approximate) MSE-minimizing specification, whereas empty cells indicate exclusion from the model. In each table, variables are listed in order of appearance. As in section 2.4.1, *capr* is the single most powerful predictor for every $h \geq 12$. Conversely, it is not included in the most parsimonious models for $h = 1$ and $h = 4$, although in both cases it enters the specification selected by the BIC.

[14]The results in this section are obtained using the R package developed by Hastie et al. (2016) on a 2-core CPU with 2.4 GHz speed, setting a time limit of 1200 seconds for each k.

[15]Note that the results are exact only for small values of k (in general, the algorithm converges to a solution for $k < 5$), while they are approximations of the true MSE-minimizing models for larger values.

The OECD composite leading indicator (*cli*) is the second variable in order of appearance for $h = 1, 4, 12, 20$. The growth rate of private residential fixed investment (*prfi*) is particularly useful for horizons up to 3 years, while the credit/GDP ratio (*cred_gdp*) features in the upper part of the table for all horizons (although it is not included in the BIC-selected model for $h = 1$). The financial condition index (*nfci*) is especially useful in parsimonious specifications for short horizons, while the 10-year Treasury rate (*gs10*) is second in order of appearance for $h = 28$ and fourth for $h = 20$.

2.5 Out-of-sample evaluation

The second part of the analysis deals with evaluating the forecasting power of the predictors out of the estimation sample. To this aim, direct and iterated forecasts are tracked over time using a recursive-window scheme.[16] Direct forecasts are made using multi-period ARDL models, while iterated forecasts are produced by bivariate and multivariate VAR models as well as by three data-intensive variants: LASSO VARs, Large Bayesian VARs and factor models using principal components. The forecast horizons are the same as in the in-sample evaluation, i.e. 1, 4, 12, 20 and 28 quarters. As a result of the recursive-window scheme, time series of forecasts for different horizons are constructed for each one of the competing models and are used for comparisons. In particular, predictors are evaluated using the mean square forecast errors (MSFE) computed over the period 1990Q1-2017Q4. Given the maximum forecast horizon of 28 quarters, this implies setting the ending point of the shortest estimation window in 1983Q1, which is accommodated by moving the starting point at an earlier date than in the previous sections, namely 1968Q2. As the maximum number of lags is set to 5, this specific start date is chosen to ensure that all the VAR and the one-step ARDL models are estimated using data as far back as 1967Q1, which is the first quarter in which data are available (after transformation) for at least 90% of the time series in the dataset. Concerning the multi-step ARDL models, the range of data used for estimation will depend on the relevant horizon, the longest range starting in 1960Q1 (28 quarters of horizon plus 5 lags before 1968Q2). In 1960Q1, 66.7% of the transformed variables have data and

[16] Rolling windows of various fixed lengths (40, 60 and 80 quarters) have also been considered. However, the lowest mean square forecast errors are generally achieved in the recursive-window case, which is therefore reported in the paper.

in 1960Q2 the percentage jumps to 85.0% (see Table 2.2).[17] Multivariate VAR models in section 2.5.2.2 are estimated using the reduced dataset of Table 2.1. In 1967Q1, all variables in the reduced dataset have data except *nfci*, which starts in 1971Q1.

Competing models are initially estimated on the shortest sample 1968Q2-1983Q1 and used to make forecasts for the interval 1983Q2-1990Q1. Then the sample is recursively expanded by one quarter at a time and the estimation and forecasting steps are repeated in each iteration. The procedure ends upon reaching the sample 1968Q2-2016Q1, while subsequent quarters are used for forecast evaluation only.[18]

2.5.1 Direct forecasts: ARDL models

The first out-of-sample evaluation procedure is based on direct forecasts produced by bivariate ARDL models as in (2.1), in which the lag lenghts p and q are selected recursively using the BIC. Let $\widehat{y}^h_{i,t+h|t}$ denote the direct out-of-sample forecasts of y^h_{t+h} made by the model with the i-th predictor, estimated on data up to time t:

$$\widehat{y}^h_{i,t+h|t} = \widehat{\beta}^{(t)}_0 + \widehat{\beta}^{(t)}_1(L)y_t + \widehat{\beta}^{(t)}_{2,i}(L)x_{i,t} \tag{2.4}$$

and let $\widehat{u}_{i,t+h|t} = y^h_{t+h} - \widehat{y}^h_{i,t+h|t}$ denote the forecast error incurred by the model at time $t+h$. The i-th predictor is ranked based on the following MSFE:

[17] Given the unbalanced nature of the dataset, the actual starting date of the sample will adjust to the availability of the time series used for estimation. In sections 2.5.1 and 2.5.2.1, which consider one predictor at a time, all predictors are used regardless of the length of the respective time series. Therefore, when interpreting the results it should be kept in mind that a fraction of the predictors have fewer observations available for estimation than others. To ensure that the MSFE is computed over the same timespan for all predictors, only variables that have sufficient data to produce forecasts for 1990Q1 should be considered. However, even if the variables with an insufficient number of observations are included in the evaluation, none of them ranks among the best performing predictors, as reported in sections 2.5.1 and 2.5.2.1. The large pooling models presented in sections 2.5.3.1 to 2.5.3.3 exclude from estimation those variables (21 in total) whose time series start after 1967Q1 (as they would lead to discard observations for all other regressors), while the 14 predictors with the shortest time series (which do not have enough data to make direct forecasts for as early as 1990Q1, at least over the longest horizon) are excluded from the forecast combinations.

[18] Accordingly, the MSFE for the 1-quarter horizon is computed using actual GDP data up to 2016Q2 and the MSFE for the 4-quarter horizon is computed using data up to 2017Q1. For the 12-, 20- and 28-quarter horizons, actual observations are used up to 2017Q4.

$$MSFE_{i,h} = \frac{1}{T_1 - h - T_0 + 1} \sum_{t=T_0}^{T_1-h} \left(\widehat{u}_{i,t+h|t}\right)^2$$

where T_0 is the end date of the shortest sample and $T_1 - h$ is the end date of the longest sample.

Table 2.13 reports the MSFE of the ARDL models relative to the benchmark AR, i.e. model (2.1) without the terms associated with x_i and with the lag length p selected recursively by the BIC. Table 2.15 shows the root mean square forecast error (RMSFE) of the benchmark AR. The log CAPR ratio is by far the best predictor over long horizons ($h > 12$) and the second best predictor for both $h = 4$ and $h = 12$. The NNBLI ratio is the most effective predictor for $h \leq 12$ and among the three best predictors for horizons longer than 3 years. Stock market valuation ratios, namely the S&P 500 dividend yield and the log CAPE ratio, are particularly useful for 1-quarter-ahead forecasts. For the shortest horizon, forecast gains over the benchmark are limited: when $h = 1$, the smallest relative MSFE is 0.89 (RMSFE of 0.54%). Conversely, for longer horizons, the log CAPR ratio and the NNBLI ratio achieve MSFE values as low as 0.27 (log CAPR for $h = 28$), 0.29 (log CAPR for $h = 20$) and 0.32 (NNBLI for $h = 12$). For $h = 4$, NNBLI has a relative MSFE of 0.53. Such results are all the more remarkable if one considers that only two variables have MSFE values lower than 0.8 for $h = 4$, only 4 variables for $h = 12, 20$ and only 3 variables for $h = 28$. The absolute RMSFE also helps appreciate the forecast performance of CAPR and NNBLI. In terms of 28-quarter cumulative GDP growth, the RMSFE of CAPR is 4.15%, corresponding to an average annual error of 0.59 percentage points of GDP growth for 7 years. For $h = 20$, the RMSFE of CAPR is 3.32%, implying an average annual error of 0.66%. For $h = 12$, the RMSFE of NNBLI is 2.50%, corresponding to an average annual error of 0.83%.

Other top performers include the new orders for durable goods (*AMDMNOx*) (for h between 1 and 12), the OECD composite leading indicator (*cli*) (for $h = 4, 12, 20$) and the unadjusted price-rent ratio (*pr*) (for $h = 4, 20, 28$).

.

2.5.2 Iterated forecasts: VAR models

The second out-of-sample approach consists in using VAR models to compute multi-step-ahead iterated forecasts of GDP. As before, the resulting MSFE is used to build rankings of predictors.

2.5.2.1 Bivariate VAR

First, evaluation of the predictors is conducted using VAR models that include only real GDP growth and one predictor at a time. The i-th VAR can be written as:

$$\widetilde{y}_t^{(i)} = a_0^{(i)} + a_1^{(i)} t + \sum_{j=1}^{p} B_j^{(i)} \widetilde{y}_{t-j}^{(i)} + \varepsilon_t^{(i)} \tag{2.5}$$

where $\widetilde{y}_t^{(i)}$ is the vector containing the year-on-year growth rate of real GDP and the i-th predictor at time t, $a_0^{(i)}$ is a 2×1 vector of constants, $a_1^{(i)}$ is a 2×1 vector of trend coefficients, $B_j^{(i)}$ is a 2×2 matrix of coefficients, $\forall j = 1, \ldots, p$, and $\varepsilon_t^{(i)}$ is a 2×1 vector of error terms. The lag length p is recursively selected by the BIC for the whole system and the maximum length is again fixed at 5.

Predictors are ranked based on the performance of the VARs in terms of forecasts of the h-period-ahead GDP level:

$$MSFE_{GDP}^{(i,h)} = \frac{1}{T_1 - h - T_0 + 1} \sum_{t=T_0}^{T_1 - h} \left(\ln(GDP_{t+h}) - \ln(\widehat{GDP}_{t+h|t}^{(i)}) \right)^2$$

where $\ln(\widehat{GDP}_{t+h|t}^{(i)})$ is the forecast of the log GDP level for period $t+h$ obtained from model i by cumulating the growth rate forecasts over time.[19]

Table 2.14 reports the MSFE of the best 30 VAR models for each horizon, relative to the benchmark AR.[20] Table 2.16 shows the RMSFE of the AR. The top positions remain largely unchanged with respect to Table 2.13. Once again, the log CAPR is the best predictor for $h = 20, 28$ and ranks second for $h = 4, 12$. The NNBLI ratio is still the most useful predictor for $h = 4, 12$, whereas its relative predictive ability is weaker for $h = 1$ compared

---

[19]Let r denote the remainder of the division of h by 4 and $\widehat{y}_{t+s|t}^{(i)}$ the forecast of GDP growth produced by model i for period $t + s$, then:

$$\ln(\widehat{GDP}_{t+h|t}^{(i)}) = \begin{cases} \ln(GDP_t) + \sum_{\tau=1}^{1+h/4} \widehat{y}_{t+\tau \cdot 4|t}^{(i)} & \text{if } h \text{ is a multiple of 4} \\ \ln(GDP_{t-4+r}) + \sum_{\tau=1}^{1+(h-r)/4} \widehat{y}_{t-4+r+\tau \cdot 4|t}^{(i)} & \text{if } h \text{ is not a multiple of 4} \end{cases}$$

[20]In this case the benchmark AR follows specification (2.5).

to Table 2.13. Regarding 1-quarter-ahead forecasts, the credit/GDP ratio now exhibits the best performance, while the S&P dividend yield is still the second best variable.

Just as in Table 2.13, the forecast gains provided by CAPR and NNBLI over the benchmark AR are substantial for every $h \geq 4$. In particular, for $h = 12$ the relative MSFE of NNBLI is 0.37, corresponding to an absolute RMSFE of 2.58% (an annual average of 0.86 percentage points of GDP growth). For $h = 20$, the log CAPR gives a relative MSFE of 0.44 and an absolute RMSFE of 4.13% (annual average: 0.83%), while for $h = 28$ it gives a relative MSFE of 0.50 and a RMSFE of 6.19% (0.88%). For $h = 4$, the relative MSFE of NNBLI is 0.54, corresponding to an absolute RMSFE of 1.3%. Only 11 bivariate VARs beat the AR for $h = 1$. Once again, the lowest relative MSFE is much higher in this case: 0.91 (RMSFE of 0.64%).

2.5.2.2 Multivariate VAR

This section deals with the evaluation of the relative importance of predictors in the context of VAR models that include more than one predictor at the same time. As already mentioned, the analysis focuses on the reduced dataset. Specifically, a collection of multivariate VAR models are estimated by taking as endogenous variables the real GDP growth rate plus all possible combinations of k predictors from the set of 26 predictors in Table 2.1, with $k = 1, \ldots, 4$. Accordingly, up to 5 endogenous variables are included in the VARs and the total number of estimated models is 17,901. Given the generic combination m of predictors, the resulting VAR model is:

$$\widetilde{y}_t^{(m)} = a_0^{(m)} + a_1^{(m)} t + \sum_{j=1}^{p} B_j^{(m)} \widetilde{y}_{t-j}^{(m)} + \varepsilon_t^{(m)} \tag{2.6}$$

where $\widetilde{y}_t^{(m)} = S^{(m)} \widetilde{y}_t$ and $S^{(m)}$ is a $\left(1 + k^{(m)}\right) \times (1 + n)$ selection matrix that extracts GDP and $k^{(m)}$ additional variables from $\widetilde{y}_t$, which is the vector containing all variables in the dataset. $a_0^{(m)}$ is the $\left(1 + k^{(m)}\right) \times 1$ vector of constants, $a_1^{(m)}$ is the $\left(1 + k^{(m)}\right) \times 1$ vector of trend coefficients, $B_j^{(m)}$ is a $\left(1 + k^{(m)}\right) \times \left(1 + k^{(m)}\right)$ matrix of coefficients. The lag selection method is the same as before.

The direct outcome of the analysis is a ranking of different *subsets* of predictors for each forecast horizon, based on the MSFE. Then, to measure the relative importance of single predictors, variables are ranked based on the frequency of inclusion in the list of best performing models. Such rankings are built using the top 5%, top 1%, top 100 and top 10 models, in addition to the single best model. As in the previous section, the relevant MSFE is the MSFE for the GDP level.

Table 2.17 indicates the single best-performing VAR for each target horizon. The table shows the composition of each model and the relative MSFE over all horizons. All models outperform the best bivariate VARs from Table 2.14 over their respective target horizons, except for the model with target horizon 4 (which is only beaten by NNBLI). What is more, the models tend to improve over bivariate VARs even for horizons that are not their direct target.

Tellingly, all models except the best one for $h = 1$ include the house price-rent ratio, either in the form of *capr* or *pr*. On top of that, other housing indicators are present in all five models: the best models for $h = 1$ and $h = 28$ include residential investment, models for $h = 12$ and $h = 20$ include housing starts and the model for $h = 4$ includes the house price growth rate. In keeping with Tables 2.5 and 2.7, money growth appears useful when included in multivariate models for longer horizons. In particular, $m2$ is included in the best models for $h = 20$ and $h = 28$. Once again the stock market is especially valuable for shorter horizons, while the OECD composite leading indicator is included in the best models for $h = 1$, $h = 12$ and $h = 20$.

Figures 2.3 to 2.7 report the predictors' frequencies of inclusion in the best-performing VARs for each horizon. In each figure, the left panel shows the frequencies computed over the top 5%, top 1%, top 100 and top 10 models, as well as the composition of the single best model. The right panel displays the histogram based on the top 1% models. The log CAPR is the variable with the highest frequency in the top 1% for $h = 20, 28$ and the second highest frequency for $h = 12$. The OECD composite leading indicator has the highest frequency for $h = 12$ and the second highest frequency for $h = 1$ and for $h = 20$. As in Table 2.3 for in-sample evaluation, residential fixed investment ranks high for short horizons, but is also the second most included variable for $h = 28$. The S&P 500 index growth in real terms is quite useful over short horizons, as it has the third highest frequency in the top 1% models for both $h = 1$ and $h = 4$, but it also ranks high for $h = 28$.

2.5.3 Pooled-information methods

The approaches presented in the previous sections make comparisons between single variables or small subsets of variables to find the best predictors. How useful is it to focus on the most powerful predictors only? This section explores methods that pool all available information, or a large portion of it, and investigates whether they can achieve forecast gains over methods that focus on a few predictors at a time. Two types of information pooling are considered:

pooling at the estimation stage and pooling at the forecasting stage. In the first case, several data-intensive models are estimated using shrinkage or factor extraction: Large Bayesian VAR (LBVAR) (Bańbura et al. 2011), LASSO VAR and post-LASSO VAR models, factor models using principal components of the predictors. Each method allows to exploit all predictors simultaneously. In the second case, the forecasts produced by different predictors are combined using standard weighting schemes.

The different pooling models are intended to reflect three different approaches to the problem of high dimensionality. The LBVAR approach retains all available predictors in the forecasting model, as it applies a shrinkage method that does not restrict any coefficient to be exactly zero. The LASSO VAR (and its post-LASSO extension) performs variable selection by setting a subset of coefficients exactly to zero. Finally, the principal component approach reduces the dimension of the model by summarizing the dataset of predictors into a small number of factors.

In addition to pooling all available predictors, each method is also implemented using large specific subsets of predictors. In particular, the methods are applied to the pool of variables representing financial-cycle indicators and to the subsets of 50 predictors that are in the upper part of the rankings in section 2.5.2.2.[21] Moreover, the LBVAR, the LASSO VAR and the post-LASSO VAR are also estimated on the reduced dataset.

2.5.3.1 Large Bayesian VAR (LBVAR)

The LBVAR model has been shown by Bańbura et al. (2011) to be a valid alternative to factor-based approaches for forecasting (see also Karlsson 2013).

Let us consider a VAR model as in equation (2.6), but specified for the entire vector $\widetilde{y}_t$.[22] Let us also express the VAR as a simultaneous equations system:

$$Y = XB + E \tag{2.7}$$

[21] Financial-cycle indicators are identified as follows: (i) variables in the FRED-QD dataset that are are classified into one of the following groups: Housing (Table 2.38), Money and Credit (excluding the 5 money stock variables) (Table 2.43), Household Balance Sheets (Table 2.44), Stock Market (Table 2.47), Non-Household Balance Sheets (Table 2.48); (ii) additional variables in FRED-QD: residential investment and residential utilities, mortgage interest rates and rents; (iii) in the reduced dataset, variables from *sp500* to the bottom in Table 2.1.

[22] Actually, 251 out of 272 variables are used, due to data availability reasons explained before (see note 17).

where $Y = (\widetilde{y}_1, \ldots, \widetilde{y}_T)'$ is the $T \times N$ stacked vector of endogenous variables and $N = n+1$ is the total number of variables in the full dataset. $X = (\widetilde{x}_1, \ldots, \widetilde{x}_T)'$ is a $T \times (Np+2)$ matrix where $\widetilde{x}_t = \left(\widetilde{y}'_{t-1}, \ldots, \widetilde{y}'_{t-p}, 1, t\right)'$, $B = (B_1, B_2, \ldots, B_p, a_0, a_1)'$ is the $(Np+2) \times N$ matrix of coefficients and $E = (\varepsilon_1, \ldots, \varepsilon_T)'$ is the $T \times N$ stacked vector of error terms. Following Bańbura et al. (2011), the normal inverted Wishart prior is considered:

$$vec(B)|\Psi \sim N(vec(B_0), \Psi \otimes \Omega_0)$$
$$\Psi \sim IW(S_0, \alpha_0)$$

where Ψ denotes the covariance matrix of the errors and the hyperparameters B_0, Ω_0, S_0 and α_0 are chosen so to match the Minnesota prior. In particular, Bańbura et al. (2011) show that the normal inverted Wishart prior can be implemented by adding dummy observations to system (2.7). Additional dummies can be included to impose a further prior on the sum of coefficients of the lagged terms.

In more detail, let Y_{d1} and X_{d1} denote the dummies that implement the Minnesota prior. These are given by:

$$Y_{d1} = \begin{pmatrix} \text{diag}\left(\delta_1\sigma_1, \ldots, \delta_N\sigma_N\right)/\lambda \\ 0_{N(p-1)\times N} \\ \cdots\cdots\cdots\cdots\cdots \\ \text{diag}\left(\sigma_1, \ldots, \sigma_N\right) \\ \cdots\cdots\cdots\cdots\cdots \\ 0_{2\times N} \end{pmatrix}, \qquad X_{d1} = \begin{pmatrix} J_p \otimes \text{diag}\left(\sigma_1, \ldots, \sigma_N\right)/\lambda & 0_{Np\times 2} \\ \cdots\cdots\cdots\cdots\cdots & \\ 0_{N\times Np} & 0_{N\times 2} \\ \cdots\cdots\cdots\cdots\cdots & \\ 0_{2\times Np} & \text{diag}\left(e_0, e_1\right) \end{pmatrix}$$

where $J_p = \text{diag}\left(1, 2, \ldots, p\right)$, $\delta_i = 1$ is the prior mean for the coefficient on the first lag of variable i, λ is an inverse measure of the tightness of the prior on the VAR coefficients, $\sigma_1, \ldots, \sigma_N$ define the prior on the covariance matrix of the errors, while e_0 and e_1 determine the tightness of the priors on the constant and the time trend, respectively (this is an adaptation of the setup considered by Bańbura et al. 2011, where there is no time trend). The additional prior on the sum of coefficients is introduced by adding the following dummies Y_{d2} and X_{d2}:

$$Y_{d2} = \frac{\text{diag}\left(\delta_1\mu_1, \ldots, \delta_N\mu_N\right)}{\tau} \qquad X_{d2} = \Big(\left(1_{1\times p}\right) \otimes \text{diag}\left(\delta_1\mu_1, \ldots, \delta_N\mu_N\right)/\tau \quad 0_{N\times 2}\Big)$$

where μ_i is the mean of variable i and τ determines the tightness of the prior.

Following Alessandri and Mumtaz (2017), the prior mean δ_i is set to the OLS estimate of an AR(1) regression for variable i and σ_i is set to the standard error from the same regression. A flat prior is imposed on both the constant and the time trend by setting $e_0 = e_1 = 10000$. μ_i is set to the mean of variable i over the estimation sample. The model is estimated using a grid of possible values for the shrinkage parameters λ and τ. In particular, the values considered for λ lie in the range [0.0001, 0.1], while τ can assume values $\tau = 10\lambda$ and $\tau = 100\lambda$. The value $\tau = 100\lambda$ generally provides better forecast performance and is used to report the results.

Eventually, adding a total of T_d dummy observations Y_d and X_d, where $Y_d = (Y'_{d1}, Y'_{d2})'$ and $X_d = (X'_{d1}, X'_{d2})'$, is equivalent to imposing $B_0 = (X'_d X_d)^{-1} X'_d Y_d$, $\Omega_0 = (X'_d X_d)^{-1}$, $S_0 = (Y_d - X_d B_0)'(Y_d - X_d B_0)$ and $\alpha_0 = T_d - \kappa$, where $\kappa = Np + 2$ is the number of coefficients in any single equation of system (2.7). The dummy-augmented system can be written as:

$$\underset{T_* \times N}{Y_*} = \underset{T_* \times \kappa}{X_*} \underset{\kappa \times N}{B} + \underset{T_* \times N}{E_*}$$

where $Y_* = (Y', Y'_d)'$, $X_* = (X', X'_d)'$, $E_* = (E', E'_d)'$ and $T_* = T + T_d$. The posterior means of the parameters coincide with the OLS estimates of the regression of Y_* on X_*, i.e.:

$$\widehat{B} = \left(X'_* X_*\right)^{-1} X'_* Y_*$$

which also correspond to the modes, as the posterior distribution for B is normal. Therefore, $\widehat{B}$ is used to compute the LBVAR point forecasts to be evaluated.

The LBVAR considered so far includes all available predictors. To evaluate the *potential* gains from information pooling, it is also worth examining how the model performs when a large subset of predictors is selected using an ex-post perspective. To this aim, five additional LBVAR models are estimated and reported, each of which includes real GDP growth plus the best 50 predictors for each horizon, selected using the VAR-based rankings of section 2.5.2.1.[23] In addition, the model is estimated using financial-cycle predictors only.

Table 2.18 shows the relative MSFE for the LBVAR using all available (251) predictors,

[23] In this case, the value of λ is selected recursively across windows by maximizing the marginal likelihood over the interval $\lambda \in [0.001, 0.1]$. Again, two possibilities are considered for τ: $\tau = 10\lambda$ and $\tau = 100\lambda$, where the ratio of τ to λ is fixed across windows. The value $\tau = 100\lambda$ still provides better forecast performance.

the 5 alternative subsets of 50 predictors and the pool of financial-cycle indicators. In general, no model outperforms the best bivariate VARs from Table 2.14. The full LBVAR performs poorly regardless of the values used for the shrinkage parameters. Conversely, the model using the best predictors for $h = 20$ exhibits a good performance: it beats all other LBVARs for every forecast horizon and would rank second in Table 2.14 for $h = 4$ (its MSFE is 0.75), 5th for $h = 12$ (0.75), 9th for $h = 20$ (0.82) and 6th for $h = 28$ (0.70).[24] The LBVAR using financial-cycle indicators performs better than the largest LBVAR for every $h \geq 12$. Over the same horizons, it also outperforms the models using the best 50 predictors for target horizons 1,4 and 28 quarters. Conversely, it shows quite a poor performance over short horizons.

Finally, the LBVAR estimated on the reduced dataset performs poorly and is not reported in order to save space.

2.5.3.2 LASSO VAR and post-LASSO VAR

The second pooling method consists in estimating a high-dimensional VAR model using the LASSO penalty (LASSO VAR). To illustrate the LASSO VAR, let us consider an additional representation of the VAR model, which is obtained by stacking the equations in system (2.7), i.e. by rewriting (2.7) as a system of seemingly unrelated regressions (SUR):

$$\widetilde{y} = (I_N \otimes X)\, \widetilde{\beta} + \varepsilon \equiv \widetilde{X}\widetilde{\beta} + \varepsilon \tag{2.8}$$

where $\widetilde{y} = \text{vec}(Y)$, $\widetilde{\beta} = \text{vec}(B)$ is a $N\kappa \times 1$ matrix and I_N is the identity matrix of size N. Also, $\varepsilon \sim \mathcal{N}(0, \Psi \otimes I_T)$, where I_T is the identity matrix of size T. This representation allows for different equations in the VAR to have different explanatory variables. The LASSO estimator of model (2.8) solves:

$$\min_{\widetilde{\beta}} \quad \|\widetilde{y} - \widetilde{X}\widetilde{\beta}\|_2^2 + \lambda \|S\widetilde{\beta}\|_1 \tag{2.9}$$

where S is a $N(\kappa - 1) \times N\kappa$ selection matrix that excludes the constants from the penalty term, so that they are included with certainty in the final model.

Tables 2.19-2.22 report the results of a LASSO VAR(1) estimated on the large dataset

[24] Out of the best 50 predictors, the number of financial-cycle indicators is 18 for $h = 1$, 21 for $h = 12$ and $h = 20$, 22 for $h = 4$ and 23 for $h = 28$.

of 252 variables and a LASSO VAR(5) for the reduced dataset.[25] Tables 2.19 and 2.21 show the relative MSFE for several values of λ. Both LASSO VAR models achieve relatively high forecast accuracy over long horizons for some values of the penalty term, while they always perform poorly over short horizons. Even for long horizons, however, the models fail to outperform the best bivariate VARs. The LASSO VAR has also been estimated using subsets of highly performing predictors as in section 2.5.3.1 and the results do not change much. Tables 2.20 and 2.22 display square matrices summarizing the degree of sparsity in the estimated LASSO VARs. In particular, each row corresponds to an equation in the system and each column corresponds to a predictor on the right-hand side of the equations. Blue cells identify predictors that are assigned non-zero coefficients in the last estimation window, which ends in 2016Q1 (using the value of λ that gives the best forecasts for $h \geq 4$).[26]

Table 2.24 shows the MSFE of the LASSO VAR estimated on GDP and financial-cycle indicators only. For the purpose of comparison with the other LASSO VARs, both the lag length of 1 and the lag length of 5 are considered. For $\lambda = 0.0025$, the model with lag length 5 outperforms the reduced LASSO VAR over all horizons except $h = 1$, while the model with lag length 1 produces almost the same results as the largest LASSO VAR (when the latter has $\lambda = 0.0005$).

Table 2.23 shows the relative MSFE of the post-LASSO VAR estimated on the reduced dataset, for different values of the penalty parameter. In this case, the LASSO penalty is only used to select the non-zero coefficients, then the system is estimated by SUR.[27] The post-LASSO tends to improve over the LASSO in terms of 1-quarter forecasts, but does not generally yield forecast gains for other horizons.

[25]The lag length of 1 is used for the largest LASSO VAR to prevent computational problems.

[26]Note that in the LASSO VAR(5) for the reduced dataset a given predictor may have zero coefficients on specific lags and non-zero coefficients on others, so that a colored cell indicates that at least one lag of the variable in column is included in the shrunk equation for the variable in row.

[27]When the full dataset is considered, the number of non-zero coefficients makes estimation by SUR infeasible, so the system is estimated using OLS equation by equation. The model becomes explosive for many values of λ and is not reported. Note however that for $\lambda = 0.001$ the model is stable and performs slightly worse than the LASSO VAR with the same penalty parameter.

2.5.3.3 Factor model: VAR with principal components

Next, I consider a factor model, in the form of a VAR process for GDP growth and a set of factors extracted from all predictors:

$$\begin{pmatrix} y_t \\ \widehat{F}_t \end{pmatrix} = \begin{pmatrix} a_{0,y} \\ a_{0,\widehat{F}} \end{pmatrix} + \begin{pmatrix} a_{1,y} \\ a_{1,\widehat{F}} \end{pmatrix} t + \sum_{j=1}^{p} B_j \begin{pmatrix} y_{t-j} \\ \widehat{F}_{t-j} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \varepsilon_{\widehat{F},t} \end{pmatrix}$$

where $\widehat{F}_t$ denotes the $v \times 1$ vector of the first v estimated principal components of the n predictors in the full dataset, $a_{0,\widehat{F}}$ and $a_{1,\widehat{F}}$ are $v \times 1$ vectors of coefficients and B_j is a $(1+v) \times (1+v)$ matrix of coefficients, $\forall j = 1, \ldots, p$. Table 2.25 presents results for $v = 1, 2, 3, 4$. No model beats all the one-predictor VARs from Table 2.14. The model with 2 principal components delivers good forecasts for the 1-quarter horizon, being outperformed by only one bivariate VAR. The same holds true for the model with 3 principal components when $h = 4$.

Interestingly, restricting attention to financial-cycle indicators lead to substantial forecast gains for longer horizons. Table 2.26 presents the results of VAR models in which the principal components are extracted from financial-cycle variables only. For $h \geq 12$, the models using 2 and 3 principal components outperform all models in Table 2.25 and would rank in the top 10 positions in Table 2.14.

2.5.3.4 Forecast combinations

While the approaches in sections 2.5.3.1-2.5.3.3 pool information prior to forecasting, an alternative pooling strategy is given by forecast combination, which finds widespread application in the forecasting literature (Elliott and Timmermann 2016) on the grounds that individual models are likely to be misspecified and that combining forecasts from different models should increase efficiency compared to individual forecasts. Measuring the performance of combined forecasts helps give a sense of how useful it is in practice to establish rankings of predictors. If combined forecasts turn out to outperform forecasts from every individual model, then the information contained in poorly-ranking predictors should not be discarded.

This section presents results on forecast combinations for both direct and iterated forecasts from one-predictor models, using two standard combination schemes. Let w_i denote the weight assigned to model i and M the number of models to be combined. The simplest approach consists in using an equal-weighted average of the forecasts:

$$w_i^{equal} = 1/M$$

There is ample empirical evidence that equal weighting performs well for point forecasts (e.g., Stock and Watson 2003) and often outperforms more sophisticated weighting strategies, giving rise to a "forecast combination puzzle" (Elliott and Timmermann 2016, Smith and Wallis 2009). The second approach is Bayesian model averaging (BMA). In particular, since the value of the BIC for model i provides an asymptotic approximation to its marginal likelihood, the BMA weights are approximated by:

$$w_i^{bma} = \frac{\exp(-0.5BIC_i)}{\sum_{i=1}^{M} \exp(-0.5BIC_i)}$$

The weights are computed recursively across estimation windows.

In analogy with the previous sections, results are presented for three cases: (i) all possible predictors are taken into account;[28] (ii) combinations are limited to an ex-post selection of predictors: for $M = 10, 100, 200$, the models considered are the best-performing ones from sections 2.5.1 and 2.5.2.1; (iii) combinations are computed over financial-cycle predictors only.

The results of the first two cases are presented in Tables 2.27 and 2.28 for direct and iterated forecasts, respectively. Let us first focus on the combinations comprising all available models. The resulting forecasts perform systematically worse than the best individual models. In the case of direct (ARDL) forecasts, they perform poorly except for the 1-quarter horizon. In the case of iterated (VAR) forecasts, they would enter the top 30 list in Table 2.14 for every horizon, the exact position ranging from 5th (for $h = 1$) to 24th (for $h = 20$). In general, the choice between equal and BMA weights has little impact on the precision of the forecasts. Averaging forecasts over ex-post selected subsets of predictors lowers the MSFE. Combining the 10 best models improves over every individual models for $h = 1$, both for ARDL and VAR models, and for $h = 28$ in the case of VAR models, whereas for all other horizons the combined forecasts fail to beat the best single-predictor models.

Finally, Tables 2.29 and 2.30 show the results obtained by averaging forecasts over

[28] As explained before (see note 17), the 14 predictors with the shortest time series are excluded from the combinations. Accordingly, the total number of models to be combined is 258. In the case of VAR models, two additional models producing explosive forecasts are excluded, so that the total number of models is 256.

financial-cycle predictors exclusively. The results for iterated forecasts are noteworthy. The combined forecasts perform better than the top-100 averages from Table 2.30 for $h = 12, 20, 28$. Moreover, they outperform all individual models in Table 2.14 for $h = 1$, while they would rank second for $h = 4$ and within the top 10 for all other horizons. Direct forecasts perform quite well for short horizons: they have a lower MSFE than the 100-predictor average forecasts for $h = 1$ and approximately the same MSFE for $h = 4$. Conversely, they fare worse than the 200-predictor combination for $h = 12$, while they rank between the 100- and the 200-predictor combinations for $h = 20, 28$. Like iterated forecasts, they beat the individual models for $h = 1$.

2.5.4 Tests of forecast accuracy and encompassing

In sections 2.5.1 and 2.5.2.1, forecasts based on the best predictors perform much better than forecasts from the benchmark AR models. To determine if the predictive content of the variables is statistically significant, I rely on the ENC-F test by Clark and McCracken (2001, 2013), which is suitable for comparing nested models and has been shown to have comparatively high power among forecast encompassing and accuracy tests (Busetti and Marcucci 2013). Under the null hypothesis, the candidate predictor has no marginal predictive power, i.e. the benchmark AR encompasses the candidate model, while under the alternative the variable has predictive power. The models in sections 2.5.1 and 2.5.2.1 cannot be used to conduct the test because the number of lags changes over time. Therefore, I compute forecasts from all models using a fixed number of lags equal to 5. Let $\widehat{c}_{t+h|t} = \widehat{\varepsilon}^{(m_1)}_{GDP,t+h|t} \left(\widehat{\varepsilon}^{(m_1)}_{GDP,t+h|t} - \widehat{\varepsilon}^{(m_2)}_{GDP,t+h|t} \right)$, where $\widehat{\varepsilon}_{GDP,t+h|t}$ denotes a forecast error, while m_1 and m_2 identify the benchmark AR model and the alternative model, respectively. The ENC-F test statistic is:

$$\text{ENC-}F = \left(\sum_{t=T_0}^{T_1-h} \widehat{c}_{t+h|t} \right) / MSFE^{(m_2)}_{GDP}$$

In addition, I use the MSE-F statistic (Clark and McCracken 2001) to test if a bivariate model and the benchmark AR have equal predictive accuracy in terms of MSFE, against the alternative hypothesis that the candidate model beats the AR. Let $\widehat{d}_{t+h} = \left(\widehat{\varepsilon}^{(m_1)}_{GDP,t+h} \right)^2 -$

$\left(\widehat{\varepsilon}^{(m_2)}_{GDP,t+h}\right)^2$. The MSE-$F$ test statistics is given by:

$$\text{MSE-}F = \left(\sum_{t=T_0}^{T_1-h} \widehat{d}_{t+h}\right) / MSFE^{(m_2)}_{GDP}$$

The critical values for the ENC-F and MSE-F tests are computed using the restricted VAR bootstrap explained in Clark and McCracken (2013).

Table 2.31 reports the results of the tests, focusing on the case of iterated forecasts provided by the log CAPR, the NNBLI ratio and the credit/GDP ratio, which are best performing predictors over the different horizons in Table 2.14. All three variables have significant predictive content based on the ENC-F test. In particular, the predictive content of CAPR and NNBLI is significant at the 1% level for all horizons. CAPR and NNBLI provide a significantly lower MSFE than the benchmark AR for every $h \geq 4$, while the difference in forecast accuracy is not significant for $h = 1$. Conversely, the credit/GDP ratio only provides a significant improvement in MSFE for $h = 1$.

Finally, I also report the results of the Diebold–Mariano (DM) test of forecast accuracy, making comparisons on a pairwise basis between the best performing direct (iterated) forecasts for any given horizon and the other 29 forecasts from Table 2.13 (Table 2.14). The results must be taken with great caution for several reasons relating to the validity of the DM assumptions. In particular, the use of a recursive-window scheme, the time-varying specification of the models due to changing lag length, and the high persistence of the forecast variable for long horizons (multi-period cumulative GDP growth) may affect the reliability of the results (see Elliott and Timmermann 2016 and Rossi 2005). The null hypothesis in this case is that the two models provide equal MSFE over the period of interest. I consider the left-tail alternative hypothesis that the best performing model has a significantly lower MSFE. Table 2.32 and 2.33 show the results. The lowest average p-values are achieved by the best performing models for $h = 4$ and $h = 12$, in which the NNBLI ratio is used as predictor. In this case, most p-values fall below 10%. Higher p-values are generally obtained for longer horizons, although not exceeding 20% in most cases (as mentioned above, however, a longer horizon may imply more severe problems for the DM test, due to the persistence of the forecast variable).[29] Finally, the highest average p-values are obtained for the 1-quarter

[29] As a hint of possible problems with the DM test, the average p-value across the 25 worst predictors for $h = 28$ (among *all* predictors in the dataset) is 0.13 for iterated forecasts and 0.12 for direct forecasts, and the minimum p-value is 0.07 in both cases, even though a number of predictors have a relative MSFE that

forecasts, both in the case of ARDL and in the case of VAR models, meaning that the forecast improvement provided by the best performing model appears less significant over a very short horizon.

2.5.5 Comparison with IMF forecasts

To assess the practical relevance of the results for forecasters, this section reports the forecast performance of a major forecasting institution, the IMF, and compares it with the models considered in this paper.[30] Table 2.34 summarizes the performance of the IMF forecasts over the period 1990-2017 using the RMSFE and the relative MSFE with respect to both the direct and the iterated forecasts by the benchmark AR models. In particular, since the IMF World Economic Outlook (WEO) provides forecasts in the form of annual growth rates up to 5 years ahead, I adopt the following approach to construct 4-quarter-ahead, 12-quarter-ahead and 20-quarter-ahead forecasts of the GDP level: (i) I focus on the Spring issues of the WEO; (ii) I use the last quarter of the year prior to each issue as the starting point for forecasting, i.e. the latest available observation of GDP; (iii) I apply the annual forecast growth rates to the starting value to compute forecasts of the GDP level and (iv) I assign each resulting value to the last quarter of the relevant forecast year. For example, the annual forecast growth rate for 1990 published in the 1990 Spring issue is used to compute the 4-quarter-ahead forecast for 1990Q4 based on data up to 1989Q4, the annual rates up to 1992 are used to compute the 12-quarter-ahead forecast for 1992Q4 based on data up to 1989Q4, and so on.

As can be seen by comparison of Table 2.34 with Tables 2.13, 2.14 and 2.17, the IMF forecasts do not outperform the best small multivariate models nor the best one-predictor models for any forecast horizon. In more detail, the relative performance of the IMF is quite high for $h = 4$ and deteriorates as the forecast horizon increases. For $h = 4$, the IMF forecasts are only outperformed by the bivariate models using the NNBLI ratio and by the

---

exceeds 2.

[30]The IMF forecasts are particularly suitable for comparisons in this context, as they cover horizons from 1 to 5 years. In the case of other major forecasters, comparisons would necessarily have more limitations in terms of horizons. For instance, the Fed Greenbook forecasts cover a maximum horizon of 2 years. The OECD publishes annual forecasts for the following year and long-term projections (currently, the horizon is 2060) in 10-year steps. The Survey of Professional Forecasters (SPF) includes quarterly forecasts up to 4 quarters ahead, annual forecasts for the next three years but only starting from the 2009Q2 survey, and 10-year annual average forecasts. However, as OECD (2014) points out, “the profile and magnitude of the errors in the GDP growth projections [over 2007-2012] of other international organizations and consensus forecasts are strikingly similar”.

best multivariate VAR model with target horizon equal to 4. For $h = 12$, they are beaten by bivariate models using the NNBLI ratio, the log CAPR ratio, the OECD CLI, the unfilled orders for nondurable goods (*AMDMUOx*) plus a few more bivariate VAR models (most notably, those using the unadjusted price/rent ratio and residential fixed investment). They are also outperformed by all VAR models in Table 2.17 except the one with target $h = 1$. Finally, for $h = 20$ the IMF forecasts fall short of the benchmark AR models and, a fortiori, perform much worse than the best performing models in this paper. In particular, the relative MSFE with respect to the direct AR forecasts is 1.082, far above the 0.2859 achieved by *capr* and higher than all top 30 ARDL models in Table 2.13. The relative MSFE with respect to the iterated AR forecasts is 1.07, once again much higher than the relative MSFE provided by *capr* (0.4381) and by the best VAR models with target $h \geq 12$ in Table 2.17, and worse than all top 30 VARs from Table 2.14.

Figure 2.8 contrasts different forecasts of U.S. GDP made on the verge of the crisis. The green line represents the 20-quarter-ahead iterated forecast from the bivariate VAR using the log CAPR ratio, estimated on data up to 2007Q2. The dashed orange line identifies the iterated forecast from the bivariate VAR using the NNBLI ratio. The dashed red line is the forecast from the Fall 2007 issue of the IMF WEO. The blue line is the realized level of real GDP. While the IMF forecast more or less follows the pre-crisis trend, both the CAPR and the NNBLI correctly anticipate the magnitude of the slump.

2.5.6 Predictive importance over time

According to Stock and Watson (2003), the instability of predictive relationships is the norm. Considering the rankings in Tables 2.13-2.14, a variable may rank high overall as a result of producing accurate forecasts only in some specific time intervals, while being a poor predictor in other periods. In this section, the forecast evaluation sample 1990Q1-2017Q4 is partitioned into three sub-samples to evaluate the stability of the rankings over time. The three sub-samples are given by the pre-crisis period 1990Q1-2007Q2, the crisis period 2007Q3-2009Q4 and the post-crisis period 2010Q1-2017Q4. The relative MSFE of direct and indirect forecasts from bivariate models are computed over each sub-sample.

For $h = 1$ and $h = 4$, no variable among the best-performing ones from Tables 2.13-2.14 is included in top 30 list for every sub-sample, except for the log housing starts (*houst*) in the case of direct forecasts with $h = 4$. For $h = 12$, only the log CAPR and the NNBLI ratio feature in the top 30 list in every sub-sample for both direct and iterated forecasts. The Fed funds rate (both in levels and in differences) is among the best 30 for all three periods only in the case of direct forecasts, while the OECD CLI is always among the top performers only for iterated forecasts. For $h = 20$, only the log CAPR and the unadjusted log price-rent ratio are among the top 30 in all periods for both types of forecasts. The NNBLI ratio and the growth rate of housing starts (*HOUST*) are always among the best predictors for direct forecasts, while residential fixed investment as a share of GDP (*prfi_gdp*) is always included in the top 30 for iterated forecasts. Finally, only the log CAPR is invariably included in the top 30 for $h = 28$. The NNBLI ratio and 10-year BAA corporate spread are always included for direct forecasts, while the growth rate of real estate assets of households and nonprofit organizations (*HNOREMQ027Sx*) and *prfi_gdp* are always included in the case of iterated forecasts.

The instability of predictive importance is widespread across predictors. For each list in Tables 2.13-2.14, the majority of predictors (on average, approximately 21 predictors out of 30, with peaks of 26-28 for short horizons) feature among the best 30 in no more than one period out of three.

Against this backdrop, the predictive importance of the log CAPR and the NNBLI ratio appears remarkably stable for horizons of 3-5 years. The CAPR is always among the best 4 predictors for $h = 20$, regardless of the period and the type of forecast. In particular, over the period 1990Q1-2007Q2 it ranks 1st in terms of iterated forecasts (relative MSFE of 0.35) and 2nd in terms of direct forecasts (relative MSFE of 0.60). During the crisis period, it is 1st for direct forecasts (0.18) and 3rd for iterated forecasts (0.35). For $h = 12$, it has a comparatively good performance throughout all periods (between the 1st and the 13th position in the rankings) and is especially accurate during the crisis period: it is 4th in terms of direct forecasts (0.27) and 1st for iterated forecasts (0.20). For $h = 28$, it improves during and after crisis, climbing from the 11th and 30th positions in 1990Q1-2007Q2, for direct and iterated forecasts respectively, to the 1st (direct) and 3rd (iterated) positions in 2010Q1-2017Q4. Its relative performance is more unstable for $h = 4$: it only ranks high during the crisis period (1st for direct and 3rd for iterated forecasts), which is enough to be 2nd in the overall rankings of Tables 2.13-2.14. The NNBLI is the best predictor for $h = 12$ both before and after the crisis, and fares only slightly worse during the crisis (5th and 3rd for direct and iterated forecasts, respectively). For $h = 4$, it is the best predictor during the period 1990Q1-2007Q2 for both types of forecasts, and in the top 4 during the crisis period, while its performance considerably deteriorates after the crisis. For $h = 20$, it ranks either 1st or 2nd both before and after the crisis, while it appears less useful during the crisis. Similar results are obtained for $h = 28$, although not as good as for $h = 20$. Overall, it appears that the predictive importance of CAPR increases during the crisis, while that of NNBLI decreases.

In addition to the log CAPR, other financial-cycle indicators dominate the rankings of MSFE during the crisis period. House price growth indicators (*USSTHPI*, *SPCS10RSA*, *hpi*), the credit-GDP ratio (*cred_gdp*) and the mortgage-income ratio (*mortg_inc*) provide good long-horizon forecasts for the period 2007Q3-2009Q4. Interestingly, during the crisis, household balance sheets also provide the best predictors for 1-quarter-ahead forecasts: the growth rates of real total assets (*TABSHNOx*) and real net worth (*TNWBSHNOx*) of households and nonprofit organizations are the best predictors in terms of both direct and iterated forecasts (with a relative MSFE ranging between 0.34 and 0.5), and the ratio of net worth to disposable income (*NWPIx*) ranks 3rd-4th.

Regarding short horizons, the high position of the S&P dividend yield in Tables 2.13 and 2.14 is largely driven by its good performance in the period 1990Q1-2007Q2, as the variable performs poorly after the crisis. The CAPE ratio is good for short-horizon direct forecasts before and during the crisis, but not after. Still, in the overall ranking of Table 2.13 it is third for $h = 1$ and 16th for $h = 4$.

2.5.7 Further remarks

To get a further indication on the robustness and stability of predictive relationships, it is worth assessing whether in-sample and out-of-sample evaluations offer consistent results. When comparing the different pieces of evidence presented in the paper, only a few results appear robust throughout. In particular, the predictive usefulness of the log CAPR ratio, the NNBLI ratio and, to a lesser extent, of the OECD CLI is a recurrent outcome of the analysis. In fact, it is indicated almost unanimously by the different evaluation methods.

A further question to be addressed is how much the results depend on the specific data transformations considered in the paper. In particular, the transformations recommended by McCracken and Ng (2015) for the FRED-QD dataset, which are applied here, entail heterogeneous treatment of the ratios included in the FRED-QD, most notably the financial-cycle ratios. For instance, while the business sector liabilities-to-income ratios are taken in levels, the ratios of business sector net worth to income are transformed to differences.[31]

[31] In some cases, the additional series from Table 2.1 provide complementary transformations to the FRED-QD series, as in the case of the public debt-to-GDP ratio. In other cases, they provide approximately complementary transformations. For instance, while the household liabilities-to-income ratio is transformed to growth rates in the FRED-QD, the reduced dataset includes the household mortgage debt-to-income ratio (which co-varies closely with the total liabilities-to-income ratio) in levels. Another example is given by the inclusion of the CAPE in levels in the reduced dataset, which complements to some extent the S&P dividend yield in differences included in the FRED-QD dataset.

As a robustness check, I consider the case in which all ratios and rates are considered in levels. The main results regarding the leading role of CAPR and NNBLI hold true in this case. In the in-sample analysis, the top positions in the R^2 ranking remain unchanged. As for the out-of-sample analysis, the top positions in the ranking of direct forecasts are also unaffected. In the case of iterated forecasts, two ratios, namely the ratios of net worth to disposable business income for the nonfinancial noncorporate and corporate business sectors (FRED-QD labels: *TNWBSNNBBDIx* and *TNWMVBSNNCBBDIx*, respectively), provide comparable results to those produced by CAPR and NNBLI. In particular, they rank in the first two positions for $h = 20$ and they are third and second respectively for $h = 12$ (between NNBLI and CAPR). Also, the corporate ratio ranks second for $h = 4$ (between NNBLI and CAPR) and for $h = 28$ (between CAPR and PR). However, unlike the CAPR and the NNBLI, the two net worth-income ratios perform poorly in direct forecasts (158th and 133th respectively for $h = 12$, 93th and 96th for $h = 28$, 100th and 129th for $h = 20$), therefore their forecasting power does not appear to be robust.

2.6 Conclusions

A new emphasis on the role of macro-financial linkages has permeated macroeconomics in recent years. Financial cycles, building up in the background of business cycles and interacting with them through a variety of mechanisms, can have profound and disruptive effects on economic activity. This paper has presented an in-depth, financial-cycle-oriented exploration of a broad range of potential predictors of real GDP in the United States, using data from 1960 to 2017. In-sample explanatory power and out-of-sample forecast ability have been measured through several criteria and compared with each other to assess the robustness of the results. Throughout, the analysis has made extensive use of techniques for high-dimensional modelling and variable selection.

The paper offers new stylized facts on financial cycles and their role in predictions of real activity, highlighting the leading role of specific housing and balance-sheet ratios, especially over long horizons. First, a log cyclically-adjusted house price-rent ratio (CAPR) has emerged as the most powerful predictor over horizons longer than 12 quarters and one of the best two predictors for horizons of 4-12 quarters. The ratio of liabilities to income in the nonfinancial noncorporate business sector (NNBLI) has been the second key predictor for horizons ranging from 4 to 28 quarters. Compared to the other candidate predictors, the performance of the log CAPR and the NNBLI has proved remarkably robust to different evaluation criteria and stable across time periods. Among the other variables, a composite leading indicator (the OECD CLI) has provided high-ranking results for some horizons across all evaluation methods.

The empirical evidence is suggestive that financial-cycle variables as a whole contain valuable predictive content. Variables such as the credit-to-GDP ratio, the household mortgage debt-to-income and net worth-to-income ratios, residential fixed investment or stock market ratios show high predictive ability in several occasions. Also, the small VAR models that provide the most accurate forecasts in the paper invariably include financial-cycle indicators. Furthermore, when information-pooling methods are used to produce forecasts, restricting attention to financial-cycle predictors appears in general a better strategy than pooling all available information, at least for horizons longer than one year. Overall, housing, credit and balance-sheet indicators appear more valuable than stock market indicators for explaining GDP fluctuations over multi-year horizons, especially during and after the crisis.

While valuable predictive content is scattered across variables, the choice of specific financial-cycle information to be used for predictions appears critical. Except for the shortest horizons, no large pooling method (nor the forecasts by a major institution such as the IMF) is able to outperform small models that include selected financial-cycle variables or even the simplest models using only the log CAPR or the NNBLI ratio as predictors.

Bibliography

[1] Alessandri, P. and H. Mumtaz (2017), "Financial conditions and density forecasts for US output and inflation." *Review of Economic Dynamics*, 24, 66–78.

[2] Bańbura, M., D. Giannone, and L. Reichlin (2010). "Large Bayesian vector auto regressions." *Journal of Applied Econometrics*, 25 (1), 71–92.

[3] Banerjee, A., M. Marcellino and T. Masten (2005). "Leading indicators for euro-area inflation and GDP growth." *Oxford Bulletin of Economics and Statistics*, 67(S1), 785–813.

[4] Bluedorn, J. C., J. Decressin and M. E. Terrones (2016), "Do Asset Price Drops Foreshadow Recessions?" *International Journal of Forecasting*, 32 (2), 518–26.

[5] Borio C. (2012). "The financial cycle and macroeconomics: what have we learnt?" BIS working paper 395.

[6] Brunnermeir, M. and M. Oehmke, (2013), "Bubbles, financial crises, and systemic risk." In *Handbook of the Economics of Finance*, vol. 2. M. H. Constantinides and R. M. Stulz (eds.), pp. 1221–1288, North Holland Elsevier, Oxford,

[7] Buiter, W. H. (2010), "Housing Wealth Isn't Wealth." *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 4, Article 2010–22.

[8] Busetti, F., and J. Marcucci (2013), "Comparing forecast accuracy: A Monte Carlo investigation." *International Journal of Forecasting*, 29, 13–27.

[9] Campbell, J. Y. and R. Shiller (1998), "Valuation ratios and the long-run stock market outlook.", *The Journal of Portfolio Management*, 24 (2), 11-26.

[10] Campbell, J. Y. (1999) "Asset Prices, Consumption and the Business Cycle." In *Handbook of Macroeconomics*, vol. 1., J. B. Taylor and M. Woodford (eds.), pp. 1231–303, Amsterdam, Elsevier.

[11] Claessens, S., M. A. Kose and M. E. Terrones (2011). "Financial Cycles: What? How? When?" IMF Working Paper 76.

[12] Claessens, S. and M. A. Kose (2017). "Asset prices and macroeconomic outcomes: a survey." BIS Working Paper No 676.

[13] Clark, T. E., and M. W. McCracken (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics*, 105, 85-110.

[14] Clark, T. E., and M. W. McCracken (2013), "Evaluating the Accuracy of Forecasts from Vector Autoregressions." Federal Reserve Bank of St. Louis Working Paper 2013-010A.

[15] Contessi S. and U. Kerdnunvong (2015), "Asset Bubbles. Detecting and Measuring Them Are Not Easy Tasks." The Regional Economist, Federal Reserve Bank of St. Louis.

[16] Davis, M. A., A. Lehnert and R. F. Martin (2008), "The Rent-Price Ratio for the Aggregate Stock of Owner-Occupied Housing." *Review of Income and Wealth*, 54 (2), 279-284.

[17] Drehmann, M., C. Borio, and K. Tsatsaronis (2012), "Characterising the Financial Cycle: Don't Lose Sight of the Medium Term!" BIS Working Paper 380.

[18] Elliott, G. and Timmermann, A. (2016), *Economic Forecasting*. Princeton, Princeton University Press.

[19] Gertler M. and S. Gilchrist (2018), "What Happened: Financial Factors in the Great Recession." *The Journal of Economic Perspectives*, 32 (3), 3-30.

[20] Hastie, T., R. Tibshirani amd R. J. Tibshirani (2017), "Extended Comparisons of Best Subset Selection, Forward Stepwise Selection, and the Lasso.", mimeo.

[21] International Monetary Fund (2000), *World Economic Outlook: Asset Prices and the Business Cycle* (May). Washington, IMF.

[22] International Monetary Fund (2003), "When Bubbles Burst." In *World Economic Outlook* (April). Washington, IMF.

[23] International Monetary Fund (2011), "IMF Performance in the Run-Up to the Financial and Economic Crisis. IMF Surveillance in 2004–07." Washington, IMF, Independent Evaluation Office.

[24] Jordà O., M. Schularick and A. Taylor (2013), "When credit bites back." Journal of Money, Credit and Banking, 45 (2), 3–28.

[25] Jordà O., M. Schularick and A. Taylor (2014), "The Great Mortgaging: Housing Finance, Crises, and Business Cycles." Federal Reserve Bank of San Francisco Working Paper 2014-23.

[26] Jordà O., M. Schularick and A. Taylor (2016), "Macrofinancial History and the New Business Cycle Facts." Federal Reserve Bank of San Francisco Working Paper 2016-23.

[27] Karlsson, S. (2013), "Forecasting with Bayesian Vector Autoregression." In *Handbook of Economic Forecasting*, vol. 2, part B, G. Elliott and A. Timmermann (eds.), pp. 791-897, Amsterdam, North Holland.

[28] Kindleberger, C. P. (1978), "Manias, Panics, and Crashes: A History of Financial Crises." New York, Basic Books.

[29] Leamer, E. E., (2007), "Housing is the Business Cycle." Proceedings of Economic Policy Symposium, Housing, Housing Finance, and Monetary Policy, Federal Reserve Bank of Kansas City, August 30–September 1, Jackson Hole, pp. 149–233.

[30] Marcellino, M., J. Stock and M. Watson (2003), "Macroeconomic forecasting in the Euro area: country specific versus Euro wide information." *European Economic Review*, 47(1), 1–18.

[31] McCracken M. and S. Ng (2015), FRED-QD Appendix to "FRED-MD: A Monthly Database for Macroeconomic Research." Federal Reserve Bank of St. Louis Working Paper 2015-012B. Available at https://research.stlouisfed.org/econ/mccracken/fred-databases/.

[32] Mian A. and A. Sufi (2013), "Household balance sheets, consumption, and the economic slump." *Quarterly Journal of Economics*, 128 (4), 1687–1726.

[33] Mian A. and A. Sufi (2014a), "House of Debt: How They (and You) caused the Great Recession, and How We Can Prevent It from Happening Again." Chicago, University of Chicago Press.

[34] Mian A. and A. Sufi (2014b), "What explains the 2007-2009 drop in employment?" Econometrica, 82 (6), 2197–2223.

[35] Mian A. and A. Sufi (2018), "Finance and Business Cycles: The Credit-Driven Household Demand Channel." *The Journal of Economic Perspectives*, 32 (3), 3-30.

[36] Minsky, H (1982): “Can ’it’ happen again? Essays on Instability and Finance.” Armonk: M. E. Sharpe.

[37] OECD (2014), “OECD forecasts during and after the financial crisis: a post mortem.” Economics Department Working Papers No. 1107.

[38] Piazzesi, M., and M. Schneider (2016), “Housing and Macroeconomics.” In *Handbook of Macroeconomics*, vol. 2, J. B. Taylor and H. Uhlig (eds.), pp. 1547–1640, Amsterdam, North-Holland.

[39] Phillips, P. C. B., S.P. Shi and J. Yu (2015), “Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500.” *International Economic Review*, 56 (4), 1043-1078.

[40] Reinhart, C. and K. Rogoff (2009), *This Time is Different: Eight Centuries of Financial Folly.* Princeton, Princeton University Press.

[41] Rossi, B. (2005), “Testing Long-Horizon Predictive Ability with High Persistence, and the Meese-Rogoff Puzzle.” *International Economic Review*, 46 (1), 61-92.

[42] Rossi, B. and T. Sekhposyan T. (2014), “Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set.” *International Journal of Forecasting*, 30 (3), 662–682.

[43] Shiller, R. (2015), *Irrational Exuberance.* Princeton, Princeton University Press.

[44] Smith, J., and K. F. Wallis (2009), “A simple explanation of the forecast combination puzzle.”, *Oxford Bulletin of Economics and Statistics*, 71, 331–55.

[45] Sommer, K., P. Sullivan and R. Verbrugge (2011), “Run-up in the House Price-Rent Ratio: How Much Can Be Explained by Fundamentals?” BLS Working Paper 441.

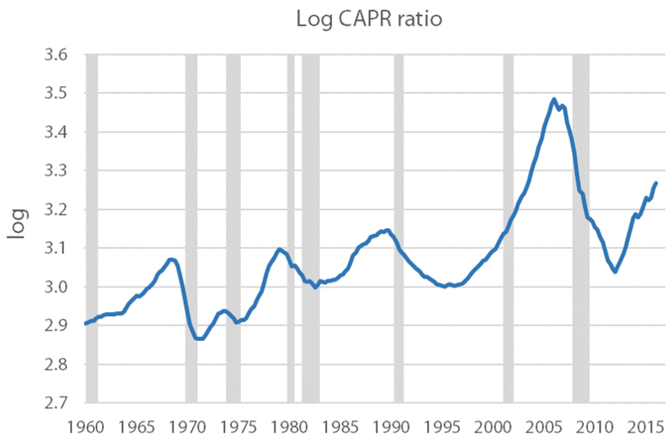
[46] Stock, J. and M. Watson (2003). “Forecasting output and inflation: The role of asset prices.” *Journal of Economic Literature*, 41 (3), 788–829.

[47] Stock, J. and M. Watson (2012) “Disentangling the Channels of the 2007-2009 Recession.” NBER WP No. 18094.

[48] Taylor, J. B. and H. Uhlig (eds.) (2016), *Handbook of Macroeconomics.* Amsterdam, North-Holland.

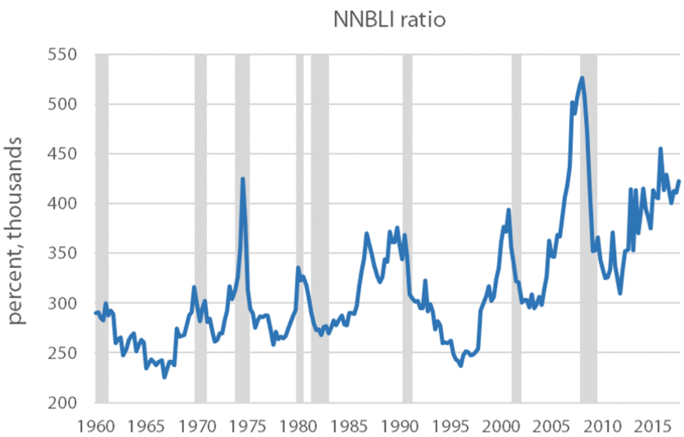
Figures and tables

Figure 2.1: Log cyclically-adjusted house price-to-rent ratio (CAPR)



Notes: Quarterly data from 1960Q1 to 2016Q1. Data on prices and rents are from the Lincoln Institute of Land Policy (https://www.lincolninst.edu/), based on Davis et al. (2008), and from the FRED website (https://fred.stlouisfed.org/). See section 2.3 in the text for more details. Shaded areas indicate NBER recessions.

Figure 2.2: Nonfinancial noncorporate business sector liabilities-to-income ratio (NNBLI)



Notes: Quarterly data from 1960Q1 to 2016Q1. The data are from the FRED-QD dataset (McCracken and Ng 2015) available at https://research.stlouisfed.org/econ/mccracken/fred-databases/. Shaded areas indicate NBER recessions.

Table 2.1: Reduced dataset

label	description	data source	source label	transformation
gdp	Real GDP growth rate	FRED	GDPC1	$\ln(x_t) - \ln(x_{t-4})$
infl	Inflation rate	FRED	CPIAUCSL	$\ln(x_t) - \ln(x_{t-4})$
fedfunds	Federal funds rate	FRED	FEDFUNDS	x_t
gs10	10-year Treasury constant maturity rate	FRED	GS10	x_t
baa10ym	Moody's BAA 10-year yield minus Treasury	FRED	BAA10YM	x_t
reer	Real narrow effective exchange rate	FRED	RNUSBIS	$\ln(x_t) - \ln(x_{t-4})$
m2	M2 money growth	FRED	M2SL	$\ln(x_t) - \ln(x_{t-4})$
unrate	Civilian unemployment rate	FRED	UNRATE	x_t
bci	OECD Business Confidence Index	OECD	—	x_t
cli	OECD Composite Leading Indicator	OECD	—	x_t
oil	Oil price inflation rate	FRED	WTISPLC	$\ln(x_t) - \ln(x_{t-4})$
sp500	Real S&P500 index growth	Shiller	—	$\ln(x_t) - \ln(x_{t-4})$
cape	Cyclically adjusted price/earnings ratio	Shiller	—	$\ln(x_t)$
pdebt	Real public debt growth	FRED	GFDEBTN/CPIAUCSL	$\ln(x_t) - \ln(x_{t-4})$
pdebt_gdp	Public debt/GDP ratio	FRED	GFDEGDQ188S	x_t
cred	Real credit growth	FRED	CRDQUSAPABIS/CPIAUCSL	$\ln(x_t) - \ln(x_{t-4})$
cred_gdp	Credit/GDP ratio	BIS	—	x_t
hpi	Real house price growth	Shiller	—	$\ln(x_t) - \ln(x_{t-4})$
houst	Housing starts	FRED	HOUST	$\ln(x_t)$
pr	Price/rent ratio	Lincoln Institute	—	$\ln(x_t)$
capr	Cyclically-adjusted price/rent ratio	Lincoln Institute; FRED	— ; CUUR0000SEHA, CPIAUCNS	$\ln(x_t)$
mortg	Households' real mortgage debt growth	FRED	HHMSDODNS/CPIAUCSL	$\ln(x_t) - \ln(x_{t-4})$
mortg_inc	Households' mortgage/income ratio	FRED	HHMSDODNS/DSPI	x_t
prfi	Private residential fixed investment growth	FRED	PRFI/CPIAUCSL	$\ln(x_t) - \ln(x_{t-4})$
prfi_gdp	Private residential fixed investment/GDP ratio	FRED	PRFI/GDP	x_t
pip_inc	Personal interest payments/income ratio	FRED	B069RC1/DSPI	x_t
nfci	Chicago Fed national financial condition index	FRED	NCFI	x_t

Notes: Shiller = http://www.econ.yale.edu/~shiller/data.htm; Lincoln Institute = Lincoln Institute of Land Policy, http://www.lincolninst.edu/resources/, based on original data by Davis et al. (2008) (using the Case-Shiller price index after 2000).

Table 2.2: Time series availability in the full dataset

	available	% total	not available	% total
1960Q1	181	66.5%	91	33.5%
1960Q2	231	84.9%	41	15.1%
1967Q1	245	90.1%	27	9.9%
1974Q1	258	94.9%	14	5.1%
1988Q1	269	98.9%	3	1.1%
2001Q1	272	100.0%	0	0.0%

Notes: The full dataset is composed of the reduced dataset (Table 2.1) and the FRED-QD dataset (Tables 2.35-2.48. Available at https://research.stlouisfed.org/econ/mccracken/fred-databases/). The table refers to data after transformations (see Table 2.1 and McCracken and Ng 2015).

Table 2.3: Regression R^2 of single-predictor ARDL models

	$h = 1$		$h = 4$		$h = 12$		$h = 20$		$h = 28$	
1	cli	0.4402	cli	0.4323	capr	0.6254	capr	0.6797	capr	0.7284
2	prfi_gdp	0.3668	capr	0.4022	pr	0.5710	pr	0.6535	pr	0.6982
3	PCESVx	0.3427	NNBTILQ027SBDIx	0.3607	NNBTILQ027SBDIx	0.5647	mortg_inc	0.5780	mortg_inc	0.6838
4	PRFIx	0.3421	prfi_gdp	0.3487	AMDMUOx	0.5544	NNBTILQ027SBDIx	0.5503	cred_gdp	0.6493
5	prfi	0.3370	pr	0.3469	cli	0.5050	cred_gdp	0.5284	NWPIx	0.5281
6	nfci	0.3339	DOTSRG3Q086SBEA	0.3326	mortg_inc	0.4545	NWPIx	0.4736	gs10	0.4810
7	houst	0.3299	nfci	0.3273	cred_gdp	0.4181	AMDMUOx	0.4419	NNBTILQ027SBDIx	0.4719
8	bci	0.3137	PRFIx	0.3245	NWPIx	0.4075	gs10	0.4318	prfi_gdp	0.4302
9	CUSR0000SAS	0.3069	AMDMUOx	0.3203	fedfunds	0.3802	cli	0.4014	REALLNx	0.4299
10	PERMIT	0.3054	prfi	0.3171	gs10	0.3473	EXCAUSx	0.3694	TLBSHNOx	0.4262
11	PCECC96	0.3040	IPCONGD	0.3098	HWIx	0.3421	fedfunds	0.3588	mortg	0.4113
12	CLAIMSx	0.2930	fedfunds	0.2969	CPIAUCSL	0.3286	bci	0.3307	NDMANEMP	0.4105
13	PERMITS	0.2914	CP3M	0.2926	bci	0.3235	GS1TB3Mx	0.2638	LIABPIx	0.4066
14	HOUSTW	0.2857	PCECC96	0.2923	CUSR0000SA0L5	0.3222	mortg	0.2566	fedfunds	0.3747
15	FPIx	0.2838	PERMITMW	0.2906	PCECTPI	0.3205	TWEXMMTH	0.2540	EXCAUSx	0.3461
16	PERMITW	0.2816	NWPIx	0.2890	prfi_gdp	0.3195	prfi_gdp	0.2415	hpi	0.3011
17	CPILFESL	0.2724	houst	0.2872	mortg	0.3193	DFSARG3Q086SBEA	0.2381	cred	0.3003
18	PERMITMW	0.2715	T5YFFM	0.2833	CPILFESL	0.3166	reer	0.2355	USINFO	0.2902
19	CPF3MTB3Mx	0.2695	USSTHPI	0.2820	fedfunds	0.3063	nfci	0.2347	AMDMUOx	0.2835
20	HWIx	0.2592	CUSR0000SAS	0.2805	CP3M	0.3040	REALLNx	0.2135	USSTHPI	0.2811
21	MANEMP	0.2559	hpi	0.2719	PCEPILFE	0.3000	LIABPIx	0.2078	cape	0.2715
22	A014RE1Q156NBEA	0.2558	PERMIT	0.2714	DFSARG3Q086SBEA	0.2998	unrate	0.2049	unrate	0.2632
23	DMANEMP	0.2476	PCESVx	0.2714	B020RE1Q156NBEA	0.2937	TLBSHNOx	0.2049	USPBS	0.2624
24	HOUST	0.2466	HOUSTW	0.2692	infl	0.2881	USSERV	0.1942	bci	0.2622
25	IMPGSC1	0.2442	PERMITW	0.2687	CUSR0000SAS	0.2771	USINFO	0.1915	houst	0.2560
26	IPMANSICS	0.2435	MORTGAGE30US	0.2590	hpi	0.2728	TB3SMFFM	0.1912	NNBTILQ027Sx	0.2520
27	fedfunds	0.2428	HWIx	0.2575	USSTHPI	0.2664	AHETPIx	0.1884	GS1TB3Mx	0.2456
28	USSTHPI	0.2420	CPILFESL	0.2555	TB3MS	0.2607	hpi	0.1853	NNBTASQ027Sx	0.2305
29	capr	0.2410	TB6M3Mx	0.2540	CPIULFSL	0.2599	cape	0.1825	MANEMP	0.2052
30	MORTG10YRx	0.2376	mortg	0.2538	PERMITMW	0.2586	HWIx	0.1749	NONREVSLx	0.2043

Notes: For each h, the dependent variable is the h-quarter cumulative GDP growth. Estimated on data from 1974Q1 to 2017Q4.

Table 2.4: Statistical significance in post-LASSO regressions: $h \leq 12$ (λ: min MSE)

$h = 1$		$h = 4$		$h = 12$	
(Intercept)	·	(Intercept)	**	(Intercept)	***
houst	·	fedfunds	*	infl	*
bci		gs10	*	sp500	***
cli		hpi		hpi	
prfi		unrate	***	houst	**
nfci		bci		cli	**
PCESVX		cli	*	capr	***
A014RE1Q156NBEA	***	prfi	*	mortg_inc	
GCEC1		nfci		Y033RC1Q027SBEAX	
A823RL1Q225SBEA		A014RE1Q156NBEA	*	A014RE1Q156NBEA	
DPIC96		A823RL1Q225SBEA		A823RL1Q225SBEA	
IPNCONGD	·	IPMAT	·	OUTNFB	**
NDMANEMP		NDMANEMP	***	IPNCONGD	**
USMINE	**	USMINE	**	IPBUSEQ	
USWTRADE		USGOVT		USEHS	
CES9091000001	*	USWTRADE		USINFO	**
HOUSTNE	*	CES9093000001		USMINE	*
ANDENOX		UNRATELTX	***	USGOVT	
GPDICTPI	*	LNS14000012		USTRADE	
DMOTRG3Q086SBEA		LNS13023557		CES9092000001	***
DFDHRG3Q086SBEA		LNS13023705	***	UNRATELTX	
DODGRG3Q086SBEA		HOANBS		LNS14000012	
DFXARG3Q086SBEA	*	HWIX		LNS14000025	**
DGOERG3Q086SBEA		HOUST5F	*	UEMPLT5	
DTRSRG3Q086SBEA		AMDMUOX		AWOTMAN	
DRCARG3Q086SBEA		ANDENOX		HWIX	***
DFSARG3Q086SBEA		INVCQRMTSPL		PCEPILFE	
DOTSRG3Q086SBEA		GDPCTPI		DSERRG3Q086SBEA	*
WPSFD4111		GPDICTPI	***	DREQRG3Q086SBEA	
WPU0531		DMOTRG3Q086SBEA		DODGRG3Q086SBEA	**
CES3000000008X	***	DREQRG3Q086SBEA		DHLCRG3Q086SBEA	**
OPHNFB	***	DODGRG3Q086SBEA	**	DTRSRG3Q086SBEA	
MORTG10YRX		DFXARG3Q086SBEA	·	DOTSRG3Q086SBEA	
GS10TB3MX	***	DCLORG3Q086SBEA		WPU0531	
CPF3MTB3MX		DGOERG3Q086SBEA		WPU0561	
AMBSLREALX		DONGRG3Q086SBEA		AHETPIX	***
M2REALX		DHUTRG3Q086SBEA	·	CES2000000008X	**
CONSUMERX	***	DHLCRG3Q086SBEA		CES3000000008X	*
REVOLSLX		DRCARG3Q086SBEA		OPHNFB	
TABSHNOX		DFSARG3Q086SBEA	**	GS10TB3MX	***
EXSZUSX	*	DOTSRG3Q086SBEA	***	AMBSLREALX	
EXUSUKX		CES2000000008X		M1REALX	**
EXCAUSX		CES3000000008X	·	MZMREALX	**
IPB51222S		OPHPBS		BUSLOANSX	·
UEMPMEAN	**	MORTGAGE30US		REVOLSLX	
TOTRESNS		MORTG10YRX	***	EXSZUSX	
CPIAPPSL		GS10TB3MX	**	UMCSENTX	*
CUSR0000SAD		AMBSLREALX	***	UEMPMEAN	·
CUSR0000SA0L2		M1REALX	***	NONBORRES	
DTCTHFNM	·	M2REALX		GS5	
INVEST		BUSLOANSX	***	TB3SMFFM	·
COMPAPFF	**	CONSUMERX	*	WPSID62	
PERMITNE		REVOLSLX	*	DTCOLNVHFNM	
PERMITMW		LIABPIX		DTCTHFNM	***
TLBSNNCBX	·	TARESAX		INVEST	·
TLBSNNCBBDIX		VXOCLSX	*	PERMITNE	
TTAABSNNCBX		EXSZUSX		PERMITS	
NNBTILQ027SBDIX		EXJPUSX	·	NASDAQCOM	***
		EXUSUKX		TTAABSNNCBX	
		IPFUELS	***	TNWMVBSNNCBX	
		NONBORRES		NNBTILQ027SBDIX	**
		PPICMM			
		CPITRNSL			
		CPIMEDSL	*		
		CUSR0000SAD	*		
		CES0600000008			
		DTCOLNVHFNM			
		DTCTHFNM	*		
		CONSPIX	**		
		NIKKEI225			
		NASDAQCOM			
		TLBSNNCBX			
		TNWMVBSNNCBX			
		NNBTILQ027SX	***		
		NNBTILQ027SBDIX			

Notes: OLS regressions estimated on data from 1974Q1 to 2017Q4 using HAC standard errors. For each h, the dependent variable is the h-quarter cumulative GDP growth. Regressors are at time $t - h$ and are selected by LASSO, using the value of λ that minimizes the cross-validated MSE. Significance levels: *** 0.1%, ** 1%, * 5%, · 10%.

Table 2.5: Statistical significance in post-LASSO regressions: $h > 12$ (λ: min MSE)

$h = 20$		$h = 28$	
(Intercept)	***	(Intercept)	***
gs10	***	fedfundsv	*
reer		gs10	·
m2	***	m2	***
houst	*	hpi	
bci	**	cred_gdp	***
cli		bci	*
capr	***	capr	**
A014RE1Q156NBEA	***	mortg	***
A823RL1Q225SBEA		pcdgx	·
SLCEX	**	PNFIX	*
EXPGSC1		A014RE1Q156NBEA	**
IPB51110SQ	*	GCEC1	
IPB51220SQ		A823RL1Q225SBEA	·
USEHS		DPIC96	·
USMINE	***	TCU	
CES9091000001	*	USFIRE	*
CES9093000001	***	USINFO	
CIVPART	**	USMINE	**
UEMP27OV	***	USGOVT	·
LNS13023705		CES9092000001	
LNS13023569	***	CES9093000001	**
HWIX	***	UEMP27OV	**
AMDMUOX		LNS13023705	***
INVCQRMTSPL	**	DODGRG3Q086SBEA	
DGDSRG3Q086SBEA	**	DFXARG3Q086SBEA	
DDURRG3Q086SBEA		DHUTRG3Q086SBEA	*
DMOTRG3Q086SBEA		WPSFD4111	*
DOTSRG3Q086SBEA		OPHPBS	*
WPSFD4111	*	UNLPNBS	
WPU0531	*	GS1	**
CES3000000008X		MORTG10YRX	
COMPRNFB	***	M1REALX	**
OPHPBS	·	MZMREALX	***
TB6M3MX	***	NONREVSLX	***
MZMREALX	·	EXCAUSX	·
REALLNX	***	B021RE1Q156NBEA	
REVOLSLX	***	NONBORRES	**
TOTALSLX	***	WPSID62	
EXSZUSX		CUSR0000SAD	***
EXJPUSX	***	CUSR0000SAS	
GFDEBTNX	·	COMPAPFF	
IPFUELS	·	PERMITS	
CPIMEDSL	*	NIKKEI225	
CES0600000008	***	NASDAQCOM	
INVEST		TLBSNNCBX	
NIKKEI225		TLBSNNCBBDIX	***
NASDAQCOM		TTAABSNNCBX	**
TNWMVBSNNCBBDIX		NNBTILQ027SBDIX	***
NNBTILQ027SX			
NNBTILQ027SBDIX	*		

Notes: OLS regressions estimated on data from 1974Q1 to 2017Q4 using HAC standard errors. For each h, the dependent variable is the h-quarter cumulative GDP growth. Regressors are at time $t - h$ and are selected by LASSO, using the value of λ that minimizes the cross-validated MSE. Significance levels: *** 0.1%, ** 1%, * 5%, · 10%.

Table 2.6: Statistical significance in post-LASSO regressions: $h \leq 12$ (λ: one-s.e. rule)

$h = 1$		$h = 4$		$h = 12$	
(Intercept)	*	(Intercept)	*	(Intercept)	***
bci		hpi	**	infl	
cli	**	unrate	**	sp500	**
PCESVX	*	bci	·	hpi	
A014RE1Q156NBEA	***	nfci	·	houst	*
USMINE	·	PCESVX		cred	
PERMIT		A014RE1Q156NBEA	***	cli	*
GPDICTPI		A823RL1Q225SBEA		capr	***
DFDHRG3Q086SBEA		IPB51220SQ		mortg_inc	
DFXARG3Q086SBEA		USMINE	*	A823RL1Q225SBEA	
MORTG10YRX	·	USGOVT		IPB51110SQ	
CPF3MTB3MX		USWTRADE	*	IPNCONGD	**
TABSHNOX	·	LNS14000012	*	IPBUSEQ	*
IPB51222S		LNS13023705	***	USEHS	
CUSR0000SAS		HWIX	*	USINFO	*
PERMITMW		HOUST5F		USMINE	
PERMITW		AMDMUOX		CES9091000001	
TLBSNNCBBDIX	*	DMOTRG3Q086SBEA	·	CES9092000001	***
NNBTILQ027SBDIX	**	DREQRG3Q086SBEA		AWOTMAN	·
		DODGRG3Q086SBEA	**	HWIX	***
		DFXARG3Q086SBEA	*	DSERRG3Q086SBEA	*
		DCLORG3Q086SBEA		DREQRG3Q086SBEA	
		DGOERG3Q086SBEA	·	DODGRG3Q086SBEA	***
		DHUTRG3Q086SBEA		DTRSRG3Q086SBEA	
		DHLCRG3Q086SBEA		DOTSRG3Q086SBEA	
		DRCARG3Q086SBEA		WPU0531	
		DFSARG3Q086SBEA		AHETPIX	·
		DOTSRG3Q086SBEA	**	OPHNFB	**
		WPSFD4111		TB3MS	**
		CES2000000008X		GS1TB3MX	
		CES3000000008X		GS10TB3MX	***
		OPHPBS		AMBSLREALX	
		MORTGAGE30US	*	M1REALX	***
		MORTG10YRX	*	MZMREALX	***
		TB6M3MX		REVOLSLX	
		GS10TB3MX	**	EXSZUSX	
		AMBSLREALX	*	EXJPUSX	
		M1REALX		UMCSENTX	*
		M2REALX		UEMPMEAN	*
		CONSUMERX	**	TB3SMFFM	
		REVOLSLX		WPSID62	*
		TARESAX		DTCOLNVHFNM	
		VXOCLSX	***	DTCTHFNM	***
		EXUSUKX	·	INVEST	
		UMCSENTX		COMPAPFF	
		IPFUELS	*	PERMITNE	·
		PPICMM	·	PERMITS	**
		CES0600000008		NASDAQCOM	**
		DTCOLNVHFNM		TTAABSNNCBX	
		DTCTHFNM	·	TNWMVBSNNCBX	
		CONSPIX	***	NNBTILQ027SBDIX	
		NIKKEI225			
		TLBSNNCBX	*		
		TNWMVBSNNCBX	*		
		NNBTILQ027SX	***		
		NNBTILQ027SBDIX	***		

Notes: OLS regressions estimated on data from 1974Q1 to 2017Q4 using HAC standard errors. For each h, the dependent variable is the h-quarter cumulative GDP growth. Regressors are at time $t - h$ and are selected by LASSO, using the one-standard-error rule for λ. Significance levels: *** 0.1%, ** 1%, * 5%, · 10%.

Table 2.7: Statistical significance in post-LASSO regressions: $h > 12$ (λ: one-s.e. rule)

$h = 20$		$h = 28$	
(Intercept)	***	(Intercept)	***
gs10	***	fedfunds	**
reer		gs10	
m2	***	m2	***
houst		hpi	
cli		cred_gdp	***
capr	***	bci	*
pip_inc	*	capr	
A014RE1Q156NBEA	***	mortg	***
A823RL1Q225SBEA		prfi_gdp	
SLCEX	***	PNFIX	
EXPGSC1	*	A014RE1Q156NBEA	***
IPB51110SQ		GCEC1	
IPB51220SQ	**	A823RL1Q225SBEA	
USEHS		TCU	·
USMINE	***	USFIRE	*
CES9091000001	·	USINFO	
CES9093000001	***	USMINE	**
CIVPART	***	USGOVT	
UEMP27OV	***	CES9092000001	·
LNS13023705		CES9093000001	**
LNS13023569	*	UEMP27OV	***
HWIX	***	LNS13023705	***
AMDMUOX		DODGRG3Q086SBEA	
DGDSRG3Q086SBEA	***	DFXARG3Q086SBEA	
DDURRG3Q086SBEA		DHUTRG3Q086SBEA	*
DMOTRG3Q086SBEA	·	WPSFD4111	**
DOTSRG3Q086SBEA		OPHPBS	***
WPSFD4111	*	UNLPNBS	
WPU0531	**	GS1	***
CES3000000008X		MORTG10YRX	
COMPRNFB	***	M1REALX	**
TB6M3MX	***	MZMREALX	***
MZMREALX	*	CONSUMERX	
REALLNX	***	NONREVSLX	***
REVOLSLX	***	EXCAUSX	*
TOTALSLX	***	GS5	·
EXJPUSX	**	CUSR0000SAD	***
GFDEBTNX	**	COMPAPFF	
CPIMEDSL	**	PERMITS	
CES0600000008	**	NIKKEI225	**
INVEST		NASDAQCOM	*
CONSPIX	*	TLBSNNCBX	
NIKKEI225		TLBSNNCBBDIX	***
NASDAQCOM		TTAABSNNCBX	**
TNWMVBSNNCBBDIX	**	NNBTILQ027SBDIX	***
NNBTILQ027SX	*		
NNBTILQ027SBDIX			

Notes: OLS regressions estimated on data from 1974Q1 to 2017Q4 using HAC standard errors. For each h, the dependent variable is the h-quarter cumulative GDP growth. Regressors are at time $t-h$ and are selected by LASSO, using the one-standard-error rule for λ. Significance levels: *** 0.1%, ** 1%, * 5%, · 10%.

Table 2.8: Best subset selection for increasing model size: $h = 1$

variable\ k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
prfi	■		■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
cli		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
cred_gdp				■					■	■																				
nfci				■					■	■																				
gs10						■	■	■			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
fedfunds							■						■	■												■	■	■	■	■
gdp								■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
infl										■					■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
trend											■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
mortg											■	■			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
capr											■	■			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
bci											■	■		■																■
prfi_gdp													■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
baa10ym													■	■					■	■	■	■	■	■	■					
pr													■	■																
houst															■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pdebt_gdp															■											■	■	■	■	
mortg_inc																■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
sp500																■	■													■
oil																	■	■	■	■	■	■	■	■	■	■	■	■	■	■
cape																		■	■	■	■	■	■	■	■	■	■	■	■	■
hpi																							■	■	■	■	■	■	■	■
pdebt																							■	■	■	■	■	■	■	■
pip_inc																													■	■
cred																														
m2																														
unrate																														
reer																														

Notes: The table shows the selected subsets of regressors for an increasing number of regressors k, based on MSE minimization. Note that for $k > 4$ the results should be taken as approximations of the true MSE-minimizing subsets (see note 14 in the text). The dependent variable is the h-quarter cumulative growth rate of GDP. Regressors are lagged variables between time $t - h - 4$ and time $t - h$. Blue cells indicate inclusion in the best subset. The red border identifies the model with the lowest BIC. Estimated over the period 1974Q1-2017Q4.

Table 2.9: Best subset selection for increasing model size: $h = 4$

variable\ k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
prfi	■	■	■	■	■	■							■	■	■	■	■				■	■	■	■	■	■	■	■	■	■
cli		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
cred_gdp				■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
nfci				■	■	■	■	■								■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pdebt					■							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
infl						■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
unrate						■	■	■	■	■	■																			
hpi							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
oil							■	■	■	■	■							■	■	■	■	■	■	■	■	■	■	■	■	■
mortg_inc								■		■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
bci									■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
gs10									■																		■	■	■	■
fedfunds									■	■	■																			
gdp												■	■	■	■	■	■	■	■	■	■	■	■	■	■					■
sp500												■	■	■	■	■	■	■	■	■	■	■								
baa10ym															■	■							■	■	■	■	■	■	■	■
m2																	■	■	■	■	■	■	■	■	■	■	■	■	■	■
cred																	■									■	■	■	■	■
capr																		■	■	■	■	■	■	■	■	■	■	■	■	■
cape																			■											
mortg																				■	■	■	■	■	■	■	■	■	■	■
reer																						■	■	■	■	■	■	■	■	■
pip_inc																									■					
prfi_gdp																										■	■	■		
houst																													■	■
pdebt_gdp																														
pr																														
trend																														

Notes: The table shows the selected subsets of regressors for an increasing number of regressors k, based on MSE minimization. Note that for $k > 4$ the results should be taken as approximations of the true MSE-minimizing subsets (see note 14 in the text). The dependent variable is the h-quarter cumulative growth rate of GDP. Regressors are lagged variables between time $t - h - 4$ and time $t - h$. Blue cells indicate inclusion in the best subset. The red border identifies the model with the lowest BIC. Estimated over the period 1974Q1-2017Q4.

Table 2.10: Best subset selection for increasing model size: $h = 12$

variable\ k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
capr	■	■	■					■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
cli		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
prfi			■											■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
cape				■	■						■	■	■	■	■	■	■							■	■	■	■	■	■	■
unrate				■	■	■												■	■	■	■	■	■	■		■	■	■	■	■
cred_gdp				■	■	■	■														■	■	■	■	■	■	■	■	■	■
infl					■	■	■																■	■	■	■	■	■	■	■
trend						■	■					■						■	■	■	■	■	■	■	■	■	■	■	■	■
bci							■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
hpi							■											■	■	■	■	■	■	■	■	■	■	■	■	■
gdp								■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pdebt								■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pdebt_gdp								■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pr								■	■	■	■	■	■	■	■	■	■				■	■	■	■	■	■	■	■	■	■
m2								■															■	■						
pip_inc										■	■	■	■	■	■	■	■													
nfci											■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
baa10ym												■	■	■	■	■	■					■								■
mortg_inc																	■	■	■	■	■	■	■	■	■	■	■	■	■	■
fedfunds																	■				■			■						
prfi_gdp																		■	■	■	■	■	■	■	■	■	■	■	■	■
sp500																		■	■	■	■	■	■	■	■	■	■	■	■	■
oil																		■	■	■										
cred																								■	■	■	■	■	■	■
gs10																										■				
houst																														
mortg																														
reer																														

Notes: The table shows the selected subsets of regressors for an increasing number of regressors k, based on MSE minimization. Note that for $k > 4$ the results should be taken as approximations of the true MSE-minimizing subsets (see note 14 in the text). The dependent variable is the h-quarter cumulative growth rate of GDP. Regressors are lagged variables between time $t - h - 4$ and time $t - h$. Blue cells indicate inclusion in the best subset. The red border identifies the model with the lowest BIC. Estimated over the period 1974Q1-2017Q4.

Table 2.11: Best subset selection for increasing model size: $h = 20$

variable\ *k*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
capr	■	■	■	■													■	■	■		■	■	■	■	■	■	■	■	■	■
cli		■		■			■	■	■					■	■	■					■	■	■	■	■	■	■	■	■	■
m2			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
gs10			■	■								■	■	■	■	■	■	■	■	■										
cred_gdp					■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
hpi					■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
trend					■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
baa10ym					■	■	■	■	■		■	■	■	■	■	■	■	■	■	■							■	■	■	■
infl							■	■	■	■	■	■	■	■	■	■	■	■	■	■										
nfci								■	■																					
mortg_inc									■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
gdp										■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
mortg										■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pdebt_gdp										■	■									■	■	■	■	■	■	■	■	■	■	■
houst												■	■	■	■	■	■	■	■											
sp500													■																	
bci														■	■	■	■	■	■	■										
prfi_gdp																■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
unrate																	■	■	■											
reer																		■	■		■	■	■	■	■	■	■	■	■	■
pdebt																				■	■	■	■	■	■	■	■	■	■	■
pr																				■	■	■	■	■	■	■	■	■	■	■
prfi																				■	■	■	■	■	■	■	■	■	■	■
oil																						■	■	■	■	■	■	■	■	■
fedfunds																								■						
cape																														
cred																														
pip_inc																														

Notes: The table shows the selected subsets of regressors for an increasing number of regressors k, based on MSE minimization. Note that for $k > 4$ the results should be taken as approximations of the true MSE-minimizing subsets (see note 14 in the text). The dependent variable is the h-quarter cumulative growth rate of GDP. Regressors are lagged variables between time $t - h - 4$ and time $t - h$. Blue cells indicate inclusion in the best subset. The red border identifies the model with the lowest BIC. Estimated over the period 1974Q1-2017Q4.

Table 2.12: Best subset selection for increasing model size: $h = 28$

variable\ *k*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
capr	■	■					■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
gs10		■		■	■	■																			■	■	■	■	■	■
cred_gdp			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pdebt_gdp			■	■	■	■								■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
fedfunds			■											■														■	■	■
gdp				■	■	■	■				■	■	■			■	■		■	■			■	■	■					
sp500					■	■							■										■	■	■	■	■	■	■	
nfci						■							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
infl							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
mortg							■	■	■	■							■	■	■	■	■	■	■	■	■	■	■	■	■	
pr							■	■	■	■	■	■	■	■	■	■														
trend							■	■	■	■	■	■	■	■	■	■														
bci								■	■	■			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
mortg_inc								■	■	■	■	■	■	■	■	■														
pdebt										■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
prfi_gdp										■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
baa10ym											■	■				■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pip_inc											■	■					■	■	■	■	■	■	■	■	■					
oil													■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
cli															■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
houst																	■									■	■	■	■	■
prfi																		■	■	■	■	■	■	■	■	■	■	■	■	■
cred																					■	■	■	■	■	■	■	■	■	■
hpi																						■	■	■	■	■	■	■	■	■
reer																									■				■	■
unrate																										■	■	■	■	■
m2																														■
cape																														

Notes: The table shows the selected subsets of regressors for an increasing number of regressors k, based on MSE minimization. Note that for $k > 4$ the results should be taken as approximations of the true MSE-minimizing subsets (see note 14 in the text). The dependent variable is the h-quarter cumulative growth rate of GDP. Regressors are lagged variables between time $t - h - 4$ and time $t - h$. Blue cells indicate inclusion in the best subset. The red border identifies the model with the lowest BIC. Estimated over the period 1974Q1-2017Q4.

Table 2.13: Direct forecasts by single-predictor ARDL models: mean squared forecast errors

	$h=1$		$h=4$		$h=12$		$h=20$		$h=28$	
1	NNBTILQ027SBDIx	0.8905	NNBTILQ027SBDIx	0.5250	NNBTILQ027SBDIx	0.3223	capr	0.2859	capr	0.2683
2	S_P_div_yield	0.8999	capr	0.7265	capr	0.5153	NNBTILQ027SBDIx	0.4293	pr	0.4318
3	cape	0.9106	AMDMUOx	0.8139	cli	0.6170	pr	0.4576	NNBTILQ027SBDIx	0.5188
4	LNS12032194	0.9216	pr	0.8349	AMDMUOx	0.7760	cli	0.7847	NASDAQCOM	0.8143
5	AMDMUOx	0.9263	cli	0.8496	NWPIx	0.8228	AHETPIx	0.9012	EXCAUSx	0.8284
6	IPCONGD	0.9264	IPCONGD	0.8574	cred_gdp	0.8588	IPCONGD	0.9125	baa10ym	0.8798
7	TLBSHNOx	0.9349	NWPIx	0.8649	IPCONGD	0.8651	reer	0.9204	IPNCONGD	0.8825
8	CUSR0000SAS	0.9379	IPNCONGD	0.8700	fedfunds	0.8755	SLCEx	0.9222	VXOCLSX	0.8858
9	PCECC96	0.9404	PCECC96	0.8728	GS1TB3Mx	0.8841	LNS13023557	0.9328	IPCONGD	0.8880
10	CMRMTSPLx	0.9427	prfi_gdp	0.8780	IPNCONGD	0.9036	ANDENOx	0.9356	WPU0531	0.8915
11	TABSHNOx	0.9551	prfi	0.8840	gs10	0.9078	IPNCONGD	0.9513	reer	0.8994
12	TNWBSHNOx	0.9586	PERMITS	0.8850	PERMITMW	0.9201	AMDMUOx	0.9535	HWIx	0.9028
13	IPMANSICS	0.9590	CUSR0000SAS	0.9027	PERMIT	0.9319	PERMITMW	0.9561	B021RE1Q156NBEA	0.9056
14	HNOREMQ027Sx	0.9631	S_P_div_yield	0.9063	CUSR0000SAS	0.9366	TNWBSNNBBDIx	0.9590	cli	0.9091
15	HOUSTMW	0.9640	REVOLSLx	0.9151	bci	0.9375	baa10ym	0.9602	TWEXMMTH	0.9209
16	prfi_gdp	0.9680	cape	0.9172	PCECC96	0.9397	LNS14000026	0.9607	TNWBSNNBBDIx	0.9213
17	NIKKEI225	0.9683	PERMIT	0.9192	DFXARG3Q086SBEA	0.9415	UNRATE	0.9618	hpi	0.9217
18	pr	0.9696	PERMITW	0.9208	PCESVx	0.9418	IPDCONGD	0.9623	LNS13023557	0.9220
19	IPNCONGD	0.9703	HOUSTW	0.9275	USMINE	0.9419	UEMP27OV	0.9636	prfi_gdp	0.9242
20	mortg_inc	0.9710	IPB51220SQ	0.9308	PERMITW	0.9432	B021RE1Q156NBEA	0.9648	TNWMVBSNNCBBDIx	0.9313
21	CPILFESL	0.9729	S_P_500	0.9353	PERMITNE	0.9436	HOUST	0.9648	LNS13023705	0.9424
22	DTRSRG3Q086SBEA	0.9736	HOUSTS	0.9363	HOUST	0.9443	HWIx	0.9651	unrate	0.9428
23	PCNDx	0.9746	NASDAQCOM	0.9411	FEDFUNDS	0.9452	HOUSTMW	0.9657	PERMITS	0.9437
24	IPDCONGD	0.9754	houst	0.9429	HOUSTW	0.9461	PERMIT	0.9679	AHETPIx	0.9523
25	cred_gdp	0.9760	hpi	0.9430	SLCEx	0.9492	UEMP15T26	0.9684	PERMIT	0.9525
26	PERMITMW	0.9761	PRFIx	0.9440	CPILFESL	0.9495	USMINE	0.9685	HOUSTS	0.9555
27	IPB51110SQ	0.9761	S_P_indust	0.9448	UEMP15T26	0.9512	DFSARG3Q086SBEA	0.9695	UNRATE	0.9595
28	S_P_indust	0.9762	CPILFESL	0.9489	UEMP5TO14	0.9531	PCECC96	0.9707	ANDENOx	0.9627
29	IPDBS	0.9783	GS1TB3Mx	0.9501	ISRATIOx	0.9534	cred_gdp	0.9712	OUTBS	0.9628
30	S_P_500	0.9789	PCNDx	0.9511	EXCAUSx	0.9544	LNS13023705	0.9730	IPDCONGD	0.9629

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.15). All models are estimated recursively (shortest sample 1967Q1-1983Q1, longest sample 1967Q1-2016Q1). MSFE computed over the period 1990Q1-2017Q4.

Table 2.14: Iterated forecasts by bivariate VAR models: mean squared forecast errors

	$h = 1$		$h = 4$		$h = 12$		$h = 20$		$h = 28$	
1	cred_gdp	0.9112	NNBTILQ027SBDIx	0.5421	NNBTILQ027SBDIx	0.3679	capr	0.4381	capr	0.5035
2	S_P_div_yield	0.9307	capr	0.8056	capr	0.6610	NNBTILQ027SBDIx	0.4870	pr	0.5465
3	CUMFNS	0.9442	CPF3MTB3Mx	0.8094	cli	0.7063	pr	0.5782	prfi_gdp	0.6669
4	CPILFESL	0.9469	prfi_gdp	0.8112	AMDMUOx	0.7275	AMDMUOx	0.7191	NNBTILQ027SBDIx	0.6928
5	TLBSHNOx	0.9634	AMDMUOx	0.8260	pr	0.7958	NWPIx	0.7344	HNOREMQ027Sx	0.6939
6	DFSARG3Q086SBEA	0.9667	S_P_div_yield	0.8438	prfi_gdp	0.8287	prfi_gdp	0.7440	cli	0.7950
7	BAA	0.9689	pr	0.8471	AAA	0.8525	cli	0.7702	AMDMUOx	0.7973
8	CUSR0000SAS	0.9708	TLBSNNCBBDIx	0.8567	USMINE	0.8588	HNOREMQ027Sx	0.7954	REALLNx	0.8111
9	CMRMTSPLx	0.9771	hpi	0.8627	PRFIx	0.8615	TNWBSNNBBDIx	0.8610	USGOVT	0.8122
10	TCU	0.9850	nfci	0.8723	NWPIx	0.8714	TNWMVBSNNCBBDIx	0.8756	PRFIx	0.8449
11	FPIx	0.9886	EXPGSC1	0.8789	GS10TB3Mx	0.8959	GS10TB3Mx	0.8765	NWPIx	0.8532
12	cape	1.0111	sp500	0.8893	USGOVT	0.8994	CNCFx	0.8807	GS10TB3Mx	0.8818
13	LNS12032194	1.0123	PCNDx	0.8972	CPF3MTB3Mx	0.9023	PRFIx	0.8862	CP3M	0.8982
14	sp500	1.0128	CUSR0000SAS	0.9006	nfci	0.9078	USGOVT	0.8980	TB3MS	0.8990
15	pr	1.0177	cape	0.9036	GS10	0.9107	CPILFESL	0.9002	PERMITS	0.9065
16	PERMITMW	1.0195	S_P_500	0.9040	houst	0.9147	TLBSHNOx	0.9019	TB6MS	0.9109
17	S_P_500	1.0306	cli	0.9047	T5YFFM	0.9168	REALLNx	0.9084	FEDFUNDS	0.9119
18	CES0600000007	1.0346	mortg	0.9087	BAA	0.9169	houst	0.9161	CPIMEDSL	0.9155
19	AAA	1.0346	S_P_indust	0.9148	GS5	0.9190	nfci	0.9223	houst	0.9156
20	bci	1.0373	PERMITMW	0.9159	hpi	0.9230	PERMITNE	0.9245	GS1	0.9180
21	GPDIC1	1.0383	OILPRICEx	0.9197	CP3M	0.9247	S_P_div_yield	0.9264	PERMITNE	0.9180
22	nfci	1.0420	OUTBS	0.9200	PERMITNE	0.9271	PERMITS	0.9288	sp500	0.9214
23	PERMITS	1.0445	CPILFESL	0.9224	HOUSTNE	0.9328	GS1TB3Mx	0.9379	TLBSHNOx	0.9263
24	UNLPNBS	1.0448	DFSARG3Q086SBEA	0.9244	S_P_div_yield	0.9335	CES0600000008	0.9435	GPDIC1	0.9314
25	DPIC96	1.0463	oil	0.9251	CPILFESL	0.9358	T5YFFM	0.9444	BAA	0.9314
26	cred	1.0477	MORTG10YRx	0.9251	CES0600000008	0.9384	S_P_500	0.9449	CES0600000008	0.9318
27	OILPRICEx	1.0478	PERMITS	0.9259	sp500	0.9418	sp500	0.9452	S_P_500	0.9331
28	CUSR0000SA0L5	1.0511	NWPIx	0.9269	MZMREALx	0.9442	UEMP27OV	0.9456	S_P_indust	0.9348
29	S_P_indust	1.0513	CES0600000008	0.9309	AAAFFM	0.9458	AWHNONAG	0.9490	NNBTASQ027Sx	0.9362
30	oil	1.0523	GS1TB3Mx	0.9450	HOUSTMW	0.9492	GDPCTPI	0.9501	nfci	0.9387

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.16). All models are estimated recursively (shortest sample 1967Q1-1983Q1, longest sample 1967Q1-2016Q1). MSFE computed over the period 1990Q1-2017Q4.

Table 2.15: Benchmark AR for direct forecasts: root mean squared forecast error (RMSFE)

	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
$RMSFE$	0.00571	0.01787	0.04402	0.06211	0.08014

Notes: Out-of-sample RMSFE for the h-period-ahead real GDP level. The model is estimated recursively (shortest sample 1967Q1-1983Q1, longest sample 1967Q1-2016Q1). The RMSFE is computed over the period 1990Q1-2017Q4.

Table 2.16: Benchmark AR for iterated forecasts: root mean squared forecast error (RMSFE)

	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
$RMSFE$	0.00679	0.01796	0.04252	0.06244	0.08718

Notes: Out-of-sample RMSFE for the h-period-ahead real GDP level. The model is estimated recursively (shortest sample 1967Q1-1983Q1, longest sample 1967Q1-2016Q1). The RMSFE is computed over the period 1990Q1-2017Q4.

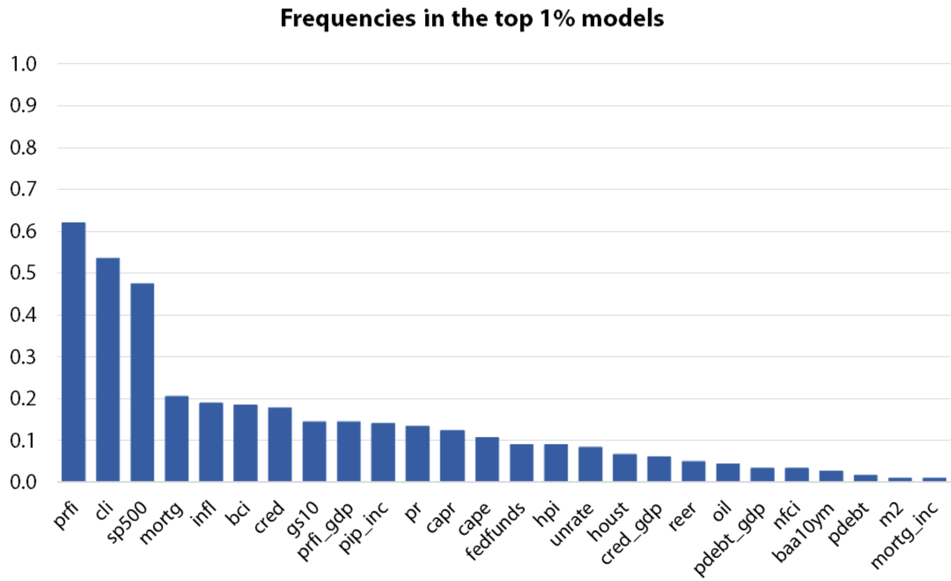
Table 2.17: Best small multivariate VAR models: mean squared forecast errors

target h	best model	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
1	fedfunds cli prfi pip_inc	0.826	0.829	0.918	2.395	3.158
4	infl sp500 hpi pr	0.927	0.587	0.660	1.001	1.024
12	houst oil cli capr	0.959	0.767	0.342	0.292	0.427
20	m2 houst cli capr	0.965	0.893	0.350	0.246	0.410
28	reer m2 pr prfi	1.218	0.905	0.762	0.388	0.262

Notes: For each target horizon h, the table shows the VAR model with the lowest out-of-sample MSFE for the h-period-ahead real GDP level, considering all possible combinations of k predictors of GDP from the reduced dataset (Table 2.1), with $k = 1, 2, 3, 4$. For each model, the MSFE is reported for all horizons, relative to the benchmark AR (Table 2.16). All models are estimated recursively (shortest sample 1967Q1-1983Q1, longest sample 1967Q1-2016Q1). The MSFE are computed over the period 1990Q1-2017Q4.

Figure 2.3: Predictors' frequencies of inclusion in best-performing VAR models: $h = 1$

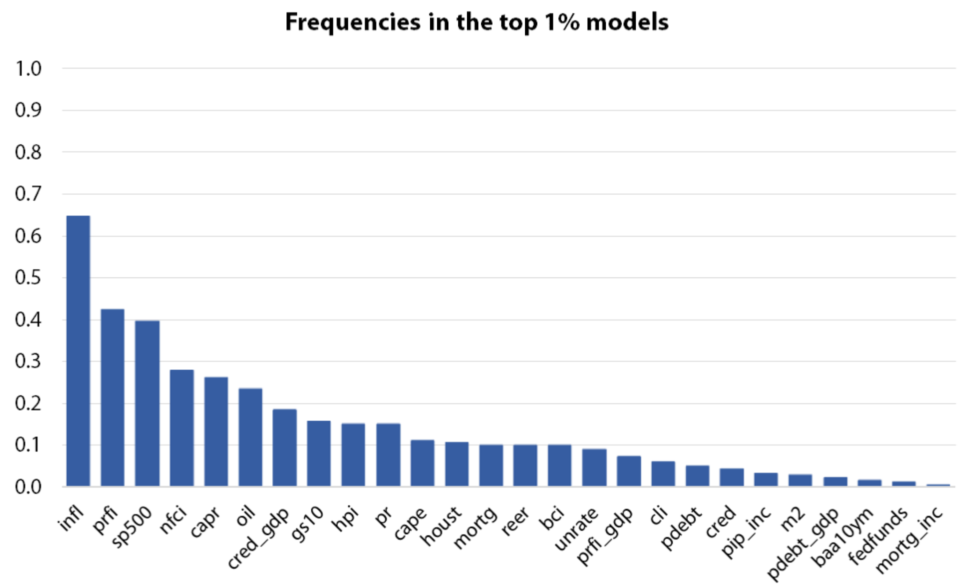
variable	frequencies top5%	top1%	top100	top10	top1
prfi	0.39	0.62	0.60	0.7	1
cli	0.48	0.54	0.53	0.7	1
sp500	0.51	0.47	0.48	0.3	0
mortg	0.20	0.21	0.23	0.3	0
infl	0.16	0.19	0.21	0.1	0
bci	0.16	0.18	0.21	0.6	0
cred	0.12	0.18	0.21	0.2	0
gs10	0.11	0.15	0.16	0.0	0
prfi_gdp	0.20	0.15	0.16	0.0	0
pip_inc	0.08	0.14	0.19	0.3	1
pr	0.19	0.13	0.10	0.1	0
capr	0.18	0.12	0.11	0.2	0
cape	0.10	0.11	0.13	0.1	0
fedfunds	0.07	0.09	0.14	0.3	1
hpi	0.11	0.09	0.08	0.0	0
unrate	0.11	0.08	0.07	0.0	0
houst	0.11	0.07	0.07	0.1	0
cred_gdp	0.09	0.06	0.03	0.0	0
reer	0.08	0.05	0.03	0.0	0
oil	0.08	0.04	0.03	0.0	0
pdebt_gdp	0.05	0.03	0.03	0.0	0
nfci	0.07	0.03	0.01	0.0	0
baa10ym	0.06	0.03	0.01	0.0	0
pdebt	0.04	0.02	0.01	0.0	0
m2	0.06	0.01	0.00	0.0	0
mortg_inc	0.04	0.01	0.00	0.0	0



Notes: The left panel shows the frequencies of inclusion in the best-performing (top 5%, top 1%, top 100, top 10 and single best) VAR models estimated on the reduced dataset (Table 2.1). The VARs include GDP growth plus all possible combinations of k predictors, for $k = 1, 2, 3, 4$ (17,901 models in total), and are ranked using the MSFE of iterated forecasts for the h-period-ahead GDP level (see also Table 2.17). The histogram in the right panel shows the frequencies of inclusion in the top 1% models.

Figure 2.4: Predictors' frequencies of inclusion in best-performing VAR models: $h = 4$

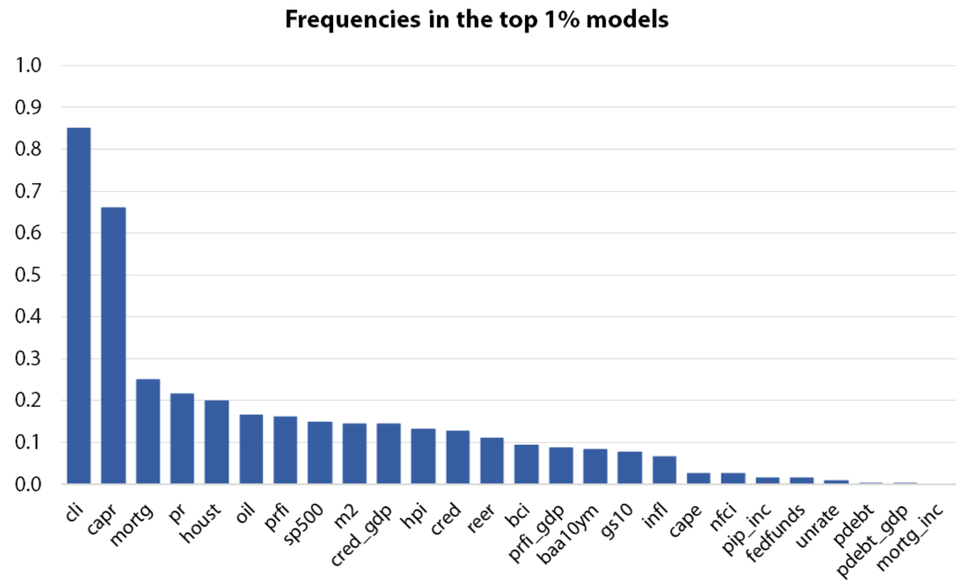
variable	frequencies				
	top5%	top1%	top100	top10	top1
infl	0.46	0.65	0.64	0.9	1
prfi	0.36	0.42	0.47	0.2	0
sp500	0.28	0.40	0.49	1.0	1
nfci	0.24	0.28	0.26	0.0	0
capr	0.30	0.26	0.26	0.5	0
oil	0.22	0.23	0.27	0.2	0
cred_gdp	0.15	0.18	0.14	0.0	0
gs10	0.11	0.16	0.13	0.1	0
hpi	0.14	0.15	0.14	0.2	1
pr	0.17	0.15	0.13	0.3	1
cape	0.12	0.11	0.10	0.0	0
houst	0.12	0.11	0.10	0.2	0
mortg	0.13	0.10	0.12	0.0	0
reer	0.14	0.10	0.11	0.0	0
bci	0.15	0.10	0.10	0.0	0
unrate	0.09	0.09	0.08	0.0	0
prfi_gdp	0.09	0.07	0.08	0.2	0
cli	0.11	0.06	0.06	0.0	0
pdebt	0.07	0.05	0.04	0.0	0
cred	0.12	0.04	0.03	0.0	0
pip_inc	0.07	0.03	0.03	0.0	0
m2	0.05	0.03	0.03	0.1	0
pdebt_gdp	0.02	0.02	0.01	0.0	0
baa10ym	0.03	0.02	0.02	0.0	0
fedfunds	0.01	0.01	0.02	0.0	0
mortg_inc	0.02	0.01	0.00	0.0	0



Notes: The left panel shows the frequencies of inclusion in the best-performing (top 5%, top 1%, top 100, top 10 and single best) VAR models estimated on the reduced dataset (Table 2.1). The VARs include GDP growth plus all possible combinations of k predictors, for $k = 1, 2, 3, 4$ (17,901 models in total), and are ranked using the MSFE of iterated forecasts for the h-period-ahead GDP level (see also Table 2.17). The histogram in the right panel shows the frequencies of inclusion in the top 1% models.

Figure 2.5: Predictors' frequencies of inclusion in best-performing VAR models: $h = 12$

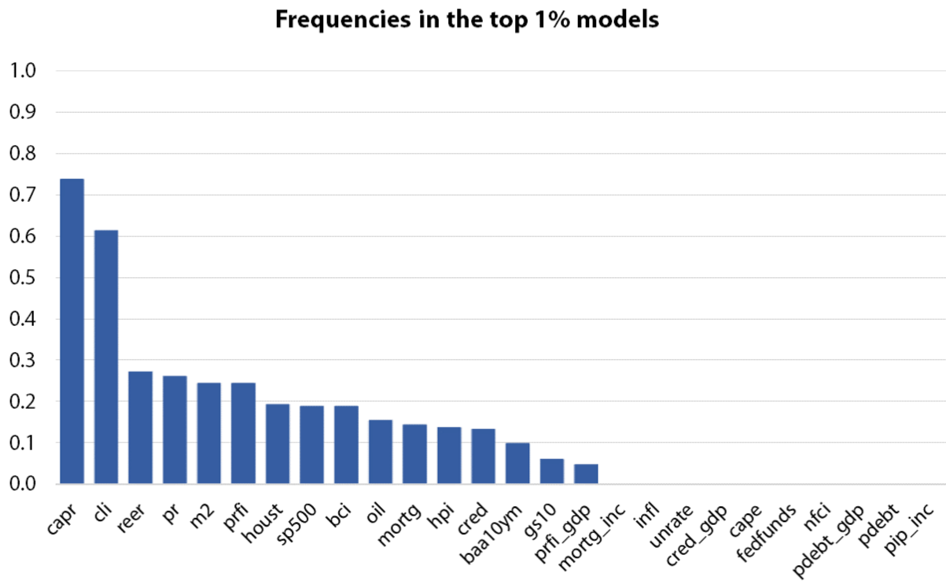
variable	frequencies top5%	top1%	top100	top10	top1
cli	0.59	0.85	0.96	1.0	1
capr	0.50	0.66	0.75	1.0	1
mortg	0.17	0.25	0.18	0.5	0
pr	0.26	0.22	0.17	0.0	0
houst	0.16	0.20	0.23	0.5	1
oil	0.16	0.17	0.15	0.2	1
prfi	0.25	0.16	0.13	0.0	0
sp500	0.16	0.15	0.17	0.2	0
m2	0.12	0.15	0.14	0.3	0
cred_gdp	0.10	0.15	0.10	0.0	0
hpi	0.19	0.13	0.12	0.0	0
cred	0.15	0.13	0.13	0.0	0
reer	0.14	0.11	0.12	0.1	0
bci	0.14	0.09	0.09	0.1	0
prfi_gdp	0.09	0.09	0.11	0.1	0
baa10ym	0.11	0.08	0.11	0.0	0
gs10	0.12	0.08	0.06	0.0	0
infl	0.09	0.07	0.07	0.0	0
cape	0.04	0.03	0.03	0.0	0
nfci	0.09	0.03	0.02	0.0	0
pip_inc	0.06	0.02	0.00	0.0	0
fedfunds	0.03	0.02	0.00	0.0	0
unrate	0.03	0.01	0.02	0.0	0
pdebt	0.02	0.01	0.00	0.0	0
pdebt_gdp	0.01	0.01	0.00	0.0	0
mortg_inc	0.02	0.00	0.00	0.0	0



Notes: The left panel shows the frequencies of inclusion in the best-performing (top 5%, top 1%, top 100, top 10 and single best) VAR models estimated on the reduced dataset (Table 2.1). The VARs include GDP growth plus all possible combinations of k predictors, for $k = 1, 2, 3, 4$ (17,901 models in total), and are ranked using the MSFE of iterated forecasts for the h-period-ahead GDP level (see also Table 2.17). The histogram in the right panel shows the frequencies of inclusion in the top 1% models.

Figure 2.6: Predictors' frequencies of inclusion in best-performing VAR models: $h = 20$

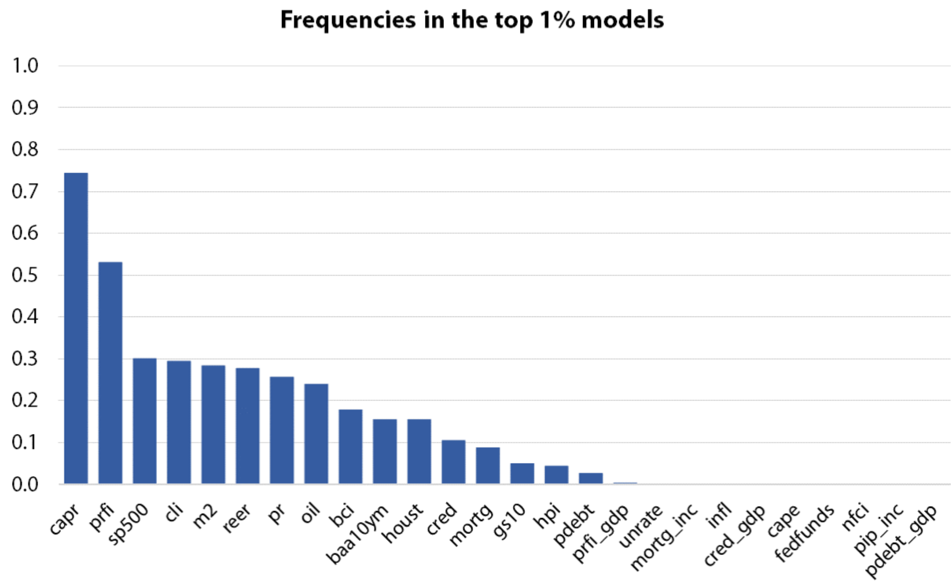
variable	frequencies top5%	top1%	top100	top10	top1
capr	0.60	0.74	0.70	0.9	1
cli	0.33	0.61	0.81	1.0	1
reer	0.22	0.27	0.25	0.2	0
pr	0.40	0.26	0.30	0.1	0
m2	0.18	0.25	0.27	0.6	1
prfi	0.20	0.25	0.22	0.0	0
houst	0.19	0.20	0.19	0.5	1
sp500	0.18	0.19	0.22	0.3	0
bci	0.20	0.19	0.11	0.0	0
oil	0.17	0.16	0.16	0.2	0
mortg	0.15	0.15	0.11	0.2	0
hpi	0.15	0.14	0.13	0.0	0
cred	0.16	0.13	0.12	0.0	0
baa10ym	0.15	0.10	0.11	0.0	0
gs10	0.13	0.06	0.04	0.0	0
prfi_gdp	0.08	0.05	0.05	0.0	0
mortg_inc	0.04	0.00	0.00	0.0	0
infl	0.02	0.00	0.00	0.0	0
unrate	0.06	0.00	0.00	0.0	0
cred_gdp	0.01	0.00	0.00	0.0	0
cape	0.04	0.00	0.00	0.0	0
fedfunds	0.02	0.00	0.00	0.0	0
nfci	0.01	0.00	0.00	0.0	0
pdebt_gdp	0.00	0.00	0.00	0.0	0
pdebt	0.06	0.00	0.00	0.0	0
pip_inc	0.00	0.00	0.00	0.0	0



Notes: The left panel shows the frequencies of inclusion in the best-performing (top 5%, top 1%, top 100, top 10 and single best) VAR models estimated on the reduced dataset (Table 2.1). The VARs include GDP growth plus all possible combinations of k predictors, for $k = 1, 2, 3, 4$ (17,901 models in total), and are ranked using the MSFE of iterated forecasts for the h-period-ahead GDP level (see also Table 2.17). The histogram in the right panel shows the frequencies of inclusion in the top 1% models.

Figure 2.7: Predictors' frequencies of inclusion in best-performing VAR models: $h = 28$

variable	frequencies				
	top5%	top1%	top100	top10	top1
capr	0.63	0.74	0.76	0.8	0
prfi	0.24	0.53	0.72	1.0	1
sp500	0.22	0.30	0.32	0.4	0
cli	0.22	0.30	0.28	0.2	0
m2	0.18	0.28	0.32	0.6	1
reer	0.22	0.28	0.27	0.2	1
pr	0.34	0.26	0.24	0.2	1
oil	0.18	0.24	0.21	0.0	0
bci	0.19	0.18	0.17	0.2	0
baa10ym	0.15	0.16	0.15	0.1	0
houst	0.17	0.16	0.12	0.0	0
cred	0.13	0.11	0.09	0.1	0
mortg	0.17	0.09	0.08	0.0	0
gs10	0.09	0.05	0.05	0.0	0
hpi	0.14	0.04	0.04	0.0	0
pdebt	0.14	0.03	0.00	0.0	0
prfi_gdp	0.13	0.01	0.00	0.0	0
unrate	0.09	0.00	0.00	0.0	0
mortg_inc	0.03	0.00	0.00	0.0	0
infl	0.02	0.00	0.00	0.0	0
cred_gdp	0.00	0.00	0.00	0.0	0
cape	0.01	0.00	0.00	0.0	0
fedfunds	0.03	0.00	0.00	0.0	0
nfci	0.03	0.00	0.00	0.0	0
pip_inc	0.00	0.00	0.00	0.0	0
pdebt_gdp	0.00	0.00	0.00	0.0	0



Notes: The left panel shows the frequencies of inclusion in the best-performing (top 5%, top 1%, top 100, top 10 and single best) VAR models estimated on the reduced dataset (Table 2.1). The VARs include GDP growth plus all possible combinations of k predictors, for $k = 1, 2, 3, 4$ (17,901 models in total), and are ranked using the MSFE of iterated forecasts for the h-period-ahead GDP level (see also Table 2.17). The histogram in the right panel shows the frequencies of inclusion in the top 1% models.

Table 2.18: Large Bayesian VAR: mean squared forecast error

predictors	λ	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
all	0.0010	4.7603	1.0798	1.3588	1.7666	1.8463
all	0.0025	2.0043	0.8035	1.0290	1.2266	1.2395
all	0.0050	1.8854	0.8409	1.0554	1.1268	1.1611
all	0.0075	1.8402	0.9053	1.1120	1.0863	1.1532
all	0.0100	4.7603	1.0798	1.3588	1.7666	1.8463
best 50 for $h = 1$	$\lambda^* \in [0.001, 0.1]$	1.0378	0.8953	1.1578	1.2574	1.1216
best 50 for $h = 4$	$\lambda^* \in [0.001, 0.1]$	1.0493	0.8246	0.9024	1.0850	0.9906
best 50 for $h = 12$	$\lambda^* \in [0.001, 0.1]$	1.0596	1.0238	0.8163	0.9203	0.7501
best 50 for $h = 20$	$\lambda^* \in [0.001, 0.1]$	0.9986	0.7456	0.7500	0.8177	0.6952
best 50 for $h = 28$	$\lambda^* \in [0.001, 0.1]$	1.0705	1.1472	1.1542	1.3102	0.9532
financial cycle (68)	$\lambda^* \in [0.001, 0.1]$	3.1697	0.8889	0.8772	0.8754	0.8428

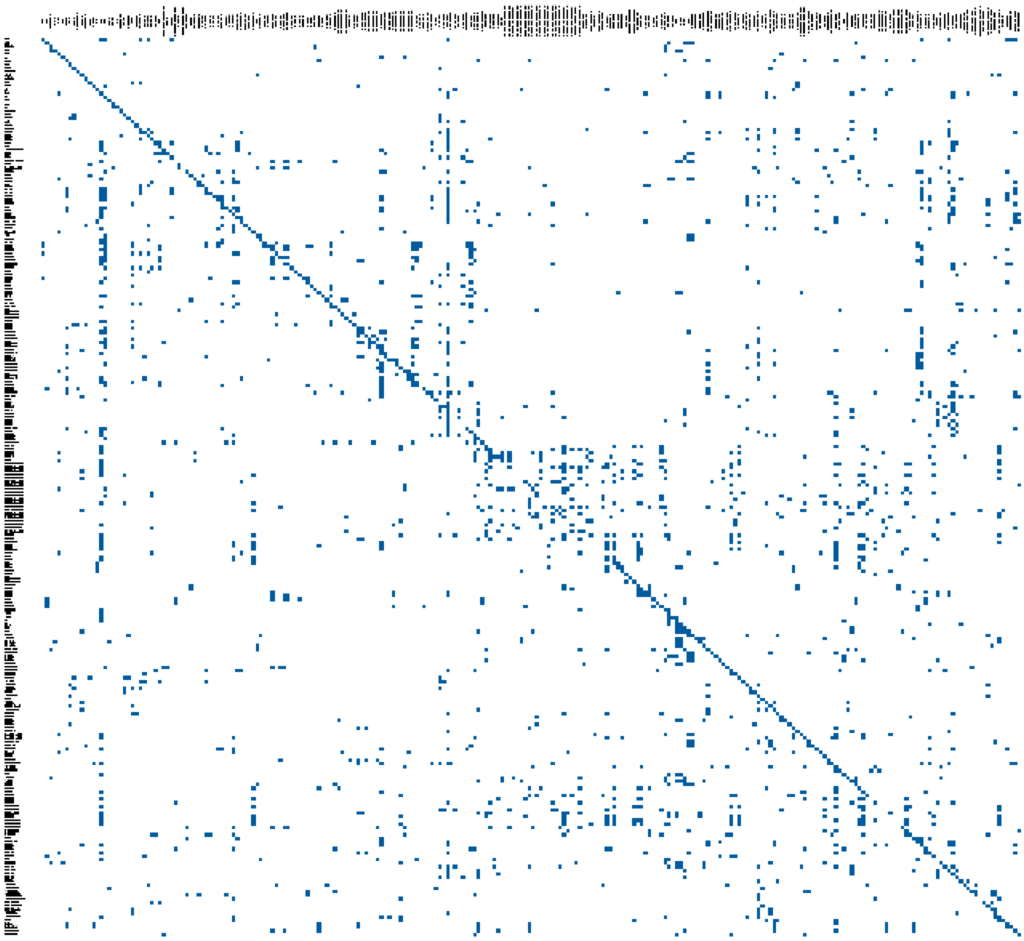
Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.16). λ is the inverse measure of tightness of the prior on VAR coefficients and λ^* denotes the value that maximizes the marginal likelihood, selected recursively. The tightness of the prior on the sum of coefficients is given by $\tau = 100\lambda$ for all models. On the selection of the best 50 predictors for different horizons, see section 2.5.2.1. See also note 21 on how financial-cycle indicators (68 variables) are identified.

Table 2.19: LASSO VAR on full dataset: mean squared forecast error

λ	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
0.00025	1.17	1.58	6.14	>10	>10
0.00050	1.32	0.88	0.80	0.76	0.65
0.00075	1.53	0.91	0.83	0.79	0.67
0.00100	1.79	0.94	0.86	0.81	0.69
0.00125	2.07	0.97	0.88	0.82	0.70
0.00150	2.42	0.99	0.90	0.83	0.71

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.16). λ denotes the penalty parameter in the LASSO VAR regression.

Table 2.20: LASSO VAR on full dataset: sparsity in 2016Q1 ($\lambda = 0.0005$)



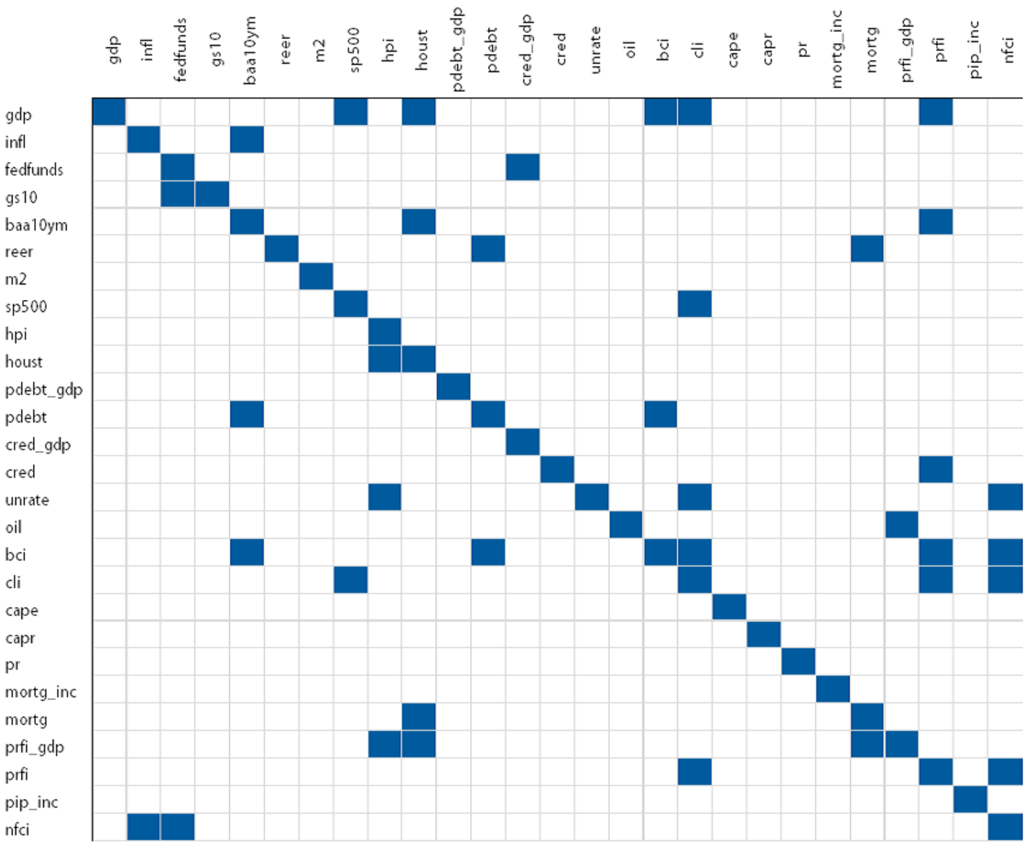
Notes: Each row represents an equation of the LASSO VAR, each column a right-hand-side variable. Colored cells identify non-zero coefficients in the model estimated in 2016Q1.

Table 2.21: LASSO VAR on reduced dataset: mean squared forecast error

λ	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
0.0010	1.2035	1.0259	>10	>10	>10
0.0025	1.3214	0.9665	0.8773	0.7984	0.6675
0.0050	1.4678	0.8834	0.8536	0.7962	0.6559
0.0075	1.6846	0.9142	0.8790	0.8232	0.6863
0.0100	1.9680	0.9562	0.8993	0.8436	0.7128
0.0125	2.2893	0.9877	0.9120	0.8535	0.7281

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.16). λ denotes the penalty parameter in the LASSO VAR regression.

Table 2.22: LASSO VAR on reduced dataset: sparsity in 2016Q1 ($\lambda = 0.005$)



Notes: Each row represents an equation of the LASSO VAR, each column a right-hand-side variable. Colored cells identify non-zero coefficients in the model estimated in 2016Q1.

Table 2.23: Post-LASSO VAR on reduced dataset: mean squared forecast error

λ	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
0.0025	1.3390	1.1249	0.9157	1.2095	1.9520
0.0050	1.2201	0.9159	0.8634	0.8830	0.7536
0.0075	1.2526	0.9596	0.9746	0.8646	0.6285
0.0100	1.2895	1.0221	1.0966	1.1087	0.7586
0.0125	1.2724	0.9803	1.0502	1.0061	0.7985

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.16). λ denotes the penalty parameter in the LASSO VAR regression.

Table 2.24: LASSO VAR using financial-cycle indicators: mean squared forecast error

λ	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
0.0025 (lag=1)	1.40	0.85	0.81	0.79	0.67
0.0025 (lag=5)	1.37	0.87	0.82	0.74	0.60
0.005 (lag=1)	2.37	0.98	0.89	0.83	0.71
0.005 (lag=5)	2.28	0.96	0.87	0.82	0.69

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.16). λ denotes the penalty parameter in the LASSO VAR regression. Only financial-cycle indicators (68 variables) are included in the LASSO VAR in addition to GDP (see note 21 in the text).

Table 2.25: VAR with principal components: mean squared forecast errors

n. PC	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
1	1.285	1.180	1.146	1.058	1.015
2	0.925	0.776	1.250	1.045	0.983
3	1.151	0.756	1.070	0.946	0.835
4	1.377	1.150	1.256	1.193	1.174

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.16). The first column indicates the number of principal components (PC) included in the VAR. The principal components are extracted from all predictors in the full dataset (Tables 2.1 and Tables 2.35-2.48).

Table 2.26: VAR with principal components of financial-cycle indicators: mean squared forecast errors

n. PC	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
1	1.099	1.040	1.117	1.211	1.180
2	1.151	0.915	0.884	0.876	0.796
3	1.079	0.986	0.792	0.800	0.773
4	0.986	1.013	1.093	1.278	1.786

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.16). The first column indicates the number of principal components (PC) included in the VAR. The principal components are extracted from the pool of financial-cycle indicators included in the dataset (see note 21 in the text).

Table 2.27: Combinations of direct (ARDL) forecasts: mean squared forecast errors

n. models	weights	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
258	equal	0.948	0.960	0.999	0.999	0.996
	bma	0.946	0.960	1.001	1.000	0.994
200	equal	0.926	0.917	0.961	0.967	0.958
	bma	0.922	0.914	0.960	0.965	0.953
100	equal	0.906	0.862	0.914	0.930	0.915
	bma	0.904	0.857	0.911	0.926	0.911
10	equal	0.817	0.578	0.632	0.687	0.655
	bma	0.817	0.578	0.627	0.690	0.662

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.15). The first column indicates the number M of single-predictor ARDL models that are combined. The total number of available models is 258 (see note 17 in the text). When $M < 258$, the models considered are those with the lowest individual MSFE. In the second column, "equal" denotes equal weighting, "bma" denotes Bayesian model averaging.

Table 2.28: Combinations of iterated (VAR) forecasts: mean squared forecast errors

n. models	weights	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
256	equal	0.960	0.896	0.927	0.941	0.919
	bma	0.956	0.893	0.924	0.941	0.917
200	equal	0.930	0.862	0.913	0.922	0.904
	bma	0.925	0.858	0.910	0.921	0.902
100	equal	0.858	0.794	0.834	0.852	0.836
	bma	0.853	0.791	0.832	0.851	0.834
10	equal	0.813	0.595	0.464	0.441	0.486
	bma	0.814	0.599	0.469	0.447	0.493

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.16). The first column indicates the number M of of single-predictor VAR models that are combined. The total number of available models is 256 (see notes 17 and 28 in the text). When $M < 256$, the models considered are those with the lowest individual MSFE. In the second column, "equal" denotes equal weighting, "bma" denotes Bayesian model averaging.

Table 2.29: Combinations of direct (ARDL) forecasts by financial-cycle predictors

weights	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
equal	0.875	0.868	0.982	0.958	0.926
bma	0.878	0.865	0.986	0.961	0.921

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.15). The combined models are single-predictor ARDL models using financial-cycle indicators (70 models). In the first column, "equal" denotes equal weighting, "bma" denotes Bayesian model averaging. See note 21 in the text.

Table 2.30: Combinations of iterated (VAR) forecasts by financial-cycle predictors

weights	$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
equal	0.893	0.798	0.827	0.826	0.828
bma	0.888	0.795	0.827	0.829	0.829

Notes: Out-of-sample MSFE for the h-quarter-ahead real GDP level, relative to the benchmark AR (Table 2.16). The combined models are single-predictor VAR models using financial-cycle indicators (70 models). In the first column, "equal" denotes equal weighting, "bma" denotes Bayesian model averaging. See note 21 in the text.

Table 2.31: Forecast encompassing and accuracy: ENC-F and MSE-F tests

		$h = 1$	$h = 4$	$h = 12$	$h = 20$	$h = 28$
capr	ENC-F	0.000	0.000	0.000	0.000	0.000
	MSE-F	0.182	0.003	0.007	0.001	0.000
NNBTILQ027SBDIx	ENC-F	0.005	0.000	0.000	0.001	0.009
	MSE-F	0.493	0.000	0.000	0.001	0.009
cred_gdp	ENC-F	0.000	0.004	0.013	0.024	0.020
	MSE-F	0.002	0.085	0.840	0.929	0.722

Notes: The table reports the p-values of the ENC-F test of forecast encompassing and the MSE-F test of equal forecast accuracy (Clark and McCracken 2001, 2013) for the h-period-ahead iterated forecasts of real GDP. Each test compares the bivariate VAR that includes GDP plus the predictor in the first column of the table with the benchmark AR. In the ENC-F test, the null hypothesis is that the AR model encompasses the VAR, while under the alternative hypothesis the candidate variable has predictive content. In the MSE-F test, the null hypothesis is that the AR and VAR models have equal MSFE, while under the alternative the VAR has a lower MSFE. For both tests, the p-values are computed using bootstrapped critical values (see Clark and McCracken 2013).

Table 2.32: Diebold-Mariano test: direct (ARDL) forecasts

	$h = 1$		$h = 4$		$h = 12$		$h = 20$		$h = 28$	
1	NNBTILQ027SBDIx	-	NNBTILQ027SBDIx	-	NNBTILQ027SBDIx	-	capr	-	capr	-
2	S_P_div_yield	0.47	capr	0.09	capr	0.02	NNBTILQ027SBDIx	0.22	pr	0.14
3	cape	0.43	AMDMUOx	0.07	cli	0.03	pr	0.12	NNBTILQ027SBDIx	0.17
4	LNS12032194	0.39	pr	0.04	AMDMUOx	0.08	cli	0.13	NASDAQCOM	0.15
5	AMDMUOx	0.33	cli	0.02	NWPIx	0.05	AHETPIx	0.13	EXCAUSx	0.15
6	IPCONGD	0.35	IPCONGD	0.11	cred_gdp	0.07	IPCONGD	0.14	baa10ym	0.15
7	TLBSHNOx	0.35	NWPIx	0.02	IPCONGD	0.11	reer	0.14	IPNCONGD	0.15
8	CUSR0000SAS	0.32	IPNCONGD	0.10	fedfunds	0.08	SLCEx	0.13	VXOCLSX	0.16
9	PCECC96	0.34	PCECC96	0.09	GS1TB3Mx	0.11	LNS13023557	0.14	IPCONGD	0.15
10	CMRMTSPLx	0.31	prfi_gdp	0.02	IPNCONGD	0.11	ANDENOx	0.13	WPU0531	0.15
11	TABSHNOx	0.33	prfi	0.02	gs10	0.10	IPNCONGD	0.14	reer	0.14
12	TNWBSHNOx	0.33	PERMITS	0.06	PERMITMW	0.07	AMDMUOx	0.14	HWIx	0.14
13	IPMANSICS	0.28	CUSR0000SAS	0.12	PERMIT	0.08	PERMITMW	0.13	B021RE1Q156NBEA	0.14
14	HNOREMQ027Sx	0.27	S_P_div_yield	0.12	CUSR0000SAS	0.11	TNWBSNNBBDIx	0.13	cli	0.14
15	HOUSTMW	0.22	REVOLSLx	0.12	bci	0.10	baa10ym	0.14	TWEXMMTH	0.13
16	prfi_gdp	0.23	cape	0.09	PCECC96	0.09	LNS14000026	0.13	TNWBSNNBBDIx	0.15
17	NIKKEI225	0.25	PERMIT	0.04	DFXARG3Q086SBEA	0.10	unrate	0.14	hpi	0.16
18	pr	0.21	PERMITW	0.03	PCESVx	0.10	IPDCONGD	0.13	LNS13023557	0.15
19	IPNCONGD	0.21	HOUSTW	0.03	USMINE	0.09	UEMP27OV	0.14	prfi_gdp	0.15
20	mortg_inc	0.21	IPB51220SQ	0.13	PERMITW	0.09	B021RE1Q156NBEA	0.14	TNWMVBSNNCBBDIx	0.16
21	CPILFESL	0.22	S_P_500	0.09	PERMITNE	0.10	HOUST	0.14	LNS13023705	0.15
22	DTRSRG3Q086SBEA	0.22	HOUSTS	0.04	HOUST	0.08	HWIx	0.12	unrate	0.15
23	PCNDx	0.21	NASDAQCOM	0.09	fedfunds	0.09	HOUSTMW	0.13	PERMITS	0.15
24	IPDCONGD	0.21	houst	0.03	HOUSTW	0.08	PERMIT	0.14	AHETPIx	0.14
25	cred_gdp	0.21	hpi	0.02	SLCEx	0.11	UEMP15T26	0.14	PERMIT	0.15
26	PERMITMW	0.21	PRFIx	0.02	CPILFESL	0.11	USMINE	0.13	HOUSTS	0.16
27	IPB51110SQ	0.22	S_P_indust	0.10	UEMP15T26	0.11	DFSARG3Q086SBEA	0.14	unrate	0.15
28	S_P_indust	0.25	CPILFESL	0.12	UEMP5TO14	0.11	PCECC96	0.14	ANDENOx	0.15
29	IPDBS	0.21	GS1TB3Mx	0.08	ISRATIOx	0.09	cred_gdp	0.06	OUTBS	0.14
30	S_P_500	0.23	PCNDx	0.09	EXCAUSx	0.10	LNS13023705	0.14	IPDCONGD	0.15

Notes: P-values of the Diebold-Mariano test of equal MSFE comparing the GDP forecasts made by the best predictor and those made by each of the following 29 predictors in the ranking of Table 2.13. Under the alternative hypothesis, forecasts by the best predictor have lower MSFE.

Table 2.33: Diebold-Mariano test: iterated (VAR) forecasts

	$h = 1$		$h = 4$		$h = 12$		$h = 20$		$h = 28$	
1	cred_gdp	-	NNBTILQ027SBDIx	-	NNBTILQ027SBDIx	-	capr	-	capr	-
2	S_P_div_yield	0.43	capr	0.06	capr	0.12	NNBTILQ027SBDIx	0.42	pr	0.20
3	CUMFNS	0.37	CPF3MTB3Mx	0.11	cli	0.06	pr	0.18	prfi_gdp	0.20
4	CPILFESL	0.38	prfi_gdp	0.14	AMDMUOx	0.07	AMDMUOx	0.17	NNBTILQ027SBDIx	0.17
5	TLBSHNOx	0.27	AMDMUOx	0.03	pr	0.11	NWPIx	0.16	HNOREMQ027Sx	0.20
6	DFSARG3Q086SBEA	0.31	S_P_div_yield	0.13	prfi_gdp	0.06	prfi_gdp	0.13	cli	0.18
7	BAA	0.32	pr	0.03	AAA	0.10	cli	0.18	AMDMUOx	0.15
8	CUSR0000SAS	0.26	TLBSNNCBBDIx	0.10	USMINE	0.03	HNOREMQ027Sx	0.15	REALLNx	0.21
9	CMRMTSPLx	0.25	hpi	0.03	PRFIx	0.06	TNWBSNNBBDIx	0.14	USGOVT	0.19
10	TCU	0.24	nfci	0.09	NWPIx	0.06	TNWMVBSNNCBBDIx	0.15	PRFIx	0.25
11	FPIx	0.26	EXPGSC1	0.10	GS10TB3Mx	0.09	GS10TB3Mx	0.14	NWPIx	0.10
12	cape	0.18	sp500	0.09	USGOVT	0.12	CNCFx	0.15	GS10TB3Mx	0.15
13	LNS12032194	0.20	PCNDx	0.08	CPF3MTB3Mx	0.10	PRFIx	0.17	CP3M	0.17
14	sp500	0.20	CUSR0000SAS	0.09	nfci	0.09	USGOVT	0.16	TB3MS	0.17
15	pr	0.16	cape	0.10	gs10	0.09	CPILFESL	0.15	PERMITS	0.19
16	PERMITMW	0.17	S_P_500	0.09	houst	0.10	TLBSHNOx	0.15	TB6MS	0.15
17	S_P_500	0.16	cli	0.01	T5YFFM	0.08	REALLNx	0.15	fedfunds	0.17
18	CES0600000007	0.12	mortg	0.02	BAA	0.09	houst	0.15	CPIMEDSL	0.19
19	AAA	0.17	S_P_indust	0.10	GS5	0.09	nfci	0.13	houst	0.16
20	bci	0.17	PERMITMW	0.05	hpi	0.06	PERMITNE	0.17	GS1	0.14
21	GPDIC1	0.11	OILPRICEx	0.09	CP3M	0.09	S_P_div_yield	0.15	PERMITNE	0.22
22	nfci	0.16	OUTBS	0.09	PERMITNE	0.11	PERMITS	0.15	sp500	0.13
23	PERMITS	0.14	CPILFESL	0.09	HOUSTNE	0.13	GS1TB3Mx	0.15	TLBSHNOx	0.15
24	UNLPNBS	0.16	DFSARG3Q086SBEA	0.08	S_P_div_yield	0.10	CES0600000008	0.14	GPDIC1	0.15
25	DPIC96	0.12	oil	0.08	CPILFESL	0.11	T5YFFM	0.15	BAA	0.12
26	cred	0.08	MORTG10YRx	0.10	CES0600000008	0.10	S_P_500	0.15	CES0600000008	0.15
27	OILPRICEx	0.14	PERMITS	0.05	sp500	0.10	sp500	0.14	S_P_500	0.13
28	CUSR0000SA0L5	0.15	NWPIx	0.05	MZMREALx	0.12	UEMP27OV	0.15	S_P_indust	0.13
29	S_P_indust	0.12	CES0600000008	0.11	AAAFFM	0.06	AWHNONAG	0.15	NNBTASQ027Sx	0.12
30	oil	0.14	GS1TB3Mx	0.09	HOUSTMW	0.07	GDPCTPI	0.14	nfci	0.09

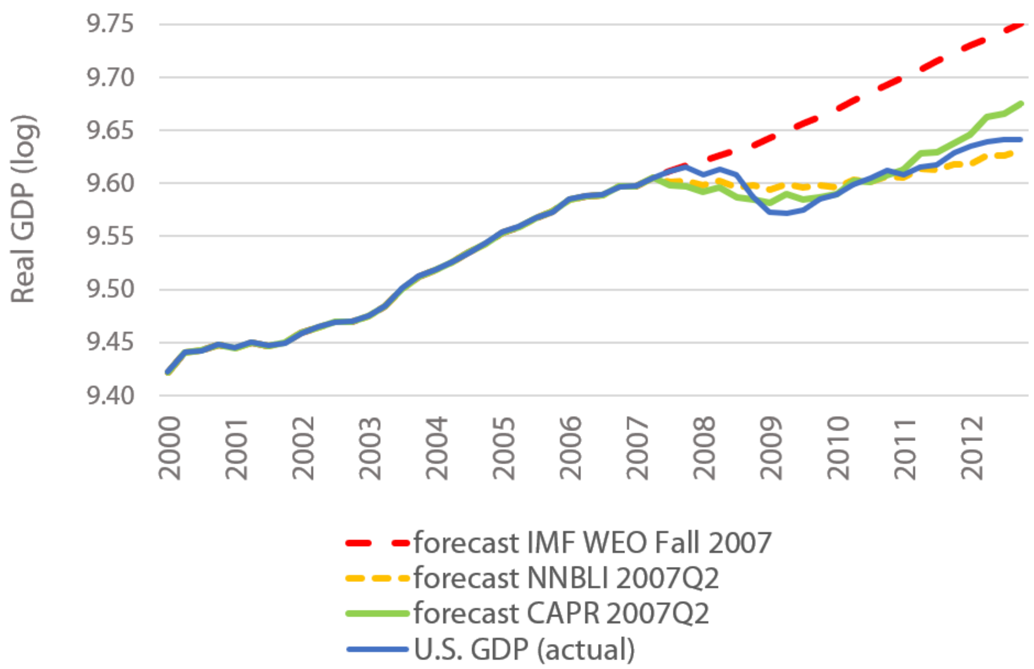
Notes: P-values of the Diebold-Mariano test of equal MSFE comparing the GDP forecasts made by the best predictor and those made by each of the following 29 predictors in the ranking of Table 2.14. Under the alternative hypothesis, forecasts by the best predictor have lower MSFE.

Table 2.34: IMF forecasts: mean squared forecast error

	$h = 4$	$h = 12$	$h = 20$
RMSFE	0.015	0.040	0.065
relative MSFE direct	0.723	0.819	1.082
relative MSFE indirect	0.715	0.877	1.070

Notes: The results are based on forecasts published in the IMF World Economic Outlook from 1990 to 2017 (Spring issues, see section 2.5.5 in the text for details). The benchmarks for the relative MSFE are the direct and iterated AR forecasts (Tables 2.15 and 2.16, respectively).

Figure 2.8: Forecasting the crisis: a comparison



Notes: The labels "forecast NNBLI 2007Q2" and "forecast CAPR 2007Q2" indicate the iterated forecasts from bivariate VAR models using NNBLI and CAPR, respectively, estimated on data up to 2007Q2.

Table 2.35: FRED-QD. Group 1: NIPA

A014RE1Q156NBEA	Shares of gross domestic product: Gross private domestic investment: Change in private inventories (Percent)
A823RL1Q225SBEA	Real Government Consumption Expenditures and Gross Investment: Federal (Percent Change from Preceding Period)
B020RE1Q156NBEA	Shares of gross domestic product: Exports of goods and services (Percent)
	Shares of gross domestic product: Imports of goods and services (Percent)
DPIC96	Real Disposable Personal Income (Billions of Chained 2009 Dollars)
EXPGSC1	Real Exports of Goods & Services, 3 Decimal (Billions of Chained 2009 Dollars)
FGRECPTx	Real Federal Government Current Receipts (Billions of Chained 2009 Dollars), deflated using PCE
FPIx	Real private fixed investment (Billions of Chained 2009 Dollars), deflated using PCE
GCEC1	Real Government Consumption Expenditures & Gross Investment (Billions of Chained 2009 Dollars)
GDPC1	Real Gross Domestic Product, 3 Decimal (Billions of Chained 2009 Dollars)
GPDIC1	Real Gross Private Domestic Investment, 3 decimal (Billions of Chained 2009 Dollars)
IMPGSC1	Real Imports of Goods & Services, 3 Decimal (Billions of Chained 2009 Dollars)
OUTBS	Business Sector: Real Output (Index 2009=100)
OUTMS	Manufacturing Sector: Real Output (Index 2009=100)
OUTNFB	Nonfarm Business Sector: Real Output (Index 2009=100)
PCDGx	Real personal consumption expenditures: Durable goods (Billions of Chained 2009 Dollars), deflated using PCE
PCECC96	Real Personal Consumption Expenditures (Billions of Chained 2009 Dollars)
PCESVx	Real Personal Consumption Expenditures: Services (Billions of 2009 Dollars), deflated using PCE
PCNDx	Real Personal Consumption Expenditures: Nondurable Goods (Billions of 2009 Dollars), deflated using PCE
PNFIx	Real private fixed investment: Nonresidential (Billions of Chained 2009 Dollars), deflated using PCE
PRFIx	Real private fixed investment: Residential (Billions of Chained 2009 Dollars), deflated using PCE
SLCEx	Real government state and local consumption expenditures (Billions of Chained 2009 Dollars), deflated using PCE
Y033RC1Q027SBEAx	Real Gross Private Domestic Investment: Fixed Investment: Nonresidential:
Y033RC1Q027SBEAx	Equipment (Billions of Chained 2009 Dollars), deflated using PCE

Table 2.36: FRED-QD. Group 2: Industrial Production

CUMFNS	Capacity Utilization: Manufacturing (SIC) (Percent of Capacity)
INDPRO	Industrial Production Index (Index 2012=100)
IPB51110SQ	Industrial Production: Durable Goods: Automotive products (Index 2012=100)
IPB51220SQ	Industrial Production: Consumer energy products (Index 2012=100)
IPB51222S	Industrial Production: Residential Utilities (Index 2012=100)
IPBUSEQ	Industrial Production: Business Equipment (Index 2012=100)
IPCONGD	Industrial Production: Consumer Goods (Index 2012=100)
IPDCONGD	Industrial Production: Durable Consumer Goods (Index 2012=100)
IPDMAT	Industrial Production: Durable Materials (Index 2012=100)
IPFINAL	Industrial Production: Final Products (Market Group) (Index 2012=100)
IPFUELS	Industrial Production: Fuels (Index 2012=100)
IPMANSICS	Industrial Production: Manufacturing (SIC) (Index 2012=100)
IPMAT	Industrial Production: Materials (Index 2012=100)
IPNCONGD	Industrial Production: Nondurable Consumer Goods (Index 2012=100)
IPNMAT	IP:Nondur gds materials Industrial Production: Nondurable Materials (Index 2012=100)
TCU	Capacity Utilization: Total Industry (Percent of Capacity)

Table 2.37: FRED-QD. Group 3: Employment and Unemployment

AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing (Hours)
AWHNONAG	Average Weekly Hours Of Production And Nonsupervisory Employees: Total private (Hours)
AWOTMAN	AWH Overtime Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing (Hours)
CE16OV	Civilian Employment (Thousands of Persons)
CES0600000007	Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing
CES9091000001	All Employees: Government: Federal (Thousands of Persons)
CES9092000001	All Employees: Government: State Government (Thousands of Persons)
CES9093000001	All Employees: Government: Local Government (Thousands of Persons)
CIVPART	Civilian Labor Force Participation Rate (Percent)
CLAIMSx	Initial Claims
DMANEMP	All Employees: Durable goods (Thousands of Persons)
HOABS	EmpHrs:Bus Sec Business Sector: Hours of All Persons (Index 2009=100)
HOAMS	Manufacturing Sector: Hours of All Persons (Index 2009=100)
HOANBS	Nonfarm Business Sector: Hours of All Persons (Index 2009=100)

HWIURATIOx	Ratio of Help Wanted/No. Unemployed
HWIx	Help-Wanted Index
LNS12032194	Employment Level - Part-Time for Economic Reasons, All Industries (Thousands of Persons)
LNS13023557	Unemployment Level - Reentrants to Labor Force (Thousands of Persons)
LNS13023569	Unemployment Level - New Entrants (Thousands of Persons)
LNS13023621	Unemployment Level - Job Losers (Thousands of Persons)
LNS13023705	Unemployment Level - Job Leavers (Thousands of Persons)
LNS14000012	Unemployment Rate - 16 to 19 years (Percent)
LNS14000025	Unemployment Rate - 20 years and over, Men (Percent)
LNS14000026	Unemployment Rate - 20 years and over, Women (Percent)
MANEMP	All Employees: Manufacturing (Thousands of Persons)
NDMANEMP	All Employees: Nondurable goods (Thousands of Persons)
PAYEMS	All Employees: Total nonfarm (Thousands of Persons)
SRVPRD	All Employees: Service-Providing Industries (Thousands of Persons)
UEMP15T26	Number of Civilians Unemployed for 15 to 26 Weeks (Thousands of Persons)
UEMP27OV	Number of Civilians Unemployed for 27 Weeks and Over (Thousands of Persons)
UEMP5TO14	Number of Civilians Unemployed for 5 to 14 Weeks (Thousands of Persons)
UEMPLT5	Number of Civilians Unemployed - Less Than 5 Weeks (Thousands of Persons)
UEMPMEAN	Average (Mean) Duration of Unemployment (Weeks)
UNRATE	Civilian Unemployment Rate (Percent)
UNRATELTx	Unemployment Rate for more than 27 weeks (Percent)
UNRATESTx	Unemployment Rate less than 27 weeks (Percent)
USCONS	All Employees: Construction (Thousands of Persons)
USEHS	All Employees: Education & Health Services (Thousands of Persons)
USFIRE	All Employees: Financial Activities (Thousands of Persons)
USGOOD	All Employees: Goods-Producing Industries (Thousands of Persons)
USGOVT	All Employees: Government (Thousands of Persons)
USINFO	All Employees: Information Services (Thousands of Persons)
USLAH	All Employees: Leisure & Hospitality (Thousands of Persons)
USMINE	All Employees: Mining and logging (Thousands of Persons)
USPBS	All Employees: Professional & Business Services (Thousands of Persons)
USPRIV	All Employees: Total Private Industries (Thousands of Persons)
USSERV	All Employees: Other Services (Thousands of Persons)
USTPU	All Employees: Trade, Transportation & Utilities (Thousands of Persons)
USTRADE	All Employees: Retail Trade (Thousands of Persons)
USWTRADE	All Employees: Wholesale Trade (Thousands of Persons)

Table 2.38: FRED-QD. Group 4: Housing

HOUST	Housing Starts: Total: New Privately Owned Housing Units Started (Thousands of Units)
HOUST5F	Privately Owned Housing Starts: 5-Unit Structures or More (Thousands of Units)
HOUSTMW	Housing Starts in Midwest Census Region (Thousands of Units)
HOUSTNE	Housing Starts in Northeast Census Region (Thousands of Units)
HOUSTS	Housing Starts in South Census Region (Thousands of Units)
HOUSTW	Housing Starts in West Census Region (Thousands of Units)
PERMIT	New Private Housing Units Authorized by Building Permits (Thousands of Units)
PERMITMW	New Private Housing Units Authorized by Building Permits in the Midwest Census Region (Thousands, SAAR)
PERMITNE	New Private Housing Units Authorized by Building Permits in the Northeast Census Region (Thousands, SAAR)
PERMITS	New Private Housing Units Authorized by Building Permits in the South Census Region (Thousands, SAAR)
PERMITW	New Private Housing Units Authorized by Building Permits in the West Census Region (Thousands, SAAR)
SPCS10RSA	S&P/Case-Shiller 10-City Composite Home Price Index (Index January 2000 = 100)
SPCS20RSA	S&P/Case-Shiller 20-City Composite Home Price Index (Index January 2000 = 100)
USSTHPI	All-Transactions House Price Index for the United States (Index 1980 Q1=100)

Table 2.39: FRED-QD. Group 5: Inventories, Orders, and Sales

ACOGNOx	Orders(ConsGoods/Mat.) Real Value of Manufacturers' New Orders for Consumer Goods Industries (Million of 2009 Dollars), deflated by Core PCE
AMDMNOx	Real Manufacturers' New Orders: Durable Goods (Millions of 2009 Dollars), deflated by Core PCE
AMDMUOx	Real Value of Manufacturers' Unfilled Orders for Durable Goods Industries (Million of 2009 Dollars), deflated by Core PCE
ANDENOx	Real Value of Manufacturers' New Orders for Capital Goods: Nondefense Capital Goods Industries (Million of 2009 Dollars), deflated by Core PCE
BUSINVx	Total Business Inventories (Millions of Dollars)
CMRMTSPLx	Real Manufacturing and Trade Industries Sales (Millions of Chained 2009 Dollars)
INVCQRMTSPL	Real Manufacturing and Trade Inventories (Millions of 2009 Dollars)
ISRATIOx	Total Business: Inventories to Sales Ratio
RSAFSx	Real Retail and Food Services Sales (Millions of Chained 2009 Dollars), deflated by Core PCE

Table 2.40: FRED-QD. Group 6: Prices

CPIAPPSL	Consumer Price Index for All Urban Consumers: Apparel (Index 1982-84=100)
CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items (Index 1982-84=100)
CPILFESL	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy (Index 1982-84=100)
CPIMEDSL	Consumer Price Index for All Urban Consumers: Medical Care (Index 1982-84=100)
CPITRNSL	Consumer Price Index for All Urban Consumers: Transportation (Index 1982-84=100)
CPIULFSL	Consumer Price Index for All Urban Consumers: All Items Less Food (Index 1982-84=100)
CUSR0000SA0L2	Consumer Price Index for All Urban Consumers: All items less shelter (Index 1982-84=100)
CUSR0000SA0L5	Consumer Price Index for All Urban Consumers: All items less medical care (Index 1982-84=100)
CUSR0000SAC	Consumer Price Index for All Urban Consumers: Commodities (Index 1982-84=100)
CUSR0000SAD	Consumer Price Index for All Urban Consumers: Durables (Index 1982-84=100)
CUSR0000SAS	Consumer Price Index for All Urban Consumers: Services (Index 1982-84=100)
CUSR0000SEHC	CPI for All Urban Consumers: Owners' equivalent rent of residences (Index Dec 1982=100)
DCLORG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Clothing and footwear (chain-type price index)
DDURRG3Q086SBEA	Personal consumption expenditures: Durable goods (chain-type price index)
DFDHRG3Q086SBEA	Personal consumption expenditures: Durable goods: Furnishings and durable household equipment (chain-type price index)
DFSARG3Q086SBEA	Personal consumption expenditures: Services: Food services and accommodations (chain-type price index)
DFXARG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Food and beverages purchased for off-premises consumption (chain-type price index)
DGDSRG3Q086SBEA	Personal consumption expenditures: Goods (chain-type price index)
DGOERG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Gasoline and other energy goods (chain-type price index)

DHCERG3Q086SBEA	Personal consumption expenditures: Services: Household consumption expenditures (chain-type price index)
DHLCRG3Q086SBEA	Personal consumption expenditures: Services: Health care (chain-type price index)
DHUTRG3Q086SBEA	Personal consumption expenditures: Services: Housing and utilities (chain-type price index)
DIFSRG3Q086SBEA	Personal consumption expenditures: Financial services and insurance (chain-type price index)
DMOTRG3Q086SBEA	Personal consumption expenditures: Durable goods: Motor vehicles and parts (chain-type price index)
DNDGRG3Q086SBEA	Personal consumption expenditures: Nondurable goods (chain-type price index)
DODGRG3Q086SBEA	Personal consumption expenditures: Durable goods: Other durable goods (chain-type price index)
DONGRG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Other nondurable goods (chain-type price index)
DOTSRG3Q086SBEA	Personal consumption expenditures: Other services (chain-type price index)
DRCARG3Q086SBEA	Personal consumption expenditures: Recreation services (chain-type price index)
DREQRG3Q086SBEA	Personal consumption expenditures: Durable goods: Recreational goods and vehicles (chain-type price index)
DSERRG3Q086SBEA	Personal consumption expenditures: Services (chain-type price index)
DTRSRG3Q086SBEA	Personal consumption expenditures: Transportation services (chain-type price index)
GDPCTPI	Gross Domestic Product: Chain-type Price Index (Index 2009=100)
GPDICTPI	Gross Private Domestic Investment: Chain-type Price Index (Index 2009=100)
IPDBS	Business Sector: Implicit Price Deflator (Index 2009=100)
OILPRICEx	Real Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma (2009 Dollars per Barrel), deflated by Core PCE 2009=100)
PCECTPI	Personal Consumption Expenditures: Chain-type Price Index (Index
PCEPILFE	Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index) (Index 2009=100)
PPIACO	Producer Price Index for All Commodities (Index 1982=100)
PPICMM	Producer Price Index: Commodities: Metals and metal products: Primary nonferrous metals (Index 1982=100)
PPIIDC	Producer Price Index by Commodity Industrial Commodities (Index

	1982=100)
WPSFD4111	Producer Price Index by Commodity for Finished Consumer Foods (Index 1982=100)
WPSFD49207	Producer Price Index by Commodity for Finished Goods (Index 1982=100)
WPSFD49502	Producer Price Index by Commodity for Finished Consumer Goods (Index 1982=100)
WPSID61	Producer Price Index by Commodity Intermediate Materials: Supplies & Components (Index 1982=100)
WPSID62	Producer Price Index: Crude Materials for Further Processing (Index 1982=100)
WPU0531	Producer Price Index by Commodity for Fuels and Related Products and Power: Natural Gas (Index 1982=100)
WPU0561	Producer Price Index by Commodity for Fuels and Related Products and Power: Crude Petroleum (Domestic Production) (Index 1982=100)

Table 2.41: FRED-QD. Group 7: Earnings and Productivity

AHETPIx	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private (2009 Dollars per Hour), deflated by Core PCE
CES0600000008	Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing (Dollars per Hour)
CES2000000008x	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Construction (2009 Dollars per Hour), deflated by Core PCE
CES3000000008x	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing (2009 Dollars per Hour), deflated by Core PCE
COMPRMS	Manufacturing Sector: Real Compensation Per Hour (Index 2009=100)
COMPRNFB	Nonfarm Business Sector: Real Compensation Per Hour (Index 2009=100)
OPHMFG	Manufacturing Sector: Real Output Per Hour of All Persons (Index 2009=100)
OPHNFB	Nonfarm Business Sector: Real Output Per Hour of All Persons (Index 2009=100)
OPHPBS	Business Sector: Real Output Per Hour of All Persons (Index 2009=100)
RCPHBS	Business Sector: Real Compensation Per Hour (Index 2009=100)
ULCBS	Business Sector: Unit Labor Cost (Index 2009=100)
ULCMFG	Manufacturing Sector: Unit Labor Cost (Index 2009=100)
ULCNFB	Nonfarm Business Sector: Unit Labor Cost (Index 2009=100)
UNLPNBS	Nonfarm Business Sector: Unit Nonlabor Payments (Index 2009=100)

Table 2.42: FRED-QD. Group 8: Interest Rates

AAA	Moody's Seasoned Aaa Corporate Bond Yield © (Percent)
AAAFFM	Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate
BAA	Moody's Seasoned Baa Corporate Bond Yield © (Percent)
BAA10YM	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity (Percent)
COMPAPFF	3-Month Commercial Paper Minus Federal Funds Rate
CP3M	3-Month AA Financial Commercial Paper Rate
CPF3MTB3Mx	3-Month Commercial Paper Minus 3-Month Treasury Bill, secondary market (Percent)
FEDFUNDS	Effective Federal Funds Rate (Percent)
GS1	1-Year Treasury Constant Maturity Rate (Percent)
GS10	10-Year Treasury Constant Maturity Rate (Percent)
GS10TB3Mx	10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)
GS1TB3Mx	1-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)
GS5	5-Year Treasury Constant Maturity Rate
MORTG10YRx	30-Year Conventional Mortgage Rate Relative to 10-Year Treasury Constant Maturity (Percent)
MORTGAGE30US	30-Year Conventional Mortgage Rate© (Percent)
T5YFFM	5-Year Treasury Constant Maturity Minus Federal Funds Rate
TB3MS	3-Month Treasury Bill: Secondary Market Rate (Percent)
TB3SMFFM	3-Month Treasury Constant Maturity Minus Federal Funds Rate
TB6M3Mx	6-Month Treasury Bill Minus 3-Month Treasury Bill, secondary market (Percent)
TB6MS	6-Month Treasury Bill: Secondary Market Rate (Percent)

Table 2.43: FRED-QD. Group 9: Money and Credit

AMBSLREALx	St. Louis Adjusted Monetary Base (Billions of 1982-84 Dollars), deflated by CPI
BUSLOANSx	Real Commercial and Industrial Loans, All Commercial Banks (Billions of 2009 U.S. Dollars), deflated by Core PCE
CONSUMERx	Real Consumer Loans at All Commercial Banks (Billions of 2009 U.S. Dollars), deflated by Core PCE
DRIWCIL	FRB Senior Loans Officer Opions. Net Percentage of Domestic Respondents Reporting Increased Willingness to Make Consumer Installment Loans
DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding Owned by Finance Companies (Millions of Dollars)
DTCTHFNM	Total Consumer Loans and Leases Outstanding Owned and Securitized by Finance Companies (Millions of Dollars)
IMFSLx	Real Institutional Money Funds (Billions of 2009 Dollars), deflated by Core PCE
INVEST	Securities in Bank Credit at All Commercial Banks (Billions of Dollars)
M1REALx	Real M1 Money Stock (Billions of 1982-84 Dollars), deflated by CPI
M2REALx	Real M2 Money Stock (Billions of 1982-84 Dollars), deflated by CPI
MZMREALx	Real MZM Money Stock (Billions of 1982-84 Dollars), deflated by CPI
NONBORRES	Reserves Of Depository Institutions, Nonborrowed (Millions of Dollars)
NONREVSLx	Total Real Nonrevolving Credit Owned and Securitized, Outstanding (Billions of Dollars), deflated by Core PCE
REALLNx	Real Real Estate Loans, All Commercial Banks (Billions of 2009 U.S. Dollars), deflated by Core PCE
REVOLSLx	Total Real Revolving Credit Owned and Securitized, Outstanding (Billions of 2009 Dollars), deflated by Core PCE
TOTALSLx	Total Consumer Credit Outstanding, deflated by Core PCE
TOTRESNS	Total Reserves of Depository Institutions (Billions of Dollars)

Table 2.44: FRED-QD. Group 10: Household Balance Sheets

CONSPIx	Nonrevolving consumer credit to Personal Income
HNOREMQ027Sx	Real Real Estate Assets of Households and Nonprofit Organizations (Billions of 2009 Dollars), deflated by Core PCE
LIABPIx	Liabilities of Households and Nonprofit Organizations Relative to Personal Disposable Income (Percent)
NWPIx	Net Worth of Households and Nonprofit Organizations Relative to Disposable Personal Income (Percent)
TABSHNOx	Real Total Assets of Households and Nonprofit Organizations (Billions of 2009 Dollars), deflated by Core PCE
TARESAx	Real HHW:TA_RESA Real Assets of Households and Nonprofit Organizations excluding Real Estate Assets (Billions of 2009 Dollars), deflated by Core PCE
TFAABSHNOx	Real Total Financial Assets of Households and Nonprofit Organizations (Billions of 2009 Dollars), deflated by Core PCE
TLBSHNOx	Real Total Liabilities of Households and Nonprofit Organizations (Billions of 2009 Dollars), deflated by Core PCE
TNWBSHNOx	Real Net Worth of Households and Nonprofit Organizations (Billions of 2009 Dollars), deflated by Core PCE

Table 2.45: FRED-QD. Group 11: Exchange Rates

EXCAUSx	EX rate:Canada Canada / U.S. Foreign Exchange Rate
EXJPUSx	Japan / U.S. Foreign Exchange Rate
EXSZUSx	Switzerland / U.S. Foreign Exchange Rate
EXUSEU	U.S. / Euro Foreign Exchange Rate (U.S. Dollars to One Euro)
EXUSUKx	U.S. / U.K. Foreign Exchange Rate
TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies (Index March 1973=100)

Table 2.46: FRED-QD. Group 12: Other

UMCSENTx	Cons. Expectations University of Michigan: Consumer Sentiment (Index 1st Quarter 1966=100)
USEPUINDXM	Economic Policy Uncertainty Index for United States

Table 2.47: FRED-QD. Group 13: Stock Markets

NASDAQCOM	NASDAQ Composite (Index Feb 5, 1971=100)
NIKKEI225	Nikkei Stock Average
S&P 500	S&P's Common Stock Price Index: Composite
S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio
S&P: div yield	S&P's Composite Common Stock: Dividend Yield
S&P: indust	S&P's Common Stock Price Index: Industrials
VXOCLSX	CBOE S&P 100 Volatility Index: VXO

Table 2.48: FRED-QD. Group 14: Non-Household Balance Sheets

CNCFx	Real Disposable Business Income, Billions of 2009 Dollars (Corporate cash flow with IVA minus taxes on corporate income, deflated by Implicit Price Deflator for Business Sector IPDBS)
GFDEBTNx	Real Federal Debt: Total Public Debt (Millions of 2009 Dollars), deflated by PCE
GFDEGDQ188S	Federal Debt: Total Public Debt as Percent of GDP (Percent)
NNBTASQ027Sx	Real Nonfinancial Noncorporate Business Sector Assets (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
NNBTILQ027SBDIx	Nonfinancial Noncorporate Business Sector Liabilities to Disposable Business Income (Percent)
NNBTILQ027Sx	Real Nonfinancial Noncorporate Business Sector Liabilities (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
TLBSNNCBBDIx	Nonfinancial Corporate Business Sector Liabilities to Disposable Business Income (Percent)
TLBSNNCBx	Real Nonfinancial Corporate Business Sector Liabilities (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
TNWBSNNBBDIx	Nonfinancial Noncorporate Business Sector Net Worth to Disposable Business Income (Percent)
TNWBSNNBx	Real Nonfinancial Noncorporate Business Sector Net Worth (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
TNWMVBSNNCBBDIx	Nonfinancial Corporate Business Sector Net Worth to Disposable Business Income (Percent)
TNWMVBSNNCBx	Real Nonfinancial Corporate Business Sector Net Worth (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
TTAABSNNCBx	Real Nonfinancial Corporate Business Sector Assets (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS

Chapter 3

Optimal Regime-Switching Density Forecasts

Abstract

This paper investigates an approach for enhancing density forecasts of non-normal macroeconomic variables using Bayesian Markov-switching models. Alternative views on the regime-switching behavior of the economy are pooled to form flexible composite forecasts, which are optimized with respect to standard objective functions, such as the sum of log predictive scores and a test of uniformity on the probability integral transforms (PITs). The optimization explores both forecast combinations and Bayesian averaging of views as pooling methods. In an application to U.S. GDP growth, the approach is shown to produce forecast distributions that are well-calibrated in absolute terms and better calibrated than those produced by a variety of alternative approaches, including popular methods for dealing with non-normality and heteroskedasticity. At the same time, it delivers good accuracy in terms of average predictive densities. The proposed framework can be used to evaluate the time-varying contribution of different views to forecast calibration and accuracy. The empirical application examines views derived from the Fed supervisory scenarios used for bank stress tests.

3.1 Introduction

In recent years, it has become essential for forecasting institutions to characterize the uncertainty around their point forecasts by assigning probabilities to a range of possible economic outcomes. Accordingly, generating economic predictions in the form of continuous probability distributions, or density forecasts, is now common practice (Elliott and Timmermann

2016). The task of forming reliable density forecasts of macroeconomic variables is a challenging one, which requires accounting for the departures from normality that are often observed empirically. In this respect, econometric research has recently shown that gains in density forecast performance can often be achieved by combining different predictive distributions (Hall and Mitchell 2007; Geweke and Amisano 2011; Elliott and Timmermann 2016; Ganics 2017). At the same time, as the recent crisis and its aftermath have highlighted, macroeconomic projections should in general allow for the possibility of discrete shifts or fundamental changes occurring in the economy, whether they be outbreaks of financial instability, political changes or other. Relatedly, while many economic agents, most notably financial institutions, routinely evaluate their potential losses as draws from continuous distributions, macroeconomic outlooks are often reduced to a limited number of distinct scenarios or regimes (e.g., Moody's 2017). This logic facilitates communication regarding economic uncertainty and finds important practical applications, for instance in the design of bank stress tests which are now integral part of the financial regulatory framework and risk management practices in major economies (e.g., Federal Reserve Board 2018). The specific characteristics of different economic regimes are themselves subject to uncertainty, and a great deal of qualitative assessments are generally required to define macro scenarios, giving rise to different views or beliefs that may be considered when producing density forecasts.

This paper investigates a method for enhancing density forecasts of macroeconomic variables using regime-switching models. In the proposed approach, composite density forecasts are constructed by pooling alternative views on the regime-switching behavior of the economy. The composition of such forecasts is optimized with respect to standard evaluation criteria for density forecasts, such as the log predictive scores and a test of uniformity based on the probability integral transform (PIT). Views differ in terms of the assumed number of unobserved regimes and/or in terms of the prior distributions on the parameters of a Bayesian Markov-switching model, resulting in different predictive densities. Two forms of pooling are explored: ex-post combinations of density forecasts from different views and Bayesian averaging of views. Based on the past performance of alternative combinations of views, an optimization procedure selects forecast weights or Bayesian prior probabilities to be used for forecasting future periods. The resulting mixture forecasts are evaluated by means of a recursive out-of-sample forecasting exercise. Empirically, the approach is illustrated using a Markov-switching autoregressive model (MSAR) for U.S. GDP growth, considering both vague views and strong views derived from the Fed macroeconomic scenarios used in the bank stress tests 2015-2018. In the application, the proposed strategy is found to be especially useful to improve the calibration of forecast distributions. In this respect, it outperforms a number of alternative approaches by generating PITs that are well-behaved according to several criteria. The optimization step appears instrumental in producing such result. At the same time, the method achieves comparatively good accuracy in terms of log scores, although along this second dimension it does not outperform the best alternative methods.

The approach is intended to deal with non-normality by producing extremely flexible predictive distributions. Such flexibility results from three key elements. First, density forecasts from any Markov-switching model are weighted averages of the different regime-specific predictive densities, where the weights are the probabilities of the economy ending up in the different regimes. In other words, forecasts allow for regime changes to occur over the forecast horizon and this alone gives rise to mixture distributions, which are in general non-normal even if their individual components are normal (on regime-switching models, see Frühwirth-Schnatter 2006 and Hamilton 2016, among many others). Second, composite predictions are formed here by averaging different views on the Markov-switching model, which means that the forecast densities will be mixtures of mixtures of normals, thereby adding a further layer of flexibility. Moreover, as a result of Bayesian estimation, density forecasts incorporate the uncertainty on the coefficients of the Markov-switching model for any given view.

This approach connects different strands of research on forecasting. First, it is similar in spirit to other Bayesian regime-switching approaches, such as those adopted by Pesaran, Pettenuzzo and Timmermann (2006), who use a break point model (a generalization of regime-switching models) with hyperparameter uncertainty, and by Bauwens, Carpantier and Dufays (2017), who estimate a Markov-switching model with an unknown and potentially infinite number of regimes. However, this paper focuses on a finite set of experts' views to elicit values for the hyperparameters of the regime-switching model, while in Pesaran et al. (2006) and Bauwens et al. (2017) the hyperparameters are random draws from statistical distributions. More generally, unlike those studies, this paper is concerned with optimizing regime-switching density forecasts, based on out-of-sample evaluation criteria.[1] Second, the paper is closely related to research on optimal density forecast combinations (Hall and Mitchell 2007; Geweke and Amisano 2011; Ganics 2017). In fact, the proposed approach can be thought as a convenient alternative to forecast combinations of different models, since

[1]The proposed methodology fixes a maximum number of regimes, whereas Bauwens et al. (2017) allow for infinite regimes using a nonparametric Dirichlet process. However, when estimating a model for U.S. GDP growth (with different breaks for the mean and variance parameters), they find that the posterior probability that the number of regimes is at most 5 lies between 98% and 100% for the mean parameters and between 74% and 100% for the variance, depending on the prior used for estimation.

it combines views on a single Markov-switching model. Given its ability to produce highly flexible approximations of unknown distributions, by means of finite mixtures of mixtures of normals, it can also be seen as a parsimonious alternative to nonparametric methods. In addition, it differs from approaches that assume non-normal errors (e.g., Hansen 1994), in that it allows for a clear economic explanation of non-normality based on different macroeconomic regimes.

While the available evidence on the point forecast performance of regime-switching models is mixed (Elliott and Timmermann 2016), the rise of density forecasting has opened up new opportunities for such models. For instance, Geweke and Amisano (2011) have shown the usefulness of hidden Markov mixtures for producing density forecasts of stock market returns. In Alessandri and Mumtaz (2017), a threshold VAR (in which changes in regime depend on financial conditions) produces good density forecasts of U.S. GDP during the Great Recession. As already mentioned, Bauwens, Carpantier and Dufays (2017) use an infinite Markov-switching autoregressive moving average (ARMA) model to produce density forecasts of U.S. GDP.

Density forecasts can be evaluated using several criteria (see Corradi and Swanson 2006, Elliot and Timmermann 2016 for reviews). This paper adopts two of most popular criteria as objective functions to build optimal composite forecasts. The first one is the log score, which measures the ability to assign high probabilities to outcomes that are truly likely to be observed. The second one is a uniformity test on the sequence of PITs, which provides a measure of the calibration of the forecasts.[2] Both measures have already been used to compute forecast combinations.[3] Moreover, to evaluate the results, two other measures of correct calibration are also considered, namely two tests of independence based on the first two moments of the PITs (Rossi and Sekhposyan 2014).

The remainder of the paper is organized as follows: Section 3.2 explains the methodology, Section 3.3 introduces the empirical application and presents the results, and Section 3.4 concludes.

[2] A well-calibrated forecast is one that does not make systematic errors: if p is the predicted probability assigned to a given random event, then that event should empirically occur with frequency p.

[3] Hall and Mitchell (2007) pioneered density forecast combinations using log scores. Geweke and Amisano (2011) use the log scores to combine five different models of stock returns. Ganics (2017) provides theoretical results on the use of PITs for optimal forecast combinations and presents an empirical application using linear autoregressive distributed lag (ARDL) models of industrial production.

3.2 Methodology

3.2.1 The Markov-switching autoregressive (MSAR) model

This section illustrates the approach using a Markov-switching autoregressive (MSAR) model in which the intercept and the variance of the error term depend on the unobserved state of the economy. Let y_t denote a macroeconomic variable of interest at time t. The MSAR can be expressed as:

$$y_t = \sum_{j=1}^{p} \alpha_j y_{t-j} + \beta_{S_t} + \varepsilon_t \tag{3.1}$$
$$\varepsilon_t \sim N(0, \sigma^2_{S_t})$$

where S_t is the unobserved state variable at time t, β_{S_t} is the intercept in regime S_t, α_j for $j = 1, \ldots, p$ is a state-independent autoregressive term,[4] p is the maximum lag, ε_t is the error term and $\sigma^2_{S_t}$ is the regime-dependent variance of the error. In particular, S_t is a Markov chain characterized by a transition matrix $\boldsymbol{\xi}$, where the element ξ_{kj} in row k and column j represents the probability of transition from state k to state j:

$$\xi_{kj} = Pr(S_t = j | S_{t-1} = k) \tag{3.2}$$

with $k, j = 1, \ldots, K$, where K is the number of regimes in the economy. Therefore, the MSAR captures the typical autocorrelation of macro variables in two ways: by means of the autoregressive coefficients in (3.1) and through the persistence in the state variable S_t as expressed by the transition matrix. Finally, let $\boldsymbol{\vartheta}$ denote the vector of parameters of the MSAR model, i.e. $\boldsymbol{\vartheta} = (\beta_1, \ldots, \beta_K, \sigma_1, \ldots, \sigma_K, \alpha_1, \ldots, \alpha_p, \boldsymbol{\xi})$, and let $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_K, \sigma_1, \ldots, \sigma_K, \alpha_1, \ldots, \alpha_p)$.

3.2.2 Bayesian estimation with multiple views

3.2.2.1 Bayesian estimation of Markov-switching models

This section summarizes the Bayesian approach to the estimation of Markov-switching models following Frühwirth-Schnatter (2006) and adopting her notation. Let us define $\mathbf{y}$ =

[4] Hamilton (1989) uses state-independent autoregressive coefficients to study U.S. GDP growth.

$(y_0, y_1, \ldots, y_T)$ and $\mathbf{S} = (S_0, S_1, \ldots, S_T)$. The posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{y})$ for model (3.1) is obtained using Bayes' theorem:

$$p(\boldsymbol{\vartheta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})$$

where $p(\boldsymbol{\vartheta})$ is the prior on the parameters and $p(\mathbf{y}|\boldsymbol{\vartheta})$ is the likelihood function, which in this case is a Markov mixture of normals. Treating $\mathbf{S}$ as data, the Markov mixture likelihood can be expressed as the sum of the complete-data likelihood $p(\mathbf{y}, \mathbf{S}|\boldsymbol{\vartheta})$ over all possible values of the state vector $\mathbf{S}$:

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\vartheta}) &= \sum_{\mathbf{S} \in S_K} p(\mathbf{y}, \mathbf{S}|\boldsymbol{\vartheta}) \\ &= \sum_{\mathbf{S} \in S_K} p(\mathbf{y}|\mathbf{S}, \boldsymbol{\theta}_1 \ldots, \boldsymbol{\theta}_K)p(\mathbf{S}|\boldsymbol{\xi}) \end{aligned} \tag{3.3}$$

As shown in Frühwirth-Schnatter (2006), expression (3.3) factors in a convenient way that simplifies estimation. In particular, it can be shown that if the prior assumes (i) the independence of the parameter vector $\boldsymbol{\theta}$ across regimes and (ii) the independence between the parameter vector $\boldsymbol{\theta}$ and the transition matrix $\boldsymbol{\xi}$, i.e.

$$p(\boldsymbol{\vartheta}) = \prod_{k=1}^{K} p(\boldsymbol{\theta}_k)p(\boldsymbol{\xi})$$

then the complete-data posterior, i.e.

$$p(\boldsymbol{\vartheta}|\mathbf{y}, \mathbf{S}) = \prod_{k=1}^{K} p(\boldsymbol{\theta}_k|\mathbf{y}, \mathbf{S})p(\boldsymbol{\xi}|\mathbf{S})$$

factors in the same way as the complete-data likelihood $p(\mathbf{y}, \mathbf{S}|\boldsymbol{\vartheta})$. This facilitates the application of conventional Markov Chain Monte Carlo (MCMC) methods used for Bayesian estimation, in a context where, due to the Markov-switching nature of the model, the prior $p(\boldsymbol{\vartheta})$ and the posterior $p(\boldsymbol{\vartheta}|\mathbf{y})$ are not conjugate and the posterior does not assume any convenient analytical form.

Finally, the posterior $p(\boldsymbol{\vartheta}|\mathbf{y})$ can be expressed as the sum of the posterior for the augmented parameter vector $(\mathbf{S}, \boldsymbol{\vartheta})$ over all possible realizations of $\mathbf{S}$:

$$p(\boldsymbol{\vartheta}|y) = \sum_{\mathbf{S} \in S_K} p(\mathbf{S}, \boldsymbol{\vartheta}|\mathbf{y})$$

In practice, Bayesian estimation samples from the joint posterior $p(\mathbf{S}, \boldsymbol{\vartheta}|\mathbf{y})$, using:

$$p(\mathbf{S}, \boldsymbol{\vartheta}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{S}, \boldsymbol{\vartheta}) p(\mathbf{S}|\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta})$$

3.2.2.2 Estimating the MSAR with multiple views

In line with the estimation framework presented so far, the MSAR (3.1) is estimated here using MCMC methods and assuming independence priors of the following form:

$$p(\alpha_1, \ldots, \alpha_p, , \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_K^2) = \prod_{j=1}^{p} p(\alpha_j) \prod_{k=1}^{K} p(\beta_k) \prod_{k=1}^{K} p(\sigma_k^2)$$

The priors follow conventional distributions, which are:

$$\begin{aligned} \beta_k &\sim \mathcal{N}(b_{0,k}, B_{0,k}) \\ \sigma_k^2 &\sim \mathcal{G}^{-1}(c_0, C_0) \\ \alpha_j &\sim \mathcal{N}(a_{j,0}, A_{j,0}) \\ j &= 1, \ldots, p \end{aligned}$$

where $\mathcal{N}$ and $\mathcal{G}^{-1}$ denote Normal and inverse Gamma distributions, respectively, and $b_{0,k}$, $B_{0,k}$, $c_0, C_0, a_{j,0}, A_{j,0}$ are hyperparameters to be selected by the researcher.

In addition, for the transition matrix $\boldsymbol{\xi}$ it is assumed that the rows are independent and each row follows a Dirichlet distribution $\mathcal{D}$:

$$\boldsymbol{\xi}_k \sim \mathcal{D}(e_{k1}, \ldots, e_{kK})$$

where $e_{k1}, \ldots, e_{kK}$ are hyperparameters, for $k = 1, \ldots, K$.

The number of regimes is also treated as unknown. Accordingly, a discrete prior is defined for K, fixing a maximum number $\overline{K}$:

$$\begin{aligned} \pi_K^0 &= Pr\,(K) \\ K &= 1, \ldots, \overline{K} \\ \sum_{K=1}^{\overline{K}} \pi_K^0 &= 1 \end{aligned} \tag{3.4}$$

Note that the letter π will be used throughout the text to denote discrete probability distributions.

Next, for any given number of states K, a number P_K of alternative priors on the MSAR parameters are considered. Each prior is identified by a specific set of values for the hyperparameters $(b_{0,1}, \ldots, b_{0,K}, B_{0,1}, \ldots, B_{0,K}, a_{0,1}, \ldots, a_{0,p}, A_{0,1}, \ldots, A_{0,p}, c_0, C_0, e_{11}, \ldots, e_{KK})$. Let $\boldsymbol{\vartheta}_{K,i}^0$ denote the generic i-th prior assuming K states. A hyperprior probability $\pi(\boldsymbol{\vartheta}_{K,i}^0|K)$ is assigned to $\boldsymbol{\vartheta}_{K,i}^0$, such that

$$\sum_{i=1}^{P_K} \pi(\boldsymbol{\vartheta}_{K,i}^0\,|K) = 1 \tag{3.5}$$

In other words, a discrete hierarchical prior is defined with respect to $\boldsymbol{\vartheta}$. The unconditional prior probability of $\boldsymbol{\vartheta}_{K,i}^0$ is equal to the joint prior probability of $\boldsymbol{\vartheta}_{K,i}^0$ and the number K of regimes, i.e. $\pi(\boldsymbol{\vartheta}_{K,i}^0) = \pi(\boldsymbol{\vartheta}_{K,i}^0, K)$. Using $\pi_{K,i}^0$ to denote this unconditional probability, we have that:

$$\pi_{K,i}^0 \equiv \pi(\boldsymbol{\vartheta}_{K,i}^0) = \pi(\boldsymbol{\vartheta}_{K,i}^0\,|K)\,\pi_K^0$$

In what follows, let us refer to $\boldsymbol{\vartheta}_{K,i}^0$ as a *view* about the regime-switching properties of the economy. Thus, defining a view implies (i) choosing the number of regimes and (ii) choosing a prior for the MSAR parameters $\boldsymbol{\vartheta}$. Also, let $\boldsymbol{\pi}^0$ denote the vector of length $\sum_{K=1}^{\overline{K}} P_K$ containing the unconditional prior probabilities of all views, i.e. $\boldsymbol{\pi}^0 = (\pi_{1,1}^0, \ldots, \pi_{\overline{K},P_{\overline{K}}}^0)$.

The posterior probabilities of the views depend on the prior $\boldsymbol{\pi}^0$ and on the marginal likelihood of the MSAR model under the different views. In particular, the unconditional posterior probability of view $\boldsymbol{\vartheta}_{K,i}^0$ is equal to the joint posterior probability of $\boldsymbol{\vartheta}_{K,i}^0$ and the number K of regimes, i.e. $\pi(\boldsymbol{\vartheta}^0_{K,i}|\mathbf{y}) = \pi(\boldsymbol{\vartheta}^0_{K,i}, K|\mathbf{y})$, and is given by:

$$\pi_{K,i} \equiv \pi(\boldsymbol{\vartheta}^0_{K,i}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\vartheta}^0_{K,i})\pi^0_{K,i}}{\sum_{K=1}^{\overline{K}}\sum_{j=1}^{P_K} p(\mathbf{y}|\boldsymbol{\vartheta}^0_{K,j})\pi^0_{K,j}} \tag{3.6}$$

where $p(\mathbf{y}|\boldsymbol{\vartheta}^0_{K,i}) = p(\mathbf{y}|\boldsymbol{\vartheta}^0_{K,i}, K) = \int p(\mathbf{y}|\boldsymbol{\vartheta}_K, \boldsymbol{\vartheta}^0_{K,i}, K)p(\boldsymbol{\vartheta}_K|\boldsymbol{\vartheta}^0_{K,i}, K)d\boldsymbol{\vartheta}_K$, with $\boldsymbol{\vartheta}_K$ denoting the parameter vector in the MSAR model with K regimes.

3.2.3 Density forecasts

Computing density forecasts from a MSAR model requires three steps (Frühwirth-Schnatter 2006). In what follows, let us add a time subscript to the vector of observations $\mathbf{y}$, so that $\mathbf{y}_t = (y_0, y_1, \ldots, y_t)$. Also, let us assume that the current time period is T and the forecast horizon is one period. The first step consists in using the MCMC algorithm to sample both the current unobserved regime S_T and the MSAR parameters $\boldsymbol{\vartheta}$ from the posterior distribution $p(\mathbf{S}, \boldsymbol{\vartheta}|\mathbf{y}_T)$. Let $(\boldsymbol{\vartheta}^{(d)}, S_T^{(d)})$ denote a generic MCMC draw. Next, each draw is used to forecast the future state of the economy. Taking $S_T^{(d)}$ as the starting value, a stochastic forecast $S_{T+1}^{(d)}$ is computed using the matrix of transition probabilities $\boldsymbol{\xi}^{(d)}$, i.e. based on (3.2). Third, $y_{T+1}^{(d)}$ is sampled from the normal predictive density $p(y_{T+1}, |\mathbf{y}_T, \boldsymbol{\vartheta}^{(d)}, S_{T+1}^{(d)})$. In particular,

$$y_{T+1}|\mathbf{y}_T, \boldsymbol{\vartheta}^{(d)}, S_{T+1}^{(d)} = k \sim \mathcal{N}\left(\sum_{j=1}^{p} \alpha_j^{(d)} y_{T+1-j} + \beta_k^{(d)}, \sigma_k^{(d)2}\right)$$

Conditional on knowing the state of the economy in the future period $T+1$, the predictive distribution of y_{T+1} is a Normal for any given parameter vector. However, since the future state of the economy is unknown, the density forecast of y_{T+1} produced by the MSAR will be a mixture of the different regime-specific normals, where the mixture weights are given by the probabilities of the economy ending up in the different possible regimes at $T+1$. As a result, the MSAR is generally able to produces highly flexible, non-normal forecast distributions.[5] In addition, Bayesian estimation incorporates the uncertainty on the parameters $\boldsymbol{\vartheta}$ into the density forecasts. What is more, considering alternative views allows for an additional degree of flexibility, as formalized below.

Assuming a known number of regimes K and a known parameter vector $\boldsymbol{\vartheta}$, the one-step-

[5] Also, the predictive densities are non-linear in y_T and heteroskedastic (Frühwirth-Schnatter 2006).

ahead density forecast at time T is the following finite mixture of K normal components:

$$p\left(y_{T+1}|\mathbf{y}_T,\boldsymbol{\vartheta}\right) = \sum_{k=1}^{K} p\left(y_{T+1}|\mathbf{y}_T,\boldsymbol{\theta}_k\right) Pr\left(S_{T+1}=k|\mathbf{y}_T,\boldsymbol{\vartheta}\right)$$

Next, as a result of Bayesian estimation, the density forecast for any given view integrates out parameter uncertainty:

$$p\left(y_{T+1}|\mathbf{y}_T,\boldsymbol{\vartheta}_{K,i}^0\right) = \int p\left(y_{T+1}|\mathbf{y}_T,\boldsymbol{\vartheta}_K,\boldsymbol{\vartheta}_{K,i}^0\right) p(\boldsymbol{\vartheta}_K|\mathbf{y}_T,\boldsymbol{\vartheta}_{K,i}^0)d\boldsymbol{\vartheta}_K$$

where, as before, $\boldsymbol{\vartheta}_K$ denotes the parameter vector $\boldsymbol{\vartheta}$ when K regimes are assumed. Finally, averaging over different views $\boldsymbol{\vartheta}_{K,i}^0$, we get:

$$p\left(y_{T+1}|\mathbf{y}_T,\boldsymbol{\pi}^0\right) = \sum_{K=1}^{\overline{K}} \sum_{i=1}^{P_K} p\left(y_{T+1}|\mathbf{y}_T,\boldsymbol{\vartheta}_{K,i}^0\right) \pi_{K,i} \tag{3.7}$$

where $\pi_{K,i}$ depends on the prior probability vector $\boldsymbol{\pi}^0$ and on the marginal likelihoods of the different views according to equation (3.6). Forecast (3.7) is a composite forecast in which the weight assigned to the view-specific forecast $p\left(y_{T+1}|\mathbf{y}_T,\boldsymbol{\vartheta}_{K,i}^0\right)$ is given by the posterior probability of the view, $\pi_{K,i}$. Therefore, (3.7) is a mixture of mixtures. If we take the set of alternative views as given, the forecast combination weights are unambiguously pinned down by the data $\mathbf{y}_T$ and by the prior vector $\boldsymbol{\pi}^0$.

In addition to the Bayesian averaging of views in (3.7), let us also consider standard non-Bayesian forecast combinations. In this case, let us express a forecast combination of different MSAR views, where the vector of combination weights is denoted by $\mathbf{w}$, as:

$$p\left(y_{T+1}|\mathbf{y}_T,\mathbf{w}\right) = \sum_{K=1}^{\overline{K}} \sum_{i=1}^{P_K} p\left(y_{T+1}|\mathbf{y}_T,\boldsymbol{\vartheta}_{K,i}^0\right) w_{K,i} \tag{3.8}$$

where $w_{K,i} \geq 0$ is the weight assigned to view $\boldsymbol{\vartheta}_{K,i}^0$ and $\sum_{K=1}^{\overline{K}} \sum_{i=1}^{P_K} w_{K,i} = 1$.

3.2.4 Optimizing density forecasts

The composite density forecasts from the MSAR with multiple views are optimized with respect to two alternative objective functions, based on statistics that are commonly used to evaluate density forecast performance: the log score and the probability integral transform (PIT).

The log score is the log of the predictive density function evaluated at the actual realization of the forecast variable. Let y^o_{t+h} (where "o" stands for "observed") denote the realization of variable y_{t+h}, which is not observed at time t, when the forecast for $t+h$ is produced. Also, let R be the length of the timespan over which forecasts are optimized. The first objective function, denoted by f_1, is given by the sum of log scores over the period of interest. For combinations using generic weights $\mathbf{w}$ as in (3.8), the sum of log scores at time τ can be expressed as:

$$f_{1,\tau}\left(\mathbf{w}\right) = \sum_{t=\tau-h-R+1}^{\tau-h} \ln\left(p\left(y^o_{t+h}|\mathbf{y}_t, \mathbf{w}\right)\right)$$

For combined forecasts using Bayesian averaging as in (3.7), the objective function can be written as:

$$f_{1,\tau}\left(\boldsymbol{\pi}^0\right) = \sum_{t=\tau-h-R+1}^{\tau-h} \ln\left(p\left(y^o_{t+h}|\mathbf{y}_t, \boldsymbol{\pi}^0\right)\right)$$

The PIT is the cumulative predictive density function evaluated at the actual realization of the variable. If the density forecast used to compute the PIT corresponds to the true distribution of the variable, then, for $h=1$, the PIT values are the realizations of independently and identically distributed (i.i.d.) Uniform $(0,1)$ variables (Diebold et al. 1998). Therefore, a uniformity test on the PITs is a test of correct specification of the density forecasts (see also Rossi and Sekhposyan 2014). Accordingly, the second objective function for forecasts of type (3.8) is given by:

$$f_{2,\tau}\left(\mathbf{w}\right) = -ks\left(\left\{\Phi\left(y^o_{t+1}|\mathbf{y}_t, \mathbf{w}\right)\right\}_{t=\tau-R}^{\tau-1}\right)$$

where $\Phi\left(\cdot\right)$ denotes the cumulative predictive density function, i.e.

$$\Phi\left(y^o_{t+1}|\mathbf{y}_t, \mathbf{w}\right) \equiv \int_{-\infty}^{y^o_{t+1}} p\left(y_{t+1}|\mathbf{y}_t, \mathbf{w}\right) dy_{t+1}$$

while function $ks(\cdot)$ represents the test statistics of a Kolmogorov-Smirnov (KS) test of uniformity. Note that maximizing $-ks(\cdot)$ is equivalent to maximizing the p-value of the KS test. Analogously, for Bayesian averaging we have:

$$f_{2,\tau}\left(\boldsymbol{\pi}^0\right) = -ks\left(\left\{\Phi\left(y^o_{t+1}|\mathbf{y}_t, \boldsymbol{\pi}^0\right)\right\}_{t=\tau-R}^{\tau-1}\right)$$

Both the optimization based on f_1 and the one based on f_2 are solved numerically. For each f_i, with $i = 1, 2$, the optimization algorithm delivers two vectors at time τ: the vector of optimal forecast weights $\mathbf{w}^*_{i,\tau}$ for the set of alternative views, i.e.:

$$\mathbf{w}^*_{i,\tau} \equiv \arg\max_{\mathbf{w}} f_{i,\tau}\left(\mathbf{w}\right) \tag{3.9}$$

and the vector of optimal prior probabilities $\boldsymbol{\pi}^{0*}_{i,\tau}$:

$$\boldsymbol{\pi}^{0*}_{i,\tau} \equiv \arg\max_{\boldsymbol{\pi}^{\mathbf{0}}} f_{i,\tau}\left(\boldsymbol{\pi}^{\mathbf{0}}\right) \tag{3.10}$$

The former represents the typical problem addressed in the literature on density forecast combination, whereas the latter can be seen as an empirical method for eliciting priors in the context of Bayesian model averaging. The optimal prior $\boldsymbol{\pi}^{0*}_{i,\tau}$ represents the discrete prior probability distribution of views such that the resulting posterior $\boldsymbol{\pi}^*_{i,\tau}$, when used as a vector of forecast weights, maximizes the density forecast performance, based on the selected objective function. In practice, the main difference between (3.9) and (3.10) is that the first problem directly delivers weights for forecast combination, while in the second case the actual forecast weights will also depend on the marginal likelihoods of all views, i.e. $p(\mathbf{y}|\boldsymbol{\vartheta}^0_{K,i})\ \forall K, i$.

3.3 Empirical application

This section assesses the empirical performance of the approach proposed in the paper. The application deals with density forecasts of U.S. real GDP growth and uses quarterly data from 1948Q1 to 2017Q2 (Figure 3.1). The growth rate considered is the year-on-year growth rate (expressed in percentage points in what follows). The lag length p is set to 5, in consideration of the quarterly frequency of the variable. The optimal weights $\mathbf{w}^*$ and optimal priors $\boldsymbol{\pi}^{0*}$ are tracked over time by means of a recursive optimization scheme. Their forecast performance is assessed on an evaluation sample, i.e. using observations of the target variable that have not been used in the optimization procedure, as described in section 3.3.2.

3.3.1 Views

A total of 13 alternative views on the regime-switching properties of U.S. GDP are considered. Eight views impose strongly informative priors derived from the scenarios of the Fed stress tests 2015-2018.[6] The remaining five views are vague views, defined by imposing a diffuse prior on the MSAR parameters under different assumptions on the number of regimes $K = 1, 2, 3, 4, 5$.

Let us first consider the Fed-based views. For each of the four stress tests under consideration, two views are constructed, one with $K = 3$ and the other with $K = 5$. In the view with $K = 3$, one of the regimes (which may be called the "normal times" regime), is derived from the Fed baseline scenario, another ("adverse regime") from the adverse scenario and the last one ("severely adverse regime") from the severely adverse scenario.[7] In particular, each regime is "centered" on the corresponding scenario using the following rule. Consider an AR(5) model where the coefficients are given by the k-state-specific hyperparameters of the prior $\boldsymbol{\vartheta}^0_{K,i}$, i.e.:

$$y_t = \sum_{j=1}^{5} a_j^{(K,i)} y_{t-j} + b_{0,k}^{(K,i)} + \varepsilon_t$$

In this model, the unconditional expectation of y_t is

$$E\left(y_t\right) = \frac{b_{0,k}^{(K,i)}}{1 - \sum_{j=1}^{5} a_j^{(K,i)}} \tag{3.11}$$

Then, after making an assumption on the state-independent $a_j^{(K,i)}$, with $j = 1, \ldots, 5$, each regime-specific $b_{0,k}^{(K,i)}$ is chosen in such a way that expectation (3.11) matches a specific value derived from the relevant scenario of the Fed stress test. For the normal times regime, this value is the average growth rate in the last 4 quarters of the baseline scenario, which is assumed to be close to the convergence value of the year-on-year growth rate in the absence

[6]See https://www.federalreserve.gov/supervisionreg/dfa-stress-tests.htm.

[7]Although the Fed stress scenarios represent hypothetical paths and not forecasts, they are intended to be plausible even when severe. Therefore, they can legitimately be assigned predictive probabilities (see e.g., Yuen 2013) and used to form density forecasts.

of shocks.[8] For both the adverse and the severely adverse regimes, the value to be matched is the average growth rate in the first 4 quarters of the corresponding scenario, as the first quarters are those when the negative shocks are assumed to occur and the growth rates are lowest.

An example may help. Let us consider the view with $K = 3$ derived from the 2018 Fed stress test. The average growth rate of GDP in the last 4 quarters of the baseline scenario is 2.1%, while the average growth rates in the first 4 quarters of the adverse and severely adverse scenarios are -2.125% and -6.275% respectively. Assuming that the prior mean for the autoregressive coefficients is 0.9 for the first lag and 0 for higher-order lags, which approximates the OLS estimate of a simple AR(1) for GDP growth over the entire sample, then $\sum_{j=1}^{5} a_j^{(K,i)} = 0.9$. Accordingly, the prior means for the regime-specific intercepts are set to $b_{0,1} = 2.1/(1-0.9) = 0.21$ for the normal times regime, $b_{0,2} = -2.125/(1-0.9) = -0.2125$ for the adverse regime and $b_{0,3} = -6.275/(1-0.9) = -0.6275$ for the severely adverse regime.

The four stress test-based views with $K = 5$ expand the views with $K = 3$ by adding two regimes: a regime which we may call “recovery from adverse shock”, designed to match the last 4 quarters of the adverse scenario, and a regime of “recovery from severely adverse shock”, which matches the last 4 quarters of the severely adverse scenario. This is done in consideration of the fact that growth rates in the last 4 quarters of the adverse and severely adverse scenarios are assumed to be higher than the baseline rates, implying a rebound of the economy after a negative shock. Of course, such regimes may be more generally interpreted as “favorable regimes” characterized by positive shocks and not necessarily as recoveries from recessions.

In the five vague views, all priors on the intercepts are centered on 0 and have a variance of 1 percentage point, while the priors on the autoregressive coefficients are centered on 0.5 for the first lag, on 0 for the higher-order lags, and have a variance of 1. The combination of these assumptions implies a large prior variance on the regime-specific means of the GDP growth rate. In the Fed-based views, the priors for both β and α are strongly informative, so as to ensure that the regime-specific prior means are tightly centered on the stress test values, based on equation (3.11). In particular, both priors are assumed to have a minimal variance, equal to 10^{-5}. For the autoregressive coefficients α, the prior mean is assumed to be 0.9 for the first lag and 0 for higher-order lags, as in the previous example.

No strong assumption is made regarding the regime-switching error variance σ_k^2. Instead, a diffuse hierarchical prior is assumed for all views. Specifically, a Gamma hyperprior is

[8] The stress test scenarios are defined in terms of annualized quarter-on-quarter growth rates, so that averaging over the last 4 quarters approximates the year-on-year growth rate in the last quarter.

defined for C_0:[9]

$$C_0 \sim \mathcal{G}\left(g_0, G_0\right)$$

To make the prior on σ_k^2 diffuse, the following values are selected for the hyperparameters: $c_0 = 3$, $g_0 = 0.5$ and $G_0 = 0.5$. These imply that σ_k^2 has a prior expected value of 0.5 percentage points of GDP and a high prior variance, equal to 1.25 percentage points (see Appendix 3.A for the derivations).

Finally, the hyperparameters for the k-th row of the transition matrix $\boldsymbol{\xi}$ are $e_{kk} = 2$ and $e_{kj} = 1/(K-1)$ if $k \neq j$, $\forall k, j$. Given the properties of the Dirichlet distribution, $\mathrm{E}(\xi_{kj}) = e_{kj}/(\sum_{l=1}^{K} e_{kl})$. Therefore, the prior expected probability of remaining in the same state k in the next period is $\mathrm{E}(\xi_{kk}) = 2/3$ regardless of the number of regimes K, while the probability of moving to a different, specific state j decreases with the number of regimes, $\mathrm{E}(\xi_{kj}) = 1/[3(K-1)]$.

The summary of the alternative views is provided in Table 3.1, where views 1-5 are the vague views while views 6-13 are those derived from the Fed stress tests 2015-2018. Table 3.2 displays the GDP scenarios of the Fed stress tests.

3.3.2 Optimization scheme

In the empirical application, a recursive-window estimation scheme is used to generate a sequence of density forecasts.[10] Next, forecasts are used to carry out the optimization of weights/priors, which is iterated over time. The procedure can be described as follows. Let us assume that the current period is T_w and the forecast horizon is h. For each view under consideration, the MSAR model is recursively estimated using observations between time t_0 and time t, with $t = T_0, T_0+1, \ldots, T_w - h$. T_0 is therefore the end period of the shortest estimation sample. Estimates at T_0 are used to make forecasts for period $T_0 + h$, estimates at $T_0 + 1$ are used to make forecasts for $T_0 + 1 + h$, and so on. At time T_w, a sequence of past forecasts is available for each view. At this point, the algorithm computes the optimal

[9] Accordingly, the independence prior of the MSAR model becomes:

$$p(\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_K^2, C_0) = \prod_{j=1}^{p} p(\alpha_j) \prod_{k=1}^{K} p(\beta_k) \prod_{k=1}^{K} p(\sigma_k^2) p(C_0)$$

[10] In this context, the choice of using expanding windows for estimation, as opposed to rolling windows, increases the probability that the variable "visits" the highest possible number of regimes within the sample.

weights/priors based on the last R forecasts, i.e. maximizes the relevant objective function between $T_w - R + 1$ and T_w. Once the optimal weights/priors are retrieved, they are used to combine the different view-specific forecasts for the future period $T_w + h$, which is out of the optimization sample. When the actual value of the variable of interest is observed, at time $T_w + h$, the performance of the composite forecast is measured. The index T_w runs from $T_0 + h + R - 1$ to $\overline{T} + h$, where $\overline{T}$ is the end of the largest estimation sample. $\overline{T} + 2h$ is the last available observation for the target variable. Therefore, the period from $T_0 + 2h + R - 1$ to $\overline{T} + 2h$ defines the evaluation sample. Figure 3.2 summarizes the procedure (cf. Ganics 2017).

More specifically, the application to U.S. GDP growth sets t_0 = 1948Q1, T_0 = 1967Q4, $R = 40$ quarters, $h = 1$ quarter and $\overline{T}$ = 2016Q4. Accordingly, the evaluation sample runs from 1978Q1 to 2017Q2. Results are also reported for $R = 20$.[11]

3.3.3 Results

Table 3.3 shows the performance of the optimal forecast weights and optimal priors over the evaluation sample and compares it with five alternative benchmark approaches. The first approach is a simple linear AR(5) model, corresponding to view no. 1 in Table 3.1. In the second approach, forecasts are produced using the individual view that exhibits the highest marginal likelihood, selected recursively across estimation windows. The third approach uses an AR model estimated on rolling windows of 80 quarters to accommodate time-varying parameters.[12] The remaining two approaches consider uniform combination schemes for the alternative views, assigning equal forecast weights and equal prior probabilities, respectively, to different values of K and, given K, equal weights/probabilities to the alternative views defined using K regimes.[13] As mentioned in section 3.2.4, weights $\mathbf{w}_1^*$ and priors $\boldsymbol{\pi}_1^*$ result from the optimization taking the sum of log scores as objective function, while $\mathbf{w}_2^*$ and $\boldsymbol{\pi}_2^*$ are obtained by maximizing the p-value of the Kolmogorov-Smirnov (KS) test of uniformity for the PITs. The table shows the average predictive density (APD) (i.e. the average of the exponential of the log scores) and the p-value of the KS test. Besides, two additional

[11] To estimate the MSAR model I use the MATLAB package `bayesf Version 2.0` by Frühwirth-Schnatter (2008). For each MSAR estimate, the MCMC algorithm uses 1000 iterations as burn-in and 1000 iterations to store the results. Starting from the sample of forecasts produced by the MCMC algorithm, a complete probability density function is fitted using standard kernel methods.

[12] Using rolling windows of 40 quarters gives similar results.

[13] For instance, in the case of equal prior probabilities, it is assumed that $\pi_K^0 = 1/\overline{K}$ for each K and that $\pi(\boldsymbol{\vartheta}_{K,i}^0|K) = 1/P_K$ for each view $\boldsymbol{\vartheta}_{K,i}^0$. See (3.4) and (3.5).

measures of correct specification of density forecasts are taken into consideration, namely the p-values of the Ljung–Box test of serial independence for the first and second moment of the PITs (see Rossi and Sekhposyan 2014). Since correct calibration implies that the PITs are realizations of i.i.d. variables, both tests should not reject the null of serial independence for forecasts to be considered well-calibrated. In the table, LB1 denotes the test on the first moment and LB2 the test on the second moment. Following Rossi and Sekhposyan (2014), in both tests the null hypothesis is serial independence over up to 4 lags.

The main result is that optimized regime-switching composite forecasts achieve well-behaved PITs, unlike all benchmarks considered. The optimization step generates substantial improvements in density forecast performance as measured by the uniformity of the PIT. As can be seen from Table 3.3, using the optimal priors $\boldsymbol{\pi}_2^*$ and the optimal weights $\mathbf{w}_2^*$ results in the highest p-values in the KS test of PIT uniformity, 0.32 and 0.21 respectively, while also ensuring that both tests of independence of the PITs do not reject the null hypothesis. By contrast, the recursively estimated linear AR, the two uniform weighting schemes and the approach using the views with the highest marginal likelihood all lead to rejection of the null of uniformity at the 5% level. The AR model estimated on a rolling window gives a p-value of 10% in the KS test, but strongly rejects serial independence in the second moment of the PITs. In general, for all MSAR-based forecasts the null of independence cannot be rejected, whereas in the case of the linear AR model the independence of the second moment is rejected regardless of the estimation scheme. Interestingly, weights $\mathbf{w}_1^*$ and priors $\boldsymbol{\pi}_1^*$ both lead to increases in the KS p-value relative to uniform combinations, even though they are optimized using the log scores as objective function.

Second, the optimization step appears less useful for producing gains in terms of log scores. The APDs of the log-score-optimized forecasts are higher than those achieved by the recursive-window AR, the rolling-window AR and equal forecast weights, but are roughly the same as those obtained by using uniform prior probabilities or by recursively selecting the view with the highest marginal likelihood. Moreover, using the sum of log scores as objective function results in small increases in APD compared to using the KS statistics. Overall, the comparatively good accuracy in terms of APDs appears to be driven more by the Markov-switching model than by the optimization procedure.

To summarize, optimizing the combinations of views enhances the calibration of density forecasts in terms of PIT uniformity, i.e. improves the specification of the predictive distribution. This, combined with the regime-switching setup, leads to PITs that are not significantly different from i.i.d uniform variables. At the same time, the approach is capable of producing results in terms of log-score accuracy that are roughly in line with the best ones across several benchmarks.

Figure 3.3 shows the evolution over time of the well-calibrated 1-quarter-ahead forecasts based on the optimal priors $\boldsymbol{\pi}_2^*$, plotting the p.d.f. of the forecasts in each period. Figure 3.4 summarizes the same density forecasts using a fan chart, where different shades of color identify different percentiles, from 0.01 to 0.99.

The approach can be used to evaluate the time-varying contribution of different views to the composite forecasts. Figures 3.5-3.8 display the evolution over time of the optimal forecast weights and of the weights resulting from the optimal priors, i.e. the optimized posterior probabilities. In each figure, the area chart in the left panel shows the time-varying weights for all views from 1978Q1 to 2017Q2. The right panel plots the cumulative weight assigned to the views derived from the Fed supervisory scenarios. Figures 3.5 and 3.7 show the results of the optimization based on log scores, while Figures 3.6 and 3.8 show the results of the optimization based on the PITs. As can be seen from Figures 3.5 and 3.7, the vague views tend to dominate in the case of log-score optimization, especially when the prior probabilities are optimized. In terms of optimal weights $\mathbf{w}_1^*$, the cumulative weight of the Fed-based views lies in the range 10%-35% between 1979 and 1990, remains flat at zero from the end of 1990 until 2006, then starts increasing in 2007 and peaks at 61% in 2010. It rapidly declines afterwards. On average, the vague views account for more than 90% of the composite forecasts. As regards the optimized posteriors, the Fed-based views only have short-lived spikes in 1984 (21%) and 2010 (100%). Overall, the results indicate a minor role of Fed-based views in boosting density forecast accuracy. This is consistent with the fact that the maximum marginal likelihood criterion (used in the second row of Table 3.3), which gives as high APDs as the log-score-optimized weights and priors, never selects any Fed-based views.

When the PIT-based optimization is considered, the contribution of the Fed-based views is much higher. On average, they account for 33% of the combined forecasts in the case of optimal weights and over 20% in the case of optimal priors. In terms of $\mathbf{w}_2^*$, their cumulative weight exceeds 60% in 1982-1983, increases quite rapidly during the period 2007-2009 and remains steadily between 75% and 100% from 2009 to 2017. The Fed-based views also dominate in terms of optimized posteriors for most of the period 2008-2017. Their cumulative posterior probability has a first peak in 1983, while it remains close to zero from 1984 to 2008. It is important to remark that using Fed-based views is not sufficient to achieve well-calibrated forecasts. None of these views, when considered individually, leads to non-rejection of the null hypothesis in the KS test. Instead, as already stressed, the combination of different views is what produces good results in terms of PIT uniformity.

Finally, Table 3.5 shows the results for $R = 20$. The main conclusions hold true in this case, except that for PIT-based optimal weights and priors the LB1 test does nor reject at the 5% level but still rejects at the 10%.

3.3.3.1 Comparison with non-normal and heteroskedastic AR models

To evaluate the approach within the broader perspective of non-normal and heteroskedastic models, this section shows the density forecast performance of three alternative models: an AR with Student-t errors, an AR with ARCH errors and an AR with GARCH errors. The models have been estimated on both recursive windows and rolling windows of 40 and 80 quarters.[14] As with the MSAR models, the lag length for the AR component is set to 5 for all three models, while the ARCH and GARCH components have a lag length of 1.

For each model, Table 3.4 shows the APDs and the p-values for the KS, LB1 and LB2 tests over the same evaluation sample as in the previous section. When estimated on recursive windows, all three models generate non-uniform PITs and lower APDs than any MSAR-based method in Table 3.3. Their performance considerably improves when rolling windows are used, which accommodate structural instabilities. In particular, the AR with t errors achieves the highest APD (0.37) and generates PITs that do not reject the hypotheses of uniformity and independence in the first moment. Regarding independence in the second moment, the LB2 test rejects the null at the 5% when estimated on 80-quarter windows, whereas it does not reject the null at the 5% but rejects it at the 10% level when estimated on 40-quarter windows. The models with ARCH/GARCH errors always reject the hypothesis of second-moment independence and are generally outperformed by the MSAR-based methods in terms of APDs.

The results suggest that, when the PIT optimization is used, the approach proposed in the paper is able to achieve a more reliable specification of the conditional predictive distribution, based on the joint indications offered by the KS, LB1 and LB2 tests. In terms of log-score accuracy, the approach produces results that are close but below the best alternative, namely the AR model with Student-t errors estimated on rolling windows.

[14] The AR-GARCH model on rolling windows of 40 quarters is not supported by the data and is therefore not reported.

3.4 Conclusions

This paper has proposed a procedure for constructing comparatively reliable density forecasts using a regime-switching model. Composite forecasts are formed by pooling alternative model assumptions (or views) and are optimized with respect to measures of calibration (probability integral transforms or PITs) and accuracy (log scores). The approach merges the well-established benefits of forecast combination with the flexibility of mixture predictive densities provided by a single, Markov-switching model. Different sources of uncertainty are incorporated into the density forecasts. First, uncertainty on the future state of the economy is dealt with by means of the Markov-switching setup. Second, as a result of Bayesian estimation, parameter uncertainty enters the predictive densities for any given view on the regime-switching behavior of the economy. Third, "disagreement" between views is taken into account through forecast combination.

The approach appears to strike a good balance between the specification of flexible distributional shapes and the accuracy of density forecasts. In an application to U.S. GDP, the optimized regime-switching forecasts achieve PITs that are not significantly different from i.i.d uniform variables, thereby complying with theoretical prescriptions on density forecast calibration. At the same time, they exhibit a good level of accuracy in terms of average predictive densities. Moreover, the forecasts appear better calibrated than those provided by a variety of competing approaches.

Unlike an AR model with non-normal errors, which in the paper turns out to be the best competitor in terms of density forecast performance, this methodology allows for flexible predictions to be constructed by incorporating different macroeconomic scenarios as defined by experts. To illustrate this possibility, the empirical application makes use of views derived from the Fed supervisory scenarios, which are adopted in the annual bank stress tests, and tracks their contribution to the optimized forecasts over time. This feature appears particularly valuable in all contexts in which tail risks have a clear economic interpretation and when predictive simulations have to comply with external, possibly judgmental views. Researchers and practitioners interested in this kind of analysis may fine-tune the approach by selecting different objective functions in the optimization step and by tailoring the range of views to be explored.

Bibliography

[1] Alessandri, P. and H. Mumtaz (2017), "Financial conditions and density forecasts for US output and inflation." *Review of Economic Dynamics*, 24, 66–78.

[2] Bauwens, L., J. F. Carpantier and A. Dufays (2017), "Autoregressive Moving Average Infinite Hidden Markov-Switching Models." *Journal of Business and Economic Statistics*, 35 (2), 162-182.

[3] Corradi, V. and N. R. Swanson (2006), "Predictive Density Evaluation." In *Handbook of Economic Forecasting*, volume 1, G. Elliott, C. W. J. Granger and A. Timmermann (eds.), pp. 197–284, Elsevier.

[4] Diebold, F. X., T. A. Gunther and A. S. Tay (1998). "Evaluating density forecasts." *International Economic Review*, 39 (4), 863–883.

[5] Elliott, G. and A. Timmermann (2016), *Economic Forecasting.* Princeton, Princeton University Press.

[6] Federal Reserve Board (2018), "2018 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule." Federal Reserve Board report.

[7] Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models.* Springer Series in Statistics, Springer, New York.

[8] Frühwirth-Schnatter, S. (2008), *Finite Mixture and Markov Switching Models. Implementation in MATLAB using the package `bayesf Version 2.0`.* Springer Series in Statistics, Springer, New York, http://statmath.wu.ac.at/~fruehwirth/monographie/.

[9] Ganics, G. (2017), "Optimal Density Forecast Combinations." Banco de España Working Paper N. 1751.

[10] Geweke, J. and G. Amisano (2011), "Optimal prediction pools." *Journal of Econometrics*, 164 (1), 130–141.

[11] Hall, S. G. and J. Mitchell (2007), "Combining density forecasts." *International Journal of Forecasting*, 23, 1–13.

[12] Hansen, B. E. (1994), "Autoregressive conditional density estimation." *International Economic Review*, 35 (3), 705–30.

[13] Hamilton, J. D. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle." *Econometrica*, 57, 357–384.

[14] Hamilton, J. D. (2018), "Macroeconomic Regimes and Regime Shifts." In *Handbook of Macroeconomics*, vol. 2, J. B. Taylor and H. Uhlig (eds.), pp. 163–201, Amsterdam, North-Holland.

[15] Moody's (2018), "U.S. Macroeconomic Outlook Alternative Scenarios." Moody's Analytics, https://www.economy.com/.

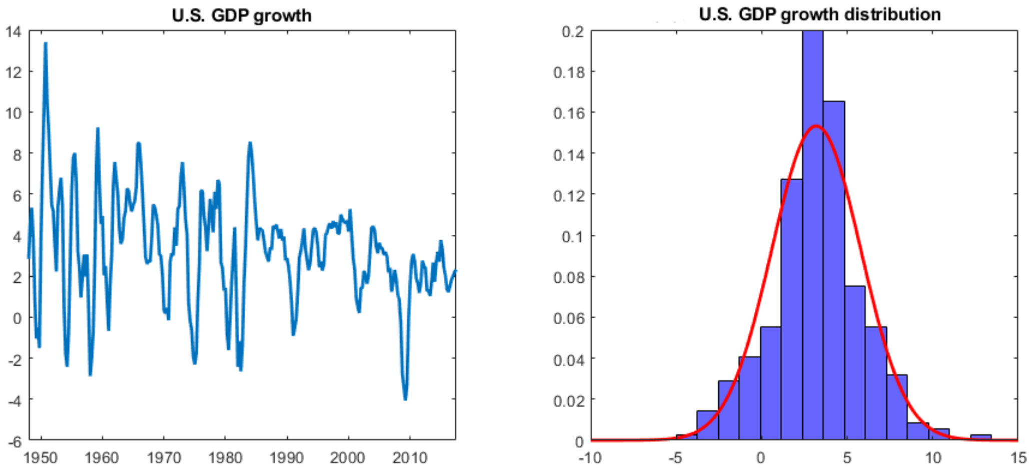
[16] Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006), "Forecasting time series subject to multiple structural breaks." *Review of Economic Studies*, 73, 1057–84.

[17] Rossi, B. and T. Sekhposyan T. (2014), "Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set." *International Journal of Forecasting*, 30 (3), 662–682.

[18] Yuen, K. (2013), "Determining the Severity of Macroeconomic Stress Scenarios." Federal Reserve Bank Supervisory Staff Reports.

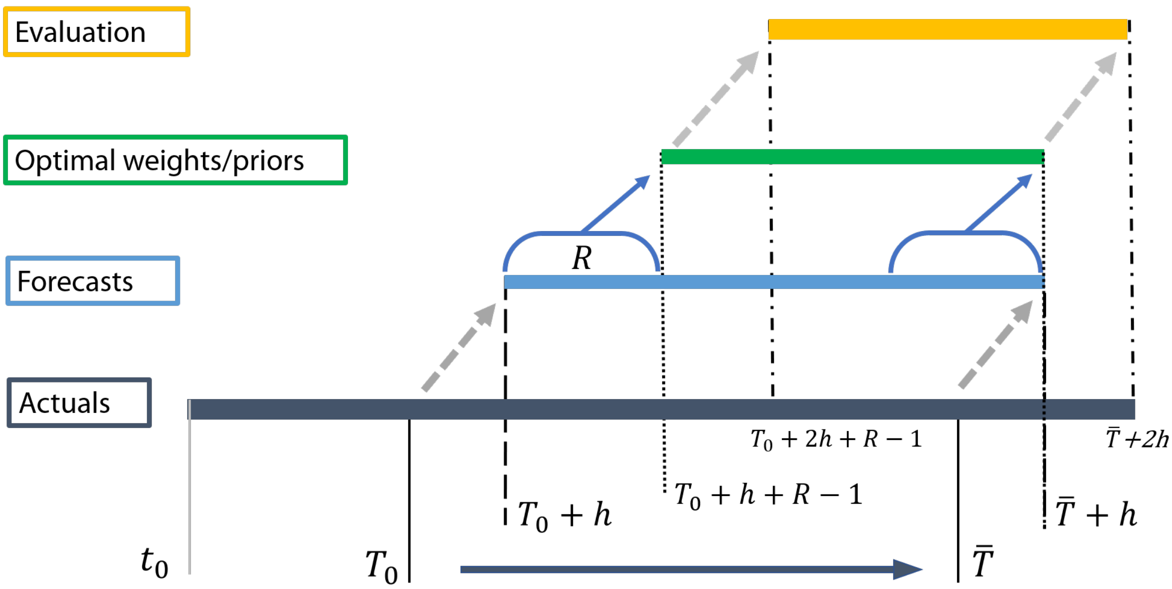
Figures and tables

Figure 3.1: U.S. real GDP growth 1948Q1-2017Q2



Notes: The left panel plots the quarterly time series of the U.S. real GDP growth rate (year-on-year) from 1948Q1 to 2017Q2. The histogram in the right panel summarizes the frequency distribution. The red line represents the normal p.d.f. with the same mean and variance as the empirical distribution. The Jarque-Bera test rejects the hypothesis of normality at the 5% level.

Figure 3.2: Optimization scheme



Notes: The figure summarizes the density forecast optimization scheme. First, the MSAR model is recursively estimated on actual GDP data (dark blue bar) using alternative views. The sample start date is denoted with t_0, the end date runs from T_0 to $\overline{T}$. For each sample window, the estimates generate density forecasts with horizon h (light blue bar). A rolling sequence of R forecasts is used to compute optimal forecast weights and prior probabilities (green bar) for the views. The optimal weights/priors obtained in each period are used to combine the view-specific forecasts for subsequent periods. The resulting composite forecasts (dark yellow bar) are evaluated by comparison with the actual data over the period from $T_0 + 2h + R - 1$ to $\overline{T} + 2h$.

Table 3.1: Alternative views for the MSAR model of U.S. GDP growth

view no.	view type	K	b_0	B_0	a_0	A_0	e	c_0	g_0	G_0
1	vague	1	0	1	(0.5, 0, 0, 0, 0)	1	2	3	0.5	0.5
2	vague	2	(0, 0)	1	(0.5, 0, 0, 0, 0)	1	2	3	0.5	0.5
3	vague	3	(0, 0, 0)	1	(0.5, 0, 0, 0, 0)	1	2	3	0.5	0.5
4	vague	4	(0, 0, 0, 0)	1	(0.5, 0, 0, 0, 0)	1	2	3	0.5	0.5
5	vague	5	(0, 0, 0, 0, 0)	1	(0.5, 0, 0, 0, 0)	1	2	3	0.5	0.5
6	Fed stress test	3	(0.265, -0.0475, -0.4275)	10^{-5}	(0.9, 0, 0, 0, 0)	10^{-5}	2	3	0.5	0.5
7	Fed stress test	3	(0.2275, -0.1850, -0.5675)	10^{-5}	(0.9, 0, 0, 0, 0)	10^{-5}	2	3	0.5	0.5
8	Fed stress test	3	(0.205, -0.1950, -0.59)	10^{-5}	(0.9, 0, 0, 0, 0)	10^{-5}	2	3	0.5	0.5
9	Fed stress test	3	(0.21, -0.2125, -0.6275)	10^{-5}	(0.9, 0, 0, 0, 0)	10^{-5}	2	3	0.5	0.5
10	Fed stress test	5	(0.39, 0.1975, 0.265, -0.0475, -0.4275)	10^{-5}	(0.9, 0, 0, 0, 0)	10^{-5}	2	3	0.5	0.5
11	Fed stress test	5	(0.39, 0.3, 0.2275, -0.1850, -0.5675)	10^{-5}	(0.9, 0, 0, 0, 0)	10^{-5}	2	3	0.5	0.5
12	Fed stress test	5	(0.39, 0.3, 0.205, -0.1950, -0.59)	10^{-5}	(0.9, 0, 0, 0, 0)	10^{-5}	2	3	0.5	0.5
13	Fed stress test	5	(0.43, 0.32, 0.21, -0.2125, -0.6275)	10^{-5}	(0.9, 0, 0, 0, 0)	10^{-5}	2	3	0.5	0.5

Notes: The table lists the 13 views considered in the empirical application. K denotes the number of regimes, $b_0, B_0, a_0, A_0, e, c_0, g_0$ and G_0 are the hyperparameters of the Bayesian MSAR model, as described in the text.

Table 3.2: Fed stress tests 2015-2018: scenarios of U.S. GDP growth

	2015			2016			2017			2018		
time	base	adv.	sev.	base	adv.	sev.	base	adv.	sev.	base	adv.	sev.
2014Q4	3	-0.6	-3.9									
2015Q1	2.9	-1.3	-6.1									
2015Q2	2.9	-0.2	-3.9									
2015Q3	2.9	0.2	-3.2									
2015Q4	2.9	0.3	-1.5									
2016Q1	2.9	0.8	1.2	2.5	-1.5	-5.1						
2016Q2	2.9	1.2	1.2	2.6	-2.8	-7.5						
2016Q3	2.9	1.7	3	2.6	-2	-5.9						
2016Q4	2.9	1.8	3	2.5	-1.1	-4.2						
2017Q1	2.7	1.8	3.9	2.4	0	-2.2	2.2	-1.5	-5.1			
2017Q2	2.7	1.9	3.9	2.5	1.3	0.4	2.3	-2.8	-7.5			
2017Q3	2.6	2	3.9	2.3	1.7	1.3	2.4	-2	-5.9			
2017Q4	2.6	2.2	3.9	2.3	2.6	3	2.3	-1.5	-5.1			
2018Q1				2.6	2.6	3	2.4	-0.5	-3	2.5	-1.3	-4.7
2018Q2				2.4	3	3.9	2.4	1	0	2.8	-3.5	-8.9
2018Q3				2.3	3	3.9	2.4	1.4	0.7	2.6	-2.4	-6.8
2018Q4				2.3	3	3.9	2.3	2.6	3	2.5	-1.3	-4.7
2019Q1				2.1	3	3.9	2	2.6	3	2.3	-0.7	-3.6
2019Q2							2.1	3	3.9	2.3	0.4	-1.3
2019Q3							2.1	3	3.9	2.1	1	-0.2
2019Q4							2	3	3.9	2	2.5	2.8
2020Q1							2	3	3.9	2.1	2.8	3.5
2020Q2										2.1	3	4
2020Q3										2.1	3.2	4.2
2020Q4										2.1	3.3	4.5
2021Q1										2.1	3.3	4.5

Notes: For each year from 2015 to 2018, the table reports the baseline, adverse and severely adverse supervisory scenarios for U.S. GDP growth (quarter-on-quarter, annualized) included in the annual stress tests conducted by the Federal Reserve.

Table 3.3: Density forecast performance of optimal pools of MSAR views vs. benchmarks

	APD	KS	LB1	LB2
AR (view no. 1)	0.27	0.00	0.61	0.01
View max marg. lik.	0.35	0.03	0.73	0.84
AR rolling window	0.33	0.10	0.59	0.00
Equal forecast weights	0.31	0.01	0.39	0.34
Equal prior prob.	0.35	0.02	0.70	0.83
Optimal weights $\mathbf{w}_1^*$	0.35	0.08	0.69	0.81
Optimal priors $\boldsymbol{\pi}_1^*$	0.35	0.06	0.74	0.75
Optimal weights $\mathbf{w}_2^*$	0.32	0.21	0.26	0.80
Optimal priors $\boldsymbol{\pi}_2^*$	0.33	0.32	0.36	0.89

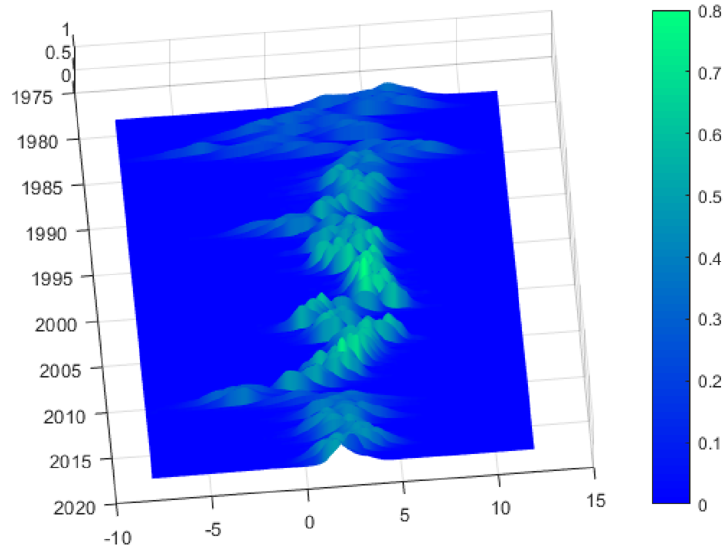
Notes: APD denotes the average predictive density, KS denotes the p-value of the Kolmogorov-Smirnov test of uniformity of the PITs. LB1 and LB2 denote the p-values of the Ljung-Box test of serial independence in the first and second moment of the PITs, respectively. All statistics are computed over the period 1978Q1-2017Q2.

Table 3.4: Density forecast performance of Student-t AR, AR-ARCH and AR-GARCH models

	APD	KS	LB1	LB2
AR(5) with t errors (recursive)	0.30	0.00	0.56	0.06
AR(5) with t errors (rolling 80)	0.37	0.50	0.76	0.04
AR(5) with t errors (rolling 40)	0.37	0.32	0.96	0.06
AR(5)-ARCH(1) (recursive)	0.27	0.00	0.57	0.00
AR(5)-ARCH(1) (rolling 80)	0.32	0.12	0.82	0.00
AR(5)-ARCH(1) (rolling 40)	0.33	0.69	0.94	0.00
AR(5)-GARCH(1,1) (recursive)	0.20	0.00	0.92	0.00
AR(5)-GARCH(1,1) (rolling 80)	0.29	0.00	0.79	0.00

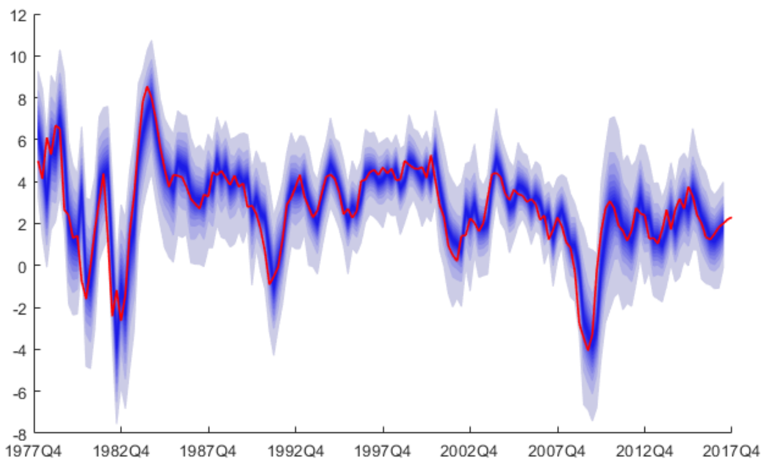
Notes: APD denotes the average predictive density, KS denotes the p-value of the Kolmogorov-Smirnov test of uniformity of the PITs. LB1 and LB2 denote the p-values of the Ljung-Box test of serial independence in the first and second moment of the PITs, respectively. The AR-GARCH model on 40-quarter rolling windows is not supported by the data and is therefore not reported. All statistics are computed over the period 1978Q1-2017Q2.

Figure 3.3: Calibrated regime-switching density forecasts: evolution of p.d.f. over time



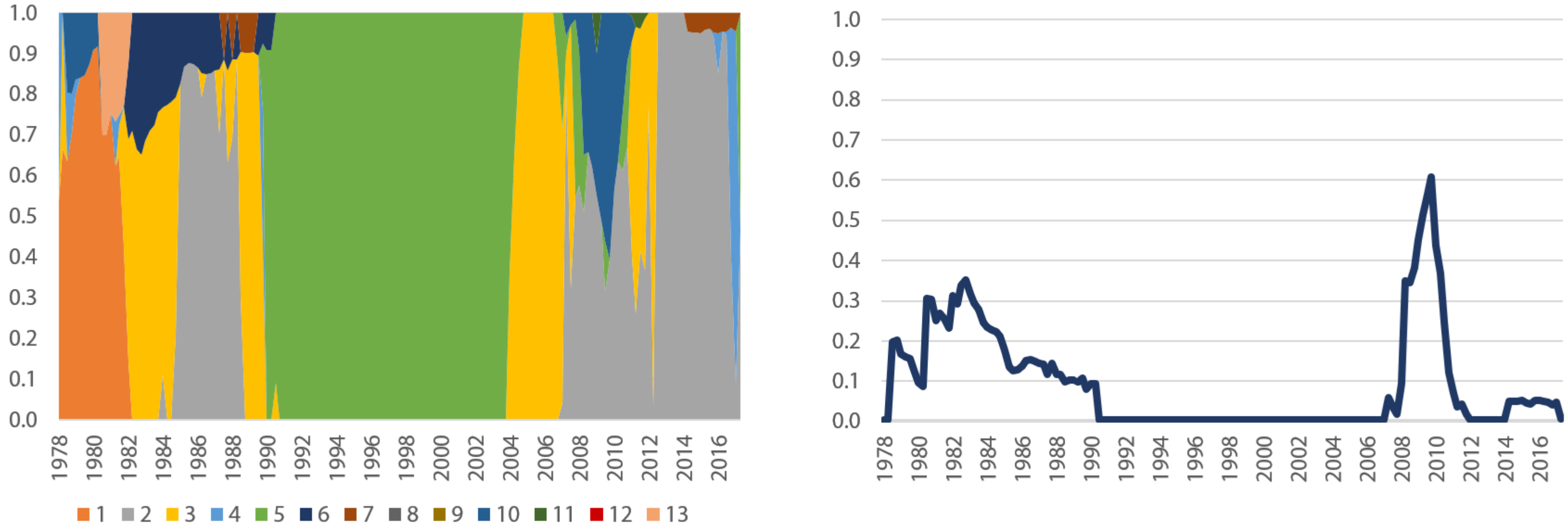
Notes: For each quarter from 1978Q1 to 2017Q2, the figure plots the p.d.f. of the 1-quarter-ahead composite density forecast produced in the previous quarter using the PIT-based optimal priors.

Figure 3.4: Calibrated regime-switching density forecasts: fan chart



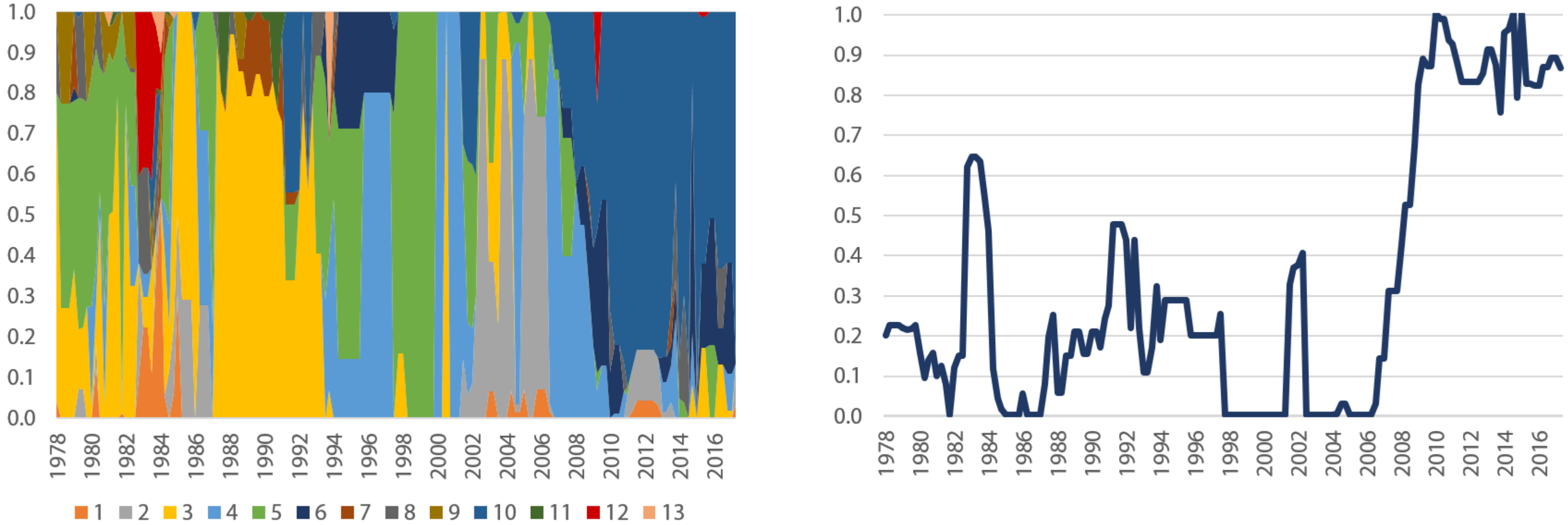
Notes: In this fan chart, different shades of color identify different percentiles (0.01, from 0.05 to 0.95 in steps of 0.05, and 0.99). For each quarter in 1978Q1-2017Q2, the chart summarizes the 1-quarter-ahead composite density forecasts produced in the previous quarter using the PIT-based optimal priors. The red line is the realized time series of GDP growth.

Figure 3.5: Optimal weights $\mathbf{w}_1^*$ over time



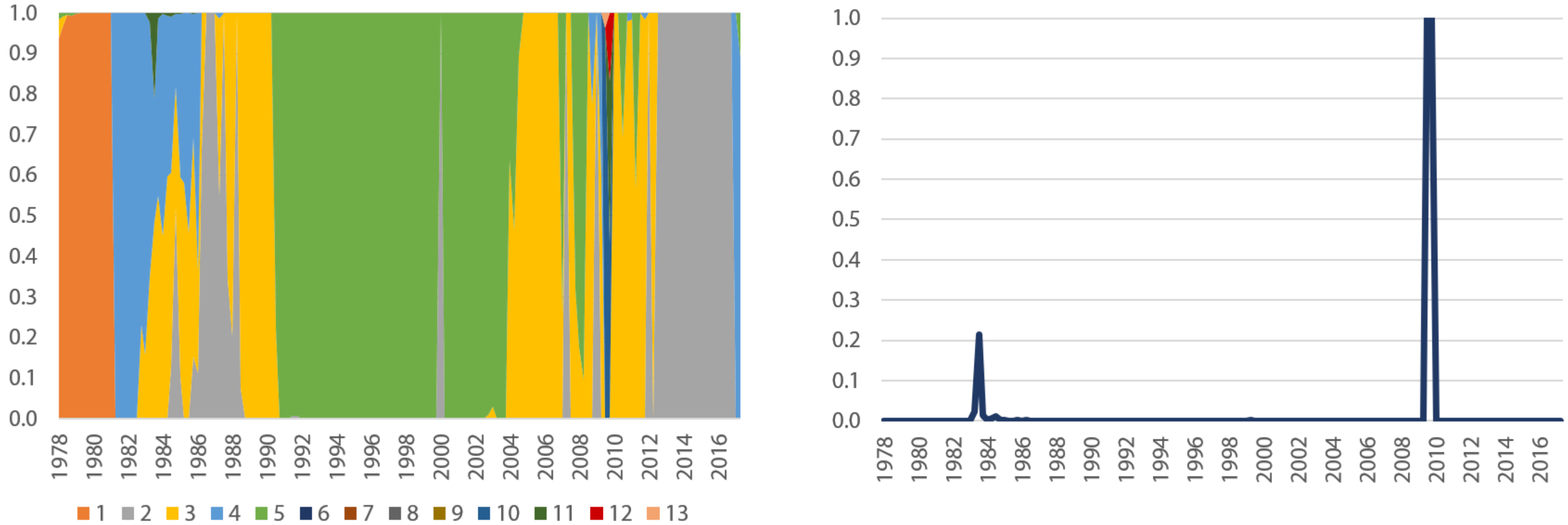
Notes: The area chart in the left panel shows the time-varying weights for all 13 views from 1978Q1 to 2017Q2. The right panel plots the cumulative weight assigned to the views derived from Fed supervisory scenarios. See Table 3.1.

Figure 3.6: Optimal weights $\mathbf{w}_2^*$ over time



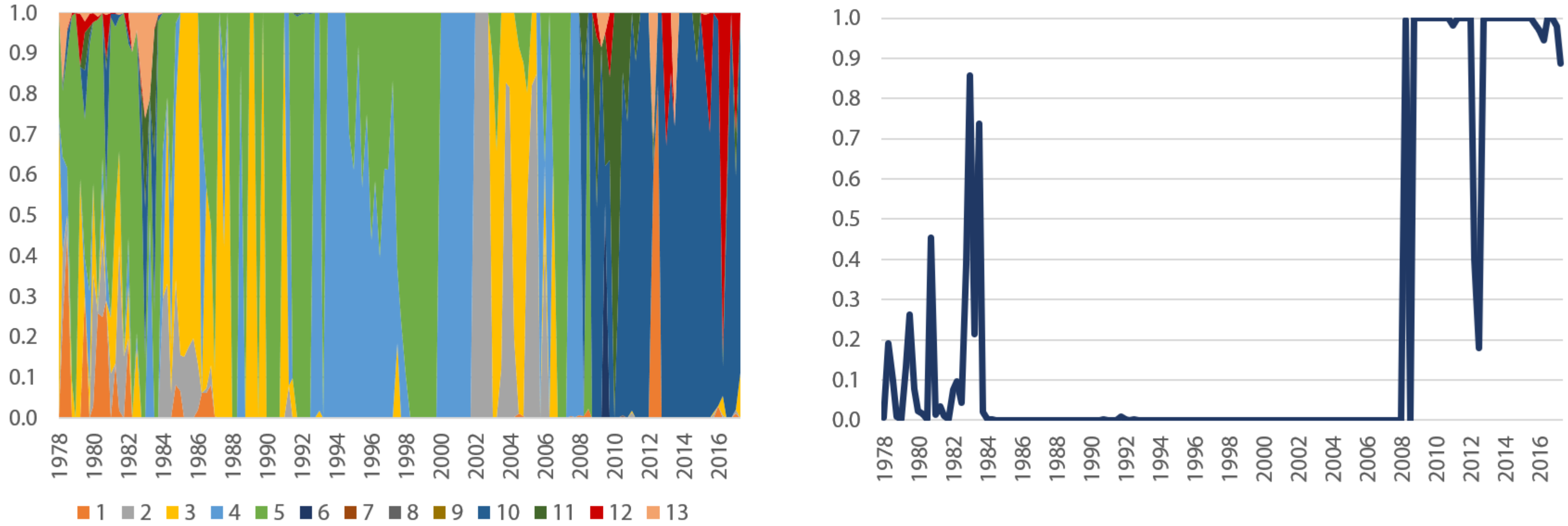
Notes: The area chart in the left panel shows the time-varying weights for all 13 views from 1978Q1 to 2017Q2. The right panel plots the cumulative weight assigned to the views derived from Fed supervisory scenarios. See Table 3.1.

Figure 3.7: Posterior probabilities from $\boldsymbol{\pi}_1^{0*}$ over time



Notes: The area chart in the left panel shows the posterior weights for all views from 1978Q1 to 2017Q2. The right panel plots the cumulative posterior weight assigned to the views derived from Fed supervisory scenarios. See Table 3.1.

Figure 3.8: Posterior probabilities from $\boldsymbol{\pi}_2^{0*}$ over time



Notes: The area chart in the left panel shows the posterior weights for all views from 1978Q1 to 2017Q2. The right panel plots the cumulative posterior weight assigned to the views derived from Fed supervisory scenarios. See Table 3.1.

Table 3.5: Density forecast performance of optimal pools of views vs. benchmarks ($R = 20$)

	APD	KS	LB1	LB2
AR (view no. 1)	0.26	0.00	0.68	0.00
View max marg. lik.	0.33	0.01	0.82	0.64
AR rolling window	0.31	0.12	0.51	0.00
Equal forecast weights	0.30	0.01	0.33	0.45
Equal prior prob.	0.33	0.01	0.80	0.59
Optimal weights $\mathbf{w}_1^*$	0.33	0.12	0.52	0.54
Optimal priors $\boldsymbol{\pi}_1^*$	0.33	0.05	0.68	0.41
Optimal weights $\mathbf{w}_2^*$	0.30	0.12	0.05	0.65
Optimal priors $\boldsymbol{\pi}_2^*$	0.31	0.16	0.09	0.42

Notes: This table shows the results obtained using a rolling sequence of $R = 20$ forecasts to compute optimal weights and priors. APD denotes the average predictive density, KS denotes the p-value of the Kolmogorov-Smirnov test of uniformity of the PITs. LB1 and LB2 denote the p-values of the Ljung-Box test of serial independence in the first and second moment of the PITs, respectively. All statistics are computed over the period 1973Q1-2017Q2 (see Figure 3.2).

Appendix 3.A Prior on the regime-switching variance

Based on the properties of the Gamma and inverted Gamma distributions, it holds that:

$$\mathrm{E}(\sigma_k^2|C_0) = \frac{C_0}{c_0 - 1}$$

$$\mathrm{Var}(\sigma_k^2|C_0) = \frac{C_0^2}{(c_0-1)^2(c_0-2)}$$

$$\mathrm{E}(C_0) = \frac{g_0}{G_0} = 1$$

$$\mathrm{Var}(C_0) = \frac{g_0}{G_0^2} = 2$$

$$\mathrm{E}(C_0^2) = \left(\frac{g_0}{G_0}\right)^2 + \frac{g_0}{G_0^2} = 3$$

Given the values of the hyperparameters, $c_0 = 3$, $g_0 = 0.5$ and $G_0 = 0.5$, it follows that:

$$\mathrm{E}(\sigma_k^2) = \frac{\mathrm{E}(C_0)}{c_0 - 1} = 0.5$$

and

$$\begin{aligned}
\mathrm{Var}(\sigma_k^2) &= \mathrm{E}(\mathrm{Var}(\sigma_k^2|C_0)) + \mathrm{Var}(\mathrm{E}(\sigma_k^2|C_0)) = \\
&= \frac{\mathrm{E}(C_0^2)}{(c_0-1)^2(c_0-2)} + \frac{\mathrm{Var}(C_0)}{(c_0-1)^2} = \\
&= \frac{3}{4} + \frac{1}{2} = 1.25
\end{aligned}$$