ESSAYS IN BAYESIAN MACROECONOMETRICS AND FORECASTING

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

vorgelegt von
M.Sc.
Philipp Hauber
aus Sindelfingen

Kiel, 2022

Acknowledgements

Contents

List of Figures

List of Tables

Introduction

Forecasting macroeconomic conditions in real-time is a crucial prerequisite for the conduct of economic policy. The interest in and importance of timely and accurate forecasts have led to active research in the areas of macroeconomics and econometrics. My dissertation makes a contribution to three strands in this literature:

i) the **improvement of macroeconometric forecasting models** by considering the question of variable selection in the context of large factor models,

ii) the **econometric methodology** by extending so-called precision samplers to applications with missing observations and

iii) the **incorporation of external information into macroeconometric models** by evaluating forecasts from a large macroeconometric model conditional on professional forecasters' view on key variables like the growth rate of the gross domestic product (GDP) and consumer price index (CPI) inflation.

Common to all my papers and in line with a growing trend in the field is the use of Bayesian estimation techniques. Moreover, the forecast evaluations are conducted in real-time, i.e. they exactly replicate the information set available at the time the forecasts were made, thus yielding a more accurate measure of the predictive performance.

In the first chapter of my dissertation - **Real-time nowcasting with sparse factor models** - I address the issue of variable selection in factor models. While factor models do not suffer from the "curse of dimensionality" like, say, vector autoregressions (VAR) and can therefore in principle be estimated with a large number of macroeconomic time series, some papers have shown that preselection of variables - using expert judgment or statistical procedures - may yield more precise estimates of the underlying factors and better forecasts. My paper addresses this issue by combining model estimation and variable selection in a single step rather than preselecting variables or relying on the complete cross-section of available macroeconomic variables. I do so in a Bayesian framework by employing so-called sparse priors on the elements of the loadings matrix that maps the observables to the factors. These priors concentrate considerably more mass at zero (compared to Normal priors) while still retaining fat tails (in contrast to shrinkage priors which merely "shove" the entire distribution towards 0 as is often done when estimating larger VAR to reduce parameter uncertainty).

In effect, this allows to filter out irrelevant variables, i.e. those that do not adhere to the factor structure and are irrelevant for forecasting. In a real-time nowcasting excercise for GDP growth in the United States and Germany, however, I find only small gains in terms of point and density accuracy from employing these sparse priors compared to Normal priors commonly used in the literature. These findings suggest that the approach in parts of the literature to estimate factor models with large cross-sections - in practice often close to or more than 100 time series - is justified.

Precision-based sampling with missing observations: A factor model application, the second chapter in my dissertation and coauthored with Dr. Christian Schumacher of the Deutsche Bundesbank, makes a methodological contribution to the literature on the Bayesian estimation of state space models. Precision samplers have been proposed in the literature as an alternative to Kalman filter-based methods to draw from the conditional distribution of unobserved components given data and parameters. These samplers, increasingly used in macroeconomics, exploit the fact that the precision matrix - the inverse of the covariance matrix - of the vector of unobserved components is typically sparse and banded in macroeconomic applications. Modern programming languages like MATLAB or `Julia` can exploit this structure to solve systems of linear equations or perform operations like the Cholesky decomposition efficiently. This makes sampling from the desired distribution feasible, often computationally more efficient and numerically stable than existing samplers based on the Kalman filter. Up to now, however, the literature on precision samplers has only considered applications involving complete datasets. Our contribution is to show how the sampler can be used when some of the observations in the data are missing, greatly increasing the range of applications of precision samplers. To this end we derive permutations of the vector of observables that preserve the bandedness of the precision matrix and thus still allow for efficient sampling. In simulations, we analyze the properties of two different approaches: joint sampling of states and missing observations as well as sequentially drawing missing observations (conditional on the states) and states (conditional on a complete dataset). We apply our sampler to the synchronization of business cycles across countries using a large, unbalanced dataset.

Since future values of time series - unconditional or conditional on some other variables in the model - can be thought of as missing values, the precision sampler outlined in Chapter II can also be applied to obtain density forecasts. In the last chapter of my dissertation - **How useful is external information from professional forecasters? Conditional forecasts in large factor models** - I propose algorithms to obtain draws from the predictive density in the context of a large factor model. Simulations show that this approach compares favorably to Kalman filter-based approaches typically used in the literature. The insights are then used to investigate to what extent the predictions of professional forecasters regarding GDP growth and CPI inflation contain information for other macroeconomic variables such as the components of gross domestic product or gross value added, the corresponding defla-

tors as well as production or labor market indicators. In a real-time forecast evaluation I find that conditioning on professional forecasts yields more accurate predictions for those variables for which the model already produces relatively accurate unconditional forecasts (real activity indicators and some price series). However, these gains in predictive accuracy are only significant when we evaluate the entire predictive distribution; for point forecasts, this is is not the case. Robustness checks reveal that while the choice of model specification has no bearing on the results, the evaluation period does: when the global financial crisis is excluded from the analysis the conditional forecasts are on average less accurate and in some cases the model forecasts do not outperform the autoregressive benchmark.

If I had to summarize my dissertation for a general, non-expert audience in a few lines, I would focus on two points - both backed up by my year's of experience as a forecaster and undoubetedly shared by many colleagues in the profession. *Firstly*, as Chapter I of my dissertation highlights, it is difficult to find a small subset from the multitude of available time series that consistently forecasts macroeconomic activity well. Indeed, my research seems to support the notion that macroeconomic data that are commonly used to forecast macroeconomic activity are dense, not sparse. Instead, analysts of macroeconomic should (and typically do!) monitor a large number of indicators and incorporate these into their models. *Secondly*, while macroeconometric models have their place in business cycle analysis, practicioners should be aware of their limits in terms of forecast performance. As the third chapter of my dissertation shows, even when we condition on the forecasts of professional forecasters for GDP growth and inflation, it is difficult to find significant improvements for other variables - at least in terms of point forecast accuracy which practicioners still predominantly focus on. As such, macroeconomic models should be viewed as a complement not a substitute to expert judgment when it comes to short-term forecasting.

Chapter I

Real-time nowcasting with sparse factor models

Abstract

Factor models feature prominently in the macroeconomic nowcasting literature, yet no clear consensus has emerged regarding the question of how many and which variables to select in such applications. Examples of both large-scale models, estimated with data sets consisting of over 100 time series as well as small-scale models based on only a few, pre-selected variables can be found in the literature. To adress the issue of variable selection in factor models, in this paper we employ sparse priors on the loadings matrix. These priors concentrate more mass at zero than those conventionally used in the literature while retaining fat tails to capture signals. As a result, variable selection and estimation can be performed simultaneously in a Bayesian framework. Using large data sets consisting of over 100 variables, we evaluate the performance of sparse factor models in real-time for US and German GDP point and density nowcasts. We find that sparse priors lead to relatively small gains in nowcast accuracy compared to a benchmark Normal prior. Moreover, different types of sparse priors discussed in the literature yield very similar results. Our findings are compatible with the hypothesis that large macroeconomic data sets typically used in now- or forecasting applications are not sparse but dense.

Keywords: factor models, sparsity, nowcasting, variable selection

JEL classification: C11, C53, C55, E37.

---

[1]Available under `http://hdl.handle.net/10419/251551`

I.1   Introduction

Professional forecasters and policy-makers require timely assessements of the current state of the macroeconomy. However, short-term forecasting or *nowcasting* of the gross domestic product (GDP) - arguably the best single indicator of macroeconomic developments and central in guiding economic policy decisions - faces numerous challenges. The GDP is subject to a considerable publication delay, with a first estimate usually provided roughly four weeks after the end of the respective quarter. At the same time, the information from a large number of potentially informative indicators is available that can be exploited in a more timely manner to improve estimates of current macroeconomic conditions. Often, however, such series are published in an asynchronous manner giving rise to an unbalanced panel or "ragged edge". For example, survey-based sentiment indicators are typically released much sooner after (or even within) the reference period than "hard" indicators such as industrial production or retail sales. Moreover, these time series are usually sampled at higher frequencies than the target variable GDP, e.g. monthly or daily. Nowcasts that reflect on the latest available information need to be able to handle this data flow in real-time, mixing information from different frequencies.

Factor models feature prominently in the nowcasting literature as they can address all the aforementioned issues in a unified modelling framework. Furthermore, they have a proven track record in terms of nowcasting GDP (Giannone et al., 2008; Schumacher and Breitung, 2008; Camacho and Perez-Quiros, 2010; Banbura and Rünstler, 2011; Kuzin et al., 2011; Aastveit et al., 2014, 2018; Marcellino et al., 2016). Notwithstanding this extensive literature, there is no firm consensus about how many and which variables to select when nowcasting GDP. From an asymptotic point of view a larger cross-section should lead to more precise estimates of the underlying factors and, hence, more accurate nowcasts. Boivin and Ng (2006), however, find both in simulations and a real-time forecasting excercise that factors extracted from smaller data sets perform as well or even better than those extracted from a much larger panel. In line with these findings, Bai and Ng (2008) advocate penalized regressions to identify a subset of variables that is closely related to the target series prior to factor extraction ("targeted predictors"). Schumacher (2010) demonstrates that such targeted predictors can improve nowcasts of German GDP when considering a large panel of international data. Similarly, Caggiano et al. (2011) provide evidence for a number of countries that pre-selection of variables can substantially improve forecast performance. In contrast, Alvarez et al. (2016) find no clear benefit from using aggregate headline series representing different economic categories as opposed to a large disaggregated data set in a forecasting excercise for the United States. Rünstler (2016) adresses the issue of variable selection by focusing on the prediction weights inherent in a factor model. Selecting those variables with the largest marginal predictive gains for GDP growth, he finds only moderate gains in forecasting accuracy for short horizons.

As a result of this on-going debate, examples of both approaches can be found in the literature: "large-scale"" factor models comprised of around 70 to 80 time series or more

(Giannone et al., 2008; Banbura and Rünstler, 2011; Kuzin et al., 2011; Aastveit et al., 2014, 2018) or "small-scale " models with up to 10 or 20 variables, pre-selected by expert judgement, statistical procedures or a combination thereof (Camacho and Perez-Quiros, 2010; Marcellino et al., 2016; Bok et al., 2018).

In this paper, the issue of variable selection is adressed by exploring the role sparsity plays in factor models. We investigate to what extent the Bayesian estimation of factor models with sparse priors on the loadings matrix can serve as an alternative to pre-selection of variables. Compared to Normal priors commonly used in the literatur, sparse priors place considerably more mass near zero while still allowing for fat tails to capture signals. Based on the work by George and McCulloch (1993, 1997) on variable selection priors in a multiple regression framework, West (2003) proposes mixtures of Normal distributions and a point mass at zero to induce sparsity in factor models. Employing these priors in a macroeconomic application, Kaufmann and Schumacher (2017) identifiy relevant variables in large panels of international GDP growth rates and disaggregate US CPI data. Global-local shrinkage priors have been proposed as a continuous approximation to discrete variable selection priors (Polson and Scott, 2010) and have been widely used in macroeconomic forecasting applications with Bayesian vector autoregressions (BVAR), e.g. Kastner and Huber (2020), Huber and Feldkircher (2019), Cross et al. (2020) or Chan (2021) in combination with conventional Minnesota-type shrinkage priors.

To assess the performance of sparse factor models, we conduct a real-time evaluation of point and density nowcasts for US and German GDP. Our contribution to the literature can be viewed in two ways: from the methodological side, we perform variable selection and estimation simultaneously in a real-time factor model setting. As such, it follows the recent contributions of Kristensen (2017) and Thorsrud (2020). The former estimates factors nonparametrically via sparse principal components and finds gains in forecasting performance while the latter induces time-varying sparsity in a Bayesian factor model using a latent threshold mechanism (Nakajima and West, 2013). A further novel aspect of our paper is a comparsion of different sparse priors that have been proposed in the literature, with a particular focus on their performance in nowcasting applications. On a conceptual level, we contribute to the current debate about the degree of sparsity in macroeconomic data and its implications for now- and forecasting (Giannone et al., 2018).

The real-time evaluation of nowcasts for US and German GDP suggests that while the factor models outperform autoregressive benchmarks, sparse priors lead to relatively small gains in nowcast accuracy compared to our benchmark Normal prior. Moreover, different types of sparse priors discussed in the literature yield very similar results. Our findings are thus compatible with the hypothesis that large macroeconomic data sets typically used in now- or forecasting applications are not sparse but dense. The remainder of this paper is structured as follows: Section 2 discusses the different sparse priors and lays out their conditional posterior distributions. The results from the real-time nowcast evaluation for US and German GDP along with a number of robustness checks are presented in Section 4. Section 5 concludes.

Notation: Let $z_{1:T} = [z_1, \ldots, z_T]'$ refer to the $T \times N$ matrix where z_t is an $N \times 1$ vector, and $z_{i,1:T} = [z_{i1}, \ldots, z_{iT}]'$ to the vector of length T corresponding to the i-th column of $z_{1:T}$. Finally, let 0_k (I_k) denote a zero column vector (identity matrix) of dimension k. Furthermore, let A_{ij} refer to the element in the i-th row and j-th column of matrix A. Conversely, $A_{i.}$ and $A_{.j}$ denote the i-th row and j-th column, respectively. Additionally, we denote by $\mathcal{N}(\mu, \sigma^2)$ the Normal distribution with mean μ and variance σ^2. Furthermore, let $\mathcal{N}(x; \mu, \sigma^2)$ denote the value of the probability density function (pdf) of $\mathcal{N}(\mu, \sigma^2)$ evaluated at x. We denote by $\mathcal{U}(a, b)$ the continuous uniform distribution on the interval a to b and by $\mathcal{G}(u, U)$ the Gamma distribution with shape and rate parameters u and U, respectively. Its pdf is given by

$$\mathcal{G}(x; u, U) = \frac{U^u}{\Gamma(u)} x^{u-1} \exp(-Ux).$$

Similarly, let $\mathcal{G}^{-1}(u, U)$ denote the Inverse Gamma distribution with pdf

$$\mathcal{G}^{-1}(x; u, U) = \frac{U^u}{\Gamma(u)} x^{-u-1} \exp\left(-\frac{U}{x}\right).$$

$\mathcal{B}(a, b)$ denotes the Beta distribution with pdf

$$\mathcal{B}(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}.$$

Also, let $\mathcal{C}^+(\mu, \gamma)$ denote the half-Cauchy distribution with location and scale parameters, μ and γ, respectively and pdf equal to

$$\mathcal{C}^+(x; \mu, \gamma) = 2\left(\pi\gamma\left[1 + ((x-\mu)\gamma)^2\right]\right)^{-1}.$$

I.2 Sparse factor models

I.2.1 Factor model

Let x_t denote an $N \times 1$ vector of time series observed at time $t = 1, \ldots, T$. The idea of a factor model is that x_t can be expressed as the sum of two orthogonal elements: a common and an idiosyncratic component, i.e.

$$x_t = \lambda f_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon) \tag{I.1}$$

$$\Psi(L) f_t = \eta_t, \quad \eta_t \sim \mathcal{N}(0, \Sigma_\eta) \tag{I.2}$$

where λ is the $N \times R$ loadings matrix linking the $R \times 1$ vector of factors, f_t to the observables in x_t. The product of loadings and factors constitutes the common component of the model while ϵ_t captures variable-specific, idiosyncratic developments. The dynamics of the factors are given by stationary vector autoregressions of order P. Σ_ε is a diagonal matrix, implying that the idiosyncratic components are independent of each other and we denote by σ_i^2 the

element corresponding to the i-th variable. Furthermore, the innovations to the common factors and the idiosyncratic components are uncorrelated, i.e. $\mathbb{E}\left[\eta_t \varepsilon_t'\right] = 0$. Lastly, to identify the scale of the factors we set $\Sigma_\eta = I_R$.

I.2.2 Sparse priors

From a Bayesian perspective, there are two main sparse prior alternatives: discrete mixtures and pure shrinkage priors. The former prior combines a point mass at zero or a continuous distribution with small variance and an absolutely continuous prior providing tail mass, whereas the latter is solely based on a continuous prior providing shrinkage towards zero (Carvalho et al., 2009).

A standard example for a discrete-mixtures prior is the variable selection prior for multiple regression proposed by George and McCulloch (1993) and George and McCulloch (1997). This prior was adopted and modified for factor models by West (2003) and has the following form:

$$\lambda_{ir}|\rho_j, \tau_r \sim (1-\rho_r)\delta_0(\lambda_{ir}) + \rho_r \mathcal{N}(0, \tau_r), \tag{I.3}$$

$$\rho_r|r_0 s_0 \sim \mathcal{B}(r_0 s_0, r_0(1-s_0)) \tag{I.4}$$

$$\tau_r|a_0, b_0 \sim \mathcal{G}^{-1}(a_0, b_0). \tag{I.5}$$

The unit point mass at zero is denoted as $\delta_0(\cdot)$, and nonzero loadings on factor j are drawn from a Normal prior with variance τ_r. Following the applications in Kaufmann and Schumacher (2017) and Kaufmann and Schumacher (2019), the probability of non-zero loading ρ_r is factor-specific as well as the variance τ_r. ρ_r is the probability of non-zero loading and follows a Beta distribution, and scale τ_r an inverse Gamma. An extension to element-wise ρ_{ir} is discussed in Carvalho et al. (2008). However, simulation results in Kaufmann and Schumacher (2017) do not show huge differences in practice. From here on, we call the prior (I.3) the point-mass normal mixture prior, or PMNM prior in brief.

Global-local shrinkage priors are typically normal distributions, with a global and a local variance component (Polson and Scott, 2010). The idea is that global shrinkage handles the noise, whereas local variances act to detect the signals. Thus, the two components solve the trade-off between shrinking the noise toward zero leaving the large signals unshrunk.

The literature on global-local shrinkage priors mainly focusses on the application for multiple regression problems, not on factor models. An exception is the global-local prior by Bhattacharya and Dunson (2011):

$$\lambda_{ir}|\phi_{ir}, \kappa_r \sim \mathcal{N}(0, \phi_{ir}^{-1}\kappa_r^{-1}), \quad \kappa_r = \prod_{l=1}^{r} \delta_l, \tag{I.6}$$

$$\delta_1|a_1 \sim \mathcal{G}(a_1, 1) \tag{I.7}$$

$$\delta_l|a_2 \sim \mathcal{G}(a_2, 1) \quad \text{for} \quad l = 2, \ldots, r, \tag{I.8}$$

$$\phi_{ir} \sim \mathcal{G}(w/2, w/2). \tag{I.9}$$

In this prior, global shrinkage is governed factor-specific through κ_r. By the multiplicative gamma structure with $\delta_1, \ldots, \delta_r$, the κ_rs are stochastically increasing under the restriction $a_2 > 1$, which favours more shrinkage as the column index of the loading matrix increases. Local shrinkage is governed by loading-specific ϕ_{ir}, which follows a Gamma distribution. From here on, we call this prior the multiplicative-gamma prior, or MG prior in brief.

The horseshoe prior proposed by Carvalho et al. (2009) and Carvalho et al. (2010) is based on a standard half-Cauchy distribution for the local and the global scales. It has the hierarchical representation

$$\lambda_{ir}|\zeta_{ir}^2, \upsilon^2 \sim N(0, \zeta_{ir}^2 \upsilon^2) \tag{I.10}$$

$$\zeta_{ir} \sim \mathcal{C}^+(1), \qquad \upsilon \sim \mathcal{C}^+(1). \tag{I.11}$$

With respect to its shrinkage properties, the Cauchy tails in ζ_{ir} allow strong signals to remain large a posteriori. At the same time, its infinitely tall spike at the origin provides severe shrinkage for the zero elements of λ_{ir}. Note that the global variance parameter υ is applied to the whole set of loadings, not to columns of the loading matrix, and also follows a half-Cauchy distribution. Note that the horseshoe prior is free of user-chosen hyper parameters (Carvalho et al., 2010). An extension to accommodate more sparsity is the horseshoe+ prior by Bhadra et al. (2017). It is defined as (I.10) and (I.11) before, but with an augmented prior for ζ_{ir} according to $\zeta_{ir}|\chi_{ir} \sim \mathcal{C}^+(0, \chi_{ir})$ and $\chi_{ir} \sim \mathcal{C}^+(0, 1)$. From here on, we abbreviate the two priors as HS for horseshoe and HS+ for horseshoe+.

A natural benchmark to the sparse priors discussed above is a Normal prior on the elements of the loadings matrix, i.e. $\lambda_{ir} \sim \mathcal{N}(0, V_{ir})$. Uninformative priors - common in the nowcasting literature with Bayesian factor models - are imposed by setting V_{ir} to a large value. A hierarchical alternative is to consider $V_{ir} = \tau_r$ with $\tau_r \sim \mathcal{G}^{-1}(g_0, G_0) \ \forall \ r = 1, \ldots, R$ and $\forall \ i = 1, \ldots, N$. We will refer to the benchmark prior as Normal-Inverse Gamma (NIG).

Discussion of the alternative priors

In the literature, discrete mixture priors play a prominent role due to their theoretical properties. In particular, point-mass mixture priors are highly appealing by allowing for separate control of the level of sparsity and the size of the signal coefficients (Bhattacharya et al., 2015). In a multivariate normal mean model context, Castillo and van der Vaart (2012) show that the point-mass mixture prior with an appropriate beta prior on the inclusion probability and suitable tail conditions on the normal component leads to a optimal rate of posterior contraction. The posterior of the location parameter concentrates most of its mass on a ball around its true value. When inferring about sparsity on the true covariance matrix implied by a factor model, Pati et al. (2014) show that the point-mass mixture prior on the factor loadings leads to a consistent estimation of the covariance matrix when $N > T$.

In Polson and Scott (2010), the point-mass mixture prior serves as a benchmark to compare alternative global-local shrinkage priors. Following Polson and Scott (2010), the global-local shrinkage priors implying a weaker concept of sparsity than the point-mass

mixture prior, since all of the entries are assumed to be nonzero, yet most of them small compared to a few large signals. In addition, as argued in Polson and Scott (2010) and Bhattacharya et al. (2015), the global-local shrinkage priors can offer computational savings over point-mass priors.

In practice, the question is which sparsity prior should be chosen. There is a huge and growing literature on sparse priors with a number of proposals regarding new default priors and different fields of applications. However, simulation results and empirical evidence seems to be not fully conclusive. In the factor model context followed in this paper, we focus on the comparison between the Normal Inverse-Gamma prior, which serves as a benchmark in the factor models literature, on the one hand, and a range of prominent sparse priors on the other hand to consider the variety of priors available in the recent literature.

In Figure I.1, we compare the alternative priors by histograms obtained from sampling from the priors. By looking at the central bin on the left panel of the Figure I.1, we can see how the sparse priors concentrate mass near zero compared to the normal prior. The PMNM and HS+ concentrate considerably more mass than the MG near zero and the Normal prior. The tail behaviour is shown in the right panel. The HS+ and HS have heavier tails than the MG, the PMNM and the Normal prior. By construction, the PMNM converges to zero quicker the Normal prior due to the mass at zero and the same specifications for the variance in the Normal distribution.

I.2.3   Estimation and conditional posteriors

In order to estimate the model in (I.1) we need to derive the joint posterior distribution of the parameters of the model and the factors. By assumption, the prior on the loadings and the remaining parameters of the model, denoted by $\Theta = \left\{\Psi, \Sigma_\eta, \Sigma_\varepsilon\right\}$, are independent of each other. The likelihood is given by

$$L\left(x_{1:T}|f_{1:T},\lambda,\Theta\right) = \prod_{t=1}^{T} p(x_t|f_t,\lambda,\Theta) \tag{I.12}$$

$$p(x_t|f_t,\lambda,\Theta) = \frac{1}{(2\pi)^{N/2}|\Sigma_\varepsilon|^{1/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{N} \frac{[(x_{it}-\lambda f_t)]^2}{\sigma_i^2}\right\}. \tag{I.13}$$

Combining the likelihood with the prior on the parameters (and implicitly the factors) yields the joint posterior

$$p(\lambda,\Theta,f_{1:T}|x_{1:T},f_{1:T}) \propto L\left(x_{1:T}|f_{1:T},\lambda,\Theta\right) p(f_{1:T}|\lambda,\Theta) p(\lambda) p(\Theta). \tag{I.14}$$

Draws from it can be obtained using Gibbs Sampling techniques, i.e. sequentially sampling the parameters conditional on the factors and then conditioning on the draw of the parameters to sample the latent factors. Specifically, at each iteration m the Gibbs Sampler cycles through the following steps or blocks:

B1: draw $f_{1:T}^{(k)}|x_{1:T},\Theta^{(k-1)},\lambda^{(k-1)}$

Figure I.1: Alternative prior distributions
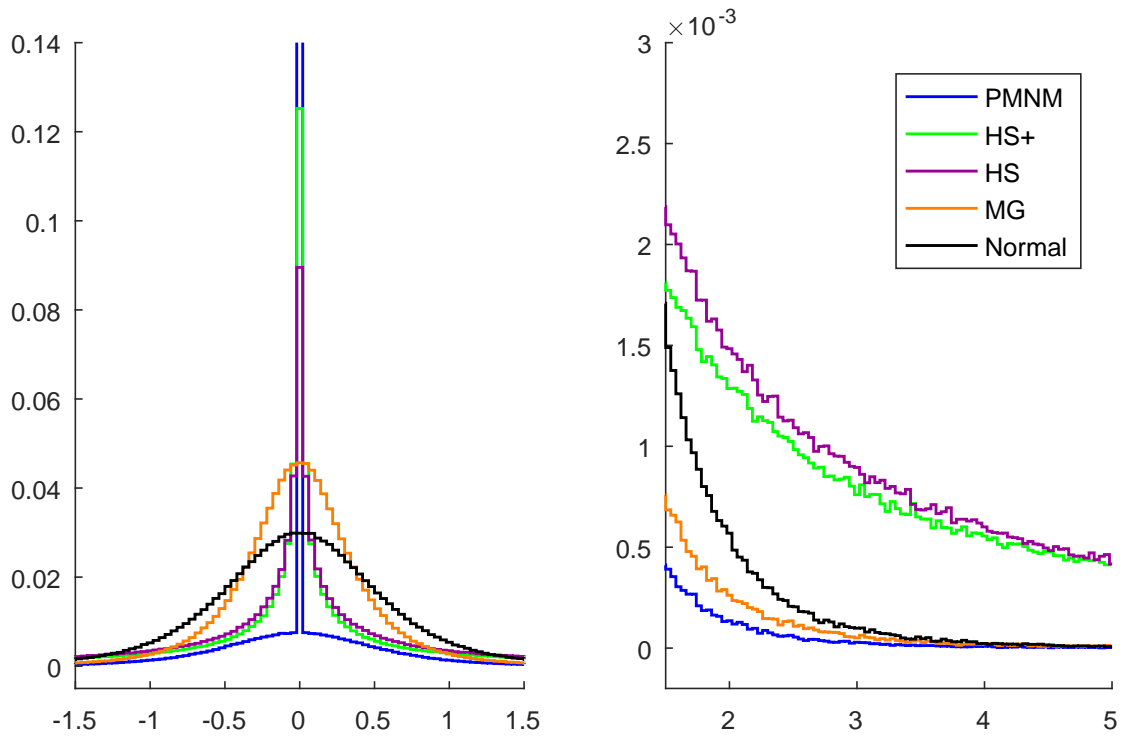
Note: The Figure shows histograms of draws from the prior distributions for 2×10^6 draws. The width of each bin is 0.04, the median bin is centered around zero. The frequency for each bin is shown on the vertical axis. PMNM is the point mass normal mixture prior with hyper parameters $r_0 = 3$, $s_0 = 0.25$, $a_0 = 2$, $b_0 = 0.5$. HS and HS+ denote the horseshoe and horseshoe+ priors, respectively. MG is the multiplicative gamma prior for the first factor with hyper parameters $a_1 = 10$, $a_2 = 2$, and $v = 3$. The Normal prior has mean zero and the variance follows an inverse Gamma distribution with $a_0 = 2$ and $b_0 = 0.5$ as in the PMNM prior. Details on the priors are provided in Section I.2.2 of the main text.

B2: draw $\Psi^{(k)}|f_{1:T}^{(k)}$ and $\Sigma_{\epsilon}^{(k)}|x_{1:T}, f_{1:T}^{(k)}, \lambda^{(k-1)}$

B3: draw $\lambda^{(k)}|x_{1:T}, f_{1:T}^{(k)}, \Sigma_{\epsilon}^{(k)}$

B4: update the hyperparameters of the prior on λ

B1 and **B2** are standard and draws from the respective posterior distributions easily obtained. We leave the details of these blocks to the Appendix and focus here on the conditional posterior distributions of the loadings matrix λ and the hyperparameters governing the different priors (**B3** and **B4**).

We start with the Normal-Inverse Gamma, Multiplicative Gamma and horseshoe(+) priors where standard Bayesian linear regression results can be employed, before turning the point mass-Normal mixture which requires a different treatment. Under a (conditionally) Normal prior, the posterior of $\lambda_{i.}$ is given by (e.g. Kroese and Chan, 2013)

$$\lambda_{i.}|\cdot \quad \sim \quad \mathcal{N}(m_i, M_i) \tag{I.15}$$

$$M_i \quad = \quad \left(D^p + \frac{1}{\sigma_i^2}\sum\nolimits_{t=J+1}^{T} f_t' f_t\right)^{-1} \tag{I.16}$$

$$m_i \quad = \quad M_i\left(\frac{1}{\sigma_i^2}\sum\nolimits_{t=J+1}^{T} f_t' x_{i,t}\right) \tag{I.17}$$

where D^p is the prior precision matrix of $\lambda_{i.}$ and depends on the hyperparameters of the respective priors $p = \{NIG, MG, HS(+)\}$.

For example, under the Normal-Inverse Gamma prior $D^{NIG} = \text{diag}(\tau_1^{-1}, \ldots, \tau_R^{-1})$. In turn, conditional on a draw of the loadings we can update the hyperparameters by sampling from their Inverse-Gamma posterior distribution, i.e.

$$\tau_r|\lambda \quad \sim \quad \mathcal{G}^{-1}\left(g_0 + 0.5N, G_0 + 0.5\sum\nolimits_{i=1}^{N}\lambda_{ir}^2\right) \tag{I.18}$$

for all $r = 1, \ldots, R$.

In the case of the MG prior, the prior precision consists of an additional term reflecting the variable-specific, local shrinkage component. Thus, $D^{MG} = \text{diag}(\kappa_1\phi_{i1}, \ldots, \kappa_R\phi_{iR})$ and updating the hyperparameters requires draws from the conditional posterior distributions given by

$$\phi_{ir}|\lambda, \kappa \quad \sim \quad \mathcal{G}\left(\frac{w+1}{2}, \frac{w + \kappa_r\lambda_{ir}^2}{2}\right) \forall\, i = 1, \ldots, N, r = 1, \ldots, R \tag{I.19}$$

$$\delta_1|\lambda, \kappa, \phi \quad \sim \quad \mathcal{G}\left(a_1 + \frac{NR}{2}, 1 + \frac{1}{2}\sum\nolimits_{l=1}^{R}\kappa_l^{(1)}\sum\nolimits_{i=1}^{N}\phi_{il}\lambda_{il}^2\right) \tag{I.20}$$

$$\delta_r|\lambda, \kappa, \phi \quad \sim \quad \mathcal{G}\left(a_2 + \frac{N}{2}(R-r+1), 1 + \frac{1}{2}\sum\nolimits_{l=1}^{R}\kappa_l^{(r)}\sum\nolimits_{i=1}^{N}\phi_{il}\lambda_{il}^2\right) \forall\, r > 1, \tag{I.21}$$

where $\kappa_l^{(r)} = \sum_{k=1, k\neq r}^{l}\delta_k \ \forall\, r = 1, \ldots, R.$

Recall that for both the horseshoe and the horseshoe+ prior, $D^{HS(+)} = v^2 \text{diag}(\zeta_{i1}^2, \ldots, \zeta_{iR}^2)$. To update the hyperparameters, we follow Makalic and Schmidt (2016), who exploit a scale-mixture representation of the half-Cauchy distribution: if $x^2|a \sim \mathcal{G}^{-1}(1/2, 1/a)$, $a \sim \mathcal{G}^{-1}(1/2, 1/A^2)$ then $x \sim \mathcal{C}^+(0, A)$ (Wand et al., 2011). Therefore, by introducing the auxiliary variables β^ζ, β^v we can rewrite the prior in (I.11) as

$$
\begin{aligned}
\zeta_{ir}^2 | \beta_{ir}^\zeta &\sim \mathcal{G}^{-1}(1/2, 1/\beta_{ir}^\zeta) \\
v^2 | \beta^v &\sim \mathcal{G}^{-1}(1/2, 1/\beta^v) \\
\beta_{11}^\zeta, \ldots, \beta_{NR}^\zeta, \beta^v &\sim \mathcal{G}^{-1}(1/2, 1).
\end{aligned}
$$

This leads to convenient posterior distributions for the hyperparameters of the form:

$$
\zeta_{ir}^2 | \lambda, \beta_{ir}^\zeta \sim \mathcal{G}^{-1}\left(1, \frac{1}{\beta_{ir}^\zeta} + \frac{\lambda_{ir}^2}{2v^2}\right) \tag{I.22}
$$

$$
v^2 | \lambda, \beta^v \sim \mathcal{G}^{-1}\left(\frac{NR+1}{2}, \frac{1}{\beta^v} + \frac{1}{2}\sum_{r=1}^{R}\sum_{i=1}^{N}\frac{\lambda_{ir}^2}{\zeta_{ir}^2}\right). \tag{I.23}
$$

Similarly, the conditional posteriors of the auxiliary variables β^v and β^ζ are also Inverse-Gamma and given by

$$
\beta_{ir}^\zeta | \zeta_{ir} \sim \mathcal{G}^{-1}\left(1, 1 + \frac{1}{\zeta_{ir}^2}\right) \forall i = 1, \ldots, N, r = 1, \ldots, R \tag{I.24}
$$

$$
\beta^v | \kappa \sim \mathcal{G}^{-1}\left(1, 1 + \frac{1}{\kappa^2}\right). \tag{I.25}
$$

For the HS+ prior, the formula for each β_{ir}^ζ needs to be adjusted to reflect the additional half-Cauchy prior on χ_{ir}, which in turn governs the prior on ζ_{ir}. To this end, we introduce another set of $\mathcal{G}^{-1}(1/2, 1)$ auxiliary variables, β^χ. In this case, the conditional posterior distribution for ζ_{ir}^2 remains unchanged as we are still only conditioning on β_{ir}^ζ. However, the rate parameter of the latter's conditional Inverse-Gamma posterior now also depends on χ_{ir}, i.e.

$$
\beta_{ir}^\zeta | \zeta_{ir}, \chi_{ir} \sim \mathcal{G}^{-1}\left(1, \frac{1}{\chi_{ir}^2} + \frac{1}{\zeta_{ir}^2}\right) \forall i = 1, \ldots, N, r = 1, \ldots, R \tag{I.26}
$$

while χ_{ir}^2 itself and the corresponding auxiliary variables can be updated at each iteration

of the Gibbs Sampler as follows:

$$\chi_{ir}^2 | \beta_{ir}^\chi \sim \mathcal{G}^{-1}\left(1, \frac{1}{\beta_{ir}^\chi} + \frac{1}{\beta_{ir}^\zeta}\right) \tag{I.27}$$

$$\beta_{ir}^\chi | \chi_{ir}^2 \sim \mathcal{G}^{-1}\left(1, 1 + \frac{1}{\chi_{ir}^2}\right), \tag{I.28}$$

again for all $i = 1, \ldots, N$, $r = 1, \ldots, R$.

The conditional posterior under the PMNM is more involved due to the mixture in (I.33). We start by defining the transformed observation $x_{it}^* = x_{i,t} - \sum_{l=1, l\neq r}^{R} \lambda_{il} f_{l,t} = \lambda_{ir} f_{r,t} + \epsilon_{it}$ which isolates the impacts of the factors other than r and the corresponding observation density $p\left(x_{it}^*|\cdot\right) = \mathcal{N}\left(\lambda_{ir} f_{rt}, \sigma_i^2\right)$. Then, combining the marginal prior on λ_{ir} with the likelihood yields

$$p(\lambda_{ir}|\cdot) = \prod_{t=J+1:T}^{T} p(x_{it}^*|\cdot)\{(1-\rho_r)\delta_0(\lambda_{ir} + \rho_r \mathcal{N}(0, \tau_r)\} \tag{I.29}$$

$$= P\left(\lambda_{ir} = 0|\cdot\right)\delta_0\left(\lambda ir\right) + P\left(\lambda_{ir} \neq 0|\cdot\right)\mathcal{N}(m_{ir}, M_{ir}) \tag{I.30}$$

with

$$M_{ir} = \left(\tau^{-1} + \frac{1}{\sigma_i^2}\sum_{t=J+1}^{T} f_{r,t}' f_{r,t}\right)^{-1} \tag{I.31}$$

$$m_{ir} = M_{ir}\left(\frac{1}{\sigma_i^2}\sum_{t=J+1}^{T} f_{r,t}' x_{i,t}^*\right). \tag{I.32}$$

In order to sample from this distribution, we need to evaluate the posterior odds ratio of a non-zero loading, i.e.

$$PO_{ir} = \frac{P\left(\lambda_{ir} \neq 0|\cdot\right)}{P\left(\lambda_{ir} = 0|\cdot\right)} = \frac{p\left(\lambda_{ir}\right)|_{\lambda_{ir}=0}}{p\left(\lambda_{ir}|\cdot\right)|_{\lambda_{ir}=0}}\frac{\rho_r}{1-\rho_r} = \frac{\mathcal{N}(0; 0, \tau_r)}{\mathcal{N}(0; m_{ir}, M_{ir})}\frac{\rho_r}{1-\rho_r} \tag{I.33}$$

A draw of λ_{ir} is then obtained by sampling from $\mathcal{N}(m_{ir}, M_{ir})$ and keeping the draw if $u \leq PO_{ir}/(1 + PO_{ir})$, where u is a draw from $\mathcal{U}(0, 1)$. Otherwise, we set $\lambda_{ir} = 0$.

I.3 Empirical application: nowcasting GDP

In this section, we use the sparse factor models outlined above to nowcast the quarterly gross domestic product (GDP) in the United States (US) and Germany. We first give an overview of the monthly indicators that are included in the dataset along with the GDP (Section I.3.1). Next, in Section I.3.2 we highlight how the factor model can be estimated while accouting for mixed frequency data. The real-time evaluation set-up and results are

discussed in Section I.3.3 and I.3.4, respectively. Lastly, we provide robustness checks to our results for different model specifications such as the transformation of survey variables or the estimation window in Section I.3.5.

I.3.1 Data

To nowcast the US GDP, we use the monthly real-time data set provided by McCracken and Ng (2016) of more than 100 time series covering a broad spectrum of macroeconomic activity and mostly ranging back until January 1959. Vintages for this dataset are available from November 1999 onwards. For the German economy we construct a dataset based on vintages from the Deutsche Bundesbank's Real-Time Database, covering series such as production, orders, turnover, prices and the labor market. These vintages are available on a broad basis from the end of 2005 onwards. To these, we add financial market indicators also sourced from the Deutsche Bundesbank. For both countries, we augment the monthly datasets with survey-based sentiment indicators which have proven useful in nowcasting applications due to their timely release.

The estimation samples for the United States and Germany start in January 1985 and 1992, respectively, and are recursively expanded. While the two datasets are similar in terms of economic categories, there are some differences regarding the importance of individual groups. For example, in the German data there is more detailed coverage of production, orders and turnover of the industrial sector, while FRED-MD contains more disaggregated labor market series. A detailed description of the variables we use and the transformations applied to the series prior to estimation can be found in Appendix I.C.

I.3.2 Bayesian estimation of mixed-frequency factor models

In order to combine the monthly indicators with our quarterly target variable - quarter-on-quarter GDP growth - we need to adjust the model in (I.1) to account for mixed frequencies. This is done by formulating the model at the highest frequency, i.e. monthly, and treating quarterly variables as monthly time series with occasionally missing observations. Thus, we essentially convert the mixed-frequency problem into a missing-value problem which can easily be handled by state space methods. Appendix I.A documents the mixed-frequency factor model and the resulting state space representation. Draws from the predictive density of GDP growth can be obtained from a Gibbs Sampler which alternately draws from the posterior distribution of the factors conditional on the parameters and then updates the parameters given the sampled factors (see Appendix I.B for details). We iteratively draw from the conditional posterior distributions 15000 times, discarding the first 5000 draws as a burn-in. The rest of the Markov chain is thinned by storing every 10-th iteration, yielding a total of $G = 1000$ draws for posterior inference.[2]

---

[2]By and large, our Gibbs Sampler shows no signs of poor mixing or non-convergence. See Appendix I.E for detailed MCMC diagnostics.

In terms of model specification, we estimate the models for all R = 1 : 10; the number of lags in the factor VAR is set to one, i.e. P = 1. Furthermore, for each prior we combine the ten individual densities into an equally-weighted pool (denoted as "pool" below).

To benchmark the performance of the different priors, we estimate a simple univariate Bayesian autoregressive model of order 1, estimated with loose priors on the autocovariance coefficient - $\mathcal{N}(0,3)$ - and the residual variance - $\mathcal{G}^{-1}(1, 0.01)$. This model is subsequently denoted as B-AR(1). Additionally, we also consider the factor model with a Normal prior on the loadings that features a relatively high and fixed variance, e.g. Amisano and Geweke (2017) and Marcellino et al. (2016). In the notation of Section I.2, the prior precision D^p is given by $\frac{1}{c}I_R$ where c is some large number, independent of any hyperparameters. We set $c = 10$ and label this "diffuse" prior in what follows as Nd.

The hyperparameters of the remaing priors are as follows: The variance of the loadings in the Normal-Inverse Gamma (NIG) is parametrized as $\mathcal{G}^{-1}(2, 1)$. For the Multiplicative-Gamma (MG) prior we choose $a_1 = 5$, $a_2 = 2$ and $w = 3$. In the case of the point mass-Normal mixture (PMNM) we set $r_0 = 5$, $s_0 = 0.5$, while the prior on τ_r is given by $\mathcal{G}^{-1}(2, 1)$. As discussed above, the horseshoe prior (HS+) requires no hyperparameters.

I.3.3   Evaluation set-up

The evaluation period ranges from the first quarter of 2000 (2000Q1) to the fourth quarter of 2018 (2018Q4) in the US. Reflecting the shorter estimation sample for Germany (and lack of available real-time vintages earlier on), we begin to evaluate nowcasts starting in 2006Q1. In both cases, we focus on nowcasts made at the end of the first month of a given quarter e.g. January for nowcasts of GDP growth in Q1. That is to say, the forecast horizon in months is h = 2.

The accuracy of the different models/priors is assessed in terms of point and density nowcasts. For the former, we compute the root mean squared forecast error (RMSFE) of model/prior $m = \{\text{B-AR}(1), \text{Nd}, \text{NIG}, \text{MG}, \text{PMNM}, \text{HS+}\}$ defined as

$$\text{RMSFE}_m = \left(\frac{1}{S} \sum_{s=1}^{S} (y_{m,T+s}^{f,Q} | \Omega_{v_s} - y_{T+s}^{Q})^2 \right)^{\frac{1}{2}}$$

for a sequence of S nowcasts, where $y_{m,T+s}^{f,Q}|\Omega_{v_s}$ is the mean of the predictive density of GDP growth at time $T+s$, conditional on the information set Ω_{v_s} available in real-time at date v_s when the nowcast for period $T+s$ was made. The corresponding realization of GDP growth which we take to be the first release is denoted by y_{T+s}^{Q}.[3]

Density nowcasts are evaluated by two scoring rules commonly used in the forecasting literature: the log score (logS) and the continuous ranked probability score (CRPS). The former is simply the predictive density of model m for period s evaluated at the realization

---

[3]Given that the mean absolute revision between first and second release are quite small for both countries, our results are robust to using the latter as the realization. Results are available upon request.

and then averaged over the evaluation period, i.e.

$$\text{logS}_m = \frac{1}{S} \sum_{s=1}^{S} \left(-\log p_{m,T+s}(y^Q_{T+s}) \right)$$

where $p_{m,T+s}(y^Q)$ is estimated from $y^{Q(1)}_{m,T+s}|\Omega_{v_s}, \ldots, y^{Q(G)}_{m,T+s}|\Omega_{v_s}$ using the theta kernel density estimator described in Botev et al. (2010). Furthermore, note that we flip the orientation of the log score to bring its interpretation in line with the RMSFE and CRPS: After premultiplying with -1, lower (higher) values of logS indicate a higher (lower) predictive accuracy.

The CRPS is calculated as (see Krüger et al., 2016):

$$\text{CRPS}_m = \frac{1}{S} \sum_{s=1}^{S} \left(\frac{1}{G} \sum_{i=1}^{G} \left| y^{Q(i)}_{m,T+s}|\Omega_{v_s} - y^Q_{T+s} \right| - \sum_{i=1}^{G} \sum_{j=1}^{G} \left| y^{Q(i)}_{m,T+s}|\Omega_{v_s} - y^{Q(j)}_{m,T+s}|\Omega_{v_s} \right| \right)$$

I.3.4 Results

The results of the nowcast evaluation for US and Germany GDP are presented in Table I.1. For the sake of readability, we only show the results for a selected number of factors as well as the pooled nowcasts.[4] Overall, the factor models in both countries perform well with root mean squared forecast errors relative to the benchmark B-AR(1) as low as 0.7 and 0.5 for US and German GDP, respectively. Density nowcasts yield similar relative gains when evaluated in terms of the CRPS while they are generally even larger under the log score. For the United States, the predictive accuracy generally increases in the number of factors R - both in terms of point and density forecasts. For Germany, the models with $R = \{5, 8\}$ also perform much better than the model with only one factor but for $R = 2$ the performance is similar to the bigger models. Noteworthy is that in both countries the equal-weight pool performs very well, often achieving relative gains almost as large as those of the best individual specification.

Turning to the question of whether sparsity-inducing priors like the MG, PMNM or HS+ generate more accurate nowcasts than convential priors like the NIG or Nd, we find mixed evidence across the two countries. For the United States, the sparse priors do not perform better as the NIG prior. Indeed, for up to $R = 5$ there are virtually no differences between the five priors we consider across both point and density nowcasts. For $R = 8$ we find that a) the PMNM, HS and the NIG do perform slightly better than the Normal-diffuse prior in terms of all three forecast accuracy measures and b) the MG prior's performance deteriorates markedly. This weaker performance carries over to the equal-weight pool where the MG prior's poor performance stands out in terms of the log score and CRPS. To some extent these results hold for the case of Germany as well. There are also no differences between the sparse prior and the NIG. They do, however, outperform the Normal-diffuse prior by a larger amount when considering point and density nowcasts. In contrast to the findings

---

[4]The results for the remaining specifications are available upon request.

Table I.1: Nowcasting results for US (top) and German (bottom) GDP

		full sample			post-crisis sample		
		RMSFE h=0	logS h=0	CRPS h=0	RMSFE h=0	logS h=0	CRPS h=0
B-AR		0.45	0.64	0.25	0.32	0.50	0.19
R=1	Nd	0.85	0.79	0.89	1.11	0.98	1.07
	NIG	0.85	0.79	0.89	1.10	0.96	1.06
	MG	0.85	0.80	0.89	1.08	0.97	1.05
	PMNM	0.86	0.79	0.89	1.10	0.97	1.06
	HS+	0.86	0.82	0.90	1.10	1.02	1.07
R=2	Nd	0.90	0.82	0.92	1.10	0.95	1.06
	NIG	0.91	0.84	0.93	1.10	0.93	1.05
	MG	0.86	0.81	0.89	1.08	0.95	1.04
	PMNM	0.90	0.82	0.92	1.13	0.96	1.07
	HS+	0.85	0.78	0.89	1.11	0.91	1.05
R=5	Nd	0.82	0.72	0.85	1.03	0.80	0.99
	NIG	0.81	0.69	0.83	0.94	0.72	0.92
	MG	0.78	0.70	0.82	0.99	0.81	0.96
	PMNM	0.79	0.68	0.82	0.99	0.76	0.95
	HS+	0.79	0.69	0.83	1.01	0.78	0.97
R=8	Nd	0.76	0.65	0.79	1.02	0.79	0.97
	NIG	0.77	0.66	0.80	1.03	0.78	0.98
	MG	0.88	0.78	0.88	0.99	0.85	0.97
	PMNM	0.73	0.62	0.77	0.95	0.70	0.91
	HS+	0.74	0.60	0.77	0.93	0.67	0.90
pool	Nd	0.76	0.68	0.80	0.96	0.80	0.95
	NIG	0.77	0.68	0.80	0.92	0.77	0.91
	MG	0.80	0.72	0.83	0.98	0.83	0.96
	PMNM	0.75	0.68	0.79	0.92	0.79	0.92
	HS+	0.76	0.66	0.79	0.94	0.74	0.92
		full sample			post-crisis sample		
		RMSFE h=0	logS h=0	CRPS h=0	RMSFE h=0	logS h=0	CRPS h=0
B-AR		0.86	1.73	0.41	0.49	0.89	0.29
R=1	Nd	0.86	0.90	0.89	0.86	0.95	0.93
	NIG	0.86	0.89	0.89	0.85	0.94	0.93
	MG	0.88	0.81	0.91	0.90	0.97	0.97
	PMNM	0.89	0.97	0.91	0.88	0.96	0.95
	HS+	0.89	1.00	0.92	0.90	0.97	0.97
R=2	Nd	0.57	0.40	0.65	0.79	0.62	0.78
	NIG	0.52	0.36	0.61	0.76	0.59	0.75
	MG	0.52	0.37	0.61	0.75	0.60	0.75
	PMNM	0.54	0.37	0.62	0.78	0.58	0.77
	HS+	0.54	0.38	0.63	0.77	0.62	0.77
R=5	Nd	0.65	0.46	0.74	0.90	0.75	0.89
	NIG	0.55	0.38	0.63	0.73	0.59	0.74
	MG	0.64	0.41	0.69	0.71	0.61	0.73
	PMNM	0.55	0.37	0.62	0.72	0.58	0.73
	HS+	0.53	0.37	0.61	0.68	0.57	0.70
R=8	Nd	0.56	0.42	0.66	0.85	0.73	0.85
	NIG	0.51	0.37	0.59	0.71	0.62	0.73
	MG	0.56	0.39	0.64	0.74	0.63	0.76
	PMNM	0.49	0.35	0.57	0.68	0.57	0.71
	HS+	0.49	0.33	0.57	0.67	0.54	0.69
pool	Nd	0.51	0.42	0.62	0.73	0.74	0.79
	NIG	0.47	0.37	0.57	0.63	0.64	0.71
	MG	0.54	0.37	0.61	0.67	0.63	0.72
	PMNM	0.48	0.36	0.56	0.62	0.61	0.69
	HS+	0.47	0.35	0.55	0.59	0.59	0.67

RMSFE is the root mean squared forecast error, logS and CRPS are the average log score and continuous ranked probability score. All entries for the factor models are relative to the B-AR benchmark (see text for details) and negatively orientated so that a value in the table below 1 corresponds to a better performance than the benchmark. The forecast horizon h is in months. For the US (Germany) the full sample period is 2000Q1 (2006Q1)-2018Q4. In both cases the post-crisis sample starts in 2010Q1 and ends in 2018Q4.

for US GDP, we do see that the nowcasts from the sparse model perform better than those from the NIG when evaluated in terms of the RMSFE and the log score or CRPS. Similar to the US, however, between the PMNM and HS+ priors there are virtually no differences in predictive accuracy while the MG's poorer performance stands out for larger R.

Lastly, we find that when excluding the Global Financial Crisis (GFC) from the evaluation the gains in predictive accuracy relative to the B-AR(1) are less pronounced for most factor models - for the United States some of the models do not outperform the autoregressive benchmark in terms of point nowcast accuracy. Our main findings with regard to the question whether sparsity matters for nowcasting, however, are unaltered. Focusing only on the post-crisis sample the sparse priors yield highly similar results in terms of all three performance measures considered. In the US, we still find that there are also no large differences between the sparse priors and the NIG or Nd in terms of predictive accuracy, while for German GDP the PMNM and HS+ priors continue to produce much better nowcasts than the Nd and somewhat better nowcasts than the NIG for $R = 8$ and the equal-weight pool.

I.3.5 Robustness analysis

The results discussed above were obtained under specific modelling/specifications choices. In our application, these concern the transformation applied to the survey indicators or the choice of estimation window. We find, however, that alternative specifications in this regard - such as not first-differencing the survey indicators or estimating the models with a rolling window - do not have a material impact on the results neither in quantitative nor qualitative terms. We discuss these choices and briefly comment on the results below. Details and tables similar to Table I.1 for the different robustness checks can be found in Appendix I.D.

First, in our baseline specification we included survey indicators in first differences. While this brings their time series behavior more in line with the rest of the dataset, an alternative is to keep the series in levels, as these indicators are by construction stationary. Examples of both approaches can be found in the literature. In our application we find that the nowcast performance overall deteriorates uniformly across models when the surveys are not first-differenced. However, the main findings are unaltered: the differences between the sparse priors and the NIG are very small. In the case of Germany the sparse priors and the NIG perform much better than the Normal-diffuse prior which for large R fails to beat the autoregressive benchmark when the surveys enter the models in levels. **Second**, while a recursively expanded estimation window leads to lower estimation uncertainty, a rolling estimation window might guard against structural instabilities in the forecasting models and therefore generate better nowcasts. However, there is no indication that these instabilities play a large role in our application as the nowcast performance only improves marginally in some cases for the United States, in particular for the smaller models, i.e. $R \leq 2$; in addition the MG prior performs much better under a rolling window. But the performance of the equal-weight pool, for example, is virtually identical. For Germany, with the exception of $R = 1$ the accuracy of point and density nowcasts is somewhat higher when the models'

parameters are estimated with a recursively expanding window. For both countries, the choice of estimation window does not have an impact on the performance of the sparse priors relative to the Normal alternatives.

I.4 Conclusion

In this paper, we have explored the role that sparsity plays in factor models. In a real-time nowcasting evaluation we find that estimating the model with sparse priors on the loadings matrix in a Bayesian framework does not lead to large gains in nowcast accuracy. Furthermore, we found very similar results for different sparse priors that have been proposed in the literature. This suggests that the practice in parts of the literature of considering large cross sections when nowcasting GDP is justified. Our findings are compatible with the hypothesis that large macroeconomic data sets typically used in now- or forecasting applications are not sparse but dense. However, we caution against generalising our findings too far, as sparsity has been shown to play a role in other macroeconomic settings (Kaufmann and Schumacher, 2017). Moreover, the recent Covid-19 pandemic has acceleted a trend in macroeconomic forecasting applications of considering new, unconvential data originating from newspapers, social media, mobile phones or internet search queries. These data sources typically provide vast numbers of time series ("big data"). We leave it to future research to address the issue of variable selection and the role that sparsity plays in such an environment.

Appendix

I.A Mixed-frequency factor model

In the following, we outline the mixed-frequency factor model used in the nowcasting application. For the reader's convenience we begin by restating the original model

$$x_t = \lambda f_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon) \tag{I.34}$$
$$\Psi(L)f_t = \eta_t, \quad \eta_t \sim \mathcal{N}(0, \Sigma_\eta)$$

where we now make it explicit that x_t is a vector of N_M stationary monthly variables. To combine the monthly time series with quarterly GDP growth, y_t^Q, we formulate the model at the highest frequency, e.g. monthly. However, we only observe y_t^Q every third period, e.g. at the last month of each quarter. Let T be the number of observations for which at least one monthly variable is available. Assuming for the sake of expostion that the sample starts in the first month a quarter and ends in the third, say January and December, we thus have $y_{1:T}^Q = \left[\text{NaN}, \text{NaN}, y_3^Q, \text{NaN}, \text{NaN}, y_6^Q, \dots, \text{NaN}, \text{NaN}, y_T^Q \right]$.

Furthermore, we assume that y_t, the unobserved monthly analogue of y_t^Q, adheres to the same factor structure as x_t. That is to say, we assume that (unobserved) month-on-month GDP growth can be expressed as

$$y_t = \lambda^y f_t + \varepsilon_t^y, \quad \varepsilon_t^y \sim \mathcal{N}(0, \Sigma_{\varepsilon^y}). \tag{I.35}$$

The unobserved y_t are linked to the quarterly observations y_t^Q via the following time aggregation rule (Mariano and Murasawa, 2003):

$$y_t^Q \approx \frac{1}{3} y_t + \frac{2}{3} y_{t-1} + \frac{3}{3} y_{t-2} + \frac{2}{3} y_{t-3} + \frac{1}{3} y_{t-4}, \forall t = 3, 6, 9, \ldots, T. \tag{I.36}$$

By putting together (I.34), (I.35) and (I.36), the model can be cast into state space form with measurement equation

$$
\begin{bmatrix} x_{1,t} \\ \vdots \\ x_{N_m,t} \\ y_t^Q \end{bmatrix} = Z \begin{bmatrix} f_t \\ \vdots \\ f_{t-4} \\ \epsilon_{1,t}^y \\ \vdots \\ \epsilon_{1,t-4}^y \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \sim \mathcal{N}(0, H) \tag{I.37}
$$

and transition equation

$$
\begin{bmatrix} f_t \\ \vdots \\ f_{t-4} \\ \epsilon_{1,t}^y \\ \vdots \\ \epsilon_{1,t-4}^y \end{bmatrix} = T \begin{bmatrix} f_{t-1} \\ \vdots \\ f_{t-5} \\ \epsilon_{1,t-1}^y \\ \vdots \\ \epsilon_{1,t-5}^y \end{bmatrix} + R \begin{bmatrix} \eta_t \\ \epsilon_t^y \end{bmatrix}, \quad \begin{bmatrix} \eta_t \\ \epsilon_t^y \end{bmatrix} \sim \mathcal{N}(0, Q) \tag{I.38}
$$

where

$$
Z = \begin{bmatrix} \lambda & 0_{N_M \times R} & \cdots & \cdots & \cdots & 0_{N_M \times 1} & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{3}\lambda^q & \frac{2}{3}\lambda^q & \frac{3}{3}\lambda^q & \frac{2}{3}\lambda^q & \frac{1}{3}\lambda^q & \frac{1}{3} & \frac{2}{3} & \frac{3}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}, \quad H = \begin{bmatrix} \Sigma_\varepsilon & 0_{N_M \times 1} \\ 0_{1 \times N_M} & 0 \end{bmatrix}
$$

and

$$
T = \begin{bmatrix} \psi_1 & 0_R & 0_{R \times 1} & \cdots \\ I_{4R} & 0_{4R \times R} & 0_{4R \times 1} & \cdots \\ 0_{1 \times R} & \cdots & 0 & 0_{1 \times 4} \\ 0_{4 \times R} & \cdots & I_4 & 0_{4 \times 1} \end{bmatrix}, \quad Q = \begin{bmatrix} \Sigma_\eta & 0_{Nm \times 1} \\ 0_{1 \times N_m} & \Sigma_{\varepsilon^y} \end{bmatrix}, \quad R = \begin{bmatrix} I_R & 0_{R \times 1} \\ 0_{4R \times R} & 0_{4R \times 1} \\ 0_{1 \times R} & 1 \\ 0_{4 \times R} & 0_{4 \times 1} \end{bmatrix}.
$$

When running the Kalman filter or smoother, missing observations can easily be dealt with by either i) removing the missings elements from the vector of observations and ad-

justing the dimensions of Z and H or ii) replacing them with arbitrary values, say 0, and setting the corresponding element in H to a very large number. We found the former to be numerically more stable in our application.

I.B Gibbs Sampler

The model is estimated by a Gibbs Sampler which alternately draws from the conditional posterior distribution of the parameters given the factors and the factors given the parameters. Draws from the predictive density of any variable of interest - in our application, quarterly GDP growth - can then be obtained conditional on these values. Let λ denote the $N \times R$ matrix of factor loadings belonging to the monthly variables while λ^y is the loading of unobserved monthly GDP. The remaining parameters of the model are collected in $\Theta = \{\Psi, \Sigma_\varepsilon, \Sigma_{\varepsilon^y}\}$. Furthermore, let φ denote the prior hyperparameters. $x_{1:T}$ is the $N \times T$ matrix of monthly observations - some of which may be missing due different publication delays or because a series has only been collected over part of the estimation sample - where T is the maximum number of periods for which at least one observation is available. Furthermore, $y^Q_{1:T}$ denotes quarterly GDP growth at the monthly frequency, where we follow the convention in the literature and assume that the quarterly observations are available in the last month of each quarter, otherwise, $y^Q_t = $ NaN; the monthly analogue of quarterly GDP growth is denoted as $y_{1:T}$ (see Appendix I.A for more details on the temporal aggregation). The forecast horizon in months is denoted by H, i.e. $T + h$ corresponds to the third month of the quarter that is being nowcasted.

The Gibbs Sampler then cycles through the following steps or blocks to draw from the predictive density $p(y^Q_{T+h}|x_{1:T}, y^Q_{1:T})$:

Step 1: $p(f_{1:T}, y_{1:T}|\lambda, \lambda^y, \Theta, x_{1:T}, y^Q_{1:T})$
Conditional on the observed monthly and quarterly data, $x_{1:T}$ and $y^Q_{1:T}$, and a draw of the parameters, we can sample from the conditional posterior distribution of $f_{1:T}$ using the state space model described in (I.37) and (I.38) and the simulation smoother in Durbin and Koopman (2002). As a by-product of the sampled state vector, we also obtain a draw of the monthly analogues of the quarterly time series.

Step 2: $p(x^+_{1:T}|f_{1:T}, \lambda, \Sigma_\varepsilon)$
As some of the elements of $x_{1:T}$ may be missing, we also sample them conditional on the factors. This yields a complete data set $x^+_{1:T}$ which is then used in the subsequent steps. Given the normality of observations and states, the posterior distribution of a generic missing observation $x_{i,t}$ conditional on the factor is also Normal with mean $\lambda_i f_t$. Furthermore, its variance does not depend on the realizations of the conditioning arguments and is simply given by the i-th diagonal element of Σ_ε.

Step 3: $p(\Psi|f_{1:T})$

Conditional on a draw of the factors, we can sample the parameters of the factor VAR. Denote by $\psi^* = \text{vec}([\psi_1, \ldots, \psi_R])$ the vectorized matrix of coefficients. Then, given a Normal prior

$$p(\psi^*) \propto \det\left(\underline{V}_{\psi^*}\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\psi^* - \underline{b}_{\psi^*})'\underline{V}_{\psi^*}^{-1}(\psi^* - \underline{b}_{\psi^*})\right\}$$

the (conditional) posterior is Normal with variance and mean given by

$$\bar{V}_{\psi^*} = \left[\underline{V}_{\psi^*}^{-1} + I_R \otimes X'X\right]^{-1}, \quad \bar{b}_{\psi^*} = \bar{V}_{\psi^*}\left\{\underline{V}_{\psi^*}^{-1}\underline{b}_{\psi^*} + I_R \otimes X'X\, b_{OLS}\right\}$$

where $X = [f_{P:T-1}, \ldots, f_{1:T-P}]$, $y = f_{P+1:T}$ and $b_{OLS} = \text{vec}\left((X'X)^{-1}X'y\right)$. We follow the literature on Bayesian VAR and impose a Minnesota type prior on $\psi*$. That is to say, the prior mean is set equal to 0 and the prior variance depends on the lag length. Specifically, for $p = 1 : P$, we have

$$Var(\psi_{r,ij}) = \begin{cases} \frac{\pi_0}{r^2}, & \text{if } i = j \\ \frac{\pi_0\pi_1}{r^2}, & \text{otherwise.} \end{cases}$$

Common values in the literature are $\pi_0 = 0.2$ and $\pi_1 = 0.1$, thus shrinking coefficients on the lags of other factors stronger towards 0.

Step 4a: $p(\lambda, \lambda^y | f_{1:T}, x_{1:T}^+, y_{1:T}, \Sigma_\varepsilon, \Sigma_{\varepsilon^y}, \varphi)$

Conditional on the factors, data, parameters and hyperparameters we can sample the loadings from their conditional posterior distribution. See Section I.2 of the main text for details.

Step 4b: $p(\varphi | \lambda, \lambda^y)$

Conditional on a draw of the loadings, we can update the hyperparameters of the different priors that govern the degree of sparsity. We again refer to the main text for the conditional posterior distributions.

Step 5: $p(\Sigma_\varepsilon, \Sigma_{\varepsilon^y} | \lambda, \lambda^y, f_{1:T}, x_{1:T}^+, y_{1:T})$

Conditional on the idiosyncratic components $\epsilon_{i,1:T}$ we can sample the diagonal elements of the covariance matrix Σ_ε by drawing from

$$\sigma_i^2 \sim \mathcal{G}^{-1}\left(\frac{u+T}{2}, \frac{U + \sum_{t=1}^T \epsilon_{i,t}^2}{2}\right)$$

where u and U are the prior shape and rate. In an analogous manner, we can sample Σ_{ε^y} given $\epsilon_{1:T}^y$. We set $u = 2$ and $U = 1$ so that the prior is centered around one but relatively diffuse.

Step 6: $p(y_{T+h}^Q | x_{1:T}, y_{1:T}^Q)$

Draws from the h-step ahead predictive density of y^Q are obtained by iterating forward

Table I.2: List of US survey indicators

mnemonic	description	transformation
gacdna	current general activity	diff
gafdna	future general activity	diff
nocdna	current new orders	diff
nofdna	future new orders	diff
shcdna	current shipments	diff
shfdna	future shipments	diff
dtcdna	current delivery time	diff
dtfdna	future delivery time	diff
ivcdna	current inventories	diff
ivfdna	future inventories	diff
uocdna	current unfilled orders	diff
uofdna	future unfilled orders	diff
ppcdna	current prices paid	diff
ppfdna	future prices paid	diff
prcdna	current prices received	diff
prfdna	future prices received	diff
necdna	current employment	diff
nefdna	future employment	diff
awcdna	current workhours	diff
awfdna	future workhours	diff
cefdfna	future capital expenditures	diff

Notes: The table lists the survey indicators from the Federal Reserve Bank of Philadelphia's Manufacturing Business Outlook Survey that are used in nowcasting US GDP growth. For the historical data and further information, see `https://www.philadelphiafed.org/research-and-data/regional-economy/business-outlook-survey/historical-data`.

(I.34) and using (I.35) and (I.36) to compute y_{T+h}^Q (Del Negro and Schorfheide, 2013).

I.C Data

This section describes the data sets used in the empirical nowcasting application. The US data set (Section I.C.1) is based on FRED-MD. Real-time vintages for Germany (Section I.C.2) are compiled from the Deutsche Bundesbank's Real-Time Database. Both datasets are augmented with survey-based sentiment indicators; to guarantee the real-time nature of the evaluation, the raw, unadjusted series are seasonally adjusted in real-time using X11-ARIMA. Furthermore, while the datasets are broadly stable in terms of available series, there are some changes as vintages for some series are added or removed over the period of the evaluation sample. Additionally, for the estimation of the models we require that every variable is available for at least half of the estimation sample. As such, the exact composition of the data sets varies slightly over time. The individual series used in each vintage are available upon request.

I.C.1 United States

Real-time vintages of GDP growth are obtained from ALFRED. Besides the target variable, our data set includes a large number of monthly time series covering various aspects of economic activity. Specifically, we use the large, monthly, real-time data set constructed by McCracken and Ng (2016) and regularly updated by the Federal Reserve Bank of St. Louis.[5] It includes series from categories such as output and income, prices, labor markets, housing and financial markets. Vintages of the "FRED-MD" data set are available from December 1999 onwards, reflecting the information available at the end of the respective month. Regarding transformations prior to estimation, we mainly follow the suggestions in McCracken and Ng (2016). However, given that our sample starts in 1985, some modifications are in order to reflect the shorter span of the time series. For example, in the original FRED-MD dataset some price series and average hourly earnings indicators are included in second (log) differences to achieve stationarity. Over our shorter, "Great Moderation" sample, the means and variances of the log differences are constant so that we do not need to difference these series twice. In contrast, housing starts and permits are included in log levels in the original dataset. Over the shorter sample used in the estimation of the nowcasting models, the log of the series exhibits large and persistent swings around the long-run mean. We therefore consider it more appropriate to difference the series to achvieve stationarity. These modifications bring our dataset more in line with the recent literature on nowcasting US GDP growth, e.g. Aastveit et al. (2018).

As the nowcasting literature has emphasized the importance of "soft", survey-based sentiment indicators, we supplement the McCracken and Ng (2016)-dataset with the Federal Reserve Bank of Philadelphia's Manufacturing Business Outlook Survey. While regional in nature, these indicators are available over a long period of time and provide potentially useful information for the gross domestic product as they are published very timely, usually in the middle of the reference month. We include all 20 series covering e.g. firms' assessement of current and future activity, orders, etc (Table I.2). All in all, this yields a monthly real-time data set containing roughly 130 variables, though as mentioned above the exact number varies slightly over the course of the evaluation period.

I.C.2 Germany

Nowcasts of German GDP growth are based on a real-time dataset comprised of over 100 monthly variables similar to that employed by Schumacher (2007). Vintages are compiled from the Deutsche Bundesbank Real-Time Database and augmented with financial market and the survey-based ifo indicators. Real-time vintages for a sufficiently large number of variables are available as of November 2005. For others, these become available over the course of the evaluation period. As a result, the size of the dataset increases somewhat from 167 monthly variables in the January 2006 vintage to 172 series as of December 2017. The

---

[5]All vintages can be downloaded from `https://research.stlouisfed.org/econ/mccracken/fred-databases/`.

Table I.3: Real-time data, Germany: production

variable/sector	adj.	trafo	1st obs.	1st vintage
industrial production	p,s,c	3	Jan 1991	6.2.1995
industrial production and construction	p,s,c	3	Jan 1991	6.2.1995
total construction	p,s,c	3	Jan 2010	3.8.2013
main construction industry	p,s,c	3	Jan 1991	6.2.1995
finishing trade	p,s,c	3	Jan 2010	3.8.2013
building construction	p,s,c	3	Jan 1991	7.11.2005
civil engineering	p,s,c	3	Jan 1991	7.11.2005
industry	p,s,c	3	Jan 1991	6.2.1995
intermediate goods	p,s,c	3	Jan 1991	6.2.1995
investment goods	p,s,c	3	Jan 1991	6.2.1995
consumption goods	p,s,c	3	Jan 1991	6.2.1995
durable goods	p,s,c	3	Jan 1991	6.2.1995
non-durable goods	p,s,c	3	Jan 1991	6.2.1995
energy	p,s,c	3	Jan 1991	6.2.1995

Source: Deutsche Bundesbank.

Table I.4: Real-time data, Germany: orders

variable/sector	adj.	trafo	1st obs.	1st vintage
industry	p,s,c	3	Jan 1991	6.9.2001
industry (domestic)	p,s,c	3	Jan 1991	6.9.2001
industry (abroad)	p,s,c	3	Jan 1991	6.9.2001
intermediate goods	p,s,c	3	Jan 1991	8.6.1995
intermediate goods (domestic)	p,s,c	3	Jan 1991	8.6.1995
intermediate goods (abroad)	p,s,c	3	Jan 1991	8.6.1995
investment goods	p,s,c	3	Jan 1991	8.6.1995
investment goods (domestic)	p,s,c	3	Jan 1991	8.6.1995
investment goods (abroad)	p,s,c	3	Jan 1991	8.6.1995
consumption goods	p,s,c	3	Jan 1991	8.6.1995
consumption goods (domestic)	p,s,c	3	Jan 1991	8.6.1995
consumption goods (abroad)	p,s,c	3	Jan 1991	8.6.1995
building construction	p,s,c	3	Jan 1991	22.11.2005
civil engineering	p,s,c	3	Jan 1991	22.11.2005
residential construction	p,s,c	3	Jan 1991	22.11.2005
construction industry (private)	p,s,c	3	Jan 1991	22.11.2005
construction industry (public)	p,s,c	3	Jan 1991	22.11.2005

Source: Deutsche Bundesbank.

starting point of our sample is January 1992.

Below we list the time-series comprising the seven groups of our dataset. The second column refers to the type of adjustment that have been applied to the time series (**p**rice-adjusted, **s**easonally adjusted, **c**alendar adjusted) while the third columns lists how each series is transformed to achieve stationarity (3 = difference of logarithm, 2 = difference, 1 = level). Furthermore, we also highlight if a series has a later starting point than January 1991 or if the vintages do not go back as far as November 2005 to indicate that the variable is only included in part of the evaluation period.

Table I.5: Real-time data, Germany: turnover

variable/sector	adj.	trafo	1st obs.	1st vintage
industry	s,c	3	Jan 1991	4.11.2005
industry (domestic)	s,c	3	Jan 1991	4.11.2005
industry (abroad)	s,c	3	Jan 1991	4.11.2005
intermediate goods	s,c	3	Jan 1991	4.11.2005
intermediate goods (domestic)	s,c	3	Jan 1991	4.11.2005
intermediate goods (abroad)	s,c	3	Jan 1991	4.11.2005
investment goods	s,c	3	Jan 1991	4.11.2005
investment goods (domestic)	s,c	3	Jan 1991	4.11.2005
investment goods (abroad)	s,c	3	Jan 1991	4.11.2005
consumption goods	s,c	3	Jan 1991	4.11.2005
consumption goods (domestic)	s,c	3	Jan 1991	4.11.2005
consumption goods (abroad)	s,c	3	Jan 1991	4.11.2005
durable goods	s,c	3	Jan 1991	4.11.2005
non-durable goods	s,c	3	Jan 1991	4.11.2005
residential construction	s,c	3	Jan 1991	22.11.2005
construction industry (private)	s,c	3	Jan 1991	20.11.2005
construction (public)	s,c	3	Jan 1991	4.11.2005
retail sales	p,s,c	3	Jan 1994	17.11.2005
retail sales excluding cars	p,s,c	3	Jan 1994	17.11.2005
retail sales: cars	p,s,c	3	Jan 1994	17.11.2005

Source: Deutsche Bundesbank.

Table I.6: Real-time data, Germany: prices

variable/sector	adj.	trafo	1st obs.	1st vintage
consumper price index (CPI)	s,c	3	Jan 1991	28.11.1995
CPI excl. energy	s	3	Jan 1991	28.1.1999
CPI excl. energy and food	s	3	Jan 1995	13.4.2017
CPI: food	s	3	Jan 1991	28.1.1999
CPI: other non-durables and durables	s	3	Jan 1995	14.11.2003
CPI: energy	s,c	3	Jan 1991	29.2.2008
CPI: services	s,c	3	Jan 2010	13.4.2017
CPI: services (excluding rents)	s	3	Jan 2000	14.11.2003
CPI: rents	s	3	Jan 1991	11.11.2005
CPI: rents excl. ancillary costs	s	3	Jan 1995	16.4.2008
produer price index (PPI): industrial products	s	3	Jan 1991	18.11.2005
PPI: industrial products excl. energy	s	3	Jan 1994	18.11.2005
PPI: agricultural products	s	3	Jan 1968	7.11.2005
export price index	s	3	Jan 1970	24.11.2005
import price index	s	3	Jan 1970	24.11.2005

Source: Deutsche Bundesbank.

Table I.7: Real-time data, Germany: labor market

variable/sector	adj.	trafo	1st obs.	1st vintage
employment	s	3	Jan 1991	8.8.1995
employment: manufacturing and mining	s	3	Jan 1991	16.11.2005
employment: main construction industry	s	3	Jan 1991	22.11.2005
hours worked	s,c	3	Jan 1991	22.11.2005
hours worked: manufacturing and mining	s,c	3	Jan 1991	16.11.2005
hours worked: main construction industry	s,c	3	Jan 1991	22.11.2005
employees subject to social security contributions	s	3	Jan 1991	30.3.2006
gross wages and salaries: manufacturing and mining	s,c	3	Jan 1991	16.11.2005
gross wages and salaries: main construction industry	s,c	3	Jan 1991	22.11.2005

Source: Deutsche Bundesbank.

27

Table I.8: Real-time data, Germany: financial markets

variable/sector	adj.	trafo	1st obs.	1st vintage
yields on debt securities issued by residents	none	2	Jan 1991	-
yields: bank debt securities	none	2	Jan 1991	-
yields: mortgage Pfandbriefe	none	2	Jan 1991	-
yields: public Pfandbriefe securities	none	2	Jan 1991	-
yields: special purpose credit institutions	none	2	Jan 1991	-
yields: other bank debt securities	none	2	Jan 1991	-
yields: corporate debt securities	none	2	Jan 1991	-
yields: public debt securities	none	2	Jan 1991	-
yields: state government securities	none	2	Jan 1991	-
government bond yields: 6 month maturity	none	2	Jan 1991	-
government bond yields: 1 year maturity	none	2	Jan 1991	-
government bond yields: 2 year maturity	none	2	Jan 1991	-
government bond yields: 3 year maturity	none	2	Jan 1991	-
government bond yields: 4 year maturity	none	2	Jan 1991	-
government bond yields: 5 year maturity	none	2	Jan 1991	-
government bond yields: 6 year maturity	none	2	Jan 1991	-
government bond yields: 7 year maturity	none	2	Jan 1991	-
government bond yields: 8 year maturity	none	2	Jan 1991	-
government bond yields: 9 year maturity	none	2	Jan 1991	-
government bond yields: 10 year maturity	none	2	Jan 1991	-
CDAX index	none	3	Jan 1991	-
Nominal effective exchange rate (narrow)	none	3	Jan 1991	-
Nominal effective exchange rate (broad)	none	3	Jan 1991	-

Source: Deutsche Bundesbank.

Table I.9: Real-time data, Germany: survey indicators

variable/sector	adj.	trafo	1st obs.	1st vintage
ifo: manufacturing, current situation	s	2	Jan 1991	-
ifo: manufacturing, climate	s	2	Jan 1991	-
ifo: manufacturing, expectations	s	2	Jan 1991	-
ifo: manufacturing, demand	s	2	Jan 1991	-
ifo: manufacturing, prices	s	2	Jan 1991	-
ifo: manufacturing, employment expectations	s	2	Jan 1991	-
ifo: manufacturing, export expectations	s	2	Jan 1991	-
ifo: manufacturing, orders from abroad	s	2	Jan 1991	-
ifo: manufacturing, inventories	s	2	Jan 1991	-
ifo: manufacturing, orders	s	2	Jan 1991	-
ifo: manufacturing, production expecations	s	2	Jan 1991	-
ifo: manufacturing, orders (m/m)	s	2	Jan 1991	-
ifo: manufacturing, production (m/m)	s	2	Jan 1991	-
ifo: manufacturing, price expectation	s	2	Jan 1991	-
ifo: wholesale, current situation	s	2	Jan 1991	-
ifo: wholesale, climate	s	2	Jan 1991	-
ifo: wholesale, expectations	s	2	Jan 1991	-
ifo: wholesale, employment expectations	s	2	Jan 1991	-
ifo: wholesale, inventories	s	2	Jan 1991	-
ifo: wholesale, order expectations	s	2	Jan 1991	-
ifo: wholesale, price expectations	s	2	Jan 1991	-
ifo: wholesale, inventories	s	2	Jan 1991	-
ifo: wholesale, prices (m/m)	s	2	Jan 1991	-
ifo: wholesale, turnover (m/m)	s	2	Jan 1991	-
ifo: retail, current situation	s	2	Jan 1991	-
ifo: retail, climate	s	2	Jan 1991	-
ifo: retail, expectations	s	2	Jan 1991	-
ifo: retail, employment expectations	s	2	Jan 1991	-
ifo: retail, inventories	s	2	Jan 1991	-
ifo: retail, order expectations	s	2	Jan 1991	-
ifo: retail, price expectations	s	2	Jan 1991	-
ifo: retail, inventories	s	2	Jan 1991	-
ifo: retail, prices (m/m)	s	2	Jan 1991	-
ifo: retail, turnover (m/m)	s	2	Jan 1991	-
ifo: construction, current situation	s	2	Jan 1991	-
ifo: construction, climate	s	2	Jan 1991	-
ifo: construction, expectations	s	2	Jan 1991	-
ifo: construction, prices	s	2	Jan 1991	-
ifo: construction, capacity utilisation	s	2	Jan 1991	-

All series are seasonally adjusted in real-time using X-11-ARIMA. *Source*: ifo-Institute.

I.D   Additional results

This section presents the results of the nowcasting evaluation for several robustness checks discussed in Section I.3.5.

United States

surveys in first differences, rolling estimation sample, P=1, first release
surveys in levels, recursive estimation sample, P=1, first release
surveys in levels, rolling estimation sample, P=1, first release

Germany

surveys in first differences, rolling estimation sample, P=1, first release
surveys in levels, recursive estimation sample, P=1, first release
surveys in levels, rolling estimation sample, P=1, first release

Table I.10: Additional results: United States (first, diff, rolling)

		full sample			post-crisis sample		
		RMSFE h=0	logS h=0	CRPS h=0	RMSFE h=0	logS h=0	CRPS h=0
B-AR		0.47	0.69	0.26	0.32	0.55	0.20
R=1	Nd	0.81	0.82	0.86	1.04	0.97	1.02
	NIG	0.92	0.89	0.93	1.05	0.94	1.02
	MG	0.84	0.82	0.88	1.09	0.97	1.05
	PMNM	0.86	0.85	0.89	1.06	0.95	1.03
	HS+	0.84	0.83	0.88	1.03	0.99	1.03
R=2	Nd	0.86	0.83	0.89	0.96	0.89	0.96
	NIG	0.89	0.85	0.91	0.99	0.87	0.98
	MG	0.84	0.80	0.87	0.98	0.88	0.97
	PMNM	0.84	0.81	0.88	1.05	0.93	1.02
	HS+	0.84	0.80	0.87	1.02	0.92	1.00
R=5	Nd	0.81	0.80	0.87	1.08	0.87	1.03
	NIG	0.79	0.75	0.84	1.12	0.86	1.05
	MG	0.79	0.72	0.83	1.12	0.85	1.04
	PMNM	0.79	0.74	0.84	1.09	0.87	1.03
	HS+	0.79	0.73	0.84	1.10	0.84	1.03
R=8	Nd	0.84	0.88	0.89	1.04	0.97	1.03
	NIG	0.80	0.75	0.83	1.05	0.82	1.00
	MG	0.75	0.70	0.80	1.02	0.78	0.98
	PMNM	0.72	0.66	0.77	0.87	0.66	0.87
	HS+	0.74	0.66	0.78	0.94	0.69	0.90
pool	Nd	0.80	0.85	0.86	1.03	0.95	1.01
	NIG	0.76	0.74	0.81	0.99	0.81	0.96
	MG	0.75	0.71	0.80	1.00	0.82	0.97
	PMNM	0.74	0.70	0.79	0.95	0.76	0.93
	HS+	0.77	0.70	0.81	1.01	0.77	0.96

RMSFE is the root mean squared forecast error, logS and CRPS are the average log score and continuous ranked probability score. All entries for the factor models are relative to the B-AR benchmark (see text for details) and negatively orientated so that a value in the table below 1 corresponds to a better performance than the benchmark. The forecast horizon h is in months. The full sample period is 2000Q1-2018Q4, the post-crisis sample starts in 2010Q1 and ends in 2018Q4.

Table I.11: Additional results: United States (first, level, rec)

		full sample			post-crisis sample		
		RMSFE h=0	logS h=0	CRPS h=0	RMSFE h=0	logS h=0	CRPS h=0
B-AR		0.45	0.64	0.25	0.32	0.50	0.19
R=1	Nd	1.04	1.13	1.05	0.99	1.07	1.03
	NIG	1.04	1.08	1.04	0.97	1.08	1.01
	MG	1.03	1.10	1.03	0.99	1.09	1.03
	PMNM	1.02	1.09	1.02	0.92	1.04	0.98
	HS+	1.01	1.10	1.01	0.91	1.04	0.98
R=2	Nd	0.93	0.89	0.96	1.20	1.04	1.14
	NIG	0.95	0.93	0.99	1.21	1.06	1.15
	MG	0.94	0.93	0.98	1.19	1.05	1.13
	PMNM	0.94	0.89	0.97	1.13	0.96	1.08
	HS+	0.93	0.88	0.96	1.15	0.98	1.09
R=5	Nd	0.83	0.75	0.86	1.07	0.86	1.01
	NIG	0.87	0.80	0.90	1.19	0.98	1.11
	MG	0.83	0.76	0.86	0.97	0.81	0.95
	PMNM	0.87	0.79	0.90	1.19	1.00	1.13
	HS+	0.86	0.77	0.89	1.19	0.97	1.12
R=8	Nd	0.81	0.72	0.85	1.02	0.80	0.98
	NIG	0.80	0.68	0.83	1.02	0.78	0.98
	MG	0.91	0.85	0.93	1.04	0.91	1.00
	PMNM	0.80	0.69	0.84	1.01	0.76	0.97
	HS+	0.78	0.67	0.82	1.03	0.78	0.98
pool	Nd	0.80	0.75	0.84	1.02	0.88	0.99
	NIG	0.81	0.76	0.85	1.03	0.87	1.00
	MG	0.89	0.84	0.91	1.05	0.95	1.02
	PMNM	0.80	0.75	0.85	1.02	0.86	0.99
	HS+	0.80	0.74	0.84	1.01	0.84	0.98

RMSFE is the root mean squared forecast error, logS and CRPS are the average log score and continuous ranked probability score. All entries for the factor models are relative to the B-AR benchmark (see text for details) and negatively orientated so that a value in the table below 1 corresponds to a better performance than the benchmark. The forecast horizon h is in months. The full sample period is 2000Q1-2018Q4, the post-crisis sample starts in 2010Q1 and ends in 2018Q4.

Table I.12: Additional results: United States (first, level, rolling)

		full sample			post-crisis sample		
		RMSFE h=0	logS h=0	CRPS h=0	RMSFE h=0	logS h=0	CRPS h=0
B-AR		0.47	0.69	0.26	0.32	0.55	0.20
R=1	Nd	0.88	0.97	0.93	0.87	1.05	0.98
	NIG	0.87	0.96	0.92	0.87	1.07	0.99
	MG	0.87	0.96	0.93	0.87	1.07	0.98
	PMNM	1.00	1.06	1.02	0.90	1.10	1.00
	HS+	0.99	1.08	1.01	0.88	1.09	0.99
R=2	Nd	0.84	0.83	0.88	0.97	0.86	0.96
	NIG	0.84	0.83	0.89	0.97	0.87	0.96
	MG	0.84	0.83	0.88	0.97	0.88	0.97
	PMNM	0.91	0.88	0.94	1.07	0.94	1.03
	HS+	0.87	0.83	0.90	1.01	0.87	0.98
R=5	Nd	0.80	0.76	0.85	0.92	0.70	0.90
	NIG	0.74	0.68	0.79	0.91	0.69	0.89
	MG	0.74	0.67	0.79	0.95	0.70	0.92
	PMNM	0.86	0.78	0.89	1.04	0.80	0.99
	HS+	0.80	0.72	0.84	1.00	0.74	0.95
R=8	Nd	0.86	0.92	0.92	1.11	0.94	1.06
	NIG	0.77	0.75	0.82	1.01	0.80	0.97
	MG	0.74	0.71	0.80	1.01	0.77	0.96
	PMNM	0.81	0.75	0.85	1.04	0.81	1.00
	HS+	0.74	0.66	0.79	0.96	0.73	0.93
pool	Nd	0.75	0.84	0.83	0.92	0.88	0.95
	NIG	0.75	0.76	0.81	0.90	0.79	0.91
	MG	0.73	0.74	0.79	0.90	0.80	0.92
	PMNM	0.80	0.78	0.85	0.97	0.86	0.96
	HS+	0.78	0.75	0.82	0.93	0.79	0.92

RMSFE is the root mean squared forecast error, logS and CRPS are the average log score and continuous ranked probability score. All entries for the factor models are relative to the B-AR benchmark (see text for details) and negatively orientated so that a value in the table below 1 corresponds to a better performance than the benchmark. The forecast horizon h is in months. The full sample period is 2000Q1-2018Q4, the post-crisis sample starts in 2010Q1 and ends in 2018Q4.

Table I.13: Additional results: Germany (first, diff, rolling)

		full sample			post-crisis sample		
		RMSFE h=0	logS h=0	CRPS h=0	RMSFE h=0	logS h=0	CRPS h=0
B-AR		0.91	1.35	0.45	0.52	1.04	0.32
R=1	Nd	0.73	0.69	0.78	0.96	0.68	0.88
	NIG	0.72	0.69	0.77	0.92	0.66	0.85
	MG	0.71	0.66	0.76	0.92	0.65	0.85
	PMNM	0.73	0.76	0.78	0.93	0.66	0.86
	HS+	0.72	0.70	0.76	0.91	0.62	0.83
R=2	Nd	0.60	0.58	0.66	0.85	0.60	0.80
	NIG	0.54	0.51	0.60	0.76	0.53	0.72
	MG	0.55	0.51	0.61	0.78	0.54	0.73
	PMNM	0.56	0.52	0.62	0.81	0.55	0.75
	HS+	0.56	0.54	0.63	0.81	0.57	0.75
R=5	Nd	0.77	0.64	0.78	0.97	0.71	0.89
	NIG	0.58	0.50	0.61	0.69	0.50	0.66
	MG	0.60	0.50	0.63	0.69	0.48	0.66
	PMNM	0.59	0.53	0.63	0.73	0.53	0.70
	HS+	0.61	0.53	0.64	0.70	0.50	0.68
R=8	Nd	0.66	0.65	0.72	1.04	0.78	0.94
	NIG	0.55	0.49	0.58	0.74	0.53	0.69
	MG	0.58	0.49	0.60	0.72	0.50	0.68
	PMNM	0.54	0.49	0.59	0.66	0.52	0.66
	HS+	0.62	0.52	0.63	0.67	0.50	0.65
pool	Nd	0.59	0.59	0.65	0.82	0.67	0.79
	NIG	0.54	0.49	0.58	0.68	0.52	0.66
	MG	0.55	0.49	0.59	0.68	0.51	0.66
	PMNM	0.57	0.52	0.61	0.71	0.55	0.69
	HS+	0.57	0.51	0.61	0.67	0.51	0.66

RMSFE is the root mean squared forecast error, logS and CRPS are the average log score and continuous ranked probability score. All entries for the factor models are relative to the B-AR benchmark (see text for details) and negatively orientated so that a value in the table below 1 corresponds to a better performance than the benchmark. The forecast horizon h is in months. The full sample period is 2006Q1-2018Q4, the post-crisis sample starts in 2010Q1 and ends in 2018Q4.

Table I.14: Additional results: Germany (first, level, rec)

		full sample			post-crisis sample		
		RMSFE h=0	logS h=0	CRPS h=0	RMSFE h=0	logS h=0	CRPS h=0
B-AR		0.86	1.73	0.41	0.49	0.89	0.29
R=1	Nd	0.88	0.66	0.98	1.10	1.01	1.09
	NIG	0.88	0.70	0.98	1.10	1.01	1.10
	MG	0.88	0.66	0.98	1.10	1.00	1.08
	PMNM	0.88	0.66	0.97	1.09	1.01	1.08
	HS+	0.88	0.69	0.97	1.06	1.01	1.07
R=2	Nd	0.87	0.65	0.92	1.02	0.99	1.01
	NIG	0.74	0.53	0.79	0.82	0.85	0.88
	MG	0.70	0.51	0.75	0.81	0.84	0.87
	PMNM	0.75	0.53	0.80	0.82	0.85	0.87
	HS+	0.76	0.53	0.80	0.81	0.85	0.87
R=5	Nd	1.01	0.86	1.14	1.18	0.96	1.08
	NIG	0.74	0.53	0.82	0.92	0.74	0.87
	MG	0.68	0.53	0.81	1.05	0.92	1.04
	PMNM	0.74	0.53	0.82	0.82	0.69	0.80
	HS+	0.71	0.50	0.79	0.85	0.70	0.82
R=8	Nd	1.17	0.87	1.17	1.28	1.04	1.21
	NIG	0.68	0.51	0.75	0.92	0.76	0.89
	MG	0.73	0.52	0.80	0.86	0.80	0.88
	PMNM	0.51	0.37	0.60	0.75	0.62	0.76
	HS+	0.51	0.36	0.59	0.72	0.59	0.73
pool	Nd	0.69	0.55	0.82	0.85	0.88	0.90
	NIG	0.55	0.45	0.66	0.73	0.76	0.80
	MG	0.59	0.48	0.70	0.82	0.83	0.87
	PMNM	0.51	0.44	0.63	0.61	0.70	0.72
	HS+	0.50	0.42	0.61	0.65	0.68	0.73

RMSFE is the root mean squared forecast error, logS and CRPS are the average log score and continuous ranked probability score. All entries for the factor models are relative to the B-AR benchmark (see text for details) and negatively orientated so that a value in the table below 1 corresponds to a better performance than the benchmark. The forecast horizon h is in months. The full sample period is 2006Q1-2018Q4, the post-crisis sample starts in 2010Q1 and ends in 2018Q4.

Table I.15: Additional results: Germany (first, level, rolling)

		full sample			post-crisis sample		
		RMSFE h=0	logS h=0	CRPS h=0	RMSFE h=0	logS h=0	CRPS h=0
B-AR		0.91	1.35	0.45	0.52	1.04	0.32
R=1	Nd	0.89	1.07	0.97	1.09	0.98	1.07
	NIG	0.88	1.11	0.97	1.08	0.99	1.07
	MG	0.89	1.12	0.96	1.07	0.96	1.04
	PMNM	0.87	1.20	0.95	1.01	0.96	1.01
	HS+	0.87	1.20	0.94	1.01	0.95	1.01
R=2	Nd	0.76	0.77	0.83	0.99	0.76	0.93
	NIG	0.65	0.68	0.71	0.89	0.76	0.87
	MG	0.65	0.69	0.71	0.88	0.77	0.86
	PMNM	0.73	0.73	0.79	0.94	0.77	0.90
	HS+	0.65	0.68	0.70	0.87	0.72	0.84
R=5	Nd	0.97	1.00	1.05	1.23	0.97	1.17
	NIG	0.62	0.64	0.68	0.77	0.64	0.76
	MG	0.60	0.61	0.66	0.73	0.62	0.73
	PMNM	0.62	0.61	0.67	0.71	0.59	0.71
	HS+	0.58	0.63	0.65	0.70	0.61	0.71
R=8	Nd	0.87	0.85	0.96	1.39	1.08	1.32
	NIG	0.58	0.55	0.64	0.82	0.63	0.78
	MG	0.61	0.56	0.66	0.76	0.58	0.73
	PMNM	0.57	0.55	0.62	0.70	0.55	0.69
	HS+	0.62	0.57	0.66	0.70	0.53	0.68
pool	Nd	0.61	0.77	0.75	0.90	0.92	0.94
	NIG	0.53	0.58	0.61	0.66	0.67	0.72
	MG	0.56	0.60	0.63	0.67	0.65	0.71
	PMNM	0.53	0.58	0.60	0.62	0.63	0.68
	HS+	0.55	0.59	0.62	0.63	0.61	0.68

RMSFE is the root mean squared forecast error, logS and CRPS are the average log score and continuous ranked probability score. All entries for the factor models are relative to the B-AR benchmark (see text for details) and negatively orientated so that a value in the table below 1 corresponds to a better performance than the benchmark. The forecast horizon h is in months. The full sample period is 2006Q1-2018Q4, the post-crisis sample starts in 2010Q1 and ends in 2018Q4.

I.E   MCMC diagnostics

To assess the performance of the Gibbs Sampler in terms of convergence and mixing, we calculate the inefficiency factors of the draws from the predictive density in the nowcasting application in Section I.3. The inefficiency factor is defined as (e.g. Chib, 2011):

$$\text{Ineff}(\hat{h}_G) = \frac{Var(\hat{h}_G)}{s^2/G} \tag{I.39}$$

where G is the length of the chain, s^2 its sample variance and $Var(\hat{h}_G)$ an estimate of the variance of the simulation error taking into account autocorrelation in the Markov chain. For independent draws from the posterior distribution the inefficiency factor is equal to 1; higher values of Ineff($\hat{h}_G$) thus signal autocorrelation in the chain as result of poor mixing or lack of convergence. An estimate of the inefficiency factors is given by the sum of the autocorrelation coefficients ρ of the posterior draws, i.e.

$$\text{Ineff}(\hat{h}_G) = 1 + 2\sum_{l}^{L}\rho_l \tag{I.40}$$

where L is some suitably chosen upper bound. Alternatively, the inefficiency factor can also be expressed as

$$\text{Ineff}(\hat{h}_G) = \frac{G}{ESS(\hat{h}_G)} \tag{I.41}$$

where $ESS(\hat{h}_G)$ is the "effective sample size" of the Markov chain. The latter can readily be computed using the R package **coda** (Plummer et al., 2006).

Figure I.2 shows boxplots of the inefficiency factors for the predictive densities from the different priors and model specifications. The majority of inefficiency factors of the predictive densities are concentrated at the lower end of the scale, indicating close to independent draws from the posterior distribution. There are a few outliers - defined as any observations that exceeds the median by 1.5 times the interquartile range and highlighted in the plot by a dot - for each specification, with some inefficiency factors as high as 50 or even 100 in one case. But overall the chains appear to mix well. An exception is the Multiplicative Gamma prior in the case of the United States where the number of outliers is much larger. We note, however, that this is only the case for models with $P = 1$, a recursive estimation sample and the survey indicators in levels. Furthermore, judging from the tables presented in Appendix I.D this does not seem to have impacted the nowcasting performance in a substantial way as the results are qualitatively and quantitatively similar to those obtained for the other specifications where the inefficiency factors are smaller.

Figure I.2: Boxplots of inefficiency factors for different model specifications

Note: The figure shows the inefficiency factors for the different model specifications. *rec/rolling*: recursive or rolling estimation window, *level/diff*: survey indicators in levels or first differences. The different priors are denoted as follows: HS+ = horseshoe plus, MG = multiplicative Gamma, NIG = Normal Inverse Gamma, Nd = Normal diffuse, PMNM = point mass normal mixture. The number of lags in the factor VAR equals $P = 1$. For details, see the main text.

Chapter II

Precision-based sampling with missing observations: A factor model application

This chapter of my dissertation is co-authored with Christian Schumacher (Deutsche Bundesbank). It has been published as Deutsche Bundesbank Discussion Paper No. 11/2021.

Abstract

We propose a new approach to sample unobserved states conditional on available data in (conditionally) linear unobserved component models when some of the observations are missing. The approach is based on the precision matrix of the states and model variables, which is sparse and banded in many economic applications and allows for efficient sampling. The existing literature on precision-based sampling is focused on complete-data applications, whereas the proposed samplers in this paper provide draws for states and missing observations by using permutations of the precision matrix. The approaches can be easily integrated into Bayesian estimation procedures like the Gibbs sampler. By allowing for incomplete data sets, the proposed sampler expands the range of potential applications for precision-based samplers in practice. We derive the sampler for a factor model, although it can be applied to a wider range of empirical macroeconomic models. In an empirical application, we estimate international factors in GDP growth in a large unbalanced data set of about 180 countries.

Keywords: Precision-based sampling, Bayesian estimation, state-space models, missing observations, factor models, banded matrices

JEL classification: C32, C38, C63, C55.

II.1 Introduction

In the recent literature, conditional samplers for unobserved states given data and parameters based on the precision matrix have received considerable attention. First applications of precision-based samplers to address economic questions have been provided by Chan and Jeliazkov (2009) and McCausland (2012), building on seminal work by Rue (2001) and Rue and Held (2005) on Gaussian Markov random fields. There is now a huge number of applications of precision-based samplers in the empirical macroeconomic literature. Recent examples of state-space models with unobserved components such as output gaps or inflation trends and time-varying parameters are Chan et al. (2013, 2016), Grant and Chan (2017), and Chan et al. (2018b) or time-varying parameter vector autoregressive (VAR) models with a vast number of applications such as Chan and Eisenstat (2018), Chan (2020), Chan et al. (2020) and references cited therein. Factor model applications are provided by Chan and Jeliazkov (2009), McCausland (2015), and Kaufmann and Schumacher (2017, 2019). These application typically employ precision-based sampling of states given data as part of a Bayesian estimation procedures like the Gibbs sampler, whose aim is to draw from the posterior density $p(\theta, \eta|x)$, where η are the unobserved states, x denotes data, and θ are model parameters. A standard Gibbs sampler iterates between drawing from the conditional posteriors $p(\eta|x, \theta)$ and $p(\theta|x, \eta)$. Drawing from $p(\eta|x, \theta)$ can be carried out efficiently by precision-based samplers, as the underlying precision matrix of states and variables is banded in many economic models and allows for the application of fast sparse matrix techniques (Rue, 2001).

The literature cited above applies precision-based samplers to complete data sets. In practice, however, observations in multivariate data sets can often be missing. In this case, an analyst has the choice of removing all time series with missing observation and using balanced data only. This, however, implies a loss of information. The alternative is to use the larger unbalanced data, but this raises the need for estimation methods that can tackle missing observations.

In this paper, we propose a precision-based sampler for unobserved states in the presence of partly missing observations. In line with the literature cited above, the state-space model is assumed to be (conditionally) linear and the disturbances follow normal distributions. If the data are completely available, the states can be sampled from a conditional distribution of a multivariate normal using fast band-matrix computation as in Rue (2001) and Chan and Jeliazkov (2009). Important alternative samplers from the literature based on the Kalman filter are provided in the seminal papers Carter and Kohn (1994), Frühwirth-Schnatter (1994), and Durbin and Koopman (2002). The main contribution of this paper is the extension of the literature on precision-based samplers by considering missing observations. We do so by implementing an efficient reordering of states, observed and unobserved variables that facilitates fast band-matrix computation. The sampler provides draws from the conditional posterior distribution $p(\eta, x_m|x_o, \theta)$ for the states and the missing observations x_m conditional on observed data x_o and parameters. Thus it can easily be integrated

into Gibbs samplers to tackle missing data as proposed by Little and Rubin (2002).

To illustrate the sampling method, we use a factor model with vector autoregressive
(VAR) dynamics for the factors and autoregressive (AR) idiosyncratic components in a
Bayesian framework (McCausland, 2015; Kaufmann and Schumacher, 2019). Alternative
missing-data approaches for factor models in the literature are Angelini et al. (2006) and
Marcellino (2007) for backdating and interpolation in a principal-components framework,
see also Bai and Ng (2019) for a recent contribution. In a Bayesian framework, Otrok and
Pourpourides (2017) interpolate data in panel data and Müller et al. (2019) interpolate in-
ternational long-run growth data. Further approaches based on the Kalman filter have been
proposed by Jungbacker et al. (2011) and Banbura and Modugno (2014), amongst others.

For the factor model, we derive alternative precision-based samplers for the factors and
missing values in the data, which differ with respect to the permutations of η, x_m and
x_o in the precision matrix. We compare the accuracy and computational efficiency of the
precision-based samplers in simulations. As an empirical application, we estimate inter-
national factors in GDP growth along the lines of the literature on international business
cycles with Bayesian techniques (Kose et al., 2003, 2008; Francis et al., 2017; Müller et al.,
2019). We compare estimation results based on balanced data for about 50 country-GDP
time series using the standard precision-based sampler and results on a larger, unbalanced
data set consisting of more than 180 GDP time series. We check whether results based on
balanced data are robust when using the larger information set.

The paper proceeds as follows. In Section II.2, we introduce the factor model, whereas
Section II.3 describes its estimation using Bayesian methods given complete data with a fo-
cus on precision-based sampling of the unobserved factors. Section II.4 provides alternative
precision-based samplers for partly missing observations. A simulation exercise to compare
the alternative precision-based samplers is provided in Section II.5. In Section II.6, we dis-
cuss the results of the empirical application. Section II.7 briefly discusses the calculation
of the marginal likelihood and extensions to other models such as time-varying parameter
(TVP) Bayesian VAR models. Section II.8 concludes.

II.2 The factor model

The factor model explains the $(N \times 1)$−dimensional vector of variables $x_t = (x_{1,t}, x_{2,t}, \ldots,$
$x_{N,t})^\mathsf{T}$ in time period t according to

$$x_t = \lambda \eta_t + \epsilon_t, \tag{II.1}$$

$$\eta_t = \phi \eta_{t-1} + u_{\eta,t}, \quad \epsilon_t = \psi \epsilon_{t-1} + u_{\epsilon,t}. \tag{II.2}$$

The $(r \times 1)$−dimensional vector of factors is denoted as η_t, and λ is the $(N \times r)$−dimensional
matrix of factor loadings. The factor representation (Equation (II.1)) holds for $t = 1, \ldots, T$.
The factors follow a VAR(1) process with the $(r \times r)$−dimensional lag parameter matrix
ϕ. The factor VAR disturbances are distributed as $u_{\eta,t} \sim \mathcal{N}(0_{r \times 1}, \omega_\eta)$. The idiosyncratic

components collected in the $(N \times 1)$–dimensional vector $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{N,t})^\mathsf{T}$ each follow AR(1) processes such that the $(N \times N)$–dimensional coefficient matrix ψ is diagonal containing the AR(1) lag parameters ψ_i for $i = 1, \ldots, N$ on the main diagonal. The one-lag specification only serves to illustrate the methods. The empirical applications later in the paper will consider a factor VAR(p) with AR(q) idiosyncratic components where $p, q > 1$. The idiosyncratic disturbances are distributed as $u_{\epsilon,t} \sim \mathcal{N}(0_{N \times 1}, \omega_\epsilon)$, where ω_ϵ is also diagonal with diagonal elements $\omega_{\epsilon,i}$ for $i = 1, \ldots, N$. We assume that $u_{\eta,t}$ and $u_{\epsilon,t}$ are mutually independent and that the VAR and AR processes in (Equation (II.2)) are stationary. In addition, the equations in (Equation (II.2)) are defined for time periods $t = 2, \ldots, T$. For $t = 1$, let $\eta_1 \sim \mathcal{N}(\eta_{1|0}, \vartheta_{\eta,1|0})$ and $\epsilon_1 \sim \mathcal{N}(\epsilon_{1|0}, \vartheta_{\epsilon,1|0})$, respectively. The means of the distributions are set equal to their unconditional mean, which is zero in our case, so $\eta_{1|0} = 0_{r \times 1}$ and $\epsilon_{1|0} = 0_{N \times 1}$, respectively. As the processes underlying (Equation (II.2)) are stationary, $\vartheta_{\eta,1|0}$ and $\vartheta_{\epsilon,1|0}$ are set equal to the unconditional covariances implied by the model equations: For the states, we define $\vartheta_{\eta,1|0} = \vartheta_\eta$, where ϑ_η is equal to the solution of the vector equation $\vartheta_\eta = \phi \vartheta_\eta \phi^\mathsf{T} + \omega_\eta$, and $\vartheta_{\epsilon,1|0}$ is a diagonal matrix with $\omega_{\epsilon,i}/(1 - \psi_i^2)$ on the main diagonal for $i = 1, \ldots, N$.

For compact notation, we stack all time periods for the variables x_t into one $(NT \times 1)$–dimensional vector according to $x = (x_1^\mathsf{T}, x_2^\mathsf{T}, \ldots, x_T^\mathsf{T})^\mathsf{T}$. In the same way, define the $(Tr \times 1)$–dimensional stacked vector of factors $\eta = (\eta_1^\mathsf{T}, \eta_2^\mathsf{T}, \ldots, \eta_T^\mathsf{T})^\mathsf{T}$ and the idiosyncratic components, $\epsilon = (\epsilon_1^\mathsf{T}, \epsilon_2^\mathsf{T}, \ldots, \epsilon_T^\mathsf{T})^\mathsf{T}$. We collect all the model parameters in the vector θ.

II.3   Precision-based sampling with complete data

For Bayesian estimation of the factor model, we first assume the data are complete and later generalize to the case when some observations are missing. Complete or balanced data means that we have one observation $x_{i,t}^o$ available for each variable explained in the model $x_{i,t} = x_{i,t}^o$ for all $i = 1, \ldots, N$ and $t = 1, \ldots, T$, or $x = x^o$ in brief.

Our aim is to sample from the posterior distribution

$$p(\eta, \theta | x) \propto L(x|\eta, \theta) p(\eta|\theta) p(\theta), \tag{II.3}$$

where the likelihood function $L(x|\eta, \theta)$ is implied by (Equation (II.1)) and (Equation (II.2)), and the priors are chosen closely in line with the existing factor model literature, see Section II.A for details. To obtain draws from the posterior distribution, we sample sequentially from the following conditional posterior distributions:

1. $p(\eta|x, \theta)$

2. $p(\theta|x, \eta)$

The precision-based sampler to draw from $p(\eta|x, \theta)$ is discussed in detail below, whereas details on the samplers for $p(\theta|x, \eta)$ and further model specifications are provided in Section II.A.

To prepare the application of conditional sampling from partitioned multivariate normals, we follow McCausland (2015) and define the joint vector

$$z = \begin{pmatrix} \eta \\ x \end{pmatrix} = \begin{pmatrix} I_{Tr} & 0_{Tr \times TN} \\ \Lambda & I_{TN} \end{pmatrix} \begin{pmatrix} \eta \\ \epsilon \end{pmatrix} \tag{II.4}$$

as a function of unobserved factors and idiosyncratic components. The matrix Λ is given by the Kronecker product $\Lambda = I_T \otimes \lambda$. According to (Equation (II.4)), the joint vector $z = (\eta^\top, x^\top)^\top$ is an affine transformation of $(\eta^\top, \epsilon^\top)^\top$, which are Gaussian and mutually independent. Hence, z also follows a multivariate normal distribution by

$$z|\theta \sim \mathcal{N}\left(0_{T(r+N) \times 1}, Q^{-1}\right), \tag{II.5}$$

where Q^{-1} is the $(T(r+N) \times T(r+N))$–dimensional covariance matrix and Q the corresponding precision matrix, which is conditional on model parameters. To facilitate efficient sampling, we have to find a tractable blocked expression for the precision matrix Q, which will allow us to apply general rules for sampling from partitioned Gaussian vectors.

We start by deriving the covariance matrix of z and write the factor VAR as $\Phi\eta = u_\eta$, where $u_\eta|\theta \sim \mathcal{N}(0_{Tr \times 1}, \Omega_\eta)$,

$$\Phi = \begin{pmatrix} I_r & & & & \\ -\phi & I_r & & & \\ & -\phi & \ddots & & \\ & & \ddots & I_r & \\ & & & -\phi & I_r \end{pmatrix}, \quad \text{and} \quad \Omega_\eta = \begin{pmatrix} \vartheta_{\eta,1|0} & & & \\ & \omega_\eta & & \\ & & \ddots & \\ & & & \omega_\eta \end{pmatrix}. \tag{II.6}$$

In stacked form, the vector of factors follows the distribution $\eta|\theta \sim \mathcal{N}(0_{Tr \times 1}, \Phi^{-1}\Omega_\eta\Phi^{-\top})$, where Φ has full rank and, hence, is invertible. Similarly, the stacked idiosyncratic components are defined as $\Psi\epsilon = u_\epsilon$ where $u_\epsilon|\theta \sim \mathcal{N}(0_{TN \times 1}, \Omega_\epsilon)$, and Ω_ϵ is a matrix containing the matrix $\vartheta_{\epsilon,1|0}$ on the first main diagonal block and the matrices ω_ϵ on the final $t = 2, \ldots, T$ main diagonal blocks. Ψ is constructed in a similar way as Φ above, but the main diagonal blocks consist of I_N matrices, and all the subdiagonal blocks are equal to $-\psi$. It follows that the vector of idiosyncratic components is distributed as $\epsilon|\theta \sim \mathcal{N}(0_{TN \times 1}, \Psi^{-1}\Omega_\epsilon\Psi^{-\top})$.

We obtain the covariance matrix Q^{-1} of the joint vector $z = (\eta^\top, x^\top)^\top$ in (Equation (II.5)) by

$$Q^{-1} = \begin{pmatrix} I & 0 \\ \Lambda & I \end{pmatrix} \begin{pmatrix} \Phi^{-1}\Omega_\eta\Phi^{-\top} & 0 \\ 0 & \Psi^{-1}\Omega_\epsilon\Psi^{-\top} \end{pmatrix} \begin{pmatrix} I & \Lambda^\top \\ 0 & I \end{pmatrix}, \tag{II.7}$$

and we can directly derive the partitioned precision matrix

$$Q = \begin{pmatrix} \Phi^\top\Omega_\eta^{-1}\Phi + \Lambda^\top(\Psi^\top\Omega_\epsilon^{-1}\Psi)\Lambda & -\Lambda^\top(\Psi^\top\Omega_\epsilon^{-1}\Psi) \\ -(\Psi^\top\Omega_\epsilon^{-1}\Psi)\Lambda & \Psi^\top\Omega_\epsilon^{-1}\Psi \end{pmatrix} = \begin{pmatrix} Q_{\eta\eta} & Q_{\eta x} \\ Q_{x\eta} & Q_{xx} \end{pmatrix}. \tag{II.8}$$

Given complete data, we can make use of the general rules for conditional sampling
from a partitioned multivariate normal distribution as in Anderson (2003), Theorem 2.5.1.
In terms of the partitioned precision matrix, Rue (2001), Section 3.1.1, provides the conditional distribution of η given $x = x^o$ defined as

$$p(\eta|x = x^o, \theta) \overset{\mathcal{D}}{=} \mathcal{N}\left(-Q_{\eta\eta}^{-1}Q_{\eta x}x^o, Q_{\eta\eta}^{-1}\right). \tag{II.9}$$

To efficiently draw a sample η^* from this distribution, Rue (2001) and Chan and Jeliazkov (2009) propose the application of fast band-matrix techniques. In the factor model, $Q_{\eta\eta} = \Phi^\mathsf{T}\Omega_\eta^{-1}\Phi + \Lambda^\mathsf{T}(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi)\Lambda$ is a block-banded matrix. Sampling proceeds as follows: Compute first the sparse Cholesky decomposition $Q_{\eta\eta} = LL^\mathsf{T}$, which implies a banded Cholesky factor L. Then, following Rue (2001), solve $Lw = -Q_{\eta x}x^o$ for w with a matrix equation solver. Afterwards, solve $L^\mathsf{T}\mu = w$ for μ. Solve $L^\mathsf{T}v = v^*$ for v, where v^* is drawn from the standard normal distribution $\mathcal{N}(0_{Tr\times 1}, I_{Tr})$. Finally, a draw of the factors is provided by $\eta^* = \mu + v$.

To compute the Cholesky decomposition and solve for the factors as outlined above, matrix programming languages such as Matlab can exploit the sparse structure of $Q_{\eta\eta}$ efficiently, see Chan and Jeliazkov (2009) and McCausland et al. (2011) for details.

II.4 Precision-based sampling with missing observations

Now consider the case when we do not observe all values in $x_{i,t}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. We assume a fraction κ - chosen such that κTN is an integer - of the observations is missing, and the missing observations can be distributed randomly across the indexes (i, t) as in Angelini et al. (2006) and Marcellino (2007). We do not model the process which generates the missing observations explicitly, and rather take the patterns of missing observations as given in the data. Following Rubin (1976), we thereby assume that the missing-data mechanism is ignorable, which is common when using macroeconomic data.

We define those model variables with missing observations as x_m, whereas variables with available observations are denoted as x_o. In the presence of missing observations, we modify the posterior sampler from Section II.3 along the lines of Little and Rubin (2002). The general Gibbs sampler by Little and Rubin (2002) starts by sampling values for missing observations from their conditional posterior distribution. In subsequent steps, these samples are combined with observed data, and enter the conditional posterior distributions for sampling the remaining model parameters.

In our case of the factor model Equations (II.1)–(II.2), we want to sample from the posterior distribution $p(\eta, x_m, \theta|x_o)$. Following Little and Rubin (2002), we do so by sequentially sampling factors, missing observations, and model parameters from their conditional posterior distributions:

1. $p(\eta|x_o, x_m, \theta)$

2. $p(x_m|x_o, \eta, \theta)$

3. $p(\theta | x_o, x_m, \eta)$

Alternatively, we provide a joint sampler for factors and missing observations conditional on parameters and observed data in a single step. In this case, the conditional distribution $p(\eta, x_m | x_o, \theta)$ replaces the first two in the sampler above.

Our general sampling strategy works as follows: Given complete data in Section II.3, we have used the variable ordering $z = (\eta^\top, x^\top)^\top$, where variables x were ordered last. In the presence of missing observations, the main idea is to permute the variables in z and obtain reordered or permuted z_{P_z} such that we can apply the same techniques for conditional sampling as in Section II.3. In particular, we move those variables with available observations x_o to the bottom of z_{P_z} and apply the same rules for conditional sampling from a Gaussian as in (Equation (II.8)) and (Equation (II.9)).

The reordering of variables in the vector z can be implemented by using a properly defined permutation matrix P_z such that $z_{P_z} = P_z z$. In general, a permutation matrix P_z is defined as a square binary matrix that has exactly one entry of 1 in each row and each column and zeros elsewhere. Permutation matrices are orthogonal matrices such that $P_z^{-1} = P_z^\top$ and $P_z P_z^\top = P_z^\top P_z = I$. The permutation by P_z implies a linear transformation of z in (Equation (II.5)), and the distribution of the transformed Gaussian z_{P_z} becomes

$$z_{P_z} | \theta \sim \mathcal{N}\left(0_{T(r+N) \times 1}, P_z Q^{-1} P_z^\top\right), \tag{II.10}$$

following standard rules for linear transformations of Gaussian vectors as in Anderson (2003), Theorem 2.4.1. Note that $P_z Q^{-1} P_z^\top$ is equal to the row- and column-permuted covariance matrix of z since

$$P_z Q^{-1} P_z^\top = ((P_z^\top)^{-1} Q P_z^{-1})^{-1} = (P_z Q P_z^\top)^{-1} = Q_z^{-1}. \tag{II.11}$$

Thus, the permuted covariance is equal to the inverse of the permuted precision matrix. A variable ordering and permutation in z is associated with a column- and row-permutation of the elements in the precision matrix. After permutation, we can partition the permuted precision matrix and apply a similar conditional sampling from a Gaussian as in (Equation (II.9)).[6]

All the permutations we consider order the variables with observed data x_o last in z_{P_z}. The reordering of variables in x into variables with missing observations x_m above the variables with available observations x_o can be implemented by using the permutation matrix P_x defined as

$$x_{P_x} = \begin{pmatrix} x_m \\ x_o \end{pmatrix} = P_x x = \begin{pmatrix} P_{x_m} \\ P_{x_o} \end{pmatrix} x. \tag{II.12}$$

---

[6]Note that we show matrix permutations in the paper only for expositional purposes. The Matlab computer codes underlying the quantitative results in the paper are based on equivalent, but more efficient index permutations (Golub and Van Loan, 2013). Let P be a permutation matrix of dimension $(K \times K)$ and p be a permutation vector defined as $p = (1, 2, \ldots, K) \times P^\top$. For a $(K \times K)$-dimensional matrix S, indexing by `S(p,:)` in Matlab is equivalent to row permutation PS, and `S(:,p)` is equivalent to column permutation $SP^\top$. To reverse the original permutation, we can use the inverse of the permutation matrix $P^\top$ or the corresponding index r defined as `r(p)=1:K;`.

The matrix P_{x_o} has TN columns corresponding to the TN elements in $x = (x_1^\top, x_2^\top, \ldots, x_T^\top)^\top$, and the number of rows is equal to $(1 - \kappa)TN$, the number of observations available for estimation. If observations were available for all variable values, P_x would equal the identity matrix. To construct P_x in the presence of missing observations, we can set P_{x_o} equal to the identity matrix and remove all those rows for which the corresponding observations are missing in the empirical data set. The matrix P_{x_m} just consists of these removed rows. Note that the position of missing observations in the data set is the only necessary information to derive the permutation matrix P_x. In particular, the permutation does not depend on model parameters.

Apart from the position of x_o, different sampling schemes can be derived depending on how the factors and variables with missing values are ordered. In this paper, we discuss two alternative samplers in the subsequent sections of the text:

1. Sequential 2-step sampler in Section II.4.1 in the spirit of Little and Rubin (2002):

    (a) $p(\eta | x_o, x_m, \theta)$: Sampling factors conditional on a sample of missing values, observed data, and parameters. No permutation is needed in this step.

    (b) $p(x_m | x_o, \eta, \theta)$: Conditionally sampling of missing observations given factors, data, and parameters using the permutation $z_{P_{2s}} = (x_m^\top, \eta^\top, x_o^\top)^\top$.

2. Joint sampling from $p(\eta, x_m | x_o, \theta)$ using the period-wise time permutation $z_{P_\tau} = (\eta_1^\top, x_{m,1}^\top, \eta_2^\top, x_{m,2}^\top, \ldots, \eta_T^\top, x_{m,T}^\top, x_o^\top)^\top$, see Section II.4.2.

Of course, both samplers aim at drawing from the same conditional posterior distribution $p(\eta, x_m | x_o, \theta)$. Differences between the samplers can arise with respect to a) convergence and mixing of the Markov chain, and b) computational efficiency. Concerning a), sequential samplers are in many cases easy to implement, because conditional distributions can generally be easier derived than joint distributions. On the other hand, sequentially sampling in two blocks using conditional distributions might lead to more correlated samples and slower convergence compared to sampling from the joint distribution in one block. Concerning b), computational efficiency, the joint sampler relies on the sparse Cholesky decomposition of one huge precision matrix, whereas the sequential sampler is based on decompositions of two smaller precision matrices for factors and missing observations, respectively. In addition, the alternative permutations of variables can influence the speed of the sparse Cholesky decomposition (McCausland et al., 2011). Furthermore, set-up costs to fill the precision matrices in each step of the Gibbs sampler vary between the samplers. Finally, there are differences between the samplers with respect to model evaluation using the marginal likelihood.

We will provide details on the samplers in the next subsections. A discussion of their differences is provided in Section II.5 by using simulation experiments. Details on how the marginal likelihood can be derived using the joint sampler are provided in Section II.7.

II.4.1   Sequential sampling of factors and missing observations

This precision-based sampler iterates between sampling factors conditional on interpolated missing values from $p(\eta|x_o, x_m, \theta)$ and, thereafter, values for missing observations conditional on factors from $p(x_m|x_o, \eta, \theta)$:

1. $p(\eta|x_o, x_m, \theta)$: Assume we have a draw for missing values $x_m = x^{m*}$. We can stack the interpolated missing data and the observed data in $x^*_{P_x} = ((x^{m*})^{\mathsf{T}}, (x^o)^{\mathsf{T}})^{\mathsf{T}}$. By reversing the data permutation from (Equation (II.12)) according to $P_x^{-1} = P_x^{\mathsf{T}}$, we can move the interpolated values to the positions of the missing observations in the original data set using $x^* = P_x^{\mathsf{T}} x^*_{P_x}$. Given the partly interpolated data, we can use the complete-data sampler from (Equation (II.9)) to draw factors conditional on the data and parameters from

$$p(\eta|x_o = x^o, x_m = x^{m*}, \theta) = p(\eta|x = x^*, \theta) \overset{\mathcal{D}}{=} \mathcal{N}\left(-Q_{\eta\eta}^{-1} Q_{\eta x} x^*, Q_{\eta\eta}^{-1}\right). \qquad \text{(II.13)}$$

   Note that the moments in (Equation (II.13)) differ from those in the complete-data case (Equation (II.9)) only with respect to x^* in the mean.

2. $p(x_m|x_o, \eta, \theta)$: We draw values for missing observations conditional on a factor sample η^* from step 1 and observed data. We use the permutation $z_{P_{2s}} = (x_m^{\mathsf{T}}, \eta^{\mathsf{T}}, x_o^{\mathsf{T}})^{\mathsf{T}}$, where the variables corresponding to missing observations in the data are ordered first, whereas factors and observed data are ordered last. We permute by $z_{P_{2s}} = P_{2s} z$ with permutation matrix

$$P_{2s} = \begin{pmatrix} 0_{\kappa TN \times Tr} & P_{x_m} \\ I_{Tr} & 0_{Tr \times TN} \\ 0_{(1-\kappa)TN \times Tr} & P_{x_o} \end{pmatrix}, \qquad \text{(II.14)}$$

   where the permutation matrices P_{x_m} and P_{x_o} decompose the model variables as in (Equation (II.12)). As (Equation (II.10)) holds for any permutation matrix, the distribution of the transformed Gaussian $z_{P_{2s}}$ is

$$z_{P_{2s}}|\theta \sim \mathcal{N}\left(0_{T(r+N)\times 1}, P_{2s} Q^{-1} P_{2s}^{\mathsf{T}}\right), \qquad \text{(II.15)}$$

   and the permuted covariance is equal to the inverse of the permuted precision matrix by $P_{2s} Q^{-1} P_{2s}^{\mathsf{T}} = (P_{2s} Q P_{2s}^{\mathsf{T}})^{-1} = Q_{2s}^{-1}$. The precision matrix can be derived as the inverse of a block matrix product by

$$\begin{aligned} Q_{2s} &= P_{2s} \begin{pmatrix} I & -\Lambda^{\mathsf{T}} \\ 0 & I \end{pmatrix} \begin{pmatrix} Q_\eta & 0 \\ 0 & Q_\epsilon \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Lambda & I \end{pmatrix} P_{2s}^{\mathsf{T}} \\ &= \begin{pmatrix} P_{x_m} Q_\epsilon P_{x_m}^{\mathsf{T}} & -P_{x_m} Q_\epsilon \Lambda & P_{x_m} Q_\epsilon P_{x_o}^{\mathsf{T}} \\ -\Lambda^{\mathsf{T}} Q_\epsilon P_{x_m}^{\mathsf{T}} & Q_\eta + \Lambda^{\mathsf{T}} Q_\epsilon \Lambda & -\Lambda^{\mathsf{T}} Q_\epsilon P_{x_o}^{\mathsf{T}} \\ P_{x_o} Q_\epsilon P_{x_m}^{\mathsf{T}} & -P_{x_o} Q_\epsilon \Lambda & P_{x_o} Q_\epsilon P_{x_o}^{\mathsf{T}} \end{pmatrix}, \qquad \text{(II.16)} \end{aligned}$$

where we have defined $Q_\epsilon = (\Psi^{-1}\Omega_\epsilon\Psi^{-\mathsf{T}})^{-1} = \Psi^{\mathsf{T}}\Omega_\epsilon^{-1}\Psi$ and $Q_\eta = (\Phi^{-1}\Omega_\eta\Phi^{-\mathsf{T}})^{-1} = \Phi^{\mathsf{T}}\Omega_\eta^{-1}\Phi$ to simplify notation from (Equation (II.7)). The variable ordering in $z_{P_{2s}} = (x_m^{\mathsf{T}}, \eta^{\mathsf{T}}, x_o^{\mathsf{T}})^{\mathsf{T}}$ is useful for conditional sampling, as the first block contains the variables with missing observations, whereas the rest contains the conditioning information, namely, factors and observed data. We thus partition the precision matrix by

$$Q_{2s} = \begin{pmatrix} Q_{x_m,x_m} & Q_{x_m,\eta x_o} \\ Q_{\eta x_o,x_m} & Q_{\eta x_o,\eta x_o} \end{pmatrix}, \tag{II.17}$$

where the top-left $(\kappa TN \times \kappa TN)$-dimensional block is defined as $Q_{x_m,x_m} = P_{x_m} Q_\epsilon P_{x_m}^{\mathsf{T}}$. Similarly to (Equation (II.9)), we can derive the conditional distribution of missing observations conditional on factors and observed data

$$p(x_m | x_o = x^o, \eta = \eta^*, \theta) \overset{\mathcal{D}}{=} \mathcal{N}\left(-Q_{x_m,x_m}^{-1} Q_{x_m,\eta x_o}((\eta^*)^{\mathsf{T}}, (x^o)^{\mathsf{T}})^{\mathsf{T}}, Q_{x_m,x_m}^{-1}\right). \tag{II.18}$$

Note that in the precision matrix Q_{x_m,x_m}, the matrix $Q_\epsilon = \Psi^{\mathsf{T}}\Omega_\epsilon^{-1}\Psi$ is the precision matrix of the idiosyncratic components' prior distribution. Since the idiosyncratic components follow a VAR(1) process, the precision matrix is block-banded (Chan and Jeliazkov, 2009). Permutation using P_{2s} just selects those idiosyncratic components corresponding to missing observations in the data and thus leaves the precision matrix $Q_{x_m,x_m} = P_{x_m} Q_\epsilon P_{x_m}^{\mathsf{T}}$ block-banded.

In the subsequent parts of the text, we call this method 'Sequential 2-step sampling'.

II.4.2 Joint sampling of factors and missing observations with time permutation

To efficiently sample from the joint distribution, we permute z such that factors and missing values are ordered together for each time period t, $(\eta_t^{\mathsf{T}}, x_{m,t}^{\mathsf{T}})^{\mathsf{T}}$. These vectors are stacked for $t = 1, \ldots, T$ and placed on top of the variables with observed data x_o. We obtain the permuted vector of variables $z_{P_\tau} = (\eta_1^{\mathsf{T}}, x_{m,1}^{\mathsf{T}}, \eta_2^{\mathsf{T}}, x_{m,2}^{\mathsf{T}}, \ldots, \eta_T^{\mathsf{T}}, x_{m,T}^{\mathsf{T}}, x_o^{\mathsf{T}})^{\mathsf{T}}$. Note that the variables in $z = (\eta^{\mathsf{T}}, x^{\mathsf{T}})^{\mathsf{T}}$ are already ordered period-wise within the blocks for factors $\eta = (\eta_1^{\mathsf{T}}, \eta_2^{\mathsf{T}}, \ldots, \eta_T^{\mathsf{T}})^{\mathsf{T}}$ and variables $x = (x_1^{\mathsf{T}}, x_2^{\mathsf{T}}, \ldots, x_T^{\mathsf{T}})^{\mathsf{T}}$ in (Equation (II.4)). To reorder the variables, we permute by $z_{P_\tau} = P_\tau z$ with

$$P_\tau = \begin{pmatrix} P_{\eta,1} & 0_{r \times NT} \\ 0_{N_{m,1} \times rT} & P_{x_m,1} \\ P_{\eta,2} & 0_{r \times NT} \\ 0_{N_{m,2} \times rT} & P_{x_m,2} \\ \vdots & \vdots \\ P_{\eta,T} & 0_{r \times NT} \\ 0_{N_{m,T} \times rT} & P_{x_m,T} \\ 0_{(N-\sum_t N_{m,t}) \times rT} & P_{x_o} \end{pmatrix}, \tag{II.19}$$

where $N_{m,t}$ denotes the number of missing values in x at time t.

The matrices $P_{\eta,t}$ for $t = 1, \ldots, T$ are equal to the rows from the identity matrix $I_{rT} = I_T \otimes I_r$ matrix corresponding to period t such that

$$P_{\eta,t} = \left(\begin{array}{ccc} 0_{r \times r(t-1)} & I_r & 0_{r \times r(T-t)} \end{array} \right). \tag{II.20}$$

Note that the matrix $P_{\eta,t}$ can be considered as a block row vector having T blocks, each consisting of r columns. The period-t block is just the $(r \times r)$ identity matrix, because all r factor values are ordered first in (Equation (II.19)) in each period.

The matrices $P_{x_m,t}$ for $t = 1, \ldots, T$ contain the block rows of the matrix P_{x_m} as defined in (Equation (II.12)) that correspond to missing observations in x_t according to

$$P_{x_m} = \left(\begin{array}{c} P_{x_m,1} \\ P_{x_m,2} \\ \vdots \\ P_{x_m,T} \end{array} \right). \tag{II.21}$$

The number of rows of $P_{x_m,t}$, $N_{m,t}$, is equal to the number of missing values in x_t at time t, whereas the number of columns is equal to TN. Note that $P_{x_m,t}$ can also be considered as a block row vector having T column blocks, each consisting of N columns, according to

$$P_{x_m,t} = \left(\begin{array}{ccc} 0_{N_{m,t} \times (t-1)N} & P_{x_m,(t,t)} & 0_{N_{m,t} \times (T-t)N} \end{array} \right), \tag{II.22}$$

where the $(N_{m,t} \times N)$ matrix $P_{x_m,(t,t)}$ selects the missing observations in period t, as implicitly defined in (Equation (II.12)), taking into account the stack of variables $x = (x_1^\mathsf{T}, x_2^\mathsf{T}, \ldots, x_T^\mathsf{T})^\mathsf{T}$ in (Equation (II.4)).

Note that all vectors and matrices above follow the convention that a block matrix with zero row or column size implies the matrix is empty, and the corresponding blocks in the matrices above represent empty placeholders. This applies to either time periods without any missing observations or time periods without data observations.

Given the permutation matrix P_τ, we can obtain the permuted vector of variables $z_{P_\tau} = P_\tau z$ and the permuted precision matrix $Q_\tau = P_\tau Q P_\tau^\mathsf{T}$. In Section II.B, we derive the precision matrix Q_τ by exploiting the period-wise block structure implied by z_{P_τ}.

Our aim is to jointly sample factors and values for missing observations from the conditional distribution $p(\eta, x_m | x_o = x^o, \theta)$. The final block in the vector z_{P_τ} is x_o and the corresponding observations serve as the conditioning set. Thus, the permuted precision matrix Q_τ can be partitioned in the following way

$$Q_\tau = \left(\begin{array}{cc} Q_{\eta x_m, \eta x_m} & Q_{\eta x_m, x_o} \\ Q_{x_o, \eta x_m} & Q_{x_o, x_o} \end{array} \right), \tag{II.23}$$

where the upper-left block $Q_{\eta x_m, \eta x_m}$ has dimensions $(rT + \sum_{t=1}^{T} N_{m,t}) \times (rT + \sum_{t=1}^{T} N_{m,t})$,

and we can directly derive the conditional distribution

$$p(\eta, x_m | x_o = x^o, \theta) \overset{\mathcal{D}}{=} \mathcal{N}(-Q_{\eta x_m, \eta x_m}^{-1} Q_{\eta x_m, x_o} x^o, Q_{\eta x_m, \eta x_m}^{-1}) \tag{II.24}$$

for factors and missing values conditional on observed data. In Section II.B, we show that $Q_{\eta x_m, \eta x_m}$ is a block-banded matrix. In particular, it has block bandwidth equal to one, thus, representing a block tridiagonal matrix (Golub and Van Loan, 2013). The reason is that the reduced form of the factor model can be written as a VAR process of order one in the factors and explained variables. Following Rue and Held (2005), precision matrices of AR processes have a bandwidth equal to the lag order of the AR process. In the appendix, we show that this result also holds in the factor model (Equation (II.1)) and (Equation (II.2)) with missing observations, where the precision matrix is block tridiagonal.

In the subsequent parts of the text, we call this method 'Joint sampling, time permutation'.

II.5 Comparing the precision-based samplers by simulations

Based on simulations, we compare the convergence properties, the computing time of the precision-based samplers as well as their equivalence in terms of mean-squared errors of simulated factors and missing observations.

II.5.1 Data-generating process and model estimation

The data-generating process (DGP) has a factor structure with $r = 2$ factors, which follow a VAR process with one lag. The idiosyncratic components each follow AR processes with one lag. The variables $x_{i,t}$ for $i = 1, \ldots, N$ are simulated according to

$$x_{i,t} = \lambda_{i.} \eta_t + \epsilon_{i,t}, \tag{II.25}$$

$$\eta_t = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.8 \end{bmatrix} \eta_{t-1} + u_{\eta,t}, \ u_{\eta,t} \sim \mathcal{N}(0, I_2), \tag{II.26}$$

$$\epsilon_{i,t} = 0.4\epsilon_{i,t-1} + u_{\epsilon,i,t}, \ u_{\epsilon,i,t} \sim \mathcal{N}(0, \omega_\epsilon). \tag{II.27}$$

For the loading matrix λ, we specify a point-mass normal mixture distribution often used for variable selection as in George and McCulloch (1993, 1997), and Geweke (1996):

$$p(\lambda_{ij}) \overset{\mathcal{D}}{=} (1 - \rho_j)\delta_0(\lambda_{ij}) + \rho_j \mathcal{N}(m_j, M), \tag{II.28}$$

$$p(\rho_j) \overset{\mathcal{D}}{=} \mathcal{B}(r_0 s_0, r_0(1 - s_0)), \tag{II.29}$$

where $\delta_0(\cdot)$ represents the Dirac function with point mass at zero and ρ_j is a factor-specific probability of a non-zero loading. The inclusion probabilities ρ_j follow a beta distribution,

$\mathcal{B}(r_0 s_0, r_0(1-s_0))$, with mean $s_0 = 0.5$ and precision $r_0 = 30$. The non-zero factor loadings are simulated out of the normal distributions $\mathcal{N}(m_j, M)$ with $m_1 = 0.60$, $m_2 = 0.40$, and $M = 0.01$. The variance $M = 0.01$ is relatively tight in order to clearly separate zero and non-zero loadings. For the variances of the idiosyncratic components, we assume an inverse gamma distribution according to $\omega_\epsilon \sim \mathcal{IG}(2, 0.5)$.

Concerning sample size, we consider $T = 100$ time-series observations and $N = 100$ variables. Given this specification, we sample factors η_{DGP} and data x_{DGP} from ($Equation$ (II.25)) and ($Equation$ (II.27)). To address missing observations, we assume that 20% of observations are missing, $\kappa = 0.20$. We randomly set $\kappa N T$ observations in the sample x_{DGP} to missing values, yielding the data set x^o used for Bayesian estimation of factors and missing observations. In the experiment, we sample $K = 100$ times from the DGP and estimate the factor model on each data set. We ran further experiments with different specifications for (T, N, κ) and alternative priors. As the results in the alternative experiments are very similar compared to the baseline case summarized below, we only report the baseline results.

In the simulation experiments, the factor model is estimated using the posterior sampler outlined in Section II.A for each data $x^{o,(k)}$ set sampled from the DGP for $k = 1, \ldots, K$. Each time, we draw $G = 10000$ times from the posterior. We obtain samples for factors, missing values, and parameters according to $\eta^{(k,g)}, x^{m,(k,g)}, \theta^{(k,g)} \sim p(\eta, x_m, \theta | x_o = x^{o,(k)})$ for $g = 1, \ldots, G$ and $k = 1, \ldots, K$. To address convergence and computing time depending on the number of draws, we consider different partitions of the raw posterior draws. In particular, we compare alternative numbers of burn-in draws $G^{\text{burn-in}}$ and numbers of posterior evaluation draws G^{eff}.

II.5.2 Comparing inefficiency factors

We compute inefficiency factors to discuss convergence and mixing of the Gibbs Sampler. The inefficiency factor can be defined as in Chib (2001) by

$$\text{IE}_{\eta,j,t,k} = 1 + 2 \sum_{m=1}^{M} (1 - \frac{m}{M}) \rho_{\eta,j,t,k}(m), \tag{II.30}$$

where $\rho_{\eta,j,t,k}(m)$ is the estimated autocorrelation at lag m of the posterior draws for model factors $\eta_{j,t}^{(k,g)} | x_o = x^{o,(k)}$ over the draws $g = G^{\text{burn-in}} + 1, \ldots, G^{\text{burn-in}} + G^{\text{eff}}$. The maximum lag order is $M = 150$. Values of $\text{IE}_{\eta,j,t,k}$ greater than one indicate autocorrelation in the chain that might be due to poor mixing or lack of convergence. We compute inefficiency factors for the posterior samples of the model factors for $j = 1, 2$, time periods $t = 1, \ldots, T$, and for $k = 1, \ldots, K$ samples from the DGP. In the left panel of Figure II.1, we show box plots for the whole distribution of inefficiency factors for model factors $\text{IE}_{\eta,j,t,k}$ across j, t, and k and for different numbers of burn-in and evaluation draws. In the right panel of Figure II.1, we show box plots for the inefficiency factors of posterior draws of values for missing observations.

The results for model factors in the left panel of Figure II.1 show that a number of burn-in

Figure II.1: Inefficiency factors for posterior samples of model factors and missing observations.
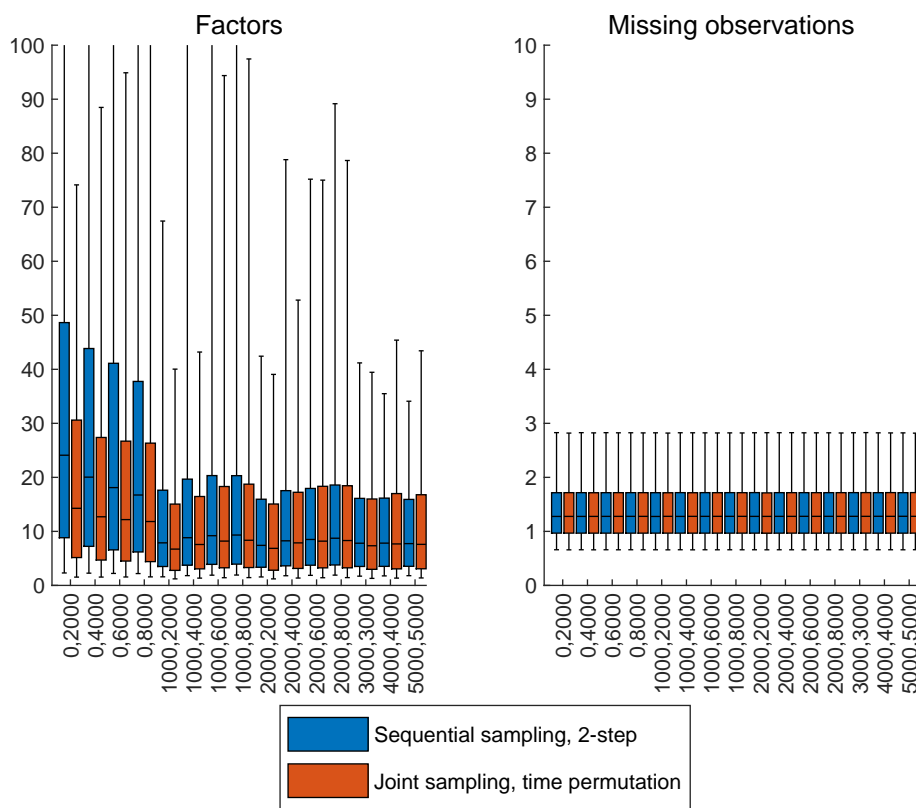
Note: In the left figure, box plots of inefficiency factors are based on the posterior samples of model factors for different numbers of burn-in and posterior evaluation draws. The first number shown in the labels of the horizontal axis refers to the number of burn-in draws $G^{\text{burn-in}}$, whereas the second number refers to the number of draws used for posterior evaluation after burn-in G^{eff}. In the right figure, box plots show inefficiency factors for the estimated values of missing observations.

draws greater or equal than $G^{\text{burn-in}} = 1000$ generally leads to a median of inefficiency factors around 10 for all three precision-based samplers. The upper bound of the interquartile ranges are in most cases below 20 for $G^{\text{burn-in}} \geq 1000$. Across different posterior sample splits, there are differences with respect to the whiskers of the boxplots, which mark the 5th and 95th percentiles. However, with increasing numbers of burn-in and evaluation draws, the bands tend to become smaller, and we find no systematic differences between the two precision-based samplers. For sample splits with zero or 1000 burn-in draws, $G^{\text{burn-in}}$, the upper bounds of the interquartile ranges for the 2-step precision-based sampler are higher than for the joint time-permutation sampler. For $G^{\text{burn-in}} > 1000$, we see no substantial differences in terms of convergence in the model factors between the two precision-based samplers.

The results for missing observations in the right panel of Figure II.1 show that convergence is very fast for all precision-based samplers. The median of the inefficiency factors is slightly greater than one, the upper bound of the interquartile range is about 1.3, and the bound of the upper whisker is slightly below 3. Note that the convergence for missing observations is substantially faster than for model factors. Convergence issues in factor models are often due to the lack of identification between factor loadings and factors, because $\lambda \eta_t = (\lambda H)(H^{-1} \eta_t)$ holds for any invertible H as documented in Lopes and West (2004); Ghosh and Dunson (2009); Bai and Wang (2014); Conti et al. (2014); Kastner et al. (2017); Chan et al. (2018a), amongst others. As discussed in parts of this literature, blocked sequential sampling of model factors conditional on loadings and subsequently loadings conditional on factors can sometimes lead to correlated draws and poor convergence, whereas joint sampling of model factors and loadings generally improves convergence, albeit making more complicated samplers necessary (Ghosh and Dunson, 2009; Chan and Jeliazkov, 2009; Conti et al., 2014; Kastner et al., 2017). This paper has a conceptually different focus on sampling missing values in the data and model factors given factor loadings. The key point is that the missing values in the data are a function of the common components, not of either the factors or the loadings alone. In Figure II.2, we show the inefficiency factors for factor loadings and the common components to address this issue. The inefficiency factors of the loadings are comparable in magnitude to the inefficiency factors of the model factors, whereas the inefficiency factors of the common components are close to one, and thus comparable to the inefficiency factors of the estimated missing values.

Overall, the simulation results indicate that convergence of estimates of missing values and common components do not to seem to be affected by any loadings-factor identification issue as mentioned above. Despite the fact that loadings and model factors show slower convergence, we see no major convergence issues for a reasonably chosen number of burn-in draws in general. The two precision-based samplers perform quite similar.

Figure II.2: Inefficiency factors for posterior samples of model factor loadings and common components.
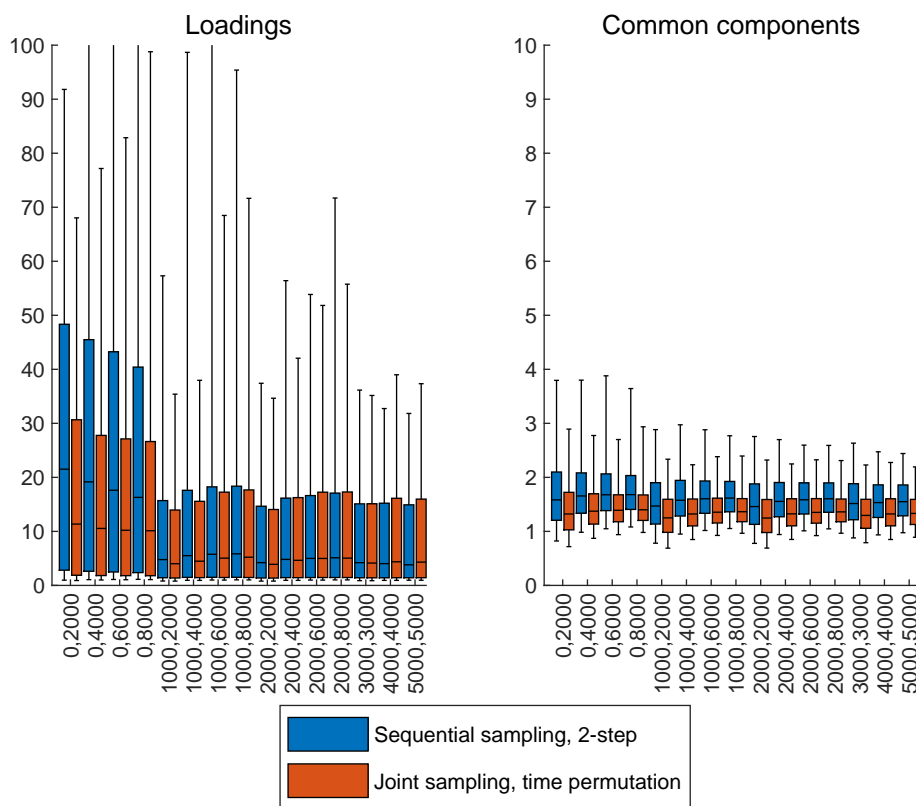
Note: In the left figure, box plots of inefficiency factors are based on the posterior samples of model factor loadings for different numbers of burn-in and posterior evaluation draws. The first number shown in the labels of the horizontal axis refers to the number of burn-in draws $G^{\text{burn-in}}$, whereas the second number refers to the number of draws used for posterior evaluation after burn-in G^{eff}. In the right figure, box plots show inefficiency factors for the common components.

II.5.3 Comparing computing time

In Figure II.3, the average computing time needed for posterior sampling is shown for different numbers of burn-in draws $G^{\text{burn-in}}$ and evaluation draws G^{eff} for the precision-based samplers. A box plot in the figure refers to the distribution of computing time across $k = 1, \dots, K$ data sets sampled from the DGP.

Figure II.3: Average computing time for posterior samples for different numbers of posterior draws.
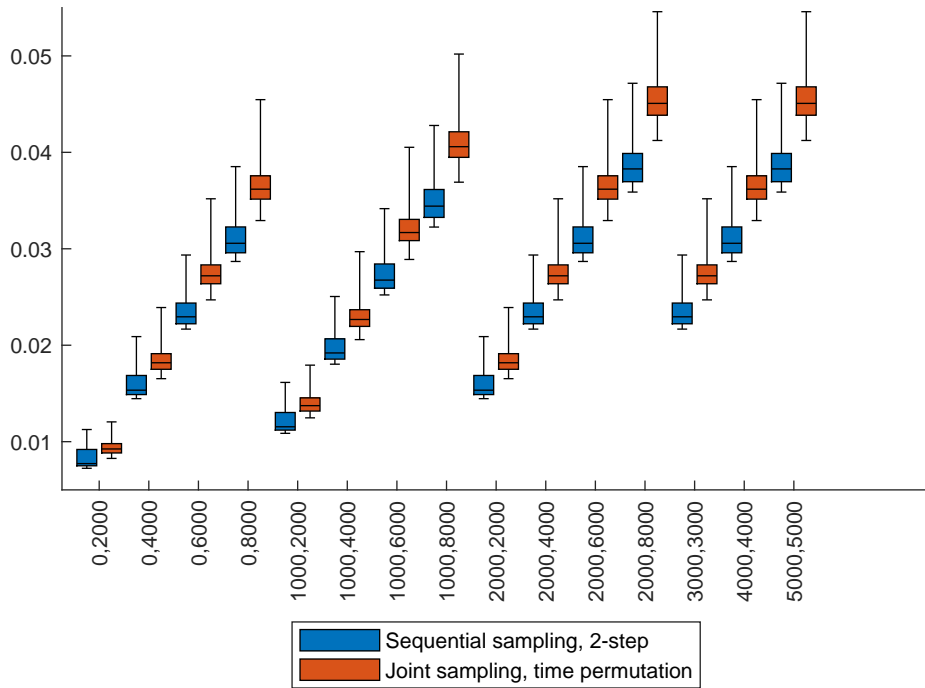
Note: Figure shows hours of computing time averaged over $K = 100$ estimation experiments. For each dataset from the DGP, we measure the elapsed time to draw from the posterior given alternative numbers of posterior draws. The first number shown in the labels of the horizontal axis refers to the number of burn-in draws $G^{\text{burn-in}}$, whereas the second number refers to the number of draws used for posterior evaluation after burn-in G^{eff}.

The results in Figure II.3 show an overall better performance of the 2-step precision-based sampler than the time permutation sampler. The median and the bounds of the interquartile ranges of the 2-step sampler are in the majority of cases smaller than those of the time permutation sampler. Despite some overlap of the 90% intervals, which are marked by the whiskers of the boxplots, the results overall indicate some computational advantages of the 2-step precision-based sampler.

To get an understanding of the better performance of the 2-step approach, we have a closer look at the flops needed for precision-based sampling as in Rue (2001); McCausland et al. (2011). In general, the Cholesky factorization $Q = LL^{\mathsf{T}}$ of a banded, symmetric, and

positive definite matrix Q with dimensions $(K \times K)$ and bandwidth b implies Kb^2 flops, where the bandwidth is the maximum number of off-diagonals, which have non-zero elements. For the forward and backward substitutions outlined at the end of Section II.3, we need $4Kb$ flops, and we need $Kb^2 + 4Kb$ flops in total. The 2-step sequential approach of Section II.4.1 requires two samples based on different precision matrices, whereas time-t permutation sampling requires one decomposition only. Both the matrix size and the bandwidth have a positive effect on the computing time according to the formulae for the overall flops. However, as the bandwidth also enters the formulae $Kb^2 + 4Kb$ squared, it has a comparatively huge impact on the overall computing time. The farther away non-zero elements are from the main diagonal, the more computing time is need for decomposing and solving. In Figure II.4, we show the precision matrices for 2-step and time-t sampling. Each entry in the figures receives a blue sign, if the corresponding entry of the precision matrix is non-zero. Panel A and B show the two precision matrices for the sequential 2-step sampler. Panel C shows the precision matrix for the time-permutation sampler. The top-left blocks highlighted by a black rectangle refer to the submatrix, which will be decomposed by sparse Cholesky factorization. The dimensions of theses submatrices are $K_{2s,\eta} = 200$ and $K_{2s,x_m} = 2000$ for the 2-step sampler, and $K_\tau = 2200$ for the time-permutation sampler. Note that the position of the missing $\kappa NT = 2000$ observations is chosen randomly in the data and thereby affects the shape of the precision matrices and bandwidths. To highlight the bandwidth near the main diagonal, a subplot in the top-right of Figure II.4 zooms the first top-left (100×100)-dimensional block of elements of the precision matrix. In the figure, the bandwidth of the precision matrix underlying the time-permutation sampler is larger than the bandwidths of the precision matrices used for 2-step sampling. In detail, the bandwidths are $b_{2s,\eta} = 3$ and $b_{2s,x_m} = 31$ for the 2-step sampler, and $b_\tau = 59$ for the time-permutation sampler. Thus, the comparatively large bandwidth of the time-permutation sampler contributes to the slower computational performance. Note, however, that the bandwidth of the precision matrix is only one source to explain the differences in computing time. There are also set-up costs to fill the precision matrix with model parameters every recursion of the sampler, which contribute to the computing time in the simulations. In the same way, permuting the moments of z at each draw is also costly. Note, however, that the permutation matrices depend only on the position of missing observations in the dataset, not the model parameters, and thus only have to be computed once for Bayesian estimation. The simulation results reflect all these different determinants of the overall computing time.

II.5.4 Comparing samples of factors and missing observations

As a final check of the accuracy of the precision-based samplers, we compare the mean-squared error $\mathrm{MSE}_{\eta,t,j}$ for the factor posterior samples defined as

$$\mathrm{MSE}_{\eta,j,t} = \frac{1}{G^{\mathrm{eff}}} \sum_{g=G^{\mathrm{burn\text{-}in}}+1}^{G^{\mathrm{burn\text{-}in}}+G^{\mathrm{eff}}} (\eta_{\mathrm{DGP},j,t} - \eta_{j,t}^{(g)}|x_o = x^o)^2, \qquad (\mathrm{II.31})$$

Figure II.4: Precision matrices after permutation

A. Sequential 2-step: Q to sample η

B. Sequential 2-step: $P_{2s}QP_{2s}^{\mathsf{T}}$ to sample x_m

C. Time permutation, $P_{\tau}QP_{\tau}^{\mathsf{T}}$
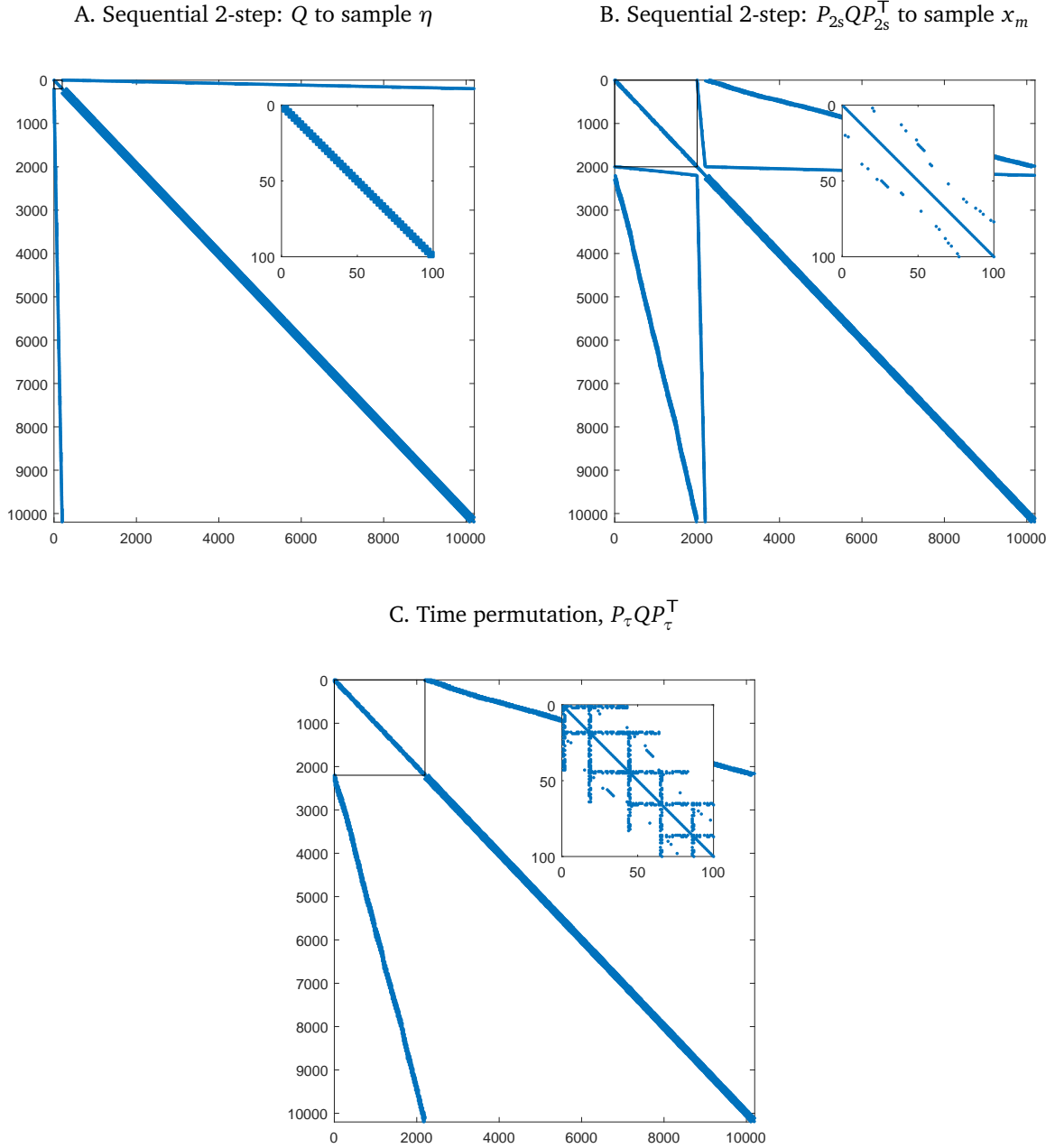
Note: The figure shows the non-zero elements in the precision matrices used in the alternative samplers. The top-left blocks highlighted by a black rectangle refer to the submatrices, which will be decomposed by Cholesky factorization. In Panel A, the highlighted part of the precision matrix is the block $Q_{\eta\eta}$ in (Equation (II.13)). In Panel B, the highlighted part of the precision matrix is Q_{x_m,x_m} in (Equation (II.18)). In Panel C, the highlighted part shows the block $Q_{\eta x_m, \eta x_m}$ under the joint time-permutation sampler in (Equation (II.24)). To show the structure and bandwidth of these submatrices near the main diagonal in more detail, a subplot in the top-right of the figure zooms the first top-left (100×100)-dimensional block of elements of the precision matrix.

for each factor indexed by j for $j = 1, 2$ and time period $t = 1, \ldots, T$. In the MSE, the posterior samples of the factor $\eta_{j,t}^{(g)} | x_o = x^o$ are subtracted from the "true" factor values sampled from the DGP $\eta_{\text{DGP}, j, t}$. As the samplers should provide draws from the same posterior distribution, we expect very similar values for large G^{eff}. Accordingly, we also compute the MSE for the missing observations by subtracting the sampled values drawn by the conditional samplers from the true values simulated from the DGP.

The results presented below are based on one data set and factors sampled from the DGP. We take $G = 10000$ raw draws from the posterior and discard the first $G^{\text{burn-in}} = 5000$ as burn-in. The MSEs are computed over the remaining $G^{\text{eff}} = 5000$ draws. The left panel of Figure II.5 provides the MSEs for the first of the two factors obtained from the different samplers for each time period. As expected, the samplers yield very similar MSEs for all $t = 1, \ldots, T$. The results show no signs of any systematic differences between the samplers. We obtain very similar results when looking at the MSEs for the missing observations in the right panel of Figure II.5, where the MSEs of the first 50 missing observations in the data are shown. The results indicate a very high similarity across the precision samplers.

Figure II.5: MSE comparison for one simulated dataset and posterior samples of first model factor and missing observations.
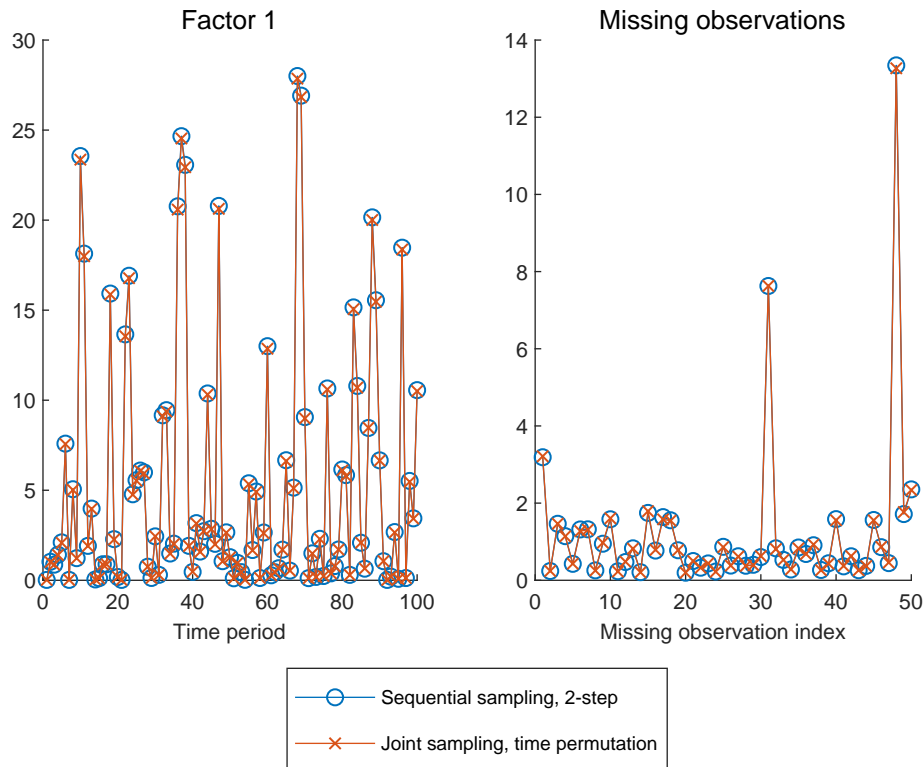
Note: In the left figure, the mean-squared errors (MSE) are computed by averaging the squared difference between $G^{\text{eff}} = 5000$ posterior draws for the first model factor and the corresponding factor sample from the DGP as defined in (Equation (II.31)). The effective samples used are obtained by taking $G = 10000$ raw samples from the posterior and discarding the first $G^{\text{burn-in}} = 5000$ as burn-in. In the right figure, the MSE is shown for posterior draws of 50 randomly chosen missing values.

To sum up the simulation results, we only find differences between the samplers with
respect to the computing time. In this regard, the 2-step sampler tends to be faster than
the time-permutation sampler. However, the precision-based samplers yield similar MSEs
for factors and missing observations, and inefficiency factors indicate no major convergence
issues for a reasonably chosen number of burn-in draws. Due to these similarities in the
simulations, we only consider the 2-step sampler in the empirical application below.

II.6 Empirical application: Bayesian estimation of international factors in GDP growth

To illustrate the precision-based sampler in the presence of missing observations in data, we
estimate the factor model with Bayesian techniques on multi-country GDP growth data. The
application follows the literature on international business cycles estimated in large factor
models (Kose et al., 2003, 2008; Francis et al., 2017; Müller et al., 2019).

II.6.1 Data and motivation for the empirical exercise

Francis et al. (2017) estimate global and regional business cycles from a large set of country-
specific annual GDP growth series. We follow these authors and choose the Penn World
Tables (PWT) to construct the data set. We use PWT version 9.1 (Feenstra et al., 2015) and
take annual real GDP (output concept) series for all available countries. Growth rates are
computed by taking first differences of the logarithm of the series in levels. We end up with
$T = 67$ time series observations for the years 1951 to 2017 and $N = 182$ countries in the
cross section. The data is unbalanced with respect to available observations per country.
Overall, 2391 of the observations are missing, which is 19.6% in relation to $TN = 12194$
potential values in the data. The number of missing observations at each period in time are
shown in Figure II.6.

In the data we find that observations are mostly missing in earlier periods of the sample:
In 1951 we have 127 observation missing out of 182, whereas no observations are missing
at the end of the sample after 2005. The decline of the number of missing observations
in the intermediate time periods can be roughly described as being step-wise across the
covered decades. From 1952 to 1960, there are more than 100 observations missing per
year, whereas the number of missing observations drops to about 70 until 1970. Between
1970 and 1990, slightly more than 20 time series observations are missing.

Compared to the existing literature by Kose et al. (2003) and Francis et al. (2017), the
data set used here contains a larger number of variables with $N_{\text{unbal}} = 182$ and thus more
cross-country information. If we remove all time series with any missing observations over
the sample period 1951 and 2017, we end up with a balanced data set of $N_{\text{bal}} = 55$ time
series, which is very close to the countries covered in Kose et al. (2003) and Francis et
al. (2017). We estimate the factor model with these two different data sets and compare

Figure II.6: Missing observations in international GDP growth data
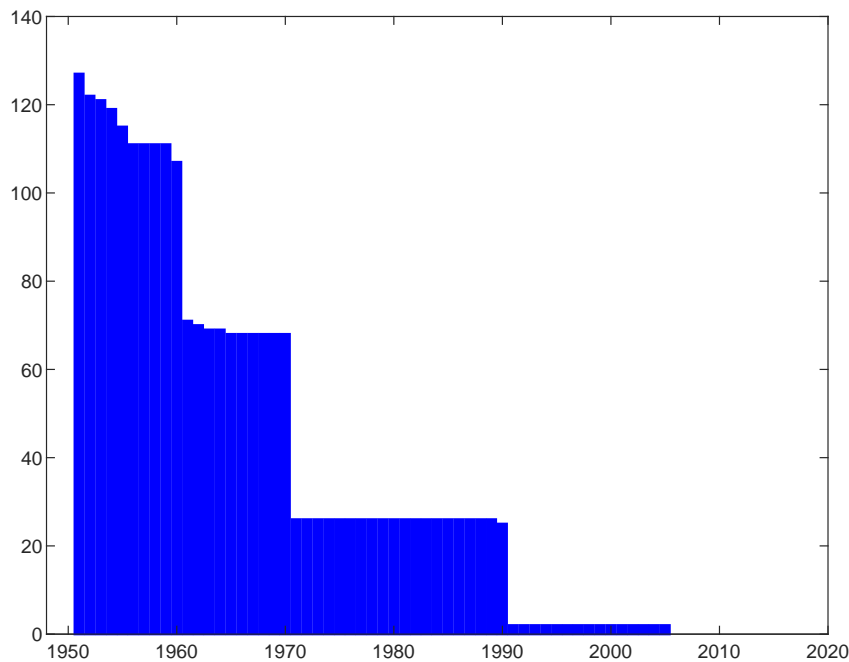
Note: The bars in the figure show the sum of missing observations in each time period. The maximum number of cross-section observations per period is equal to $N_{\text{unbal}} = 182$.

the results. In general, time-series data with partially missing observations encompasses balanced time-series data where all time series with any missing observations have been removed. Thus, estimation methods, which can tackle missing observations, can take into account larger data sets and use more information to estimate models. Note that in general we cannot expect that larger data always improves factor estimation. Whether additional information is beneficial for factor estimation, depends - amongst other things - on the number of missing observations as well as the information content in the additional data. Boivin and Ng (2006) show that noisy data can even deteriorate accuracy of factor estimates.

The empirical comparison proceeds along several dimensions: We compare the factor estimates obtained from using the two different information sets. We also identify the relevant and irrelevant variables in both data sets along the lines of Kaufmann and Schumacher (2017). Irrelevant variables are defined as not being related to factors via the loading matrix and generally cannot contribute to estimating the factors in a state-space model (Koopman and Harvey, 2003). We also investigate how variables, which are part of both information sets, are explained by the different factor estimates. In particular, we estimate the common and idiosyncratic components and investigate how the variance contributions of the common components differ for the two different information sets.

II.6.2   Model specification and Bayesian estimation

In the model for the empirical application, the loading matrix has a group structure with
$r = 6$ factors. Following the literature on international business cycles like Kose et al. (2003)
and Francis et al. (2017), we define the first factor in the model as the global factor, such that
all variables can load on this factor. Accordingly, there are no zero restrictions in the first
column of the loading matrix. The other five factors are continental factors for Africa, Asia,
Europe, North America, and South America. The continental group structure is imposed by
zero restrictions: Each country GDP variable can load on only one of the continental factors
in addition to the global factor. For each continental factor, those country GDP variables not
belonging to this particular continent receive a zero loading element in the corresponding
column of the loading matrix. Given these zero restrictions from the continental factor
structure, the factor model is identified according to the Bekker criterion outlined in Bai
and Wang (2014).

In the application, we use a more general model than specified in (Equation (II.1)) and
(Equation (II.2)). We use $p = 2$ lags in the factor VAR and $q = 2$ lags in the AR equations
for idiosyncratic components. Details on the extended model and the posterior sampler
can be found in Section II.A. In the empirical application, we take 100000 draws from the
posterior density, discard the first 50000 draws as burn-in, and take the remaining draws
for final posterior evaluation. We estimate the factor model on balanced and unbalanced
data and compare the results.

II.6.3   Results

The estimated factors using the balanced and unbalanced data set can be compared in Fig-
ure II.7. We show the median of the posterior samples together with 90% posterior bands.
The majority of the factors look similar when estimating the model on balanced and unbal-
anced data: The global factor, the European factor, and the two American factors do not
change considerably with respect to the median and the posterior intervals. The African
and Asian factors, however, are different when comparing balanced and unbalanced data.
Based on balanced data, the African and Asian factors are hardly different from zero in
terms of 90% posterior intervals across time periods. Based on the larger unbalanced data,
both factors are more precisely estimated and exhibit more pronounced cyclical swings.

Given the posterior draws of the loadings, we investigate the relevance or irrelevance of
variables in the model. Following Kaufmann and Schumacher (2017), we define irrelevant
variables as having only zero factor loadings, whereas relevant variables are related to at
least one factor. Irrelevant variables are not explained by the factors and, vice versa, do not
contain information for factor estimation. To distinguish irrelevant from relevant variables,
we compute 90% highest posterior density (HPD) intervals for each element in the loading
matrix to check whether the loading element is different from zero. We define relevant vari-
ables as having at least one non-zero factor loading in the corresponding row of the loading
matrix. In Table II.1, the number of relevant variables is shown for different continents.

Figure II.7: Factors estimated on balanced and unbalanced data
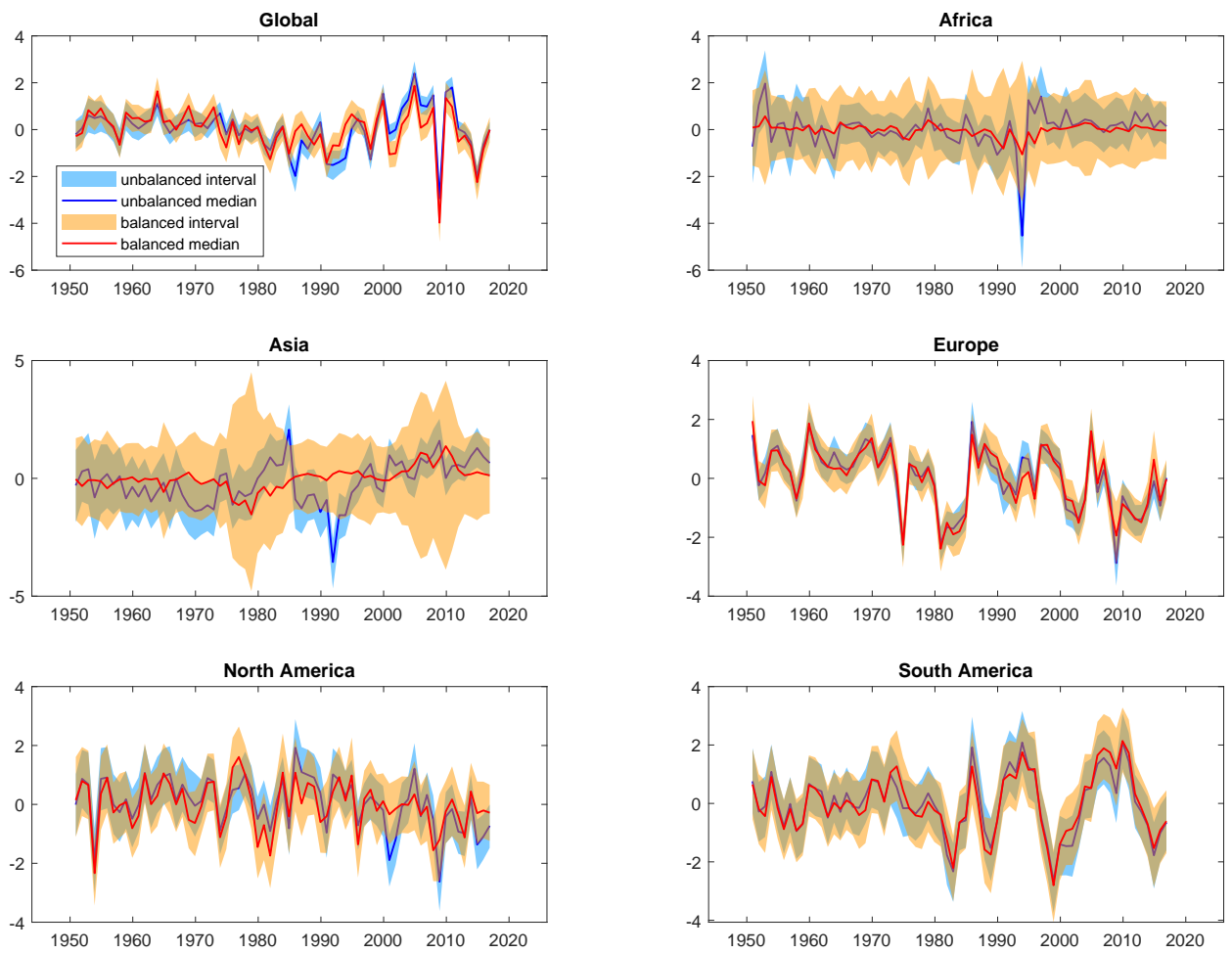
Note: The straight lines refer to the median of the posterior distribution of the factors. The shaded ares cover the 90% posterior intervals.

Within geographic regions, we observe some heterogeneity: Among African GDP growth series in the balanced data set, we find only one out of nine (11%) relevant variables. In the unbalanced data set, we have 50 African GDP growth time series, and the proportion of relevant variables is 18% (9 of 50). In the balanced data, the proportion of relevant Asian GDP series in the model is 20% and thus a bit higher compared to African GDP series. Using unbalanced data implies an increase of the number of Asian GDP time series from 10 to 50. The proportion of relevant variables compared to balanced data increases to 38% (19 of 50). For Europe and North America, we observe a relatively large proportion of relevant variables compared to the other continents when using balanced data. However, there are very few relevant time series in the additional unbalanced data. For Europe, for example, the proportion of relevant variables decreases from 89% to 63% when adding unbalanced data. The results are similar for American GDP growth data. We summarize by looking at results for all variables in the two models: Overall we find 55% relevant variables in the balanced data, whereas we have only 37% relevant variables in the unbalanced data. This indicates that only some of the added unbalanced GDP growth time series provide additional information for factor estimation on top of the balanced data only.

Table II.1: Relevant variables

Data	Unbalanced data	Balanced data
Method	Sequential 2-step	Standard precision-based
Africa	0.18 (9 of 50)	0.11 (1 of 9)
Asia	0.38 (19 of 50)	0.20 (2 of 10)
Europe	0.63 (25 of 40)	0.89 (16 of 18)
North America	0.23 (7 of 31)	0.50 (5 of 10)
South America	0.64 (7 of 11)	0.75 (6 of 8)
All variables	0.37 (67 of 182)	0.55 (30 of 55)

Note: Entries in the table are equal to the proportion of relevant variables in the geographical region. We compute highest posterior density (HPD) intervals for each element in the loading matrix to check whether the loading element is significantly different from the zero. We define relevant variables as having at least one significant non-zero factor loading in the corresponding row of the loading matrix.

We also look at the role of the common and idiosyncratic components for each time series in the data set. Figure II.8 and Figure II.9 show variance shares of common components, defined as the variance of the common component of a country GDP growth time series divided by the overall variance of the time series. The difference between one and the variance shares of common components provides the variance share of country-specific idiosyncratic components (Kose et al., 2003). Thus, the larger the variance share of common components, the more variance of GDP growth is explained by common factors and co-movements with other countries' GDP growth. The smaller the number, the more important are country-specific sources of GDP fluctuations. To summarize the posterior variability of variance shares of common components, we provide box plots. Green boxes refer to the

results based on balanced data, whereas yellow boxes refer to the results based on unbalanced data. Figure II.8 shows the variance share of the common components for those countries which are part of both the balanced an unbalanced data sets. Generally, European countries (abbreviated by EU in the table) have a comparatively high share of common component variances, followed by South American (SA) countries. African (AF) and Asian (AS) countries' GDP growth have a comparatively low variance share of common components on average. When comparing estimates for balanced and unbalanced data, the box plots are often quite similar. For a large number of GDP growth series, there is no clear advantage from using larger unbalanced data with respect to the variance share of common components. This holds, for example, for the big developed countries USA, Canada, France, and Germany, amongst others. In Figure II.9, we look at the variance share of the common components for those countries which are only part of the unbalanced data set. These are $N_{\text{unbal}} - N_{\text{bal}} = 182 - 55 = 127$ countries. We can see that there are some country GDP series with a considerable variance share of the common components, for example, Brunei, Georgia, and several Asian countries. The majority of GDP series, however, shows quite small numbers only up to 0.2, indicating a big role of idiosyncratic country-specific movements for many countries in the unbalanced data set.

To sum up the empirical results, many GDP series for Asian and African countries are not related to the factors. However, the proportion of Asian and African countries related to the factors is larger when using the unbalanced data rather than the balanced data. In the results based on unbalanced data, we generally find that the idiosyncratic components seem to dominate for the majority of countries, despite some country GDP growth series having a high variance share of the common components. The additional unbalanced data also has no strong effect on the commonality of the variables that are part of both the balanced and unbalanced data set. If an analyst were mostly interested in results for G7 countries, it might suffice to look at the posterior results from the smaller balanced data and using the simpler estimation methods. The use of the factor model based on the larger unbalanced data and the more demanding sampling methods to tackle missing observations may provide relevant insights, if the countries, which are exclusively part of the unbalanced data, are interesting in themselves to an analyst.

II.7   Extensions: Integrated likelihood, other state-space models

II.7.1   Integrated likelihood

The methods discussed in this paper can also be employed for model evaluation using the marginal likelihood. Following Chan and Grant (2016), the integrated likelihood function, obtained by integrating the unobserved states out of the conditional or complete likelihood function, $p(x|\theta) = \int p(x, \eta|\theta) d\eta = \int p(x|\eta, \theta) p(\eta|x, \theta) d\eta$, is the key quantity to derive

Figure II.8: Variance share of the common components, countries covered in both balanced
and unbalanced data
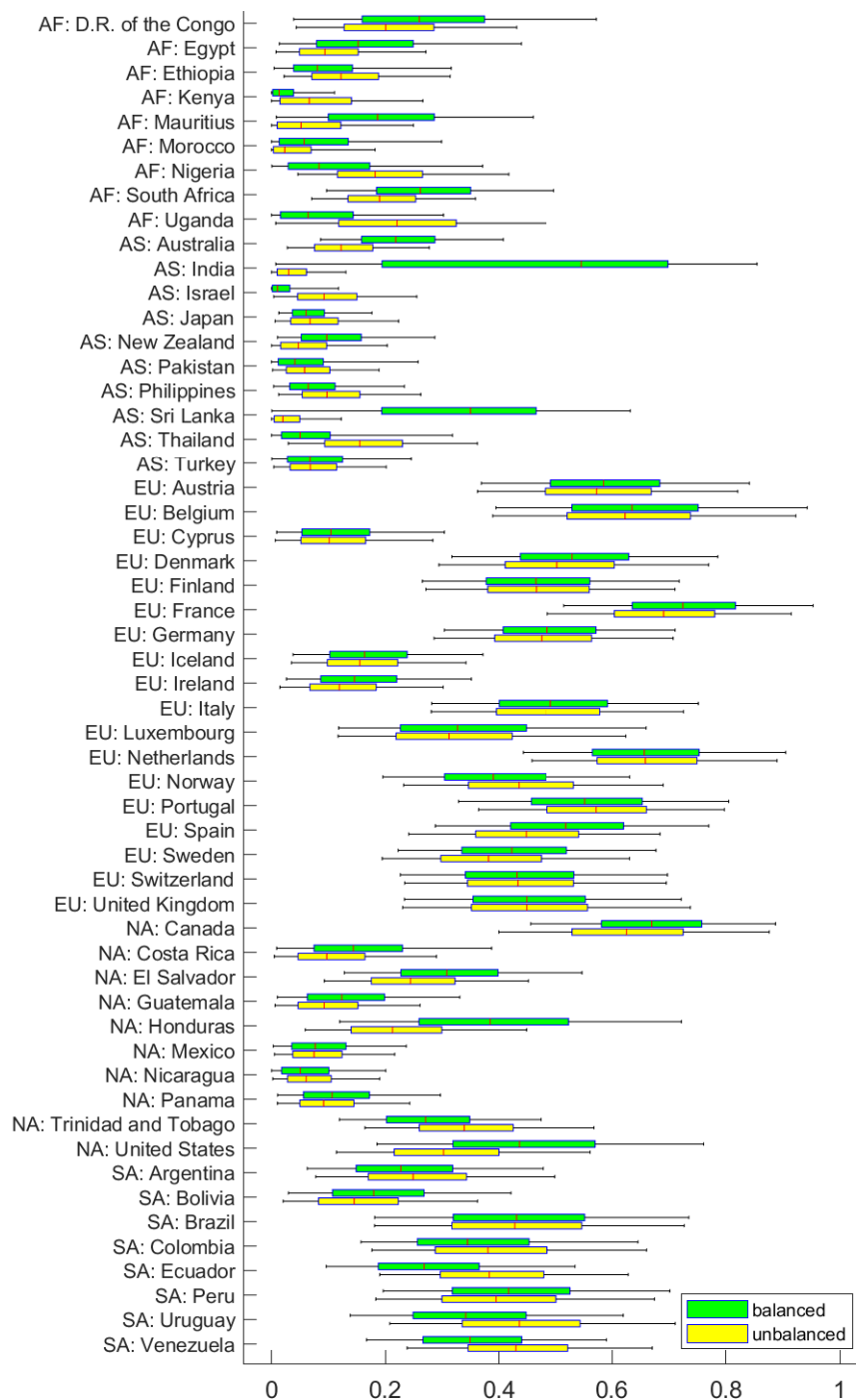
Note: The figure shows box plots reflecting the posterior sample variability of the of the common components'
variance divided by the overall variance of each GDP growth time series. The whiskers refer to 90% posterior
interval bounds. Abbreviations for continents: AF Africa, AS Asia, EU Europe, NA North America, SA South
America.

Figure II.9: Variance share of the common components, countries only covered in unbalanced data
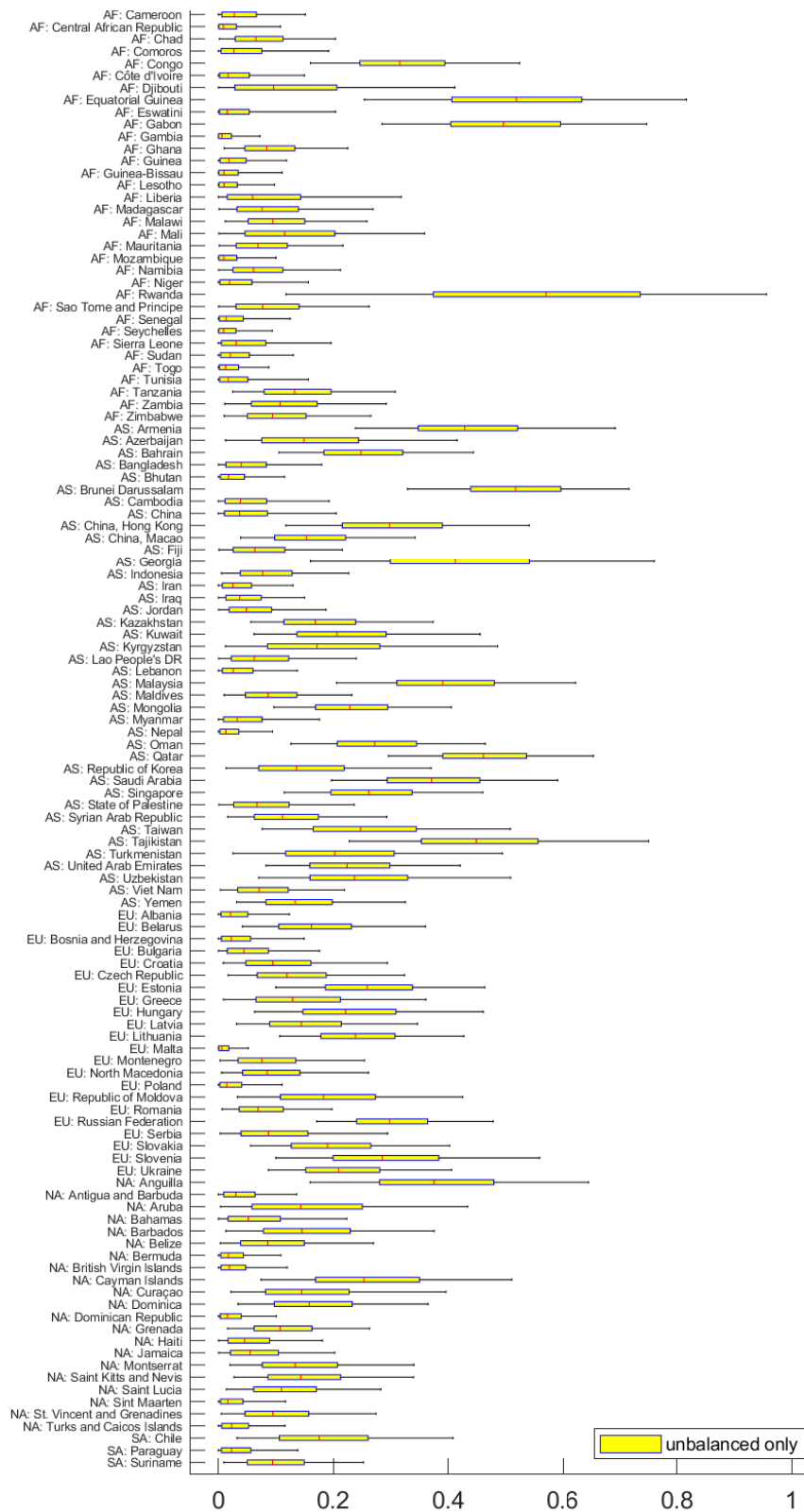
Note: The figure shows box plots reflecting the posterior sample variability of the of the common components' variance divided by the overall variance of each GDP growth time series. The whiskers refer to 90% posterior interval bounds. Abbreviations for continents: AF Africa, AS Asia, EU Europe, NA North America, SA South America.

the marginal likelihood. For factor models and given complete data for estimation, Chan
and Grant (2016) and McCausland (2012) show that the integrated likelihood function has
a normal distribution function. Given the quantities derived in this paper, we can show that
the model (Equation (II.1)) and (Equation (II.2)) given missing observations also implies a
normal density for the observed data.

In Section II.4, we have used joint normals for factors, variables with missing observa-
tions, and variables with observed data to derive conditional samplers. The permutations
of variables $z_{P_{2s}} = (x_m^\mathsf{T}, \eta^\mathsf{T}, x_o^\mathsf{T})^\mathsf{T}$ and $z_{P_\tau} = (\eta_1^\mathsf{T}, x_{m,1}^\mathsf{T}, \eta_2^\mathsf{T}, x_{m,2}^\mathsf{T}, \ldots, \eta_T^\mathsf{T}, x_{m,T}^\mathsf{T}, x_o^\mathsf{T})^\mathsf{T}$ have all in
common that the variables with observed data x_o are ordered last.

According to (Equation (II.10)), the permuted vector z_P is distributed as

$$z_{P_z}|\theta \sim \mathcal{N}\left(0_{T(r+N)\times 1}, Q_z^{-1}\right),$$ (II.32)

with $Q_z = P_z Q P_z^\mathsf{T}$ and the corresponding covariance matrix $\Omega_z = Q_z^{-1}$. For later use, we
define the partitions

$$Q_z = \begin{pmatrix} Q_{\eta x_m, \eta x_m} & Q_{\eta x_m, x_o} \\ Q_{x_o, \eta x_m} & Q_{x_o, x_o} \end{pmatrix}, \quad \Omega_z = \begin{pmatrix} \Omega_{\eta x_m, \eta x_m} & \Omega_{\eta x_m, x_o} \\ \Omega_{x_o, \eta x_m} & \Omega_{x_o, x_o} \end{pmatrix}.$$ (II.33)

If variables jointly follow a normal distribution, the marginal distributions are also normal
distributions, and the moments of the marginal distributions can be taken from the partitions
of the joint moments (Anderson, 2003, Theorem 2.4.3). In our case, we obtain the marginal
distribution of the variables with observed data $p(x_o|\theta)$ according to

$$p(x_o|\theta) \overset{\mathcal{D}}{=} \mathcal{N}\left(0_{(1-\kappa)TN\times 1}, \Omega_{x_o, x_o}\right).$$ (II.34)

It implies the log integrated likelihood function

$$\log p(x_o = x^o|\theta) = -\frac{(1-\kappa)TN}{2}\log(2\pi) - \frac{T}{2}\log\left|\Omega_{x_o, x_o}^{-1}\right| - \frac{1}{2}(x^o)^\mathsf{T}\Omega_{x_o, x_o}^{-1} x^o.$$ (II.35)

To evaluate the integrated likelihood function, we can make use of the time-permutation
sampler from Section II.4.2. From the definition of the covariance and precision matrix

$$\begin{pmatrix} \Omega_{\eta x_m, \eta x_m} & \Omega_{\eta x_m, x_o} \\ \Omega_{x_o, \eta x_m} & \Omega_{x_o, x_o} \end{pmatrix} \begin{pmatrix} Q_{\eta x_m, \eta x_m} & Q_{\eta x_m, x_o} \\ Q_{x_o, \eta x_m} & Q_{x_o, x_o} \end{pmatrix} = I$$ (II.36)

we obtain four matrix equations. From them, we can derive Ω_{x_o, x_o}^{-1} as an expression of the
components of the precision matrix according to

$$\Omega_{x_o, x_o}^{-1} = Q_{x_o, x_o} - Q_{x_o, \eta x_m} Q_{\eta x_m, \eta x_m}^{-1} Q_{\eta x_m, x_o},$$ (II.37)

and the log integrated likelihood function becomes

$$\log p(x_o = x^o | \theta) = -\frac{(1-\kappa)TN}{2}\log(2\pi) - \frac{T}{2}\log\left|Q_{x_o,x_o} - Q_{x_o,\eta x_m}Q_{\eta x_m,\eta x_m}^{-1}Q_{\eta x_m,x_o}\right|$$

$$-\frac{1}{2}(x^o)^\mathsf{T}\left(Q_{x_o,x_o} - Q_{x_o,\eta x_m}Q_{\eta x_m,\eta x_m}^{-1}Q_{\eta x_m,x_o}\right)x^o. \tag{II.38}$$

Evaluating the quantities is straightforward using the results from the time-permutation sampler. In Section II.B, we show that $Q_{\eta x_m, \eta x_m}$ is block-banded, so the methods outline before can be directly applied in line with McCausland (2012) and Chan and Grant (2016).

Note that these results also allow to estimate a factor model with stochastic volatility as proposed in Chan and Eisenstat (2018).

II.7.2 Other state-space models

The methods in the paper were so far applied to a factor model. However, they can be easily adopted to other state-space models in applications with missing data.

If we set $\psi = 0_{N \times N}$ in (Equation (II.2)), we obtain a general state-space model, where η_t represents the state vector. In this case, we have $\Psi = I_{TN}$, which simplifies the precision matrix in (*Equation* (II.8)). We can also consider cases where the number of disturbances in the state equation is smaller to the number of states, for example, if $\eta_t = \phi \eta_{t-1} + R_t u_{\eta,t}$ with the number of columns in R_t and the length of $u_{\eta,t}$ smaller than r, as in the general state-space models discussed in Durbin and Koopman (2002).

In a similar way, we can generalize the TVP-BVAR models as in Chan and Eisenstat (2018) to applications with missing observations in the data. Consider the model

$$B_{0,t}y_t = \mu_t + B_{1,t}y_{t-1} + \cdots + B_{p,t}y_{t-p} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \Sigma_t), \tag{II.39}$$

where y_t is an N-dimensional vector of model variables, μ_t is a vector of time-varying intercepts, $B_{1,t}, \ldots, B_{p,t}$ are time-varying VAR coefficient matrices, $B_{0,t}$ is a lower triangular matrix with ones on the diagonal, and time-varying volatilities Σ_t for $t = 1, \ldots, T$. In the literature such as Chan and Eisenstat (2018); Chan (2020); Chan et al. (2020), estimation of the VAR coefficients using precision-based samplers given complete data proceeds by conditioning on lags of y_t on the right-hand side. If some of the observations are missing, this conditioning is not feasible. In this case, the methods developed in this paper can be adopted to obtain a Gibbs sampler for TVP-BVAR models in the presence of partly missing observations along the lines of Little and Rubin (2002). In particular, we can provide a sampling step for missing observations in y_t. Given augmented data, the samplers as proposed in the papers cited above can be applied for Bayesian estimation of the model parameters in (Equation (II.39)).

To derive the conditional posterior distribution for missing observations, rewrite the model (Equation (II.39)) in stacked form with $y = (y_1^\mathsf{T}, y_2^\mathsf{T}, \ldots, y_T^\mathsf{T})^\mathsf{T}$ and $\tilde{\mu}_y = (\mu_1^\mathsf{T}, \mu_2^\mathsf{T}, \ldots,$

$\mu_T^\mathsf{T})^\mathsf{T}$ as

$$\Phi_y y = \tilde{\mu}_y + \epsilon_y \tag{II.40}$$

with

$$\Phi_y = \begin{pmatrix} B_{0,1} & & & & & & \\ -B_{1,1} & B_{0,2} & & & & & \\ -B_{2,1} & -B_{1,2} & B_{0,3} & & & & \\ \vdots & \ddots & \ddots & \ddots & & & \\ -B_{p,1} & & \ddots & \ddots & \ddots & & \\ & \ddots & & \ddots & \ddots & B_{0,T-1} & \\ & & -B_{p,T-p} & \cdots & -B_{2,T-2} & -B_{1,T-1} & B_{0,T} \end{pmatrix}, \tag{II.41}$$

$\epsilon_y \sim \mathcal{N}(0, \Sigma)$, and $\Sigma = \operatorname{diag}(\Sigma_1, \Sigma_2, \ldots, \Sigma_T)$.

Given model parameters θ_y, we have $y|\theta_y \sim \mathcal{N}(\mu_y, \Phi_y^{-1}\Sigma\Phi_y^{-\mathsf{T}})$ with $\mu_y = \Phi_y^{-1}\tilde{\mu}_y$ and precision matrix $Q_y = (\Phi_y^{-1}\Sigma\Phi_y^{-\mathsf{T}})^{-1} = \Phi_y^{\mathsf{T}}\Sigma^{-1}\Phi_y$. We can now reorder the variables in y into variables with missing observations y_m above the variables with available observations y_o as in (Equation (II.12)) by

$$y_{P_y} = \begin{pmatrix} y_m \\ y_o \end{pmatrix} = P_y y = \begin{pmatrix} P_{y_m} \\ P_{y_o} \end{pmatrix} y \tag{II.42}$$

such that the transformed vector y_{P_y} is distributed as

$$y_{P_y}|\theta_y \sim \mathcal{N}\left(P_y\mu_y, (P_y Q_y P_y^{\mathsf{T}})^{-1}\right). \tag{II.43}$$

We can now partition the mean and the precision matrix as in (Equation (II.42)) and obtain

$$P_y\mu_y = \begin{pmatrix} \mu_m \\ \mu_o \end{pmatrix}, \qquad P_y Q_y P_y^{\mathsf{T}} = \begin{pmatrix} Q_{y_m,y_m} & Q_{y_m,y_o} \\ Q_{y_o,y_m} & Q_{y_o,y_o} \end{pmatrix}, \tag{II.44}$$

which provides us with the moments of the distribution for the missing observations y_m conditional on the observations $y_o = y^o$ and the model parameters. The conditional distribution of the missing values is given by

$$p(y_m|y_o = y^o, \theta_y) \overset{\mathcal{D}}{=} \mathcal{N}\left(\mu_m - Q_{y_m,y_m}^{-1}Q_{y_m,y_o}(y^o - \mu_o), Q_{y_m,y_m}^{-1}\right). \tag{II.45}$$

Note that it is not straightforward to jointly sample missing values and time-varying VAR parameters in the TVP-BVAR model. The reason is that products of missing values and time-varying VAR parameters are present on the right-hand side of the TVP-BVAR model, rendering the resulting state-space system non-linear.

II.8 Conclusions

We propose a simple and efficient precision-based sampler for unobserved states and missing observations conditional on available data. The approach extends the existing literature on precision-based sampling, which typically considers complete-data applications. By allowing the investigation of incomplete data sets, the sampler proposed here expands the range of potential applications for precision-based samplers in practice.

The approach can be applied to a wide range of state-space models such as time-varying parameter BVARs, as their corresponding precision matrix has a block-banded structure. In this paper we apply the sampler to a large dynamic factor model. To facilitate sampling in the presence of missing observations, we reorder the variables in the precision matrix by alternative permutations. Based on the permuted precision matrices with small bandwidth, we can employ fast band-matrix computing techniques to draw from the conditional distributions of factors and missing observations given available data. In the simulations, a 2-step precision-based sampler, which sequentially samples factors and missing observations, turns out to be computationally efficient. On the other hand, a joint sampler based on time permutation is slower, but facilitates computing the integrated likelihood for model comparison more easily. Both can be directly integrated into Bayesian estimation procedures like the Gibbs sampler.

Appendix

II.A Posterior sampler

A more general specification of the factor model than (Equation (II.1)) and (Equation (II.2)) has $p \geq 1$ lags in the factor VAR and $q \geq 1$ AR lags in the equations for idiosyncratic components. We obtain the model equations

$$x_t = \lambda \eta_t + \epsilon_t, \tag{II.46}$$

$$\phi(L)\eta_t = u_{\eta,t}, \quad \psi(L)\epsilon_t = u_{\epsilon,t}. \tag{II.47}$$

with polynomials $\phi(L) = I_r - \phi_1 L - \ldots - \phi_p L^p$ and $\psi(L) = I_N - \psi_1 L - \ldots - \psi_q L^q$ in the lag operator $Ly_t = y_{t-1}$. The polynomial matrices in $\psi(L)$ are assumed to be diagonal. The factor VAR disturbances are distributed as $u_{\eta,t} \sim \mathcal{N}(0_{r \times 1}, \omega_\eta)$, and we assume for simplicity $\omega_\eta = I_r$. With respect to the idiosyncratic components we assume $u_{\epsilon,t} \sim \mathcal{N}(0_{N \times 1}, \omega_\epsilon)$, where ω_ϵ is also diagonal with blocks $\omega_{\epsilon,i}$ for $i = 1, \ldots, N$ on the main diagonal.

Concerning the prior specifications of the model, we follow closely the existing factor model literature. We use sparse priors for the elements in the loading matrix as in George and McCulloch (1993, 1997); Geweke (1996); Carvalho et al. (2008); Kaufmann and Schumacher (2017, 2019). The hierarchical prior for the loadings is denoted as $p(\lambda|\theta_\lambda)$ with a prior for hyper-parameters $p(\theta_\lambda)$. The factors and missing observations follow multivariate

normal priors given model parameters, $p(\eta|\theta_+)$ and $p(x_m|\theta_+)$, where further model parameters are collected in θ_+: It contains the factor VAR polynomial parameters $\phi(L)$ and the idiosyncratic components AR polynomial parameters $\psi(L)$, which follow normal distributions truncated to the stationary region (Litterman, 1986). θ_+ also contains the variances of the innovations in the idiosyncratic components AR models $\omega_{\epsilon,i}$ for $i = 1, \ldots, N$. We use an inverse gamma distribution as the prior. Details regarding the prior specifications can be found in the subsections below.

Given available data x_o, we want to obtain samples from the posterior distribution

$$p(\eta, x_m, \lambda, \theta_\lambda, \theta_+|x_o) \propto L(x_o|\eta, x_m, \lambda, \theta_\lambda, \theta_+)p(x_m|\theta_+)p(\lambda|\theta_\lambda)p(\eta|\lambda, \theta_+)p(\theta_+)p(\theta_\lambda).$$
(II.48)

The sampler closely follows the Gibbs sampler proposed by Little and Rubin (2002) to tackle missing data. In a first step, Little and Rubin (2002) sample missing observations for x_m, and combine these samples with observed data. In subsequent steps, complete-data conditional posterior distributions are used for sampling the remaining model parameters. The use of the complete-data conditional posterior distributions in these steps is valid because the product of $L(x_o|\eta, x_m, \lambda, \theta_\lambda, \theta_+)$ and $p(x_m|\theta_+)$ in (Equation (II.48)) equals the complete-data likelihood $L(x_o, x_m|\eta, \lambda, \theta_\lambda, \theta_+)$, as implied by the model equations (Equation (II.46)) and (Equation (II.47)).

In our case, we expand on Little and Rubin (2002) by also estimating unobserved factors in the first block of the Gibbs sampler along with the missing observations. The sampler proceeds with the following three steps:

1. $p(\eta, x_m|x_o, \lambda, \theta_\lambda, \theta_+)$: To sample factors and missing observations, we have two options:

   (a) When applying time permutation precision-based sampling, we provide a joint draw of $\eta, x_m|x_o, \lambda, \theta_\lambda, \theta_+$ in a single step.

   (b) When applying the 2-step precision-based sampling, we sample sequentially from $p(\eta|x_o, x_m, \lambda, \theta_\lambda, \theta_+)$ and $p(x_m|x_o, \eta, \lambda, \theta_\lambda, \theta_+)$.

2. $p(\lambda|x_o, x_m, \eta, \theta_\lambda, \theta_+)$, $p(\theta_\lambda|x_o, x_m, \lambda)$: Sampling the loadings λ, and hyper-parameters related to the hierarchical prior for the loadings θ_λ given loadings λ.

3. $p(\theta_+|x_o, x_m, \eta, \lambda, \theta_\lambda)$: Sampling the rest of the model parameters.

II.A.1 Factors and missing observations: $p(\eta, x_m|x_o, \lambda, \theta_\lambda, \theta_+)$

With lag orders $p, q > 1$ in the factor model, the precision-based samplers as described in the main text for $p = q = 1$ have to be modified only in terms of the precision matrix Q. With

respect to the stacked factors, we have $\eta|\phi(L), \omega_\eta, \vartheta_{\eta,1|0} \sim \mathcal{N}(0_{Tr\times1}, \Phi^{-1}\Omega_\eta\Phi^{-\top})$, where

$$\Phi = \begin{pmatrix} I_r & & & & & & \\ -\phi_1 & I_r & & & & & \\ -\phi_2 & -\phi_1 & I_r & & & & \\ \vdots & \ddots & \ddots & \ddots & & & \\ -\phi_p & & \ddots & \ddots & \ddots & & \\ & \ddots & & \ddots & \ddots & I_r & \\ & & -\phi_p & \cdots & -\phi_2 & -\phi_1 & I_r \end{pmatrix}, \quad \Omega_\eta = \begin{pmatrix} \vartheta_{\eta,1|0} & & & \\ & \omega_\eta & & \\ & & \ddots & \\ & & & \omega_\eta \end{pmatrix}. \quad \text{(II.49)}$$

The matrix entries for the VAR lag polynomial $\phi(L)$ are taken from the previous Gibbs draw. We set $\omega_\eta = I_r$. As the initial covariance of factors $\vartheta_{\eta,1|0} = \vartheta_\eta$, we use the covariance matrix of factors and their lags implied by the VAR in companion form

$$\underbrace{\begin{pmatrix} \eta_t \\ \eta_{t-1} \\ \vdots \\ \eta_{t-p} \end{pmatrix}}_{\tilde{\eta}_t} = \underbrace{\begin{pmatrix} \phi_1 & \cdots & \cdots & \phi_p \\ I_r & & & 0 \\ & \ddots & & \vdots \\ & & I_r & 0 \end{pmatrix}}_{\tilde{\phi}} \begin{pmatrix} \eta_{t-1} \\ \eta_{t-2} \\ \vdots \\ \eta_{t-(p+1)} \end{pmatrix} + \underbrace{\begin{pmatrix} u_{\eta,1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\tilde{u}_{\eta,t}}. \quad \text{(II.50)}$$

The VAR innovations in the companion form $\tilde{u}_{\eta,t}$ have the $(pr \times pr)$ covariance matrix $\tilde{\omega}_\eta = \text{Var}(\tilde{u}_{\eta,t})$, which contains ω_η in the top-left block and zeros elsewhere. The companion form implies the $(pr \times pr)$ second-moment matrix of stacked factors lags $\tilde{\vartheta}_\eta = \text{Var}(\tilde{\eta}_t)$ from solving the equation $\text{vec}(\tilde{\vartheta}_\eta) = [I_{(pr)^2} - \tilde{\phi} \otimes \tilde{\phi}^\top]^{-1} \times \text{vec}(\tilde{\omega}_\eta)$. As the initial covariance of factors $\vartheta_{\eta,1|0} = \vartheta_\eta$, we use the first $(r \times r)$ block-diagonal element in $\tilde{\vartheta}_\eta$.

With respect to the stacked idiosyncratic components, we have $\epsilon|\psi(L), \omega_\epsilon, \vartheta_{\epsilon,1|0} \sim \mathcal{N}(0_{TN\times1}, \Psi^{-1}\Omega_\epsilon\Psi^{-\top})$ where Ψ and Ω_ϵ are constructed along the lines of (Equation (II.49)). The matrix entries for the lag polynomial $\phi(L)$ and the innovation variances ω_ϵ are taken from the previous Gibbs draw. To initialize the second moments of the idiosyncratic components by $\vartheta_{\epsilon,1|0}$, we use the second moments implied by the AR models for the idiosyncratic components, which can be computed using a companion form in the same way as for the factors.

Given Φ, Ω_η, Ψ, and Ω_ϵ, we define the precision matrix Q in the same way as in (Equation (II.7)) and (Equation (II.8)) in the main text and can apply the alternative precision-based samplers accordingly.

II.A.2 Loadings: $p(\lambda|x_o, x_m, \eta, \theta_\lambda, \theta_+)$, $p(\theta_\lambda|x_o, x_m, \lambda)$

Following George and McCulloch (1993, 1997); Geweke (1996); Kaufmann and Schumacher (2019), we use a sparse point-mass normal mixture prior on the loadings with a common

probability of non-zero loading on factor j across variables according to

$$p(\lambda_{ij}) \overset{\mathcal{D}}{=} (1 - \rho_j)\delta_0(\lambda_{ij}) + \rho_j \, \mathcal{N}(0, \tau_j), \qquad (\text{II}.51)$$

$$p(\rho_j) \overset{\mathcal{D}}{=} \mathcal{B}(r_0 s_0, r_0(1 - s_0)), \qquad (\text{II}.52)$$

where the Dirac delta function $\delta_0(\cdot)$ assigns all probability mass to zero. To capture potential factor-specific scaling of loadings, we specify an inverse gamma distribution for $\tau_j \sim \mathcal{IG}(g_0, G_0)$. The expected probability of non-zero factor loading, s_0, and precision r_0 are hyperparameters. In the empirical application, we specify $s_0 = 0.5$, $r_0 = 3.0$, $g_0 = 2$, and $G_0 = 0.5$. We define the hyper-parameters by $\theta_\lambda = \{\rho_j, \tau_j\}$ for $j = 1, \ldots, r$.

Given complete data x, which combines samples of missing data and the observed data, we can sample from $p(\lambda | x, \eta, \theta, \theta_\lambda)$ in the following way. The posterior odds of a non-zero factor loading in (Equation (II.51)) are given by

$$\frac{P(\lambda_{ij} \neq 0 | x, \cdot)}{P(\lambda_{ij} = 0 | x, \cdot)} = \frac{p(\lambda_{ij})|_{\lambda_{ij}=0}}{p(\lambda_{ij}|\cdot)|_{\lambda_{ij}=0}} \frac{\rho_j}{1 - \rho_j} = \frac{\mathcal{N}(0; 0, \tau_j)}{\mathcal{N}(0; m_{ij}, M_{ij})} \frac{\rho_j}{1 - \rho_j}, \qquad (\text{II}.53)$$

where the moments m_{ij} and M_{ij} are

$$M_{ij} = \left(\frac{1}{\sigma_i^2} \sum_{t=q+1}^{T} \left(\psi_i(L)\eta_{jt} \right)^2 + \frac{1}{\tau_j} \right)^{-1}, \quad m_{ij} = M_{ij} \left(\frac{1}{\sigma_i^2} \sum_{t=q+1}^{T} \left(\psi_i(L)\eta_{jt} \right) x_{it}^* \right), \qquad (\text{II}.54)$$

and x_{it}^* is a transform of the variables by

$$x_{it}^* = \psi_i(L)x_{it} - \sum_{l=1, l \neq j}^{k} \lambda_{il}\psi_i(L)\eta_{lt} = \lambda_{ij}\psi_i(L)\eta_{jt} + \epsilon_{it},$$

which isolates the effect of factor j on variable i. Note that the conditional sampler in the empirical application is applied only to those elements in λ, which are not fixed to zero in the continental group factors.

We choose $\lambda_{ij} \neq 0$ if $U \leq PO_{ij}/(1 + PO_{ij})$, where U is a draw from the uniform distribution over $[0, 1]$. If we choose $\lambda_{ij} \neq 0$, λ_{ij} is drawn from $\mathcal{N}(m_{ij}, M_{ij})$, otherwise it is set equal to zero.

Given λ_{ij}, we can update the hyper-parameters θ_λ. The conditional posterior of ρ_j is $p(\rho_j | x, \cdot) \overset{\mathcal{D}}{=} \mathcal{B}(r_{1j}, r_{2j})$ with $r_{1j} = r_0 s_0 + S_j$, $r_{2j} = r_0(1 - s_0) + N_j - S_j$, and N_j is the number of loading elements not fixed to zero a-priori in the continental factors for $j = 2, \ldots, 6$, whereas for the global factor indexed by $j = 1$, we have $N_1 = N$. We also define $S_j = \sum_{i=1}^{N} I\{\lambda_{ij} \neq 0\}$. In the simulation experiments in Section II.5, we do not set zero restrictions in the loadings. In that case, we have to use $r_{2j} = r_0(1 - s_0) + N - S_j$.

The conditional posterior of the τ_j is $p(\tau_j | x, \cdot) \overset{\mathcal{D}}{=} \mathcal{IG}(g_j, G_j)$ with $g_j = g_0 + 0.5 N_j$ and $G_j = G_0 + 0.5 \sum_{i=1}^{N} \lambda_{ij}^2$.

II.A.3 Further parameters: $p(\theta_+|x_o, x_m, \eta, \lambda, \theta_\lambda)$

In the VAR polynomial for the factors and the AR polynomial for the idiosyncratic components, we employ a standard Minnesota prior specification (Litterman, 1986). The prior variance for the first autoregressive lag is set to $\vartheta_0^2 = 0.09$, whereas the shrinkage factor for the prior variance on lags of other variables is equal to $\vartheta_1^2 = 0.03$. The normal prior moments for the VAR polynomial parameters are mean zero and variance $\mathrm{Var}(\phi_{l,i,j}) = \vartheta_0^2/l^2$ for $i = l$ at lag l, and $\mathrm{Var}(\phi_{l,i,j}) = \vartheta_1^2 \vartheta_0^2/l^2$ for $i \neq l$ at lag l.

With respect to the variance of the innovations in the idiosyncratic components AR model, we choose $p(\omega_{\epsilon,i}) \stackrel{\mathcal{D}}{=} \mathcal{IG}(2.0, 1.0)$ for $i = 1, \ldots, N$. The conditional posteriors are standard in the literature and not reported here. More details can be found in Appendix A.3 in Kaufmann and Schumacher (2019).

II.B Precision matrix of joint time-permutation sampler

Given the permutation matrix P_τ, we obtain the permuted precision matrix $Q_\tau = P_\tau Q P_\tau^\mathsf{T}$ for sampling factors and missing observations. Below we show that Q_τ is a block-banded matrix. In particular, it is a block tridiagonal matrix with varying block dimensions, having blocks in the lower diagonal, main diagonal and upper diagonal, and all other blocks being zero matrices (Golub and Van Loan, 2013, Chapter 4.5).

To derive Q_τ, first note that the precision matrix Q from (Equation (II.8)) is a function of the prior precision matrices Q_η and Q_ϵ according to

$$Q = \begin{pmatrix} \Phi^\mathsf{T}\Omega_\eta^{-1}\Phi + \Lambda^\mathsf{T}(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi)\Lambda & -\Lambda^\mathsf{T}(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi) \\ -(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi)\Lambda & \Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi \end{pmatrix} = \begin{pmatrix} Q_\eta + \Lambda^\mathsf{T}Q_\epsilon\Lambda & -\Lambda^\mathsf{T}Q_\epsilon \\ -Q_\epsilon\Lambda & Q_\epsilon \end{pmatrix}. \quad \text{(II.55)}$$

As factors and idiosyncratic components follow VAR(1) processes, both precision matrices Q_η and Q_ϵ are block tridiagonal (Chan and Jeliazkov, 2009). In particular, we have

$$Q_\eta = \Phi^\mathsf{T}\Omega_\eta^{-1}\Phi$$
$$= \begin{pmatrix} \vartheta_{\eta,1|0}^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi & -\phi^\mathsf{T}\omega_\eta^{-1} & & & \\ -\omega_\eta^{-1}\phi & \omega_\eta^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi & -\phi^\mathsf{T}\omega_\eta^{-1} & & \\ & \ddots & \ddots & \ddots & \\ & & -\omega_\eta^{-1}\phi & \omega_\eta^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi & -\phi^\mathsf{T}\omega_\eta^{-1} \\ & & & -\omega_\eta^{-1}\phi & \omega_\eta^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi \end{pmatrix},$$
$$\text{(II.56)}$$

and $Q_\epsilon = \Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi$ can be written accordingly.

Given Q_η and Q_ϵ, we can show that the four submatrices of Q in (Equation (II.55)),

$$Q = \begin{pmatrix} Q_\eta + \Lambda^\mathsf{T}Q_\epsilon\Lambda & -\Lambda^\mathsf{T}Q_\epsilon \\ -Q_\epsilon\Lambda & Q_\epsilon \end{pmatrix} = \begin{pmatrix} Q_{\eta\eta} & Q_{\eta x} \\ Q_{x\eta} & Q_{xx} \end{pmatrix}, \quad \text{(II.57)}$$

are also block tridiagonal matrices each. Note that Λ is a block-diagonal matrix. Pre- and post-multiplying the block tridiagonal matrix Q_ϵ by Λ leads to a block tridiagonal matrix without altering the block bandwidth. The sum of two block tridiagonal matrices $Q_\eta + \Lambda^\mathsf{T} Q_\epsilon \Lambda$ is also a block tridiagonal matrix. In particular, we obtain

$$
Q_{\eta\eta} = \begin{pmatrix}
Q_{\eta\eta,(1,1)} & Q_{\eta\eta,(1,2)} & & & & & \\
Q_{\eta\eta,(2,1)} & Q_{\eta\eta,(2,2)} & Q_{\eta\eta,(2,3)} & & & & \\
& Q_{\eta\eta,(3,2)} & Q_{\eta\eta,(3,3)} & Q_{\eta\eta,(3,4)} & & & \\
& & \ddots & \ddots & & \ddots & \\
& & & Q_{\eta\eta,(T-1,T-2)} & Q_{\eta\eta,(T-1,T-1)} & Q_{\eta\eta,(T-1,T)} \\
& & & & Q_{\eta\eta,(T,T-1)} & Q_{\eta\eta,(T,T)}
\end{pmatrix} \tag{II.58}
$$

with

$$
Q_{\eta\eta,(1,1)} = \vartheta_{\eta,1|0}^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi + \lambda^\mathsf{T}(\vartheta_{\epsilon,1|0}^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi)\lambda, \tag{II.59}
$$

$$
Q_{\eta\eta,(t,t)} = \omega_\eta^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi + \lambda^\mathsf{T}(\omega_\epsilon^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi)\lambda \qquad \text{for} \quad t = 2,\ldots,T, \tag{II.60}
$$

$$
Q_{\eta\eta,(t,t+1)} = -\phi^\mathsf{T}\omega_\eta^{-1} - \lambda^\mathsf{T}(-\psi^\mathsf{T}\omega_\epsilon^{-1})\lambda \qquad \text{for} \quad t = 1,\ldots,T-1, \tag{II.61}
$$

$$
Q_{\eta\eta,(t+1,t)} = -\omega_\eta^{-1}\phi - \lambda^\mathsf{T}(\omega_\epsilon^{-1}\psi)\lambda \qquad \text{for} \quad t = 1,\ldots,T-1. \tag{II.62}
$$

Accordingly, the non-zero submatrices of $Q_{\eta x} = \left(Q_{\eta x,(t_1,t_2)}\right)_{t_1,t_2=1,\ldots,T}$ are given by

$$
Q_{\eta x,(1,1)} = -\lambda^\mathsf{T}(\vartheta_{\epsilon,1|0}^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi), \tag{II.63}
$$

$$
Q_{\eta x,(t,t)} = -\lambda^\mathsf{T}(\omega_\epsilon^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi) \qquad \text{for} \quad t = 2,\ldots,T, \tag{II.64}
$$

$$
Q_{\eta x,(t,t+1)} = \lambda^\mathsf{T}(\psi^\mathsf{T}\omega_\epsilon^{-1}) \qquad \text{for} \quad t = 1,\ldots,T-1, \tag{II.65}
$$

$$
Q_{\eta x,(t+1,t)} = \lambda^\mathsf{T}(\omega_\epsilon^{-1}\psi) \qquad \text{for} \quad t = 1,\ldots,T-1, \tag{II.66}
$$

and all other submatrices are zero blocks, $Q_{\eta x,(t_1,t_2)} = 0$ for $|t_1-t_2| > 1$. Finally, the non-zero submatrices of $Q_{x\eta} = \left(Q_{x\eta,(t_1,t_2)}\right)_{t_1,t_2=1,\ldots,T}$ are given by

$$
Q_{x\eta,(1,1)} = -(\vartheta_{\epsilon,1|0}^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi)\lambda, \tag{II.67}
$$

$$
Q_{x\eta,(t,t)} = -(\omega_\epsilon^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi)\lambda \qquad \text{for} \quad t = 2,\ldots,T, \tag{II.68}
$$

$$
Q_{x\eta,(t,t+1)} = (\psi^\mathsf{T}\omega_\epsilon^{-1})\lambda \qquad \text{for} \quad t = 1,\ldots,T-1, \tag{II.69}
$$

$$
Q_{x\eta,(t+1,t)} = (\omega_\epsilon^{-1}\psi)\lambda \qquad \text{for} \quad t = 1,\ldots,T-1, \tag{II.70}
$$

and $Q_{x\eta,(t_1,t_2)} = 0$ for $|t_1 - t_2| > 1$.

The precision matrix $Q_\tau = P_\tau Q P_\tau^\mathsf{T}$ can be obtained by block multiplication

$$
Q_\tau =
\begin{pmatrix}
P_{\eta,1} & 0 \\
0 & P_{x_m,1} \\
P_{\eta,2} & 0 \\
0 & P_{x_m,2} \\
\vdots & \vdots \\
P_{\eta,T} & 0 \\
0 & P_{x_m,T} \\
0 & P_{x_o}
\end{pmatrix}
\begin{pmatrix}
Q_{\eta\eta} & Q_{\eta x} \\
Q_{x\eta} & Q_{xx}
\end{pmatrix}
\begin{pmatrix}
P_{\eta,1}^\mathsf{T} & 0 & P_{\eta,2}^\mathsf{T} & 0 & \cdots & P_{\eta,T}^\mathsf{T} & 0 & 0 \\
0 & P_{x_m,1}^\mathsf{T} & 0 & P_{x_m,2}^\mathsf{T} & \cdots & 0 & P_{x_m,T}^\mathsf{T} & P_{x_o}^\mathsf{T}
\end{pmatrix}
\quad \text{(II.71)}
$$

with $P_{\eta,t} = \begin{pmatrix} 0_{r\times r(t-1)} & I_r & 0_{r\times r(T-t)} \end{pmatrix}$ and $P_{x_m,t} = \begin{pmatrix} 0_{N_{m,t}\times(t-1)N} & P_{x_m,(t,t)} & 0_{N_{m,t}\times(T-t)N} \end{pmatrix}$. Note
that the permuted variables z_{P_τ} permutation matrix P_τ can be considered as having $T+1$
row blocks, where T row blocks contain model factors η_t and variables corresponding to
missing observations $x_{m,t}$ for each period t, and the $(T+1)$-st row block contains variables
corresponding to observed data x_o. This blocking is implied by the variable ordering in
$z_{P_\tau} = (\eta_1^\mathsf{T}, x_{m,1}^\mathsf{T}, \eta_2^\mathsf{T}, x_{m,2}^\mathsf{T}, \ldots, \eta_T^\mathsf{T}, x_{m,T}^\mathsf{T}, x_o^\mathsf{T})^\mathsf{T}$. Given the blocking in P_τ, the corresponding
precision matrix can be partitioned in the same way. We obtain Q_τ as

$$
Q_\tau =
\begin{pmatrix}
Q_{\tau,(1,1)} & Q_{\tau,(1,2)} & & & & & & Q_{\tau,(1,T+1)} \\
Q_{\tau,(2,1)} & Q_{\tau,(2,2)} & Q_{\tau,(2,3)} & & & & & Q_{\tau,(2,T+1)} \\
& \ddots & \ddots & \ddots & & & & \vdots \\
& & Q_{\tau,(T-2,T-3)} & Q_{\tau,(T-2,T-2)} & Q_{\tau,(T-2,T-1)} & & & Q_{\tau,(T-2,T+1)} \\
& & & Q_{\tau,(T-1,T-2)} & Q_{\tau,(T-1,T-1)} & Q_{\tau,(T-1,T)} & & Q_{\tau,(T-1,T+1)} \\
& & & & Q_{\tau,(T,T-1)} & Q_{\tau,(T,T)} & & Q_{\tau,(T,T+1)} \\
Q_{\tau,(T+1,1)} & Q_{\tau,(T+1,2)} & \cdots & & Q_{\tau,(T+1,T-1)} & Q_{\tau,(T+1,T)} & & Q_{\tau,(T+1,T+1)}
\end{pmatrix},
$$
$$\text{(II.72)}$$

with $(T+1)\times(T+1)$ blocks. Note that the dimensions of the blocks in Q_τ are time-varying,
depending on the number of missing observations in each period, $N_{m,t}$.

The submatrices of Q_τ are given by

$$
Q_{\tau,(t,t)} =
\begin{pmatrix}
Q_{\eta\eta,(t,t)} & Q_{\eta x,(t,t)} P_{x_m,(t,t)}^\mathsf{T} \\
P_{x_m,(t,t)} Q_{x\eta,(t,t)} & P_{x_m,(t,t)} Q_{\epsilon,(t,t)} P_{x_m,(t,t)}^\mathsf{T}
\end{pmatrix}
\quad \text{(II.73)}
$$

for $t = 1, 2, \ldots, T$ on the main diagonal, and

$$
Q_{\tau,(t,t+1)} =
\begin{pmatrix}
Q_{\eta\eta,(t,t+1)} & Q_{\eta x,(t,t+1)} P_{x_m,(t+1,t+1)}^\mathsf{T} \\
P_{x_m,(t,t)} Q_{x\eta,(t,t+1)} & P_{x_m,(t,t)} Q_{\epsilon,(t,t+1)} P_{x_m,(t+1,t+1)}^\mathsf{T}
\end{pmatrix},
\quad \text{(II.74)}
$$

$$
Q_{\tau,(t+1,t)} =
\begin{pmatrix}
Q_{\eta\eta,(t+1,t)} & Q_{\eta x,(t+1,t)} P_{x_m,(t,t)}^\mathsf{T} \\
P_{x_m,(t+1,t+1)} Q_{x\eta,(t+1,t)} & P_{x_m,(t+1,t+1)} Q_{\epsilon,(t+1,t)} P_{x_m,(t,t)}^\mathsf{T}
\end{pmatrix},
\quad \text{(II.75)}
$$

for $t = 1, \ldots, T-1$ on the upper and lower block diagonal, respectively. Furthermore,

$Q_{\tau,(t_1,t_2)} = 0$ for $|t_1 - t_2| > 1$ and $t_1, t_2 = 1, \ldots, T$. Concerning the $(T+1)$-st row and column, we obtain

$$Q_{\tau,(T+1,T+1)} = P_{x_o} Q_\epsilon P_{x_o}^\mathsf{T}, \tag{II.76}$$

$$Q_{\tau,(T+1,t)} = P_{x_o} \begin{pmatrix} Q_{x\eta} P_{\eta,t}^\mathsf{T} & Q_\epsilon P_{x_m,t}^\mathsf{T} \end{pmatrix} \qquad \text{for} \quad t = 1, \ldots, T, \tag{II.77}$$

$$Q_{\tau,(t,T+1)} = \begin{pmatrix} P_{\eta,t} Q_{\eta x} \\ P_{x_m,t} Q_\epsilon \end{pmatrix} P_{x_o}^\mathsf{T} \qquad \text{for} \quad t = 1, \ldots, T. \tag{II.78}$$

To sample from $p(\eta, x_m | x_o = x^o, \theta)$, the precision matrix Q_τ will be partitioned according to

$$Q_\tau = \begin{pmatrix} Q_{\eta x_m, \eta x_m} & Q_{\eta x_m, x_o} \\ Q_{x_o, \eta x_m} & Q_{x_o, x_o} \end{pmatrix}, \tag{II.79}$$

where the upper-left block $Q_{\eta x_m, \eta x_m}$ contains the first top-left $T \times T$ sub-blocks from (Equation (II.72)) and has dimensions $(rT + \sum_{t=1}^{T} N_{m,t}) \times (rT + \sum_{t=1}^{T} N_{m,t})$, whereas the lower-right block Q_{x_o, x_o} has dimensions $(NT - \sum_{t=1}^{T} N_{m,t}) \times (NT - \sum_{t=1}^{T} N_{m,t})$. The conditional distribution becomes $p(\eta, x_m | x_o = x^o, \theta) \overset{\mathcal{D}}{=} \mathcal{N}(-Q_{\eta x_m, \eta x_m}^{-1} Q_{\eta x_m, x_o} x^o, Q_{\eta x_m, \eta x_m}^{-1})$. From (Equation (II.72)), we can see that $Q_{\eta x_m, \eta x_m}$ is block tridiagonal and thus facilitates precision-based sampling efficiently.

Chapter III

How useful is external information from professional forecasters? Conditional forecasts in large factor models

Abstract

This paper evaluates forecasts from a factor model estimated with a large real-time dataset of the German economy. The evaluation focuses on a broad cross-section of variables such as activity series including components of the gross domestic product and gross value added, deflators and other price measures as well as several labor market indicators. In addition to unconditional forecasts for these variables, we also investigate to what extent the forecast accuracy improves when we condition on professional forecasters' view on GDP growth and CPI inflation. We find that over the period from 2006 to 2017 the model's unconditional forecasts are broadly in line with autoregressive benchmarks for the majority of the 37 series that we focus on in the evaluation, in some cases performing somewhat better and in others somewhat worse. For a few variables capturing real activity and some price indicators, however, we find large gains in predictive accuracy that persist for forecast horizons of up to two quarters ahead. Conditioning on external information tends to improve the forecast accuracy in some instances but typically only for those series where the unconditional forecasts are already quite accurate. For around a third of the variables under consideration, the differences in forecast accuracy between conditional and unconditional forecasts are statistically significant for density forecasts; for point forecasts on the other hand we find no significant differences. From a methodological point of view, this paper proposes precision-based sampling algorithms to draw from the predictive density - unconditional or conditional

---

on a subset of the system variables - in factor models and other models with unobserved components. Simulations show that these algorithms perform favorably compared to Kalman filter-based alternatives typically used in the literature.

Keywords: factor models, conditional forecasting, precision-based sampling

JEL classification: C11, C53, C55, E37.

III.1 Introduction

Factor models feature prominently in macroeconomic now- and forecasting applications. Typically, however, forecast evaluations in this literature focus only on a small subset of the large number of time series that are used to estimate these models. For example, Stock and Watson (2002) evaluate the forecast performance of industrial production, personal income, manufacturing and trade sales and (nonagricultural) employment in the United States. The factor-model based nowcasting literature has largely focused on predicting the real gross domestic product (GDP) (see Giannone et al., 2008; Bok et al., 2018; Schumacher and Breitung, 2008; Kuzin et al., 2013). While arguably the single most important measure of economic activity, analysts or policy-makers are typically also interested in forecasts for many other variables, e.g. concerning the labor market, price measures including deflators, components of GDP or activity measures like industrial production or turnover. Moreover, forecasts of these variables given a path of future real GDP growth and CPI inflation may be used in scenario analysis or as consistency checks to other model-based forecasts. Such conditional forecasts have received little attention in the factor model literature.

This paper therefore extends the literature on factor models by broadening the cross-sectional dimension of the forecast evaluation to include series that are not commonly featured. We consider a total of 37 time series comprising labor market indicators, price measures such as the CPI or PPI as well components of gross domestic product such as consumption, investment and exports as well gross value added of different sectors like construction or manufacturing. In addition to an evaluation of unconditional forecasts for such variables, we also investigate how the predictive accuracy changes when external information in the form of professional forecasters' views on activity and inflation is incorporated. Such an evaluation captures both how well the model describes the joint dynamics of the data as well as the accuracy of the conditioning information and thus gives a complete assessment of the forecast accuracy for the variables under consideration in this study.

Specifically, in the empirical application we rely on external information from the Reuters Poll of professional forecasters which collects the forecasts from private sector institutions and research institutes on a quarterly basis. We supplement these historical forecasts with a large real-time dataset consisting of roughly 80 time series that allows us to exactly replicate the information available at the time the forecast survey was conducted. We then estimate factor models with Bayesian methods to obtain forecasts both unconditional as well as conditional on the GDP growth and CPI inflation forecasts from the Reuters Poll. This allows us to assess how predictive accuracy improves both in terms of point and density forecast accuracy when external information is included.

As such, we also contribute to the literature on conditional forecasting which has mostly focussed on Bayesian vectorautoregressions - another class of popular macroeconmetric methods. Banbura et al. (2015) consider conditional forecasts by conditioning on future values they shed light on the stability of dynamic relationships in the Euro area. Similarly, in their empirical application Clark and McCracken (2017) condition on realizations when

testing for bias, efficiency and MSE accuracy of conditional versus unconditional forecasts.

Closer to the questions guiding our research are Krüger et al. (2017), Tallman and Zaman (2020) and Ganics and Odendahl (2021) who combine model-based density forecasts with external information in the form of forecast surveys. Using entropic tilting, these authors tend to find gains in predictive accuracy even for those variables that are not directly tilted to match the survey forecasts. However, common to both papers is the focus on only a few time series within a Bayesian vector autoregression context.[8] Our analysis differs in that we work with factor models rather than vector autoregressions, condition on quarterly point forecasts for GDP growth and CPI inflation and consider the impact on a much broader and more disaggregated range of variables.

Besides the empirical contribution, this paper also proposes precision-based sampling algorithms for conditional forecasting and scenario analysis. Precision-based sampling algorithms (Chan and Jeliazkov, 2009; McCausland, 2012) that build on the seminal work by Rue (2001) on Gaussian Markov random fields, are increasingly used in macroeconomic applications. Hauber and Schumacher (2021) generalize these precision-samplers to applications with missing data. Since future values of variables can simply be considered as missing, we show how draws from the predictive density - unconditional or conditional on future values of a subset of the system's variables - can be obtained. In a simulation study we compare the performance of these algorithms to the Kalman-filter based implementations proposed by Banbura et al. (2015).

The results of the forecast evaluation show that for the majority of time series under consideration the factor model produces forecasts that are as accurate as an autoregressive benchmarks, with relative root mean squared forecast errors (RMSFE) and continous ranked probability scores (CRPS) ranging from 0.9 to 1.1. For a few variables we find much larger gains compared to the benchmark and these persist as the forecast horizon increases. When we incorporate external information in the form of GDP growth and CPI inflation forecasts from a survey of professional forecasters, we generally find some improvements in forecast accuracy for the variables of interest from categories such as real activity, prices and the labor market. Only for a smaller number of series, however, do we find that conditioning produces large gains and these are typically the series for which the model already produces accurate unconditional forecasts. While the point and density forecast accuracy is broadly comparable in the sense that relative gains compared to an autoregressive benchmark for a given variable are similar, we find that when evaluating the entire predictive distribution the loss from the condtional forecasts is significantly lower than that of the model's unconditional predictions. For point forecasts, there is virtually no such evidence.

The remainder of this paper is structured as follows: Section III.2 illustrates the precision-based sampling algorithms for conditional forecasting, while in Section III.3 we discuss the datatset including the forecast survey used to condition the factor model forecasts as well as the evaluation set-up. Section III.4 discusses the results and several robustness checks

---

[8]The latter two also consider medium-sized VAR with 10 and 14 variables, respectively as a robustness check and find similar results to their smaller baseline models.

while Section III.5 concludes.

III.2 Precision-based sampling algorithms for conditional forecasting

In this section, we lay out the precision-based sampling algorithms to obtain conditional forecasts. To fix notation and ideas, we first review the general precision-based sampler in Section III.2.1 before turning to forecasting applications (Section III.2.2). Lastly, we also discuss how repeated draws from the predictive density given parameters can be obtained (Section III.2.3).

III.2.1 Precision-based sampling

We outline the precision-based algorithms through the lens of a dynamic factor model but the analysis can be generalized to other state space models with unobserved components. Factor models feature prominently in the macroeconomic literature. For an overview of their methodological developments and applications, see Stock and Watson (2016). Precision-based sampling applications include Chan and Jeliazkov (2009); McCausland (2015); Kaufmann and Schumacher (2017, 2019).

A *dynamic factor model* is defined as

$$\mathbf{y}_t = \sum_{k=0}^{K} \lambda_k \mathbf{f}_{t-k} + \mathbf{e}_t \tag{III.1a}$$

$$\mathbf{f}_t = \sum_{p=1}^{P} \phi_p \mathbf{f}_{t-p} + v_t \tag{III.1b}$$

$$\mathbf{e}_t = \sum_{j=0}^{J} \psi_j \mathbf{e}_{t-j} + \epsilon_t \tag{III.1c}$$

where $\mathbf{f}_t$ denotes an $R \times 1$ vector of unobserved factors which summarizes the co-movement of the observables $\mathbf{y}_t$. The dynamics of the factors are modelled as a vector autoregression of order P. The idiosyncratic components, $\mathbf{e}_t$ are modeled as independent autoregressive processes of order J, i.e. $\psi_1, \dots, \psi_J$ are diagonal matrices. The $N \times R$ loadings matrices λ_k capture the dynamic relationships between observables and factors. Setting $K = 0$, so that the variables only load contemporaneously on the factors, yields a *static factor model*. The innovations ϵ_t, v_t are Normal, uncorrelated at all leads and lags and their covariance matrices given by Ω and Σ, respectively.

Stacking the observables, factors and idiosyncratic components over t yields

$$\mathbf{y} = \mathbf{M}\mathbf{f} + \mathbf{e} \tag{III.2}$$

where

$$\mathbf{M} = \begin{bmatrix} \lambda_0 & & & & & & & \\ \lambda_1 & \lambda_0 & & & & & & \\ \lambda_2 & \lambda_1 & \lambda_0 & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & \lambda_K & \dots & \lambda_1 & \lambda_0 & & & \\ & & \ddots & \ddots & \ddots & \ddots & & \\ & & & \lambda_K & \dots & \lambda_1 & \lambda_0 \end{bmatrix}.$$

The vector of idiosyncratic components can be written as $\mathbf{H}_e\,\mathbf{e} = \varepsilon$ with

$$\mathbf{H}_e = \begin{bmatrix} I_N & & & & & \\ \text{-}\psi_1 & I_N & & & & \\ \text{-}\psi_2 & \text{-}\psi_1 & I_N & & & \\ & & & & & \\ \text{-}\psi_J & \dots & \dots & \text{-}\psi_1 & I_N & \\ & & & & & \\ & & \text{-}\psi_J & \dots & \dots & \text{-}\psi_1 & I_N \end{bmatrix}$$

and $\varepsilon \sim \mathcal{N}(0_{NT}, I_T \otimes \mathbf{\Omega})$. It follows that $\mathbf{e}$ is Normal, with mean 0 and covariance matrix $V_e = (\mathbf{H}_e^\mathsf{T}(I_T \otimes \mathbf{\Omega}^{-1})\mathbf{H}_e)^{-1}$.

In a similar fashion, the vector of factors is distributed as

$$\mathbf{f} \sim \mathcal{N}\left(0, (\mathbf{H}_\mathsf{f}^\mathsf{T}(I_T \otimes \mathbf{\Sigma}^{-1})\mathbf{H}_\mathsf{f})^{-1}\right)$$

with

$$\mathbf{H}_\mathsf{f} = \begin{bmatrix} I_N & & & & & \\ \text{-}\phi_1 & I_N & & & & \\ \text{-}\phi_2 & \text{-}\phi_1 & I_N & & & \\ & & & & & \\ \text{-}\phi_P & \dots & \dots & \text{-}\phi_1 & I_N & \\ & & & & & \\ & & \text{-}\phi_P & \dots & \dots & \text{-}\phi_1 & I_N \end{bmatrix}.$$

Define

$$\mathbf{z} := \begin{bmatrix} \mathbf{f} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} I_{TR} & 0 \\ \mathbf{M} & I_{TN} \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \mathbf{e} \end{bmatrix}. \tag{III.3}$$

It follows that $\mathbf{z}$ is Normal with mean 0 and covariance matrix $\mathbf{V}_z$. The corresponding precision matrix is the inverse of the covariance matrix:

$$
\begin{aligned}
\mathbf{Q}_z &= \mathbf{V}_z^{-1} \\
&= \mathrm{Var}\left(\begin{bmatrix} \mathbf{f} \\ \mathbf{y} \end{bmatrix}\right)^{-1} \\
&= \left(\begin{bmatrix} I & 0 \\ \mathbf{M} & I \end{bmatrix}\begin{bmatrix} \mathbf{H}_f^{-1}\mathbf{V}_v\mathbf{H}_f^{-1\mathsf{T}} & 0 \\ 0 & V_e \end{bmatrix}\begin{bmatrix} I & \mathbf{M}^\mathsf{T} \\ 0 & I \end{bmatrix}\right)^{-1} \\
&= \begin{bmatrix} I & \text{-}\mathbf{M}^\mathsf{T} \\ 0 & I \end{bmatrix}\begin{bmatrix} \mathbf{H}_f^{\mathsf{T}}\mathbf{V}_v\mathbf{H}_f & 0 \\ 0 & V_e \end{bmatrix}\begin{bmatrix} I & 0 \\ \text{-}\mathbf{M} & I \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{H}_f{}^\mathsf{T}\mathbf{V}_v\mathbf{H}_f + \mathbf{M}^\mathsf{T}V_\varepsilon\mathbf{M} & \text{-}\mathbf{M}^\mathsf{T}V_e \\ \text{-}V_e\mathbf{M} & V_e \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{Q}_f & \mathbf{Q}_{fy} \\ \mathbf{Q}_{fy}^\mathsf{T} & \mathbf{Q}_y \end{bmatrix}
\end{aligned}
$$

Bayesian estimation of Equations (III.1a)–(III.1c) requires a draw from the distribution of the factors conditional on the observations and parameters Θ, $p(\mathbf{f}_1,\ldots,\mathbf{f}_T|\mathbf{y}_1,\ldots,\mathbf{y}_T,\Theta)$. "Simulation smoothing" methods based on the Kalman filter/smoother have been proposed by Carter and Kohn (1994) and Durbin and Koopman (2001) and are widely used in the literature. Precision-based sampling algorithms (Chan and Jeliazkov, 2009; McCausland, 2012) provide an alternative to these methods.

Using standard results for multivariate Normal distributions (e.g. Bishop, 2006, pp. 86-87), we can derive the desired conditional distribution:

$$
\begin{aligned}
p\left(\mathbf{f} \mid \mathbf{y},\Theta\right) &\sim \mathcal{N}\left(\mathbf{Q}_f^{-1}\left(\mathbf{Q}_f\boldsymbol{\mu}_f - \mathbf{Q}_{fy}\left(\mathbf{y} - \boldsymbol{\mu}_y\right)\right),\mathbf{Q}_f^{-1}\right) \\
&:= \mathcal{N}\left(\boldsymbol{\mu}_{f|y},\mathbf{Q}_{f|y}^{-1}\right)
\end{aligned}
$$

with (see also Chan and Jeliazkov, 2009, eqn. 6-8):

$$
\begin{aligned}
\mathbf{Q}_{f|y} &= \mathbf{Q}_f \\
&= \mathbf{H}^\mathsf{T}\mathbf{V}_v^{-1}\mathbf{H} + \mathbf{M}^\mathsf{T}\mathbf{V}_\epsilon^{-1}\mathbf{M} \\
\boldsymbol{\mu}_{f|y} &= \mathbf{Q}_f^{-1}\mathbf{Q}_f\tilde{\mathbf{c}} + \mathbf{M}^\mathsf{T}\mathbf{V}_\epsilon^{-1}\mathbf{y}
\end{aligned}
$$

Draws from this distribution can be obtained efficiently since the $T \cdot S \times R$ matrix $\mathbf{Q}_f$ is banded and its inversion not required. Specifically, calculate the conditional mean following Rue and Held (2005, Algorithm 2.1). This requires the computation of the lower Cholesky factor $\mathbf{L}$ of $\mathbf{Q}_f$. Then draw $\mathbf{v} \sim \mathcal{N}(0, I_{TR})$, solve $\mathbf{L}^\mathsf{T}\mathbf{x} = \mathbf{v}$ for $\mathbf{x}$ and set $\boldsymbol{\mu}_{f|y} + \mathbf{x}$, yielding a draw of $\mathbf{f}$ conditional on $\mathbf{y}$ (Rue and Held, 2005, Algorithm 2.4).

III.2.2 Forecasting

In addition to a draw of the factors conditional on the data, macroeconomic applications may require forecasts - both unconditional and conditional on future paths of a subset of the observables - from the model in Equations (III.1a)–(III.1c). Examples include projections of inflation which take assumptions regarding the macroeconomic environment into account (Giannone et al., 2014) or external nowcasts as "jump-off" points for longer-term forecasts in reduced form or structural models (Faust and Wright, 2009; Wolters, 2015; Del Negro and Schorfheide, 2013); Knotek and Zaman (2019) show how macroeconomic forecasts can be improved by conditioning on nowcasts for financial variables.

Let $\mathbf{y}^o = [\mathbf{y}_1^\mathsf{T}, \ldots, \mathbf{y}_T^\mathsf{T}]^\mathsf{T}$ denote the available observations through T. For the forecast periods $h = 1 : H$, $\mathbf{y}_{c,t} \ \forall t = T + 1 : T + H$ are the observations which are being conditioned on. The entire conditioning set is given by

$$\mathbf{y}^c = [\mathbf{y}_{c,T+1}^\mathsf{T}, \ldots, \mathbf{y}_{c,T+H}^\mathsf{T}]^\mathsf{T}$$

with $N^c = \dim(\mathbf{y}^c)$. Conversely, $\mathbf{y}_{-c,t} = \mathbf{y}_t \notin \mathbf{y}_{c,t}$ denotes the observations for which no conditioning information is available at time t and which are being forecast: $\mathbf{y}^f = [\mathbf{y}_{-c,T+1}^\mathsf{T}, \ldots, \mathbf{y}_{-c,T+H}^\mathsf{T}]^\mathsf{T}$.

In what follows, we will not be concerned with the estimation of model parameters and assume that random draws from $p(\Theta|\mathbf{y}^o)$ are readily available. Also, for the sake of exposition, assume that between periods $t = 1$ and $t = T$ all observations are available. Situations where this is not the case, e.g. because some series start at a later date in the sample or because outlying observations have been removed by setting them to missing, can naturally be accommodated.

From a Bayesian perspective, forecasting boils down to sampling from the predictive density given by (see e.g. Geweke and Whiteman, 2006)

$$p(\mathbf{y}^f|\mathbf{y}^o, \mathbf{y}^c) \propto \int_\Theta p(\mathbf{y}^f|\mathbf{y}^o, \mathbf{y}^c, \Theta)p(\Theta|\mathbf{y}^o, \mathbf{y}^c) \, d\Theta. \tag{III.4}$$

Note that this density takes into account that the conditioning set $\mathbf{y}^c$ contains information about which Θ are more likely a posteriori. Draws from this density can be obtained by applying the insights in Hauber and Schumacher (2021) who generalize the precision sampler outlined above to applications with missing data. Since future values of the variables can simply be considered as missing, draws from the joint distribution of states and forecasts conditional on the data (and possibly the future paths of a subset of the variables) can be sampled efficiently.

To evaluate Equation III.4, derive the joint distribution of states and observations from $t = 1 : T + H$:

$$\mathbf{z} = [\mathbf{f}^\mathsf{T}, \mathbf{y}^\mathsf{T}]^\mathsf{T}$$
$$= [\mathbf{f}_1, \ldots, \mathbf{f}_{T+H}, \mathbf{y}_1, \ldots, \mathbf{y}_{T+H}]^\mathsf{T}$$

It follows that the distribution of $\mathbf{z}$ is Normal with mean $\boldsymbol{\mu}_z = 0$ and precision matrix $\mathbf{Q}_z$ as in Section III.2.1. To sample from the conditional distribution $p\left(\mathbf{y}^f, \mathbf{f} \mid \mathbf{y}^o, \mathbf{y}^c, \Theta\right)$ requires a permutation of $\mathbf{z}$. To this end, define the permutation matrix $\mathcal{P}$ that reorders $\mathbf{z}$ such that the states and elements of $\mathbf{y}^f$ - ordered by time period t - are placed first. That is to say,

$$\mathbf{z}_{\mathcal{P}} := \mathcal{P}\mathbf{z} = [\mathbf{f}_1^\mathsf{T}, \ldots, \mathbf{f}_T^\mathsf{T}, \mathbf{f}_{T+1}^\mathsf{T}, \mathbf{y}_{\text{-}c,T+1}^\mathsf{T}, \ldots, \mathbf{f}_{T+H}^\mathsf{T}, \mathbf{y}_{\text{-}c,T+H}^\mathsf{T}, \mathbf{y}_1^\mathsf{T}, \ldots, \mathbf{y}_{T+H}^\mathsf{T}, \mathbf{y}_{c,T+1}^\mathsf{T}, \ldots, \mathbf{y}_{c,T+H}^\mathsf{T}]^\mathsf{T}.$$

Note that the conditioning arguments $\mathbf{y}^o, \mathbf{y}^c$ are now ordered last. The permuted vector of states and observations is also Normal with (permuted) moments given by

$$\mathbf{z}_{\mathcal{P}} \sim \mathcal{N}\left(\mathcal{P}\boldsymbol{\mu}_z, \left(\mathcal{P}\mathbf{Q}_z\mathcal{P}^\mathsf{T}\right)^{-1}\right)$$

where we have made use of the fact that the inverse of a permutation matrix is equal to its transpose. The precision matrix of $\mathbf{z}_{\mathcal{P}}$ can be partitioned as follows:

$$\begin{aligned}
\mathbf{Q}_{z_{\mathcal{P}}} &:= \mathcal{P}\mathbf{Q}_z\mathcal{P}^\mathsf{T} \\
&= \begin{bmatrix} \mathbf{Q}_{\mathrm{fy^f}} & \mathbf{Q}_{\mathrm{fy^f},y^o} \\ \mathbf{Q}_{\mathrm{fy^f},y^o}^\mathsf{T} & \mathbf{Q}_{y^o} \end{bmatrix}
\end{aligned}$$

where $\mathbf{Q}_{\mathrm{fy^f},y^o}$ is a $T \cdot R + H \cdot N \times T \cdot N$ matrix and $\mathbf{Q}_{\mathrm{fy^f}}$ and $\mathbf{Q}_{y^o}$ are square with dimensions $T \cdot R + H \cdot N$ and $T \cdot N$, respectively.

Similarly, the mean of $\mathbf{z}_{\mathcal{P}}$ can be partitioned as

$$\boldsymbol{\mu}_{z_{\mathcal{P}}} = \begin{bmatrix} \boldsymbol{\mu}_{\mathrm{fy^f}} \\ \boldsymbol{\mu}_{y^o} \end{bmatrix}$$

Denote by $\mathbf{z}^*$ a draw from the conditional distribution (surpressing the dependence on the parmeters Θ for the sake of readability)

$$\mathbf{z}_{\mathrm{fy^f}} \mid \mathbf{y}^o, \mathbf{y}^c \sim \mathcal{N}\left(\mathbf{Q}_{\mathrm{fy^f}}^{-1}\left(\mathbf{Q}_{\mathrm{fy^f}}\boldsymbol{\mu}_{\mathrm{fy^f}} - \mathbf{Q}_{\mathrm{fy^f},y^o}\left(\mathbf{y}^o - \boldsymbol{\mu}_{y^o}\right)\right), \mathbf{Q}_{\mathrm{fy^f}}\right).$$

Then by reversing the permutation, i.e

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{T+H} \end{bmatrix} = \mathcal{P}^{-1} \begin{bmatrix} \mathbf{z}^* \\ \mathbf{y}^o \\ \mathbf{y}^c \end{bmatrix}$$

we can back out the draws of $\mathbf{f}$ and $\mathbf{y}^f = [\mathbf{y}_{\text{-}c,T+1}, \ldots, \mathbf{y}_{\text{-}c,T+H}]$.

Note that by having placed those components of $\mathbf{z}$ which are conditionally dependent close to each other, the bandwidth of $\mathbf{Q}_{\mathrm{fy^f}}$ is kept small. Figure III.1 shows the precision matrix of $\mathbf{z}_{\mathcal{P}}$ for a factor model with $T = 100$, $N = 100$ and $R = 10$, a forecast horizon of

$H = 20$ and conditioning on the first 10 variables[9], i.e.

$$\mathbf{y}^c = [y_{1,T+1}, \ldots, y_{10,T+1}, \ldots, y_{1,T+20}, \ldots, y_{10,T+20}].$$

The precision matrix $\mathbf{Q}_{f y^f}$ is highlighted by the square.

Figure III.1: Precision precision matrix in the case of a dynamic factor model
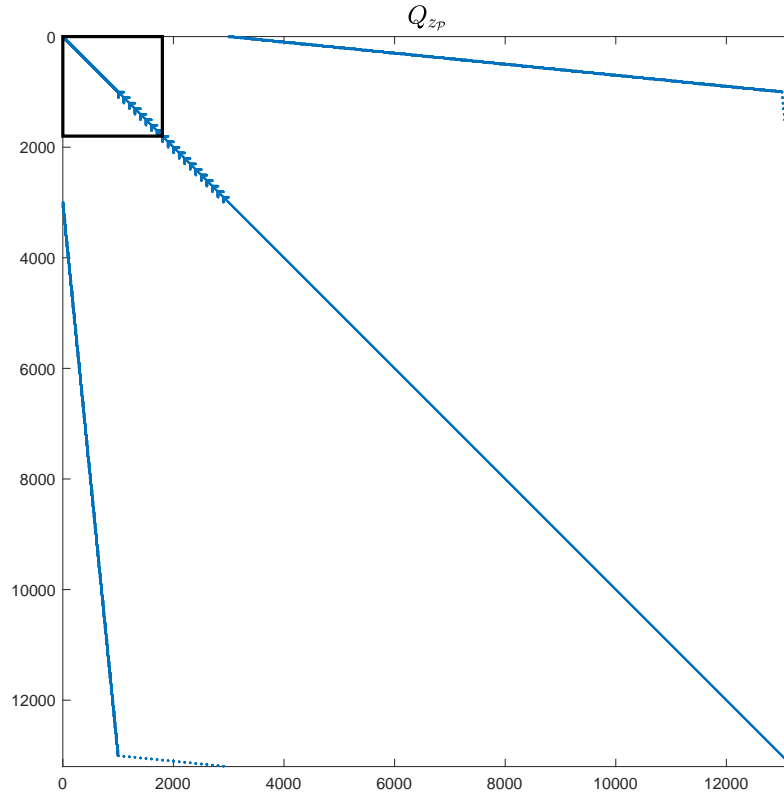
Note: The figure shows the precision matrix of the permuted vector of observations and states $\mathbf{z}_{\mathcal{P}}$ for a factor model with $T = 100$, $N = 100$ and $R = 10$. Highlighted is the $(T + H) \cdot R + N^c \times (T + H) \cdot R + N^c$ partition corresponding to $\mathbf{z}_{f y^f}$.

Similar to Waggoner and Zha (1999, Algorithm 1) and Banbura et al. (2015, p.745), draws from the predictive density $p(\mathbf{y}^f | \mathbf{y}^o, \mathbf{y}^c)$ can then be obtained by the following algorithm:

---

[9]For the sake of illustration we set $P = 1$, $J = 0$, $S = 0$.

Algorithm 1: Draws from the predictive density $p(\mathbf{y}^f|\mathbf{y}^o, \mathbf{y}^c)$

Initialize Θ^0 from $p(\Theta|y^o)$. For $g = 1 : (G_0 + G)$

1. Construct the system matrices $\mathbf{H}, \mathbf{M}, \mathbf{V}_\varepsilon, \mathbf{V}_\upsilon, \tilde{\mathbf{c}}, \mathbf{d}$ given the parameters $\Theta^{(g-1)}$

2. Compute the mean $\boldsymbol{\mu}_z$ and precision matrix $\mathbf{Q}_z$ of the joint distribution of $\mathbf{z} = [\boldsymbol{\alpha}, \mathbf{y}]$

3. Permute $\mathbf{z}$, yielding $\mathbf{z}_{\mathcal{P}'} \sim \mathcal{N}(\mu_{z'_\mathcal{P}}, Q_{z'_\mathcal{P}}^{-1})$ with partitions $[\mathbf{z}_{\alpha y^f}^\mathsf{T}, \mathbf{y}^o, \mathbf{y}^c]^\mathsf{T}$

4. Sample from the conditional distribution $\mathbf{z}_{\alpha y^f}|\mathbf{y}^o, \mathbf{y}^c, \Theta^{(g-1)}$ using (Rue and Held, 2005, Algorithm 2.1 and 2.4) and reverse the permutation to back out the draws of $\mathbf{y}^f$ and $\boldsymbol{\alpha}$

5. Draw $\Theta^{(g)}$ from $p(\Theta|\mathbf{y}^o, \mathbf{y}^c, \mathbf{y}^{f(g)}, \alpha^{(g)})$

Discard the first G_0 draws as burn-in.

If the sample information and prior dominate the information in $\mathbf{y}^c$, the impact of the conditioning arguments on the posterior distribution of the parameters is likely to be negligible and $p(\Theta|\mathbf{y}^o, \mathbf{y}^c) \approx p(\Theta|\mathbf{y}^o)$ (Del Negro and Schorfheide, 2013; Banbura et al., 2015). In such applications, draws from Equation III.4 can then be obtained by skipping step 5 at each iteration in **Algorithm 1**. Similarly, when $\mathbf{y}^c$ is empty, Step 5 can obviously be skipped as well and the algorithm returns unconditional forecasts, i.e. draws from the density $p(\mathbf{y}^f|\mathbf{y}^o)$.

While the discussion above has focussed on dynamic factor models, it can easily be extended to general state space models of the form:

$$\mathbf{y}_t = Z\boldsymbol{\alpha}_t + \varepsilon_t, \ \varepsilon_t \sim \mathcal{N}(0, H) \tag{III.5a}$$

$$\boldsymbol{\alpha}_{t+1} = T\boldsymbol{\alpha}_t + \eta_t, \ \eta_t \sim \mathcal{N}(0, Q) \tag{III.5b}$$

by setting $P = 1, J = S = 0$.

Lastly, note that Hauber and Schumacher (2021) also propose a sequential sampler that first draws the missing observations conditional on the states and then the states given a complete data set. This approach performs somewhat better in terms of runtime and could in principle be used in Algorithm 1, which in any case requires a Gibbs sampling step if the parameters are updated given the conditioning set. However, the gains in performance from faster sampling of states and forecasts may not outweigh the longer burn-in required due to the additional block in the Gibbs Sampler.

III.2.3 Soft conditions and repeated samples from $p(\mathbf{y}^f|\mathbf{y}^o, \Theta^{(g)})$

So far we have considered conditional forecasts that fix the future values of some of the endogenous variables at single points. Waggoner and Zha (1999) labels these "hard" conditions to distinguish them from soft conditions where each element of $\mathbf{y}^c$ is merely restricted to lie within a pre-specified range, i.e. $\mathbf{y}_l^f \leq \mathbf{y}^f \leq \mathbf{y}_u^f$. Draws from the unconditional predictive density $p(\mathbf{y}^f|\mathbf{y}^o, \Theta^{(g)})$ can be obtained until the conditions are satisfied.[10]

The precision-sampling algorithms outlined above are particularly suited to this task since the bottleneck in the calculations is the Cholesky factorization of the band matrix $\mathbf{Q}_{f\mathbf{y}^f}$ which requires $((T + H)R + NH) \cdot b_w^2$ floating point operations where b_w is the lower bandwidth of $\mathbf{Q}_{f,\mathbf{y}^f}$ (Golub and Van Loan, 2013, 4.3.5). However, it only needs to be computed once, irrespective of the number of desired draws (Rue, 2001). Repeated samples thus come at a significantly reduced computational burden, consisting only of $(T + H)S + NH$ independent draws from the standard Normal distribution, band backward substitution - which requires $2((T + H)S + NH) \cdot b_w$ floating point operations (Golub and Van Loan, 2013, 4.3.2)- and a vector addition.

III.2.4 Comparison with Kalman filter-based simulation smoothers

In simulations to determine the computational efficiency of the algorithms outlined above in comparison to Kalman filter-based simulation smoothers, we find that the former is computationally more efficient in terms of runtime for both large and small factor models and forecast horizons. The gains in computational efficiency are particularly pronounced for soft conditioning, where repeated draws from the predictive density are required. See Appendix III.C for details of the simulation study.

III.3 Real-time evaluation of unconditional and conditional forecasts

In this section, we present the data used in the forecast evaluation and provide details on the Bayesian estimation. Furthermore, we discuss the set-up of the real-time evaluation.

III.3.1 Data

Real-time dataset for the German economy

To estimate the factor model, we construct a large quarterly real-time dataset covering different aspects of the macroeconomy. Beyond real gross domestic product, the main building

---

[10]Similarly, hard and soft conditions can be combined, e.g. if inflation forecasts are produced conditional on an oil price path and subject to the restriction that the Federal Funds rate is positive. In such a scenario, draws from $p(\mathbf{y}^f|\mathbf{y}^o, \mathbf{y}^c, \Theta^{(g)})$ are obtained until the soft restrictions are satisfied.

block of the dataset are other series from the national accounts such as the entire expenditure-side components, e.g. private and public consumption, gross fixed capital formation, export and imports. From the production side of GDP we can only include gross value added in the industrial and construction sector as well as trade, transport and hospitality as real-time vintages for other sectors are only available from 2011 onwards.[11]

In addition to the chained volume indices we also include the corresponding deflators of the expenditure and production side components. Furthermore, our dataset contains series on real activity such as industrial production, turnover or orders as well as construction, headline CPI and PPI indices as well as the corresponding core indices and a measure of prices in the construction sector as well as labor market indicators such as economy-wide employment and wages and hours worked in the industrial and construction sector.

In addition, data are supplemented with financial indicators such as interest rates, exchange rates and stock market indices. Lastly, we also include a few survey-based indicators provided by the European Commission (ESI) and covering the industrial, construction and services sectors as well as a measure of firms' employment expectations (EEI). Note that for the latter two sources, there is no real-time problem, as financial indicators are not revised subsequent to their original publication. Similarly, the European Commission's survey indicators are only revised due the seasonal adjustment procedure. That is to say, the survey responses for May 2006 as reported in 2018 might differ (slightly) from those originally published as the seasonal factors have since been updated. While this violates the real-time assumption, such revisions are likely to be small. We therefore abstract from these subsequent changes in what follows and rely on the seasonally adjusted indices.

Altogether, our dataset contains 57 series. All of them are transformed to stationarity prior to estimaton. In most cases, this involves taking the first difference of the logarithm of the original series, though in some cases, e.g. interest rates or survey indicators, we take simple differences.[12] Details of the time series, their sources as well as the transformation employed can be found in Appendix III.A.

Professional forecasts from the Reuters Poll

As conditioning information, we rely on the Reuters Poll of forecasts for German GDP and CPI inflation. The survey compiles the views of around twenty different professional forecasters from the private sector and research institutes. It is conducted once a quarter during the first month, i.e. January, April, July and October. Until mid-2014, the respondents submit their views during the first few days of the month; from then on the survey date is around the middle of the month.[13]

For GDP, the forecasts are given in terms of quarter-on-quarter (q/q) growth rates and we can thus use the median forecast directly to obtain conditional forecasts from the model out-

---

[11]The three mentioned sectors comprise roughly 50 percent of total gross value added.

[12]The Deutsche Bundesbank's real-time database includes vintages of chained volume indices and nominal values for the national accounts. The implicit deflators are calculated by dividing the nominal series with the chained volume index and taking the first differenc of the logarithm of the resulting series.

[13]A complete list of the dates when the forecast polls were conducted is available upon request.

lined above. The inflation forecasts, however, are reported as year-over-year (y/y) growth rates of the consumer price index. To bring the forecasts in line with the treatment of CPI inflation in the model, we have to transform the individual forecasts. Specifically, we interpret the y/y forecasts of inflation, $x_t^{y/y}$, as describing the four-quarter change in the quarterly consumper price index X, i.e. $x_t^{y/y} = log(X_t/X_{t-4})$. Thus, given a sequence of y/y forecasts we can back out the implicit forecasts of the (log of the) consumer price index and derive the corresponding q/q forecasts, i.e. $x_t^{q/q} = log(X_t/X_{t-1})$ for each participant in the survey. We then use the median of these transformed values in the conditional forecasting exercise.

Figure III.2: GDP growth and CPI inflation forecasts from the Reuters Poll of professional forecasters
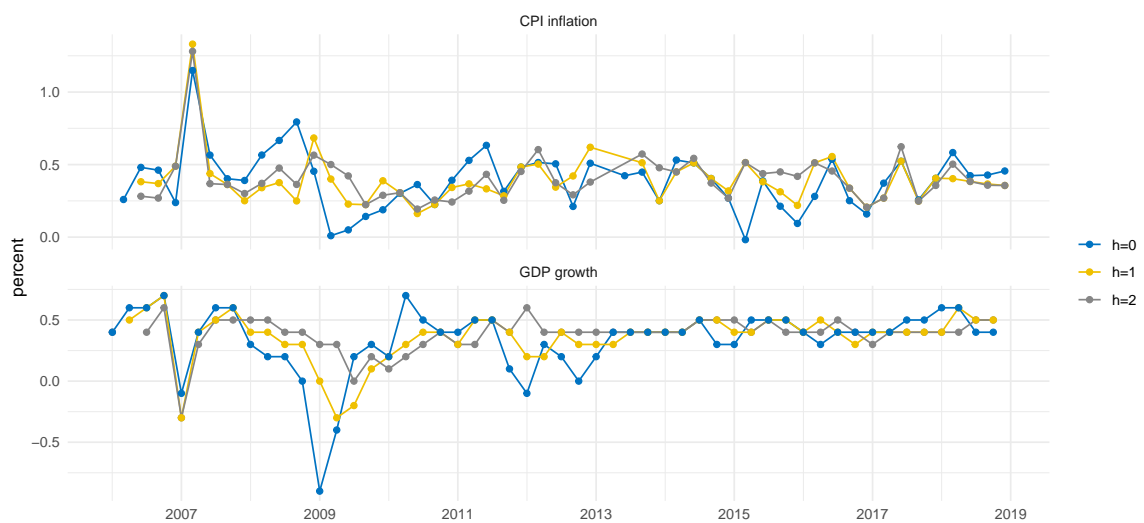
Note: The figure shows the median forecast from the Reuters Poll of professional forecasters for quarter-on-quarter change in the comsumer price index (top panel) and quarter-on-quarter GDP growth. The forecast horizon h is in quarters and relative to the reference period. For details, see the main text. *Source*: Thomson Reuters, author's calculations.

Figure III.2 plots the median forecast of q/q CPI inflation (top row) and GDP growth (bottom row). The lines correspond to different forecast horizons relative to the reference period and are given in quarters. That is to say, the $h = 1$ forecast for the first quarter in 2010 was made at the start of October 2009.

III.3.2 Estimation

The parameters of the factor model outlined above are estimated with Bayesian methods. Draws from the joint posterior distribution of factors and parameters are obtained via a Gibbs Sampler which alternately draws from the conditional posterior of the factors given the parameters and *vice versa*. For the estimation, we set $J = 1$ and $S = 0$ and $R = 2$ and run the Gibbs Sampler for a total of 20000 iterations. We discard the first 10000 as burn-in and of the remaining, store every second draw, yielding 5000 from the posterior distribution of the parameters, $p(\Theta|y)$. As the priors used in the estimation as well as the resulting

conditional posterior distributions are standard in the literature, we relegate a discussion to Appendix III.B.

III.3.3 Evaluation set-up

Given a vintage reflecting the data available to forecasters at the time the Reuters Poll was conducted and corresponding draws from the posterior distribution $p(\Theta|y)$, we generate draws from the predictive density as described in Section III.2.1 for all national accounts, price, labor market and activity indicators. These forecasts are unconditional as well as conditional on the values from the Reuters Poll for GDP growth and CPI inflation for the current and following two quarters, i.e. $h = \{0, 1, 2\}$. The evaluation sample starts in 2006 from which point on real-time vintages are available on a broad basis. It ends in the fourth quarter of 2017 as afterwards the number of missing entries in the Reuters Poll increases considerably.

Within the sample, there are also a few data irregularities. For example, in the third quarter of 2013 no CPI inflation forecasts were recorded in the Reuters Poll. We therefore exclude 2012Q3, 2012Q4 and 2013Q1 from the evaluation. Note also that for two labor market series there are some vintages which are released outside of the regular publication calendar. For example, the employment vintage scheduled for release at the end of December was not published until the first week of January, most likely due to the original publication date falling on a bank holiday. As such, when the professional forecasters were asked by Reuters to submit their forecasts by January 1st, the latest available information for this series was published at the end of November. For hours worked in the manufacturing sector, all vintages between March 3rd 2009 and March 4th 2010 are missing. As these irregularities alter the information set and effective forecast horizon compared to the other quarters, we do not consider the affected vintages for these two series. For the remaining series, however, we have a total of 41 quarters to be evaluated for all horizons.

To benchmark the performance of the different priors, we consider a simple univariate Bayesian autoregressive model of order 2 for all the series. i.e.

$$y_{i,t} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 y_{i,t-2} + \epsilon_t; \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{III.6}$$

The parameters $\beta = [\beta_0, \beta_1, \beta_2]$ and σ^2 are estimated via a Gibbs Sampler with diffuse priors given by $\beta \sim \mathcal{N}(0, 10 \cdot I_3)$ and $\sigma^2 \sim \mathcal{G}^{-1}(3, 0.01)$. Draws from the h-step ahead predicitve density are obtained by plugging in draws from the posterior distribution of the coefficients and iterating Equation III.6 forward.

Both the point and density forecast performance is evaluated. For the former, we compute the root mean squared forecast error (RMSFE) defined as

$$\text{RMSFE} = \left(\frac{1}{S} \sum_{s=1}^{S} (\hat{y}_{i,T+s}|\Omega_v - y'_{i,T+s})^2 \right)^{\frac{1}{2}}$$

for a sequence of S nowcasts, where $\hat{y}_{i,T+s}|\Omega_v$ is the mean of the predictive density of variable i at time $T+s$, conditional on the information set Ω_{v_s} available in real-time at date v_s when the forecast for period $T+s$ was made. The corresponding realization, given by the first release, is denoted by $y'_{i,T+s}$. Density forecasts are evaluated by the continuous ranked probability score (CRPS) which is given by (see Krüger et al., 2016):

$$\text{CRPS} = \frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{G}\sum_{k=1}^{G}\left|y_{i,T+s}^{(k)}|\Omega_v - y'_{i,T+s}\right| - \sum_{k=1}^{G}\sum_{j=1}^{G}\left|y_{i,T+s}^{(k)}|\Omega_v - y_{i,T+s}^{(j)}|\Omega_v\right|\right).$$

To calculate the CRPS given draws from the predictive density as well a realization, we use the R package `scoringRules` (Jordan et al., 2019).

Finally, to compare the accuracy of the conditional and unconditional forecasts more formally, we follow Ganics and Odendahl (2021) and run Diebold-Mariano tests (Diebold and Mariano, 1995). Let $L_t(y_{i,t+h}^c)$ denote the loss associated with the conditional forecasts made at time for variable i and horizon h. In the case of point forecasts, this is simply given by the squared forecast error; for density forecasts by the corresponding CRPS. Similarly, define $L_t(y_{i,t+h}^u)$ as the loss for the unconditional forecast and $d_t = L_t(y_{i,t+h}^c) - L_t(y_{i,t+h}^u)$ the loss differential. The test statistic can then be written as

$$DM_{i,h} = \frac{\frac{1}{T}\sum_{t=1}^{T}d_t}{\sqrt{\hat{\sigma}_d^2}}$$

where $\hat{\sigma}_d^2$ is an estimate of the variance of d_t. We test the null hypothesis of equal predicitve ability against the one-sided alternative that the conditional forecasts yield a lower loss.

III.4   Results

In this section we present the results of the forecast evaluation. We begin with point forecast performance before looking at how the evaluation changes when we focus on the entire predictive distribution. Lastly, we consider several robustness checks like the evaluation period as well as the model size (i.e. the number of factors).

III.4.1   Point forecast accuracy

Turning to the peformance of the unconditional forecasts first, we find that for all horizons and the majority of variables, the factor model produces forecasts that perform similar to the autoregressive benchmark with relative RMSFE between 0.9 and 1.1 (Figure III.3). Outliers in this context are the headline and "core", i.e. excluding energy, producer price indices at $h = 0$ where the performanc is considerably worse than the benchmark. Conversely, for a few series there are noteworthy gains over the simple univariate AR(2) model and these persist or even increase for larger horizons. For example, for $h = 0$ the relative RMSFE for the construction price index and gross value added in the industrial sector is around

0.6. Smaller but still sizeable improvements can also be found for two series from the expenditure side of the national accounts: equipment investment as well as the deflator for residential investment. For larger h, these relative gains even increase: the relative RMSFE for the construction subcomponent of the PPI reaches 0.35; for gross value added in the industrial sector is 0.4. In addition to these variables, at $h = 2$ there also noteworthy improvements for series that are likely highly correlated to gross industrial value added like production in the industrial and construction sector as well as industrial turnover. Note, however, that these improvements are partly due to the fact that the performance of the benchmark generally worsens as the forecast horizon increases.

For comparison, the model's unconditional forecasts for GDP growth lead to substantial relative gains of up to 20 percent relative to the autoregressive benchmark. For $h = 0$, the professional forecasters are considerably more accurate (relative RMSFE of 0.65) but for larger horizons the performance is similar to the unconditional forecasts. The picture for CPI inflation is similar: the professional forecasters are more accurate than the model's unconditional forecast at $h = 0$ but by a much smaller margin; for $h = \{1, 2\}$, the forecasts from the Reuters Poll are considerably (slightly) worse than the autoregressive benchmark (the model's unconditonal) forecasts.

Focusing on how the forecast performance increases when we condition on professional forecasters' views on GDP growth and CPI inflation, we find that there are only large differences in the relative RMSFE between the unconditional and conditional forecasts for $h = 0$. Perhaps not surprisingly gains to conditioning arise for those series for which the models already produce decent forecasts (relative to autoregressive benchmarks) and are mainly concentrated to series from the national accounts group and - to some extent - indicators capturing real activity (but not part of the national accounts). For the labor market indicators and CPI and PPI inflation, we find much smaller gains from conditioning. Filled entries in the above plot correspond to those variables for which the null hypothesis of the Diebold-Mariano test can be rejected at the 5 percent level in favor of the altnerative that the loss of the conditional forecasts is lower. This is almost never the case for any of the 37 variables we consider irrespective of the forecast horizon (Table III.1). At $h = 0$, i.e. forecasts that are made in the reference quarter, we reject the null for the gross value added in the trade, transport and hospitality sector; at $h = 2$ for gross value added in the trade, transport and hospitality sector and employment.

III.4.2 Density forecast accuracy

By and large, the results obtained in terms of point forecast accuracy carry over when we evaluate the entire predictive density. In particular, the performance relative to the benchmark as well as the gains from conditioning as measured by the average CRPS are very similar (Figure III.4). Contrary to the findings for point forecasts, however, we do find significant differences between the conditional and unconditional forecast performance, at least for $h = \{0, 1\}$, again highlighted in the figure by the filled dots. For example, at $h = 0$

Figure III.3: Point forecast evaluation
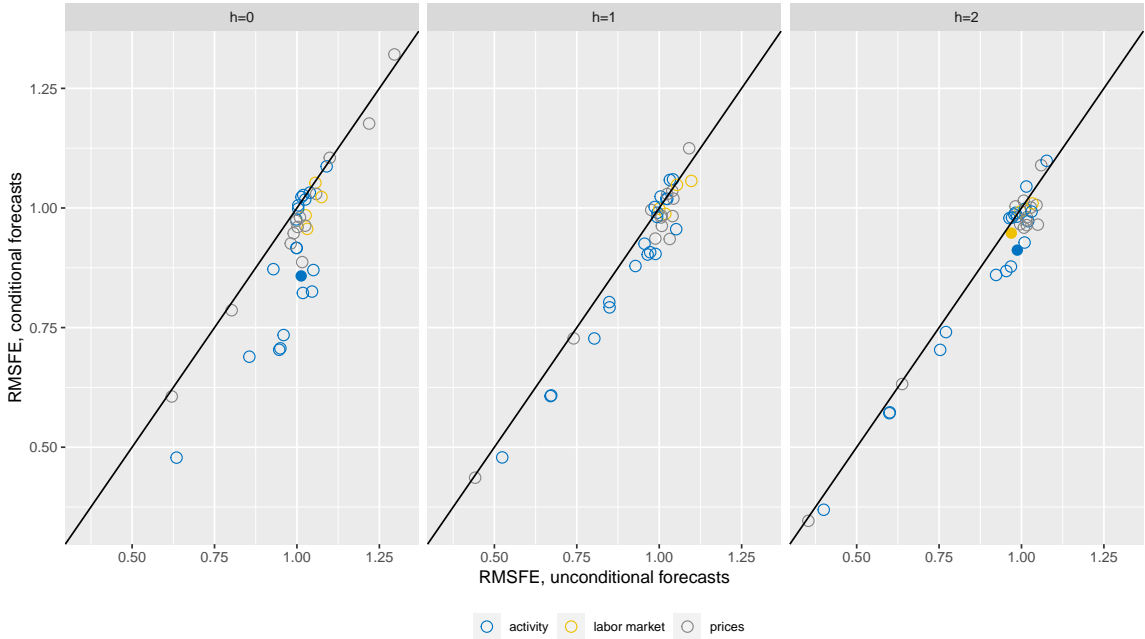
Note: The figure shows the root mean squared forecast errors (RMSFE) corresponding to unconditional forecasts (x-axis) and forecasts conditional on professional forecasters' view on GDP growth and CPI inflation (y-axis) for different time series. For each series, the RMSFE is relative to an autoregressive benchmark. Entries above (below) the 45-degree line indicate that conditional forecasts perform worse (better) than the unconditional ones. Filled points correspond to those variables for which the null hypothesis of the Diebold-Mariano test can be rejected at the 5 percent level. For details, see the main text.

Table III.1: Point forecast evaluation

category/horizon	h=0	h=1	h=2
activity	1/20	0/20	1/20
labor market	0/4	0/4	1/4
prices	0/13	0/13	0/13
total	**1/37**	**0/37**	**2/37**

The table shows the fraction of times the null hypothesis of the Diebold-Mariano test is rejected at the 5 percent significance level against the one-sided alternative that the conditional forecasts yield a lower loss. For details, see the main text.

the null of equal predictive accuracy is rejected for a third of all the variables that we consider, mainly concentrated in the activity group. For one-quarter ahead forecasts the number decreases slightly but for many activity series we still find significant gains (Table III.2).

Figure III.4: Density forecast evaluation
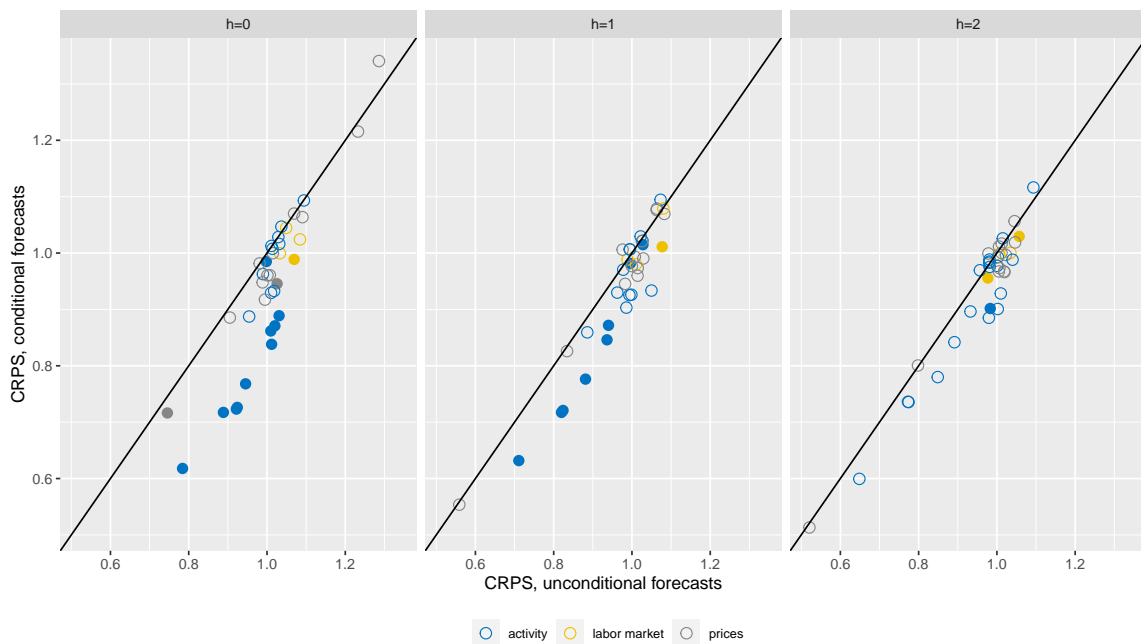
Note: The figure shows the average continuous ranked probability score (CRPS) corresponding to unconditional forecasts (x-axis) and forecasts conditional on professional forecasters' view on GDP growth and CPI inflation (y-axis) for different time series. For each series, the CRPS is relative to an autoregressive benchmark. Entries above (below) the 45-degree line indicate that conditional forecasts perform worse (better) than the unconditional ones. Filled points correspond to those variables for which the null hypothesis of the Diebold-Mariano test can be rejected at the 5 percent level. For details, see the main text.

III.4.3 Robustness checks

As a first robustness check, we drop the global financial crisis of 2008-09 from the evaluation sample, to see if the gains in predictive accuracy for point and density forecasts are also apparent in more tranquil times. Figure III.5 presents the results for the evaluation

Table III.2: Density forecast evaluation

category/horizon	h=0	h=1	h=2
activity	10/20	8/20	1/20
labor market	1/4	1/4	2/4
prices	2/13	0/13	0/13
total	**13/37**	**9/37**	**3/37**

The table shows the fraction of times the null hypothesis of the Diebold-Mariano test is rejected at the 5 percent significance level against the one-sided alternative that the conditional forecasts yield a lower loss. For details, see the main text.

subsample ranging from 2011Q1 to 2017Q4.

Figure III.5: Robustness check: post-crisis evaluation sample
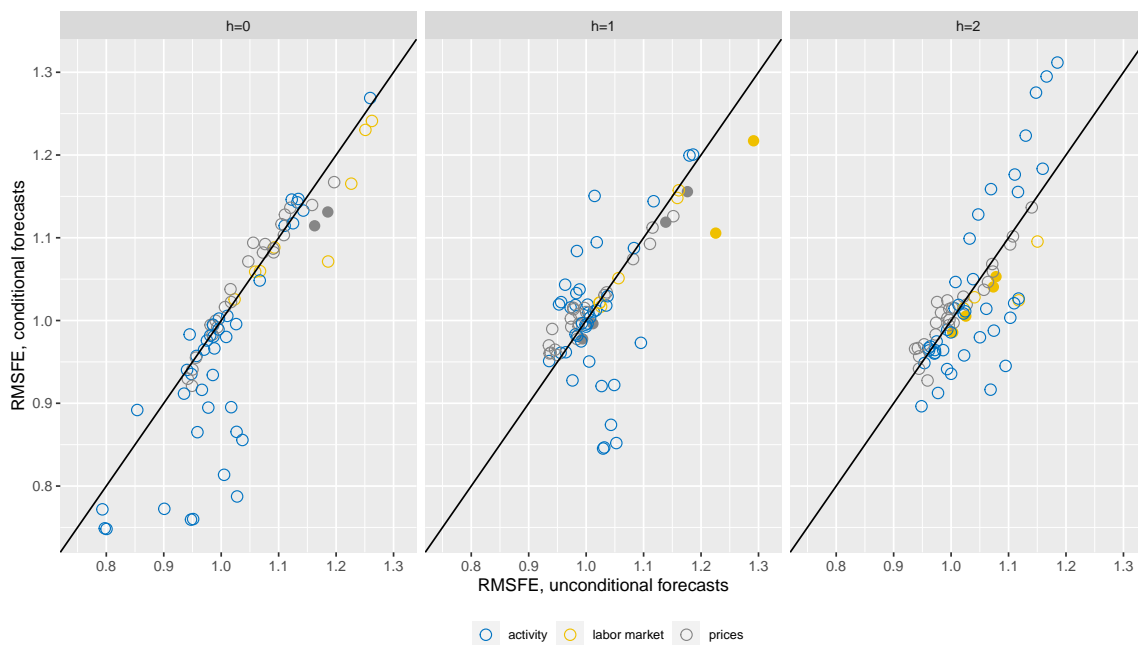
Note: The figure shows the root mean squared forecast errors (RMSFE) over the evaluation sample 2011Q1-2017Q4 corresponding to unconditional forecasts (x-axis) and forecasts conditional on professional forecasters' view on GDP growth and CPI inflation (y-axis) for different time series. For each series, the RMSFE is relative to an autoregressive benchmark. Entries above (below) the 45-degree line indicate that conditional forecasts perform worse (better) than the unconditional ones. Filled points correspond to those variables for which the null hypothesis of the Diebold-Mariano test can be rejected at the 5 percent level. For details, see the main text.

Overall, the forecast performance declines markedly relative to the autoregressive benchmarks but for some series, gains from conditioning still arise. However, for $h = \{1, 2\}$ we also find that for many series - particularly from the activity group - conditional forecasts are on average less accurate. Rejection of the null of the Diebold-Mariano test still occurs but with a fex exceptions, this occurs in cases where the autoregressive benchmark actually performs better than the factor model. As such, it is fair to say that developments during the financial crisis exert some influence on the results presented above for the entire evaluation

sample.

Besides the evaluation period, the results presented above may also be sensitive to the model specification. In particular, they were obtained under a given number of factors, $R = 2$. However, differently specified models yield similar forecast performance. Figure III.6 shows the relative RMSFE of the model with $R = 2$ on which the results above are based as well as both smaller ($R = 1$) and larger ($R = 5, 8$) models. There is little indication that model specification systematically alters the forecast performance. Conditional forecasts at $h = 0$ for the change in the producer index improve substantially when $R = 8$ with the relative RMSFE decreasing by 0.25. To a lesser extent this also holds for the unconditional forecasts with the relative RMSFE dropping by around 0.1. However, this is the exception and for other horizons and variables we find much smaller differences in either way. Consequently, the points corresponding to the relative RMSFE for $R = 1, 5, 8$ in Figure III.6 all hug the 45-degree line, which indicates identical forecast performance, quite closely.

Figure III.6: Robustness check: different R
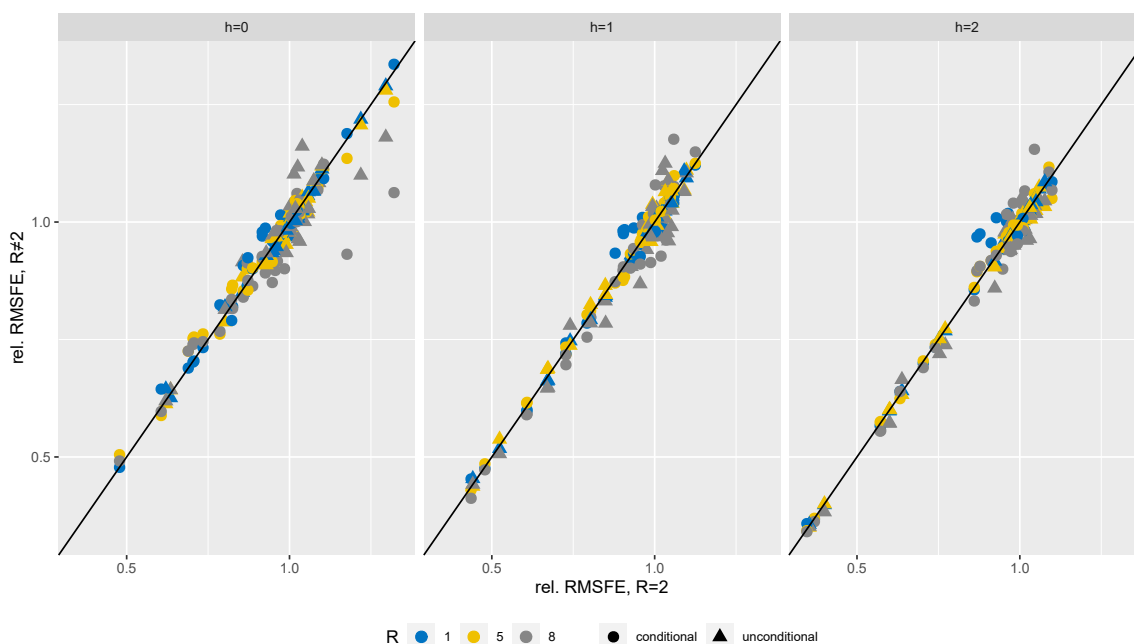
 Note: The figure shows the root mean squared forecast error (RMSFE) relative to the autoregressive benchmarks for different number of factors. On the x-axis are the values of the RMSFE for the model with 2 factors on which the main results are based. On the y-axis are the RMSFE for alternative model specifications, differentiated by colors. Entries above (below) the 45-degree line indicate that the model with 2 factors performs better (worse) than the alternative models.

III.5   Conclusion

In this paper we assess the forecast performance of factor models in real-time for Germany covering the period 2006-2017. We contribute to the literature by broadening the horizon of the forecast evaluation to include a large number of the variables that are used to estimate

the model. In addition, we also investigate to what extent the forecast performance of the model improves when we incorporate external information in the form of professional forecasters' views on GDP growth and CPI inflation. Conditioning on forecasts rather than actual realized values gives an accurate assessement of the uncertainty around conditional forecasts that is relevant to policy-makers by capturing both how accurate the model can summarize the comovement of the time series as well as the accuracy of the forecasts on which we condition.

Our results show that the factor model produces forecasts for most of the series and horizons that are as accurate as those from autoregressive benchmarks. In some cases they outperform the benchmark by a large margin. When conditioning on the forecasts of professional forecasters for GDP growth and CPI inflation, we generally find some improvements in forecast accuracy for the variables of interest from categories such as real activity, prices and the labor market. Only for a smaller number of series, however, do we find that conditioning produces large gains. Moreover, for point forecasts the differences are statistically insignificant for virtually all series. When evaluating the predictive densities with the CRPS, we do find significant gains at $h = 0$ ($h = 1$) for around a third (a quarter) of the variables under consideration. To some extent the results appear to be driven by the large drop in output in the wake of the Global Financial Crisis. as the forecast performance deteriorates relative to the benchmark when we shorten the evaluation sample to 2011-2017. On the other hand, the results are robust to model specification in the sense that increasing or decreasing the number of factors from $R = 2$ as in the baseline specification does not systematically alter the forecast performance.

Lastly, this paper proposes precision-based sampling algorithms for conditional and unconditional forecasts in factor models and more broadly state space models with unobserved components. Treating forecasts as missing observations, the insights in Hauber and Schumacher (2021) can be obtained directly to obtain draws from the relevant predictive densities. Simulations document that these algorithms can be more efficient from a computational point of view than simulation smoothers based on the Kalman filter. While the exact quantitative results of the excercise may vary depending on platform and implementation, the qualitative results indicate that precison-based sampling algorithms are a viable alternative in such applications.

Appendix

III.A   Data

This Appenix provides detail on the real-time dataset used in the empirical application. The 57 time series can be grouped in to the following categories: **activity** which includes the gross domestic product as well as expenditure- and production-side components of output. In addition, series like industrial production, turnover and orders also form part of this cat-

egory. The **prices** category contains the (implicit) deflators corresponding to the chained volume indices of the national accounts mentioned above as well as consumer and producer price series. The **labor market** is covered by series on total employment, wages as well as hours worked in the industrial and construction sector. These are the series which we focus on in the forecast evaluation and their vintages are all sourced from the Deutsche Bundesbank's Real-Time Database (RTD). Moreover, to capture the **financial** side of the economy, we include series such as interest rates, commodity prices and the CDAX stock market index. These series are also sourced from the Deutsche Bundesbank's website. Lastly, **survey** indicators are taken from the European Commission's Business and Consumer Surveys. We focus on the components of the Economic Sentiment Indicator (ESI) which capture confidence in the industrial, consumer, services, retail trade and construction sectors. In addition, we also include the employment expectations index (EEI).[14] Table III.3 lists the data series used in the empirical application, the group the indicators belong to, their source as well as any transformation applied to the series prior to estimation, e.g. whether series are in logs and/or differenced (diff).

---

[14]The series were downloaded on April 27th, 2021 from `https://ec.europa.eu/economy_ finance/db_indicators/surveys/documents/series/nace2_ecfin_2106/main_indicators_ sa_nace2.zip`

Table III.3: Description of the dataset

Series	Mnem.	Group	Transf.	Source	Notes
gross domestic product, chain index[S,C]	gdp	activity	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.Y.A.AG1.CA010.A.I
gross domestic product, deflator[S,C]	p_gdp	prices	log, 1st diff	Bundesbank RTD	implicit deflator, i.e. nominal series divided by chain index
private consumption expenditure, chain index[S,C]	c_priv	activity	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.Y.A.CA1.BA100.A.I
private consumption expenditure, deflator[S,C]	p_c_priv	prices	log, diff	Bundesbank RTD	implicit deflator, i.e. nominal series divided by chain index
government consumption expenditure, chain index[S,C]	c_gov	activity	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.Y.A.CA1.BA100.A.I
government consumption expenditure, implicit deflator[S,C]	p_c_gov	prices	log, diff	Bundesbank RTD	nominal series divided by chain index
equipment investment, chain index[S,C]	gfcf_equip	activity	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.Y.A.CE1.CA010.A.I
equipment investment, implicit deflator[S,C]	p_gfcf_equip	prices	log, diff	Bundesbank RTD	implicit deflator, i.e. nominal series divided by chain index
construction investment, chain index[S,C]	gfcf_constr	activity	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.Y.A.CF1.CA010.A.I
construction investment, implicit deflator[S,C]	p_gfcf_constr	prices	log, diff	Bundesbank RTD	implicit deflator, i.e. nominal series divided by chain index
other investment, chain index[S,C]	gfcf_other	activity	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.Y.A.CI1.CA010.A.I
other investment, implicit deflator[,C]	p_gfcf_other	prices	log, diff	Bundesbank RTD	nominal series (BBKRT1Q.DE.Y.A.CI1.CA010.V.A) divided by chain index
exports (of goods and services), chain index[S,C]	x	activity	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.Y.A.CX1.CA010.A.I
exports (of goods and services), implicit deflator[S,C]	p_x	prices	log, diff	Bundesbank RTD	nominal series (BBKRT1Q.DE.Y.A.CX1.CA010.V.A) divided by chain index
imports (of goods and services), chain index[S,C]	m	activity	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.Y.A.CM1.CA010.A.I
imports (of goods and services), implicit deflator[S,C]	p_m	prices	log, diff	Bundesbank RTD	nominal series (BBKRT1Q.DE.Y.A.CM1.CA010.V.A) divided by chain index
gross value added, industry, chain index[S,C]	gva_ind	activity	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.Y.A.AU1.AA020.A.I
gross value added, industry, implicit deflator[S,C]	p_gva_ind	prices	log, diff	Bundesbank RTD	nominal series (BBKRT1Q.DE.Y.A.CM1.CA010.V.A) divided by chain index
gross value added, construction, chain index[S,C]	gva_constr	activity	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.Y.A.AU1.AA030.A.I
gross value added, construction, implicit deflator[S,C]	p_gva_constr	prices	log, diff	Bundesbank RTD	nominal series (BBKRT1Q.DE.Y.A.AU1.AA030.V.A) divided by chain index
gross value added, trade, transport & hospitality, chain index[S,C]	gva_tth	activity	log, diff	Bundesbank RTD	nominal series (BBKRT1Q.DE.Y.A.AU1.AA040.V.A) divided by chain index
gross value added, trade, transport & hospitality, implicit deflator[S,C]	p_gva_tth	prices	log, diff	Bundesbank RTD	nominal series (BBKRT1Q.DE.Y.A.AU1.AA040.V.A) divided by chain index
consumer prices, index[S,C]	cpi	prices	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.Y.PPC1.PC100.R.I
consumer prices excl. food & energy, index[S,C]	cpi_core	prices	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.S.PPC1.PC110.R.I
producer prices, index[S]	ppi	prices	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.S.PPP1.PP100.R.I
producer prices excl. energy, index[S]	ppi_core	prices	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.S.PPP1.PP200.R.I
producer prices, construction, index	ppi_constr	prices	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.N.PPP1.PP300.R.I
producer prices, agriculture, index[S]	ppi_agri	prices	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.N.PPP1.PP400.R.I
production, industry, index[S,C]	prod	activity	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.Y.I.IP1.ACM01.C.I
orders, industry, index[S,C]	ord	activity	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.Y.I.IO1.ACM01.C.I
turnover, industry, index[S,C]	to	activity	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.Y.I.IT1.ACM01.VI
production, construction, index[S,C]	prod	activity	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.Y.I.IP1.AA020.C.I
orders, construction, index[S,C]	ord	activity	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.Y.I.IO1.AA031.C.I
turnover, construction, index[S,C]	to	activity	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.Y.I.IT1.AA031.V.A
turnover, retail, index[S,C]	to	activity	log, diff	Bundesbank RTD	price adjusted, Code: BBKRT1M.DE.Y.I.IT1.AGA01.C.I
employment[S,C]	emp	labor market	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.S.A.BF1.CA010.P.A
hours worked, industry[S,C]	h_ind	labor market	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.Y.L.BE2.AA022.H.I
hours worked, construction[S,C]	h_constr	labor market	log, diff	Bundesbank RTD	Code: BBKRT1M.DE.Y.L.BE2.AA031.H.A
wages and salaries[S,C]	w	labor market	log, diff	Bundesbank RTD	Code: BBKRT1Q.DE.S.A.DE2.CA010.V.A

CDAX, index	cdax	financial	log, diff	Bundesbank	Code: BBK01.WU001A
Nominal effective exchange rate, narrow	neer19	financial	log, diff	Bundesbank	Vis-a-vis 19 trading partners, code: BBEE1.M.I8.AAA.XZE012.A.AABAN.M00
Nominal effective exchange rate, broad	neer38	financial	log, diff	Bundesbank	Vis-a-vis 38 trading partners, code: BBEE1.M.I8.AAA.XZE021.A.AABAN.M00
Government bond yields (1-year)	ily	financial	diff	Bundesbank	Code: BBK01.WZ9808
Government bond yields (5-year)	ily	financial	diff	Bundesbank	Code: BBK01.WZ9816
Government bond yields (10-year)	ily	financial	diff	Bundesbank	Code: BBK01.WZ9826
Commodity prices, energy, index	comm_en	financial	log, diff	HWWI	in €, sourced via Deutsche Bundesbank (BBDG1.M.HWWI.N.EURO.ENERGY00.I15.EUR.A)
Commodity prices, excl. energy, index	comm_exen	financial	log, diff	HWWI	in €, sourced via Deutsche Bundesbank (BBDG1.M.HWWI.N.EURO.TOTNXNGY.I15.EUR.A)
industrial confidence indicator[s]	survey_ind	surveys	diff	Eurp. Commission	—
services confidence indicator[s]	survey_serv	surveys	diff	Eurp. Commission	—
consumer confidence indicator[s]	survey_cond	surveys	diff	Eurp. Commission	—
retail trade confidence indicator[s]	survey_retail	surveys	diff	Eurp. Commission	—
construction confidence indicator[s]	survey_constr	surveys	diff	Eurp. Commission	—
employment expectation index[s]	survey_eei	surveys	diff	Eurp. Commission	—

[s] seasonally adjusted
[c] calendar adjusted

III.B   Gibbs Sampler

This Appendix details how to sample from the posterior distribution of the parameters of the factor model, $p(\Theta|\mathbf{y})$. For the reader's convenience, I repeat the model description before discussing the blocks of the Gibbs Sampler. Setting $J = 1$, $P = 2$ and $S = 0$, the following equations constitute the model:

$$\mathbf{y}_t = \lambda\,\mathbf{f}_t + \mathbf{e}_t \tag{III.7a}$$

$$\mathbf{f}_t = \phi_1\,\mathbf{f}_{t-1} + \phi_2\mathbf{f}_{t-2} + \boldsymbol{v}_t \tag{III.7b}$$

$$\mathbf{e}_t = \psi\,\mathbf{e}_{t-1} + \boldsymbol{\varepsilon}_t \tag{III.7c}$$

The Gibbs Sampler then cycles through the following blocks to draw from the conditional posterior distributions[15]:

Step 1: $p(\mathbf{f}|\mathbf{y}, \Theta)$
Conditional on the parameters, the factors are drawn from their posterior distribution following the precision sampler outlined in Section III.2.1 of the main text.

Step 2: $p(\lambda|\mathbf{y}, \mathbf{f}, \psi, \Sigma)$
Given the factors, the observation equations of the model are independent and the loadings can be sampled variable by variable. Let the prior on the loadings be given by $\mathcal{N}(0, \mathbf{P}_\lambda^{-1})$ and let $\mathbf{P}_{e_i} = (\mathbf{H}_\psi^\mathsf{T}\sigma_i^{-2}\mathbf{H}_\psi)^{-1}$ denote the $T \times T$ precision matrix of $\mathbf{e}_i$, where

$$
\begin{bmatrix}
I_N & & & & & & \\
-\psi_i & I_N & & & & & \\
& -\psi_i & I_N & & & & \\
& & & \ddots & & & \\
& \cdots & \cdots & -\psi_i & I_N & & \\
& & & & & \ddots & \\
& & & \cdots & \cdots & -\psi_i & I_N.
\end{bmatrix}
$$

Using standard linear regression results (see Kroese and Chan, 2013), the conditional posterior for the i-th equation is given by

$$\lambda_i|\cdot \;\; \sim \;\; \mathcal{N}\left(\mathbf{m}_i^\lambda, \mathbf{M}_i^\lambda\right) \tag{III.8}$$

$$\mathbf{M}_i^\lambda \;\; = \;\; (\mathbf{P}_\lambda + \mathbf{f}^\mathsf{T}\mathbf{P}_{e_i}\mathbf{f})^{-1} \tag{III.9}$$

$$\mathbf{m}_i^\lambda \;\; = \;\; \mathbf{M}_i^\lambda(\mathbf{f}^\mathsf{T}\mathbf{P}_{e_i}\mathbf{y}_i). \tag{III.10}$$

We set $\mathbf{P}_\psi = \mathbf{I}_R$ for all variables.

---

[15]Initial values for the coefficients Θ are obtained from principal component estimates of the factors.

Step 3: $p(\psi|\mathbf{y}, \mathbf{f}, \lambda, \Sigma)$

Conditional on a draw of the loadings, for each series we can back out the idiosyncratic components $\mathbf{e}$ and update ψ. Given a Normal prior for ψ_i with mean 0 and precision matrix P_ψ, the resulting posterior is given by

$$\psi_{i.}|\cdot \;\sim\; \mathcal{N}\left(m_i^\psi, M_i^\psi\right) \tag{III.11}$$

$$M_i^\psi \;=\; \left(P_\psi + \frac{1}{\sigma_i^2}\sum\nolimits_{t=1}^{T-1} e_{i,t}^2\right)^{-1} \tag{III.12}$$

$$m_i^\psi \;=\; M_i^\psi\left(\frac{1}{\sigma_i^2}\sum\nolimits_{t=2}^{T} e_{i,t-1}' e_{i,t}\right). \tag{III.13}$$

For all variables, we set $P_\psi = 1$.

Step 4: $p(\Sigma|\mathbf{y}, \mathbf{f}, \lambda, \psi)$

Each diagonal elements of Σ can be sampled by drawing from

$$\sigma_i^2 \sim \mathcal{G}^{-1}\left(\frac{u+T-1}{2}, \frac{U+\sum_{t=2}^{T}\varepsilon_{i,t}^2}{2}\right)$$

where u and U are the prior shape and rate. We set $u = 3$, $U = 0.5$ so that the prior mean and standard deviation equal 0.25.

Step 5: $p(\phi_1, \phi_2|\mathbf{f}, \Omega)$

Let $\mathbf{X}_t = I_R \otimes [\mathbf{f}_{t-1}^\mathsf{T}, \mathbf{f}_{t-2}^\mathsf{T}]$ denote the lags of the factors and $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_t]$. Then given a Normal prior $p(\phi) \sim \mathcal{N}(0, \mathbf{P}_\phi^{-1})$, the coefficients of the factor VAR can be sampled by drawing from $\phi|\cdot \sim \mathcal{N}(\mathbf{m}^\phi, \mathbf{M}^\phi)$ where

$$\mathbf{M}^\phi = (\mathbf{P}_\phi + X^\mathsf{T}(\mathbf{I}_T \otimes \Omega^{-1})\mathbf{X})^{-1}, \; \mathbf{m}^\phi = \mathbf{M}^\phi(\mathbf{X}^\mathsf{T}(\mathbf{I}_T \otimes \Omega^{-1})\mathbf{f})$$

In line with the literature on Bayesian vectorautoregressions, we impose a Minnesota-type prior on the VAR coefficients. The prior mean is set equal to 0 and the prior precision increases in the lag length. Specifically, the prior variance of the elements of ϕ for $p = 1 : P$ is given by:

$$Var(\phi_{p,ij}) = \begin{cases} \frac{\pi_0}{p^2}, & \text{if } i = j \\ \frac{\pi_1}{p^2}, & \text{otherwise.} \end{cases}$$

Common values in the literature are $\pi_0 = 0.2$ and $\pi_1 = 0.1$, thus shrinking coefficients on the lags of other factors stronger towards 0.

III.C   Computational efficiency analysis

This Appendix compares the computational efficiency of the precision-sampling algorithms outlined in the main text to simulation smoothers such as Carter and Kohn (1994) and Durbin and Koopman (2002) that rely on the Kalman-filter and are commonly used in the literature. Previous work in this area has analyzed the computational advantages of such simulation smoothers compared to the algorithms for conditional forecasting given in Waggoner and Zha (1999) in the context of a vector autoregressions (Banbura et al., 2015); McCausland et al. (2011) analyzes the performance of precision-based samplers in the context of a time-varying parameter regression and dynamic factor models and compares the performance to the Durbin and Koopman (2002) simulation smoother.

The focus lies on both small factor models with $N = 20$ variables and $R = 2$ factors as well as two variants of larger factor models with $N = 100, R = 2$ (*large N factor model*) and $N = 100, R = 2$ (*large factor model*). We evaluate the runtime of obtaining 100 draws from the predictive density $p(y^f | y^o, y^c)$ given a sample size $T = 100$ and forecast horizons $H = \{5, 20, 50\}$ conditional on a share $\kappa = 0.1$ of the observables (*hard conditioning*).[16] Furthermore, we also consider the costs of producing repeated samples from the predictive density given the parameters as would be the case when on top of the conditioning set

$$\mathbf{y}^c = [y_{1,T+1}, \ldots, y_{\kappa \cdot N, T+1}, \ldots, y_{1,T+H}, \ldots, y_{\kappa \cdot N, T+H}]$$

restrictions are imposed on the remaining observables to lie within a prespecified range (*soft conditioning*).

The dynamics of the factor model outlined in Section III.2 are given by a VAR(1) with parameters $\phi_1 = 0.7 \cdot I_R$ and $\Sigma = I_R$. The elements of the loadings on the contemporaneous factors λ_0 are drawn from independent Normal distributions with mean 0.5 and variance 0.1; all remaining loadings are set to 0, i.e. $K = 0$. Setting $J = 0$, the idiosyncratic components are i.i.d and ω_i is chosen such that the common component - $\lambda_{0,i}.\mathbf{f}_t$ - explains ⅔ of the total variation in the i-th variable.

Details on the implementation of the Carter and Kohn (1994) and Durbin and Koopman (2002) simulation smoothers are provided in Appendix III.D.

Each model is simulated 10 times.[17] Figure III.7 shows the relative runtime for hard (left column) and soft (right column) conditioning as boxplots. The precision-based algorithms performs compared to the Kalman-filter based simulation smoothers for all horizons, with gains in computational efficiency up to a factor of 10. When repeated draws from the predictive density are required as in the case of soft conditioning, the gains in computational efficiency are much larger.

---

[16]The size of the conditioning set does not have a large impact on the relative runtime and in the interest of space I do not report results for $\kappa = \{0.5, 0.75\}$. They are available upon request.

[17]The simulations were run using Matlab2020a on an Intel Core i5 2.3 Ghz with 8 GB of RAM. To time the execution of the code, we used the function `timeit()`.

Figure III.7: Computational efficiency analysis of different simulation smoothers
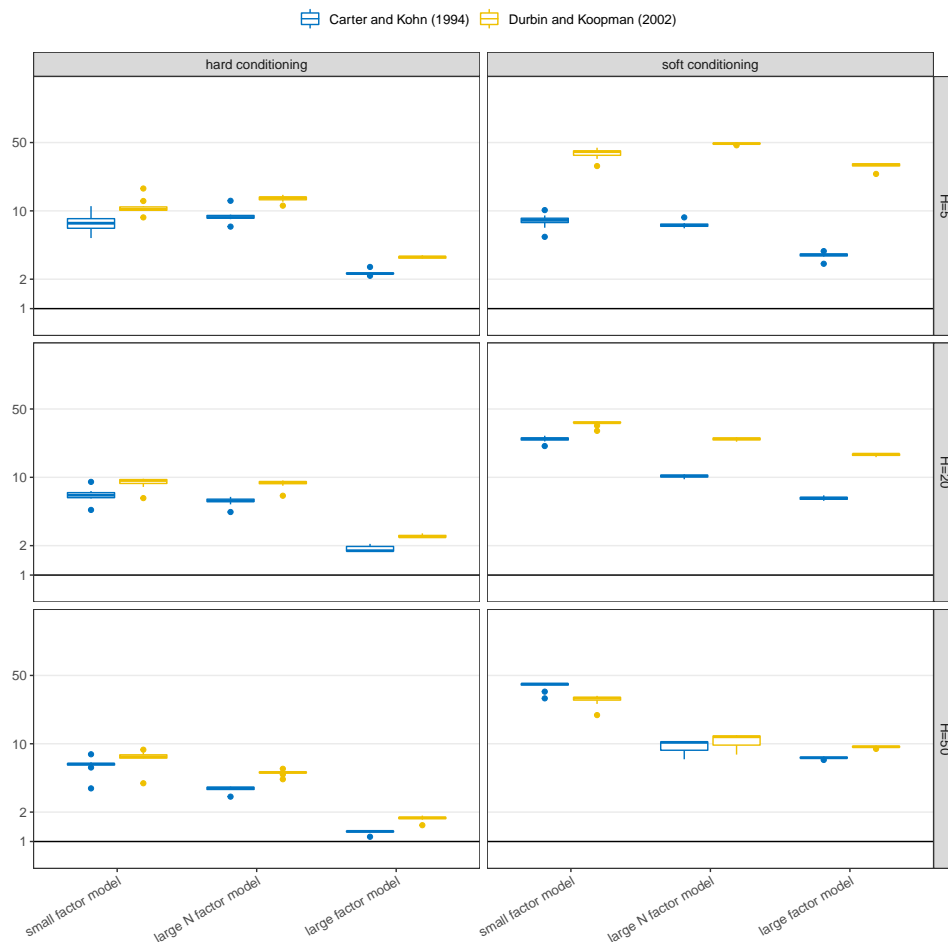
Note: The figure shows the time it takes to generate draws from the predictive density $p(\mathbf{y}^f|\mathbf{y}^o, \mathbf{y}^c)$ for different forecast horizons H using the Carter and Kohn (1994) and Durbin and Koopman (2002) simulation smoothers relative to the precision sampler outlined in the main text. The conditioning set $\mathbf{y}^c$ consists of the first $0.1 \cdot N$ variables (left column). For the small factor model, the number of variables is $N = 20$ and the number of static factors is $R = 2$; for the large N factor model, the number of variables is $N = 100$, while the large factor model also has $R = 10$. The runtime measures the entire procedure of generating a draw from the predictive density, including building the system matrices. For the precision-sampler this includes the costs of setting up the permutation matrices which is incurred only once. In the case of soft conditioning (right column), the runtime of producing 100 draws given the same parameters is reported. For details, see the main text.

III.D   Kalman-filter based simulation smoothers

This appendix provides details on the Carter and Kohn (1994) and Durbin and Koopman (2002) simulation smoothers that are used in the simulations in Appendix III.C. Although both the Carter and Kohn (1994) and Durbin and Koopman (2002) simulation smoothers rely on the Kalman filter and smoother to produce a draw from the conditional distribution of the states, there are important conceptual differences between the two.

Let $a_{t|s} = E[\alpha_t|\mathbf{y}_1,\ldots,\mathbf{y}_s]$ and $P_{t|s} = Var[\alpha_t|\mathbf{y}_1,\ldots,\mathbf{y}_s]$ denote the conditional mean and variance of the state at time t conditional on information up to time s which can be obtainted from the Kalman filter or smoother. In addition, let a_t denote a draw from the conditional distribution $p(\alpha_t|\mathbf{y}_1,\ldots,\mathbf{y}_{T+H})$.

Carter and Kohn (1994) simulation smoother

Given the parameters of the state space model Θ and $a_{t|t}, P_{t|t} \,\forall t = 1,\ldots,T+H$, the algorithm in Carter and Kohn (1994) generates a_{T+H} from

$$p(\alpha_{T+H}|\mathbf{y}_1,\ldots,\mathbf{y}_{T+H})$$

with

$$E[\alpha_{T+H}|\mathbf{y}_1,\ldots,\mathbf{y}_{T+H}] = a_{T+H|T+H}$$
$$Var[\alpha_{T+H}|\mathbf{y}_1,\ldots,\mathbf{y}_{T+H}] = P_{T+H|T+H}.$$

For $t = T + H - 1, \ldots, 1$ a_t is generated from $p(\alpha_t|\mathbf{y}_t, a_{t+1})$. For details on how to derive the moments of the conditional distributions given the output of the Kalman filter, see the original paper as well as the textbook treatment in Kim and Nelson (1999).

Durbin and Koopman (2002) simulation smoother

The Durbin and Koopman (2002) simulation smoother produces a draw from the conditional distribution of the states by first simulating the state space model in Equation III.5, producing a draw from the *joint* distribution of states and observables

$$p(\alpha_1,\ldots,\alpha_{T+H},\mathbf{y}_1,\ldots,\mathbf{y}_{T+H}).$$

Denote this joint draw by $a_1^+,\ldots,a_{T+H}^+,y_1^+,\ldots,y_{T+H}^+$. Running the Kalman smoother recursions yields $a_{t|T+H}$ and $a_{t|T+H}^+$ where $a_{t|T+H}^+ = E[\alpha_t^+|y_1^+,\ldots,y_{T+H}^+]$. By setting $a_t = a_{t|T+H} - a_{t|T+H}^+ + a^+$ adjusts the mean of a_t^+ to yield a draw from the desired conditional distribution. In practice, it is more efficient to first construct $y^* = y - y^+$ and run the Kalman smoother only once, yielding $a_{t|T+H}^* = E[a_t^*|y^*]$. A draw from $p(\alpha_1,\ldots,\alpha_{T+H}|\mathbf{y}_1,\ldots,\mathbf{y}_{T+H})$ is then obtainted by setting $a_t = a_{t|T+H}^* + a_t^+$. Note that since only an estimate of the smoothed

mean of the states is required, more efficient smoothing recursions can be employed which do not require the storage of $a_{t|t}$ and $P_{t|t}$ (see Durbin and Koopman, 2002, ch. 4.4.2).

Drawing from the (conditional) predictive density $p(y^f|I,\Theta)$

Under the common assumption of uncorrelated measurement errors, draws from the predictive density $p(y_t^{\mathrm{f}}|a_t,y^o,y^c,\Theta)$ can be obtained by sampling the measurement errors independently from $\mathcal{N}(0,\omega_i)\,\forall i$ such that $y_{i,t}\notin y_{c,t}$ and adding them to the draw of the state vector (see Banbura et al., 2015, for the general case of a full covariance matrix Ω). For the Durbin and Koopman (2002) simulation smoother the generated artifical observations y^+ can be re-used, since $Fa_{t|T+H}^* + y_t^+$ is a draw of the observations conditional on the sampled states.

Multiple draws from $p(y^f|I,\Theta)$

Additional draws from the (conditional) predictive density given the same set of parameters Θ can be produced recursively from the Carter and Kohn (1994) simulation smoother by independently sampling a_t as many times as required and conditioning on the draw to generate y_t^{f}. In the case of the Durbin and Koopman (2002) simulation smoother, each additional draw requires new simulated values of the states and observations, $a_1^+,\ldots,a_{T+H}^+, y_1^+,\ldots,y_{T+H}^+$. Given these, the entire filtering and smoothing recursions need to be re-run. However, computational savings still arise as the state covariances $P_{t|t}$ and functions thereof that are calculated during the recursions do not depend on the observations. As such they only need be calculated once and can be passed on to subsequent runs of the Kalman filter and smoother (Durbin and Koopman, 2002, p. 606).

Initialisation

The Kalman filter is initialized by setting $a_{1|0}$ and $P_{1|0}$ equal to the unconditional mean and variance, respectively, of the factor VAR.

Bibliography

Aastveit, Knut Are, Francesco Ravazzolo, and Herman K. van Dijk, "Combined Density Now-casting in an Uncertain Economic Environment," *Journal of Business & Economic Statistics*, 2018, *36* (1), 131–145.

⎯ , Karsten R. Gerdrup, Anne Sofie Jore, and Leif Anders Thorsrud, "Nowcasting GDP in Real Time: A Density Combination Approach," *Journal of Business & Economic Statistics*, January 2014, *32* (1), 48–68.

Alvarez, Rocio, Maximo Camacho, and Gabriel Perez-Quiros, "Aggregate versus disaggregate information in dynamic factor models," *International Journal of Forecasting*, 2016, *32* (3), 680–694.

Amisano, Gianni and John Geweke, "Prediction Using Several Macroeconomic Models," *The Review of Economics and Statistics*, 2017, *99* (5), 912–925.

Anderson, Theodore W., *An Introduction to Multivariate Statistical Analysis* Wiley Series in Probability and Statistics, 3rd ed., Wiley, 2003.

Angelini, Elena, Jerome Henry, and Massimiliano Marcellino, "Interpolation and backdating with a large information set," *Journal of Economic Dynamics and Control*, 2006, *30* (12), 2693 – 2724.

Bai, Jushan and Peng Wang, "Identification theory for high dimensional static and dynamic factor models," *Journal of Econometrics*, 2014, *178* (2), 794 – 804.

⎯ and Serena Ng, "Forecasting economic time series using targeted predictors," *Journal of Econometrics*, 2008, *146* (2), 304–317. Honoring the research contributions of Charles R. Nelson.

⎯ and ⎯ , "Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data," Working Paper, New York University 2019.

Banbura, Marta and Gerhard Rünstler, "A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP," *International Journal of Forecasting*, 2011, *27* (2), 333–346.

___ and Michele Modugno, "Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data," *Journal of Applied Econometrics*, 2014, *29* (1), 133–160.

___, Domenico Giannone, and Michele Lenza, "Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections," *International Journal of Forecasting*, 2015, *31* (3), 739–756.

Bhadra, Anindya, Jyotishka Datta, Nicholas G. Polson, and Brandon Willard, "The horseshoe+ estimator of ultra-sparse signals," *Bayesian Analysis*, 2017, *12* (4), 1105–1131.

Bhattacharya, Anirban and David B. Dunson, "Sparse Bayesian infinite factor models," *Biometrika*, 2011, *98*, 291–306.

___, Debdeep Pati, Natesh S. Pillai, and David B. Dunson, "Dirichlet–Laplace priors for optimal shrinkage," *Journal of the American Statistical Association*, 2015, *110* (512), 1479–1490.

Bishop, Christopher M., *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Berlin, Heidelberg: Springer-Verlag, 2006.

Boivin, Jean and Serena Ng, "Are more data always better for factor analysis?," *Journal of Econometrics*, 2006, *132* (1), 169–194.

Bok, Brandyn, Daniele Caratelli, Domenico Giannone, Argia Sbordone, and Andrea Tambalotti, "Macroeconomic Nowcasting and Forecasting with Big Data," *Annual Review of Economics*, 08 2018, *10*.

Botev, Zdravko I., Joseph F. Grotowski, and Dirk P. Kroese, "Kernel density estimation via diffusion," *Annals of Statistics*, 10 2010, *38* (5), 2916–2957.

Caggiano, Giovanni, George Kapetanios, and Vincent Labhard, "Are more data always better for factor analysis? Results for the euro area, the six largest euro area countries and the UK," *Journal of Forecasting*, 2011, *30* (8), 736–752.

Camacho, Maximo and Gabriel Perez-Quiros, "Introducing the euro-sting: Short-term indicator of euro area growth," *Journal of Applied Econometrics*, 2010, *25* (4), 663–694.

Carter, C. K. and R. Kohn, "On Gibbs sampling for state space models," *Biometrika*, 1994, *81* (3), 541–553.

Carvalho, Carlos M., Jeffrey Chang, Joseph E. Lucas, Joseph R. Nevins, Quanli Wang, and Mike West, "High-dimensional sparse factor modeling: applications in gene expression genomics," *Journal of the American Statistical Association*, 2008, *103* (484), 1438–1456.

___, Nicholas G. Polson, and James G. Scott, "Handling sparsity via the horseshoe," *Journal of Machine Learning Research*, 2009, *5*, 73–80.

— , — , and — , "The horseshoe estimator for sparse signals," *Biometrika*, 2010, *97* (2), 465–480.

Castillo, Ismaël and Aad van der Vaart, "Needles and straw in a haystack: Posterior concentration for possibly sparse sequences," *The Annals of Statistics*, 2012, *40* (4), 2069–2101.

Chan, Joshua C. C., Gary Koop, and Simon M. Potter, "A New Model of Trend Inflation," *Journal of Business & Economic Statistics*, 2013, *31* (1), 94–106.

— , — , and — , "A Bounded Model of Time Variation in Trend Inflation, NAIRU and the Phillips Curve," *Journal of Applied Econometrics*, 2016, *31* (3), 551–565.

Chan, Joshua C.C., "Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure," *Journal of Business & Economic Statistics*, 2020, *38* (1), 68–79.

— , "Minnesota-type adaptive hierarchical priors for large Bayesian VARs," *International Journal of Forecasting*, 2021.

— and Angelia L. Grant, "Fast computation of the deviance information criterion for latent variable models," *Computational Statistics and Data Analysis*, 2016, *100*, 847 – 859.

— and Eric Eisenstat, "Bayesian model comparison for time-varying parameter VARs with stochastic volatility," *Journal of Applied Econometrics*, 2018, *33* (4), 509–532.

— and Ivan Jeliazkov, "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 2009, *1* (1-2), 101–120.

— , Eric Eisenstat, and Rodney W. Strachan, "Reducing the state space dimension in a large TVP-VAR," *Journal of Econometrics*, 2020, *218* (1), 105 – 118.

— , Roberto Leon-Gonzalez, and Rodney W. Strachan, "Invariant Inference and Efficient Computation in the Static Factor Model," *Journal of the American Statistical Association*, 2018, *113* (522), 819–828.

— , Todd E. Clark, and Gary Koop, "A New Model of Inflation, Trend Inflation, and Long-Run Inflation Expectations," *Journal of Money, Credit and Banking*, 2018, *50* (1), 5–53.

Chib, Siddhartha, "Markov Chain Monte Carlo Methods: Computation and Inference," in James J. Heckman and Edward Leamer, eds., *Handbook of Econometrics*, Vol. 5, Elsevier, 2001, pp. 3569 – 3649.

— , "Introduction to Simulation and MCMC Methods," in "The Oxford Handbook of Bayesian Econometrics" 2011, pp. 183–217.

Clark, Todd E. and Michael W. McCracken, "Tests of Predictive Ability for Vector Autoregressions Used for Conditional Forecasting," *Journal of Applied Econometrics*, 2017, *32* (3), 533–553.

Conti, Gabriella, Sylvia Frühwirth-Schnatter, James J. Heckman, and Rémi Piatek, "Bayesian exploratory factor analysis," *Journal of Econometrics*, 2014, *183* (1), 31–57.

Cross, Jamie L., Chenghan Hou, and Aubrey Poon, "Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity," *International Journal of Forecasting*, 2020, *36* (3), 899–915.

Del Negro, Marco and Frank Schorfheide, "DSGE Model-Based Forecasting," in Graham Elliott and Allan Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 2 of *Handbook of Economic Forecasting*, Elsevier, 2013, pp. 57–140.

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

Durbin, James and Siem Jan Koopman, "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 2002, *89* (3), 603–616.

Faust, Jon and Jonathan H. Wright, "Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset," *Journal of Business & Economic Statistics*, 2009, *27* (4), 468–479.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer, "The Next Generation of the Penn World Table," *American Economic Review*, October 2015, *105* (10), 3150–82.

Francis, Neville, Michael T. Owyang, and Ozge Savascin, "An endogenously clustered factor approach to international business cycles," *Journal of Applied Econometrics*, 2017, *32* (7), 1261–1276.

Frühwirth-Schnatter, Sylvia, "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 1994, *15* (2), 183–202.

Ganics, Gergely and Florens Odendahl, "Bayesian VAR forecasts, survey information, and structural change in the euro area," *International Journal of Forecasting*, 2021, *37* (2), 971–999.

George, Edward I. and Robert E. McCulloch, "Variable selection via Gibbs sampling," *Journal of the American Statistical Association*, 1993, *88* (423), 881–889.

＿ and ＿ , "Approaches for Bayesian variable selection," *Statistica Sinica*, 1997, *7*, 339–373.

Geweke, John, "Variable selection and model comparison in regression," *Bayesian Statistics*, 1996, *5*, 339–373.

＿ and Charles Whiteman, "Bayesian Forecasting," in G. Elliott, C.W.J. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 1, Elsevier, 2006, pp. 3–80.

Ghosh, Josh and David B. Dunson, "Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis," *Journal of Computational and Graphical Statistics*, 2009, *18*, 306 – 320.

Giannone, Domenico, Lucrezia Reichlin, and David Small, "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 2008, *55* (4), 665–676.

_ , Michele Lenza, and Giorgio E. Primiceri, "Economic predictions with big data: the illusion of sparsity," Staff Reports 847, Federal Reserve Bank of New York April 2018.

_ , _ , Daphne Momferatou, and Luca Onorante, "Short-term inflation projections: A Bayesian vector autoregressive approach," *International Journal of Forecasting*, 2014, *30* (3), 635–644.

Golub, Gene H. and Charles F. Van Loan, *Matrix Computations* Johns Hopkins Studies in the Mathematical Sciences, 4ᵗʰ ed., Johns Hopkins University Press, 2013.

Grant, Angelia L. and Joshua C.C. Chan, "Reconciling output gaps: Unobserved components model and Hodrick-Prescott filter," *Journal of Economic Dynamics and Control*, 2017, *75*, 114–121.

Hauber, Philipp and Christian Schumacher, "Precision-based sampling with missing observations: A factor model application," Discussion Papers 11/2021, Deutsche Bundesbank 2021.

Huber, Florian and Martin Feldkircher, "Adaptive Shrinkage in Bayesian Vector Autoregressive Models," *Journal of Business & Economic Statistics*, January 2019, *37* (1), 27–39.

Jordan, Alexander, Fabian Krüger, and Sebastian Lerch, "Evaluating Probabilistic Forecasts with scoringRules," *Journal of Statistical Software, Articles*, 2019, *90* (12), 1–37.

Jungbacker, B., S. J. Koopman, and M. van der Wel, "Maximum likelihood estimation for dynamic factor models with missing data," *Journal of Economic Dynamics and Control*, 2011, *35* (8), 1358–1368.

Kastner, Gregor and Florian Huber, "Sparse Bayesian vector autoregressions in huge dimensions," *Journal of Forecasting*, November 2020, *39* (7), 1142–1165.

_ , Sylvia Frühwirth-Schnatter, and Hedibert Freitas Lopes, "Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models," *Journal of Computational and Graphical Statistics*, 2017, *26* (4), 905–917.

Kaufmann, Sylvia and Christian Schumacher, "Identifying relevant and irrelevant variables in sparse factor models," *Journal of Applied Econometrics*, 2017, *32* (6), 1123–1144.

_ and _ , "Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification," *Journal of Econometrics*, 2019, *210* (1), 116 – 134.

Kim, Chang-Jin and Charles R. Nelson, *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, Vol. 1 of *MIT Press Books*, The MIT Press, September 1999.

Knotek, Edward S. and Saeed Zaman, "Financial nowcasts and their usefulness in macroeconomic forecasting," *International Journal of Forecasting*, 2019, *35* (4), 1708–1724.

Koopman, Siem Jan and Andrew Harvey, "Computing observation weights for signal extraction and filtering," *Journal of Economic Dynamics and Control*, 2003, *27* (7), 1317–1333.

Kose, M. Ayhan, Christopher Otrok, and Charles H. Whiteman, "International Business Cycles: World, Region, and Country-Specific Factors," *American Economic Review*, September 2003, *93* (4), 1216–1239.

_ , _ , and _ , "Understanding the evolution of world business cycles," *Journal of International Economics*, 2008, *75* (1), 110 – 130.

Krüger, Fabian, Todd E. Clark, and Francesco Ravazzolo, "Using Entropic Tilting to Combine BVAR Forecasts With External Nowcasts," *Journal of Business & Economic Statistics*, 2017, *35* (3), 470–485.

Kristensen, Johannes Tang, "Diffusion Indexes With Sparse Loadings," *Journal of Business & Economic Statistics*, 2017, *35* (3), 434–451.

Kroese, Dirk P. and Joshua C.C. Chan, *Statistical Modeling and Computation*, Springer Publishing Company, Incorporated, 2013.

Krüger, Fabian, Sebastian Lerch, Thordis L. Thorarinsdottir, and Tilmann Gneiting, "Predictive Inference Based on Markov Chain Monte Carlo Output," *arXiv e-prints*, Aug 2016, p. arXiv:1608.06802.

Kuzin, Vladimir, Massimiliano Marcellino, and Christian Schumacher, "MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro Area," *International Journal of Forecasting*, 2011, *27* (2), 529–542.

_ , _ , and _ , "Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries," *Journal of Applied Econometrics*, 2013, *28* (3), 392–411.

Litterman, Robert B., "Forecasting With Bayesian Vector Autoregressions: Five Years of Experience," *Journal of Business & Economic Statistics*, 1986, *4* (1), 25–38.

Little, Roderick J.A. and Donald B. Rubin, *Statistical analysis with missing data*, 2nd ed., John Wiley & Sons, 2002.

Lopes, Hedibert Freitas and Mike West, "Bayesian model assessment in factor analysis," *Statistica Sinica*, 2004, *14* (1), 41–67.

Makalic, Enes and Daniel F. Schmidt, "A Simple Sampler for the Horseshoe Estimator," *IEEE Signal Processing Letters*, 2016, *23* (1), 179–182.

Marcellino, Massimiliano, "Pooling-Based Data Interpolation and Backdating," *Journal of Time Series Analysis*, 2007, *28* (1), 53–71.

_ , Mario Porqueddu, and Fabrizio Venditti, "Short-Term GDP Forecasting With a Mixed-Frequency Dynamic Factor Model With Stochastic Volatility," *Journal of Business & Economic Statistics*, 2016, *34* (1), 118–127.

Mariano, Roberto S. and Yasutomo Murasawa, "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 2003, *18* (4), 427–443.

McCausland, William J., "The HESSIAN method: Highly efficient simulation smoothing, in a nutshell," *Journal of Econometrics*, 2012, *168* (2), 189–206.

_ , "Dynamic Factor Models with stochastic volatility," 2015. University of Montreal, Department of Economics, mimeo.

_ , Shirley Miller, and Denis Pelletier, "Simulation smoothing for state-space models: A computational efficiency analysis," *Computational Statistics & Data Analysis*, 2011, *55* (1), 199–212.

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

Müller, Ulrich K., James H. Stock, and Mark W. Watson, "An Econometric Model of International Long-run Growth Dynamics," Working Paper 26593, National Bureau of Economic Research 2019.

Nakajima, Jouchi and Mike West, "Bayesian Analysis of Latent Threshold Dynamic Models," *Journal of Business & Economic Statistics*, 2013, *31* (2), 151–164.

Otrok, Christopher and Panayiotis M. Pourpourides, "On the cyclicality of real wages and wage differentials," *The B.E. Journal of Macroeconomics*, 2017, *19* (1), 1–18.

Pati, Debdeep, Anirban Bhattacharya, Natesh S. Pillai, and David B. Dunson, "Posterior contraction in sparse Bayesian factor models for massive covariance matrices," *The Annals of Statistics*, 2014, *42* (3), 1102–1130.

Plummer, Martyn, Nicky Best, Kate Cowles, and Karen Vines, "CODA: Convergence Diagnosis and Output Analysis for MCMC," *R News*, 2006, *6* (1), 7–11.

Polson, Nicholas G. and James G. Scott, "Shrink globally, act locally: Sparse Bayesian regularization and prediction," in J M Bernardo, M J Bayarri, J O Berger, A P Dawid, D Heckerman, A F M Smith, and M West, eds., *Bayesian statistics*, Vol. 9, Oxford University Press, 2010, pp. 501–538.

Rünstler, Gerhard, "On the Design of Data Sets for Forecasting with Dynamic Factor Models," in Eric Hillebrand and Siem Jan Koopman, eds., *Dynamic Factor Models*, Vol. 35 of *Advances in Econometrics*, Emerald Publishing Ltd, 2016, pp. 629–662.

Rubin, Donald B., "Inference and Missing Data," *Biometrika*, 1976, *63* (3), 581–592.

Rue, H. and L. Held, *Gaussian Markov Random Fields: Theory and Applications* Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press, 2005.

Rue, Håvard, "Fast sampling of Gaussian Markov random fields," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2001, *63* (2), 325–338.

Schumacher, Christian, "Forecasting German GDP using alternative factor models based on large datasets," *Journal of Forecasting*, 2007, *26* (4), 271–302.

_ , "Factor forecasting using international targeted predictors: The case of German GDP," *Economics Letters*, May 2010, *107* (2), 95–98.

_ and Jörg Breitung, "Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data," *International Journal of Forecasting*, 2008, *24* (3), 386–398.

Stock, James H. and Mark W. Watson, "Macroeconomic forecasting using diffusion indexes," *Journal of Business & Economic Statistics*, 2002, *20* (2), 147–162.

_ and _ , "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," in John. B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Vol. 2 of *Handbook of Macroeconomics*, Elsevier, 2016, chapter 0, pp. 415–525.

Tallman, Ellis W. and Saeed Zaman, "Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy," *International Journal of Forecasting*, 2020, *36* (2), 373–398.

Thorsrud, Leif Anders, "Words are the New Numbers: A Newsy Coincident Index of the Business Cycle," *Journal of Business & Economic Statistics*, 2020, *38* (2), 393–409.

Waggoner, Daniel F. and Tao Zha, "Conditional Forecasts in Dynamic Multivariate Models," *The Review of Economics and Statistics*, 1999, *81* (4), 639–651.

Wand, Matthew P., John T. Ormerod, Simone A. Padoan, and Rudolf Frühwirth, "Mean field variational Bayes for elaborate distributions," *Bayesian Analysis*, 2011, *6* (4), 847–900.

West, Mike, "Bayesian factor regression models in the large p, small n paradigm," in "Bayesian Statistics," Vol. 7 Oxford University Press 2003, pp. 723–732.

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

Erklärung zum selbständigen Verfassen der Arbeit:

Ich erkläre hiermit, dass ich meine Doktorarbeit *Essays in Bayesian Macroeconometrics and Forecasting* selbstständig und ohne fremde Hilfe angefertigt habe und dass ich als Koautor maßgeblich zu den weiteren Fachartikeln beigetragen habe. Alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehnenden Ausführungen der aufgeführten Beiträge wurden besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert.

[Philipp Hauber]