

Sparse dynamic factor modelling in macroeconometrics

Sylvia Kaufmann

Working Paper 26.05

This discussion paper series represents research work-in-progress and is distributed with the intention to foster discussion. The views herein solely represent those of the authors. No research paper in this series implies agreement by the Study Center Gerzensee and the Swiss National Bank, nor does it imply the policy views, nor potential policy of those institutions.

Sparse dynamic factor modelling in macroeconometrics

Sylvia Kaufmann[*†]

November 2, 2026

Abstract

Factor models condense the common dynamics of high-dimensional data into a few factors. Inducing sparsity in the factor loading matrix identifies strong factor-series associations, so that factors load on distinct groups of series and can be interpreted. The paper reviews an order-invariant Bayesian approach to sparse factor estimation. The model is estimated without pre-imposing rotational identification restrictions, and factor position and sign are identified by post-processing the MCMC output. Since extracted factors are distinct, the restriction of independent factor innovations can be relaxed. The resulting reduced-form state equation, also in a factor augmented VAR with observed factors, allows for further structural analysis. To handle the recent pandemic period, we introduce a Covid factor that loads on the series only from 2020 onwards, while we assume the remaining factors and their dynamics to stay the same. An application to FRED-QD data, spanning the period 2000Q1 to 2024Q1, identifies interpretable group-specific factors in addition to the pervasive ones, and the Covid factor captures the transitory, extreme volatility at the onset of the pandemic across many series. Impulse responses to a policy shock identified by maximizing the forecast error variance of the term spread are consistent with a monetary policy shock, although uncertainty remains considerable. Allowing for weakly correlated idiosyncratic innovations is a promising extension that would bring parametric sparse factor models closer to approximate factor models.

Keywords: Bayesian, correlation, Covid-19, FAVAR, MCMC, order-invariance, parameter expansion, structural analysis
JEL: C32, C38, C55, E30, E44, E52

[*]Study Center Gerzensee and University of Basel
[†]sylvia.kaufmann@szgerzensee.ch

1 Introduction

A factor model is an appropriate and often applied framework to analyze high-dimensional data. First versions were applied to data sets of moderate size, mostly originating from one study or driven by a small number of latent pervasive factors. For examples related to psychometrics see e.g. Jöreskog (1971), Jöreskog (1969) or Lawley and Maxwell (1971) and Stock and Watson (1989), Kim and Nelson (1998) or Geweke and Zhou (1996) for applications in econometrics. These econometric studies formulated parametric factor models and applied maximum likelihood to extract factors, based on an estimation-maximization algorithm. In econometrics, factor models have since been applied to increasingly large datasets, including heterogeneous data, including as many series as possible to represent specific economic concepts of interest, such as production, prices (inflation), and trade, and financial markets.

In contrast to confirmatory factor analysis, where non-zero factor loadings are usually pre-specified for factor identification, exploratory factor analysis leaves the factor loading matrix unrestricted to estimate underlying factor-series association. One established estimation approach extracts factors by non-parametric principal component analysis (Bai 2003; Stock and Watson 2002b; Stock and Watson 2002a). An alternative approach extracts factors by capturing common frequency components (Forni et al. 2000; Forni et al. 2001). When the size of data is increasingly high-dimensional, a drawback of these approaches is that underlying factors get increasingly blurred as they target the average variance explained across all data. In the worst case, non-informative additional data would only add uncertainty surrounding extracted factors and additional post-processing may be needed to obtain factor interpretation (Bai and Ng 2002; Boivin and Ng 2006; Bai and Ng 2013).

Sparse factor modelling can address these issues. By inducing sparsity in the factor loading matrix, strong factor-series associations are discriminated from weaker ones by shrinking the factor loadings of the latter ones towards zero. Thus, factor loadings of non-informative data are shrunk to zero and eventually, the non-zero loadings will identify interpretable factors, when factors load on distinct groups of series (West 2003; Carvalho et al. 2008; Despois and Doz 2023; Freyaldenhoven 2022). In parametric factor modelling, the usual approach for factor identification has been to pre-impose rotational identification restrictions on the factor loading matrix. However, such restrictions are not order-invariant and pre-imposing them without caring about the variables ordered first in the dataset may seriously distort inference (Millsap 2001; Lopes and West 2004; Chan et al. 2018).

The paper reviews a flexible and general order-invariant approach to parametric, sparse factor estimation and identification (Aßmann et al. 2016; Chan et al. 2018; Kaufmann and Schumacher 2019). To induce sparsity in the factor loading matrix, we can specify

either a hierarchical shrinkage or sparse prior distribution. We obtain draws from the posterior distribution without pre-imposing rotational identification constraints. Eventually, the factor basis is determined by the sparse posterior distribution. The orientation of the basis (rotational identification), including factor sign and position, is determined by post-processing factor draws or factor and factor loading draws in an automated way. Moreover, various simulation studies have documented that the post-processing procedure is useful in determining the number of underlying factors (Kaufmann and Schumacher 2019; Kaufmann and Pape 2026).

The paper reviews extensions of the state equation that are advantageous for further posterior evaluations, such as structural analysis or scenario-based forecasting. One extension relaxes the restriction of independent factor innovations. When a sparse factor loading matrix identifies distinct factors, we can allow for correlated factor innovations. The state equation becomes a full reduced-form system that can be used for structural evaluations, of the factor system itself or targeted variables as well. Any out of the myriad of structural identification procedure can be applied. Including so-called observed factors in the state equation, we obtain a reduced-form system corresponding to a factor augmented vector autoregression (FAVAR), see Beyeler and Kaufmann (2021). Combining a sparse factor loading matrix with an unrestricted factor innovation matrix provides a generalized version of the specification originally proposed in Bernanke et al. (2005).

The paper contributes an extended specification that helps address the econometric challenges of the recent pandemic period. Various approaches have been proposed to account for the transitory, extreme volatility in time series. In VAR modelling, Lenza and Primiceri (2022) propose to dummy out the extreme observations, assuming that the underlying dynamics remain unchanged. In factor modelling, outliers could be taken into account explicitly in each series, as proposed by Antolin-Diaz et al. (2024). However, extreme observations may show up synchronously across many data series. We therefore introduce a Covid factor that loads on series only from the start of the pandemic onwards. We obtain posterior inference in the spirit of Stock and Watson (2025), by assuming that underlying factors and factor dynamics remain relevant and the same throughout the Covid period.

In the following section, we specify the econometric model and discuss identification issues. Section 3 outlines the Bayesian framework. We discuss a hierarchical shrinkage and a sparse prior that induce sparsity in the factor loading matrix. Posterior inference is obtained by iterating over standard Gibbs sampling steps. Section 4 outlines the post-processing algorithms with which we obtain factor identification, including factor position and sign. Section 5 outlines generalizations to a full factor innovation correlation matrix and a reduced-form FAVAR. We discuss the extended specification including the Covid factor and give details to the estimation approach. In a final subsection, we mention briefly extensions to extract trending factors and give suggestions for future research, such as modelling cointegration in an encompassing framework or allowing weak correlation across idiosyncratic error terms. An application to FRED-QD data, spanning the period

2000, first quarter, to 2024, first quarter, provides an illustration in Section 6. Section 7 concludes.

2 Specification

2.1 The econometric model

Let y_t denote the period-t vector of N observed units y_{it}, $i = 1, \ldots, , N$ and $t = 1, \ldots, T$. Individual time series $y_i = (y_{i1}, \ldots, y_{iT})'$ are non-trending, mean-adjusted and standardized. To capture common dynamic features underlying the series-specific processes, we specify a factor model:

$$
\begin{aligned}
y_t &= \Lambda f_t + \varepsilon_t, && (1)\\
f_t &= \Phi f_{t-1} + \eta_t, \eta_t \sim i.i.d.\ N\left(0, I_K\right), && (2)\\
\varepsilon_t &= \Psi \varepsilon_{t-1} + \epsilon_t,\ \epsilon_t \sim i.i.d.\ N\left(0, \Sigma\right), && (3)
\end{aligned}
$$

where f_t is a $K \times 1$, $K \ll N$, vector of common latent factors, which follow a full vector autoregressive process characterised by Φ. Series-specific factor loadings are collected in Λ, which stacks series-specific factor loadings $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{iK})$. For expositional convenience we restrict the dynamic processes to order 1 for factors f_t and idiosyncratic components ε_t. This is not restrictive as every (vector) autoregressive ((V)AR) process of order p can be cast into a companion (V)AR process of order 1 by appropriately expanding the state vector and re-parameterizing the system matrices. The baseline specification (1)-(3) represents an exact factor model, where we assume that all common dynamics are captured by factors. This implies that idiosyncratic components ε_{it} are independent across units and independent of factor innovations, i.e. $E(\varepsilon_{it}\varepsilon_{jt}) = 0$, $i \neq j$, and $E(\varepsilon_{it}\eta_{kt}) = 0$, $\forall i, k$. Hence, the matrices Ψ and Σ are diagonal with elements ψ_i and σ_i^2, respectively.

To discuss identification, we use the geometric representation of a factor model (Lawley and Maxwell 1971), where series-specific factor loadings $\lambda_i = (\lambda_{i1}, \ldots \lambda_{iK})$ each represent a point in a Cartesian coordinate system and the covariance matrix of factor innovations (the conditional covariance of factors), $E(\eta_t\eta_t') = E(f_t f_t' | f_{t-1})$, provides a (factor) basis to that coordinate system, see e.g. Figure 2. As such, the identity for the innovations' covariance matrix in (2) defines an orthogonal factor basis and normalizes the factor scale, which allows us to leave the factor loadings unrestricted. It is obvious that (1)-(2) needs additional restrictions to fix the orientation of the factor basis, i.e. rotational identification, if the interpretation of factors is of interest. A more detailed and illustrative discussion on identification is provided in the next subsection, where we also motivate obtaining rotational identification after estimation, rather than imposing it beforehand as done traditionally.

In exploratory factor analysis, factor loadings usually remain unrestricted, in order to estimate rather than pre-impose factor-series association. Interpreting Λ as a sparse matrix and inducing sparsity when estimating the model, we will obtain an estimate of Λ which will reflect the series-specific importance of factors. Important factors, i.e. factors explaining a considerable share of data variation, will load strongly on respective series, while loadings of unimportant factors will be weak or shrunk towards zero. Eventually, the non-zero loadings in columns of Λ will determine factor interpretation. For example, a factor loading strongly on a group of series related to production, prices, countries or states may be interpreted as a business cycle, inflation or regional factor, respectively.

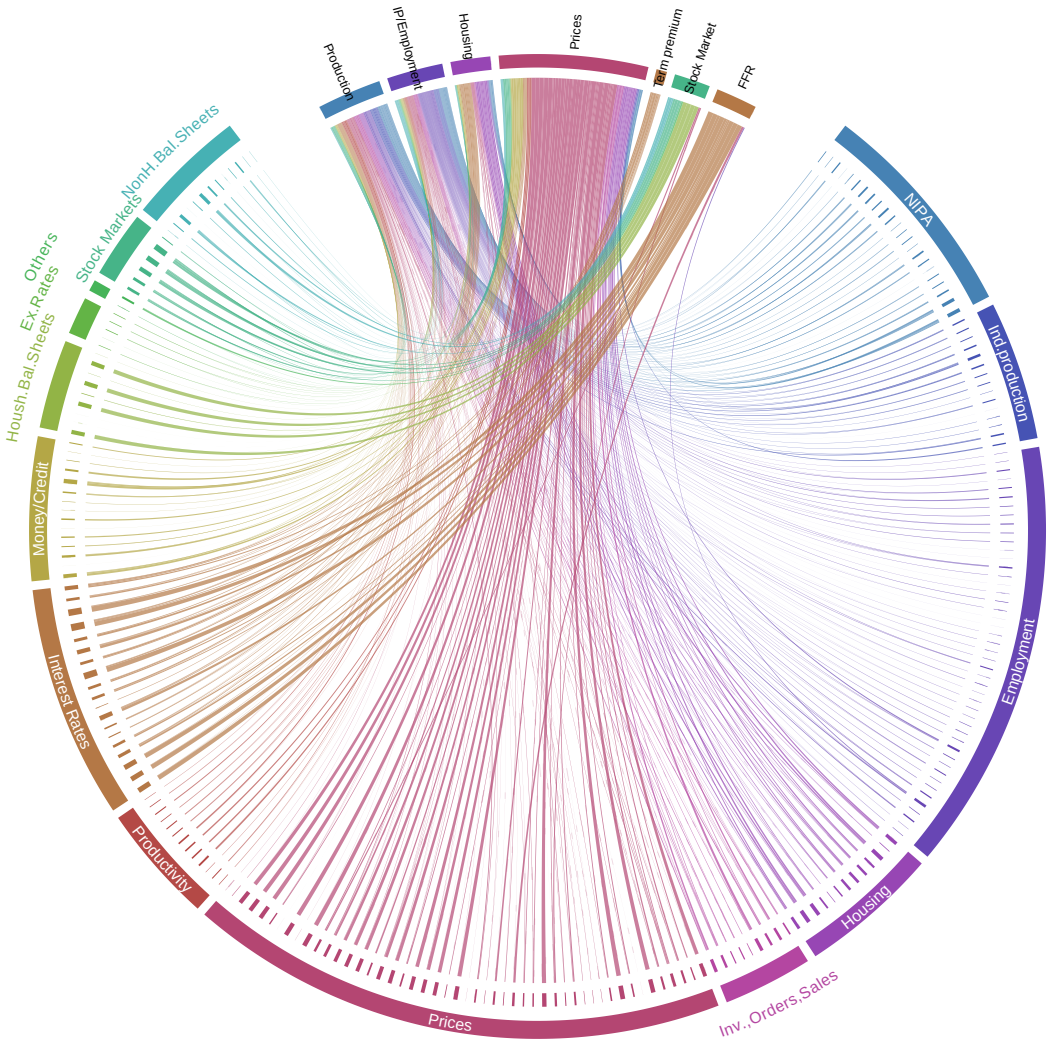
For example, Figure 1 displays a circos graph,[1] based on a sparse factor augmented VAR (FAVAR) model estimated for the FRED-QD dataset (McCracken and Ng 2021), consisting of 242 series spanning the period 2000, first quarter, to 2024, first quarter, representing various real and financial aspects of the US economy. The estimate includes a Covid factor, which loads only from the start of the Covid period onwards in 2020Q1 (Stock and Watson 2025). The factor loads on most series, and therefore is not included in the circos graph. The specification of the FAVAR is detailed in Section 6 and is similar to the one used in Beyeler and Kaufmann (2021). The thickness of chords in the graph corresponds to the posterior absolute mean of non-zero factor loadings. The length of each factor arch (top middle) is proportional to the number and the strength of non-zero loadings, and the color corresponds to the color of the group of series mainly loaded by the factor. The sample period includes two challenging periods, the financial crisis and the Covid pandemic. The sparse estimate of the factor loading matrix identifies distinct and economically interpretable unobserved factors. Factors 1, 2 and 4 represent the main pervasive factors usually extracted from this dataset, corresponding to economic (Production and employment) and price (inflation) conditions, respectively. We are able to interpret additional, weaker factors, which capture underlying dynamics of group-specific series. In particular Factors 3, 5 and 6 load most strongly on housing, interest rate spreads and stock market indices, respectively. The last factor is the federal funds rate, i.e. the observed factor included in the FAVAR. We observe that it mainly loads on interest rates. Given the relative clear interpretation of the factors, the model estimate can be used for further structural analysis or scenario-based forecasting.

2.2 Identification

Assuming the identity for the conditional covariance matrix of factors in (1)-(2), defines an orthogonal factor basis with unit length basis vectors. Nevertheless, representation (1)-(2) needs additional restrictions to determine the orientation of the factor basis, i.e. rotational identification, as for any $K \times K$ orthogonal matrix H there is an observationally

[1] Krzywinski et al. (2009)

Figure 1: FRED-QD: Circos graph. Factor-series association. The thickness of chords corresponds to the absolute posterior mean of non-zero factor loadings, $\hat{\lambda}_{ik} = |E(\lambda_{ik}|\boldsymbol{y}, \beta_{ik} \geq 0.5, \cdot)|$. The arch lengths at the factor origin and the series end corresponds to, respectively, $\sum_i \hat{\lambda}_{ik}$ and $\sum_k \hat{\lambda}_{ik}$.



equivalent representation:

$$y_t = \Lambda H' H f_t + \varepsilon_t = \tilde{\Lambda} \tilde{f}_t + \varepsilon_t, \quad (4)$$

$$\tilde{f}_t = H \Phi H' \tilde{f}_{t-1} + H\eta_t, \; H\eta_t \sim i.i.d.\; N(0, I_K). \quad (5)$$

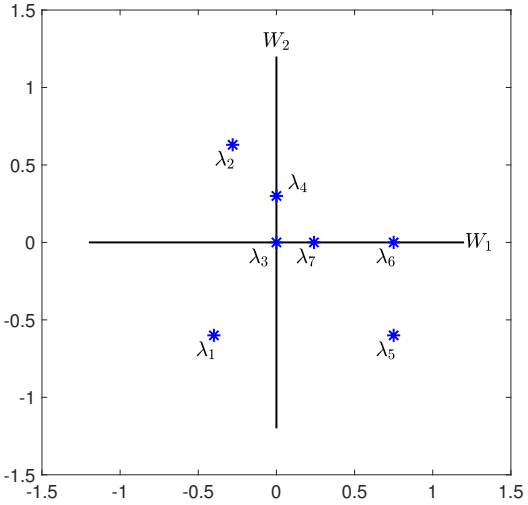
To illustrate intuitively the effects of identification restrictions, we borrow the example in Aßmann et al. (2025) and represent the factor model geometrically, for the case of seven ordered units, $N = 7$, and two factors, $K = 2$. The left panel in Figure 2 displays Λ corresponding to an orthogonal factor basis spanned by axes W_1 and W_2, or $E(f_t f_t' | f_{t-1}) = E(\eta_t \eta_t') = I_K$ as assumed in (2). The right panel displays the factor loadings in the Cartesian plane spanned by $\{W_1, W_2\}$, and illustrates that the current orientation of the basis induces a sparse structure in Λ. That is, some units, Units 1, 2 and 5, are loaded by both factors such that points in the Cartesian plane corresponding to factor loadings lie in-between axes. Some other units, Units 4, 6 and 7, are loaded by only one factor and the points corresponding to their factor loadings lie on one of the axes. Finally, Unit 3 is not loaded by any factor and the corresponding point lies at the origin of the basis. By choosing this orientation, the first unit is loaded negatively by both factors. The orientation also identifies factor position and sign, where W_1 (W_2) identifies the first (second) factor, loading uniquely and positively on Units 6 and 7 (Unit 4). Permuting axis labels W_1 and W_2 would correspond to permuting columns in Λ. Reflecting an axis about zero would induce a sign switch in the corresponding column of Λ. The current orientation induces the first factor to load positively on a majority of units, whereas the sign of the second factor is chosen to load positively on Units 2 and 4. Eventually, the current orientation would yield a factor interpretation.

The structure of Λ in Figure 2 does not correspond to restrictions usually imposed before estimation to obtain rotational identification. Figure 3 illustrates the effects of pre-imposing such restrictions on the factor basis and the loading matrix in our example. One set of restrictions imposes positive and zero coefficients on, respectively, diagonal and upper-diagonal elements of the leading $K \times K$ sub-matrix of Λ, $\lambda_{kk} > 0$, $\lambda_{ik} = 0$, $i > k$, $i, k = 1, \ldots, K$. The induced effects are illustrated in the first line of Figure 3. The restrictions induce a rotation of the factor basis from $\{W_1, W_2\}$ to $\{\widetilde{W}_1, \widetilde{W}_2\}$, obtained by a counterclockwise rotation (67.34 degrees) and a reflection of the first axis around zero. After rotation, factor loadings λ_i in $\Lambda_{\{\widetilde{W}_1, \widetilde{W}_2\}}$ reflect coordinates with respect to the basis spanned by $\{\tilde{W}_1, \tilde{W}_2\}$. Note that the restriction could be fulfilled by permuting units, e.g. $\rho = (6, 2, 1, 3, 4, 5, 7)$, without the need to rotate axes. On the other hand, Unit 3 should not be used to identify any factor, i.e. should not be ranked in the leading two rows of Λ, as the series does not lie in the factor space.

Another set of restrictions usually defines a $K \times K$ leading, diagonal matrix, $\lambda_{kk} > 0$ and $\lambda_{ik} = 0$, for $i \neq k$, $i, k = 1 \ldots K$. The second line in the figure displays the induced effects of these restrictions. They induce a negatively correlated factor basis $\{\widehat{W}_1, \widehat{W}_2\}$, with unit

Figure 2: A factor loading matrix and its representation in a Cartesian plane. The axis labels $\{W_1, W_2\}$ indicate the positive half of the axes.

$$\Lambda_{\{W_1,W_2\}} = \begin{bmatrix} -0.40 & -0.60 \\ -0.28 & 0.63 \\ 0 & 0 \\ 0 & 0.30 \\ 0.75 & -0.60 \\ 0.75 & 0 \\ 0.24 & 0 \end{bmatrix}$$

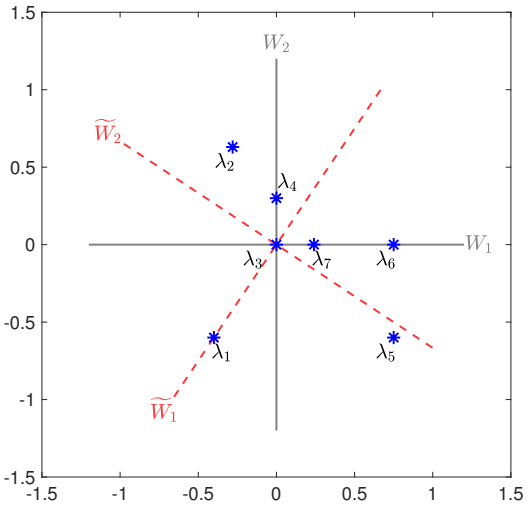


length basis vectors. In our example, the observationally equivalent representation implies correlated factor innovations, i.e. $E(\eta_t \eta_t') = \Omega = \begin{bmatrix} 1 & -0.54 \\ -0.54 & 1 \end{bmatrix}$, where the correlation -0.54 corresponds to the cosine of the angle between the axes. Note that again, the restriction could be fulfilled by permuting units in $\Lambda_{\{W_1,W_2\}}$, e.g. $\rho = (6, 4, 1, 2, 3, 5, 7)$, without the need to rotate axes. Finally, defining leading units as the factors, $\lambda_{kk} = 1$ and $\lambda_{ik} = 0$, for $i \neq k$, $i, k = 1 \dots K$ induces a correlated basis with basis vectors of different lengths. In Figure 4, the basis vectors extend to λ_1 and λ_2. The observationally equivalent representation implies an unrestricted factor innovation covariance matrix $\hat{\Omega} = \begin{bmatrix} 0.52 & -0.27 \\ -0.27 & 0.48 \end{bmatrix}$.

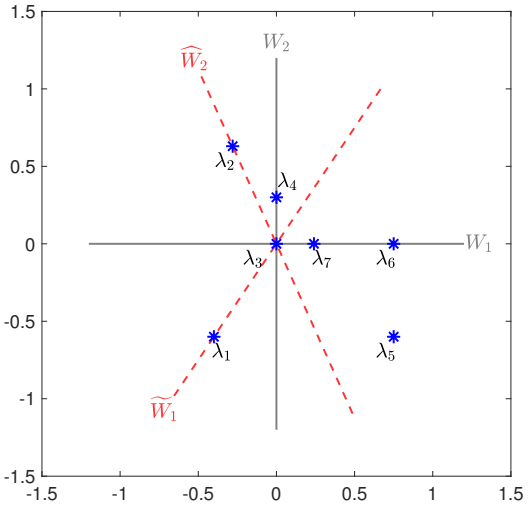
In exploratory sparse factor analysis, shrinkage is usually induced to discriminate between strong and weak factor-unit association. Pre-imposing rotational restrictions in this context can seriously bias inference. Imposing the lower positive triangular restriction to infer the sparse structure underlying the original matrix $\Lambda_{\{W_1,W_2\}}$ induces a matrix $\Lambda_{\{\widetilde{W}_1,\widetilde{W}_2\}}$ with blurred underlying sparsity (Figure 3, first line). Potentially, additional shrinking of small coefficients to zero will substantially bias the inference on the underlying sparsity and affect factor interpretations. The second line in Figure 4 shows that Unit 4 would be loaded by Factor 1, whereas it originally is loaded by Factor 2. Unit 5 would be shrunk to the space of Factor 2 only, and Unit 7 apparently would not be affected by any factor,

Figure 3: Identification restrictions. The axis labels indicate the positive half of the axes. First line: $\lambda_{kk} > 0$, $\lambda_{ik} = 0$, $i < k$. Induces a rotated factor basis. Second line: $\lambda_{kk} > 0$, $\lambda_{ik} = 0$, $i \neq k$, $i = 1, \ldots, K$. Induces a rotated and (negatively) correlated factor basis.

$$\Lambda_{\{\widetilde{W}_1, \widetilde{W}_2\}} = \begin{bmatrix} 0.72 & 0 \\ -0.37 & 0.58 \\ 0 & 0 \\ -0.25 & 0.17 \\ 0.08 & -0.96 \\ -0.42 & -0.62 \\ -0.13 & 0.20 \end{bmatrix}$$



$$\Lambda_{\{\widetilde{W}_1, \widehat{W}_2\}} = \begin{bmatrix} 0.72 & 0 \\ 0 & 0.69 \\ 0 & 0 \\ -0.14 & 0.20 \\ -0.52 & -1.13 \\ -0.81 & -0.74 \\ -0.26 & -0.24 \end{bmatrix}$$



while originally it is loaded by Factor 1.

We draw two conclusions from the example. First, the normalization of the factor scale and the assumption of orthogonal factor innovations, $E(\eta_t \eta_t') = I_K$, defines an orthogonal factor basis, without pre-determining its orientation. In this sense, it is an unrestrictive assumption that can be pre-imposed without pre-determining factor identification or interpretation. Second, pre-imposed rotational identification restrictions are not order-invariant, in the sense that different orderings in leading variables pre-determine potentially very different factor interpretations. This effect is exacerbated in sparse factor analysis.

Non order-invariance of rotational identification restrictions has been pointed out by several authors, such as Aguilar and West (2000), Millsap (2001) or Lopes and West (2004). Imposing such restrictions needs a careful selection of the "factor founding" K series ranked first in the data set, requiring them to jointly span the factor space, or represent factor interpretations of interest. Carvalho et al. (2008) designed an approach where the sequential identification of so-called factor founders was an integral part of the estimation process. Re-ordering them into leading positions, eventually induced upper-triangular zero and diagonal positive identification restrictions.

Advances in sparse factor modelling (Ročková and George 2016; Frühwirth-Schnatter et al. 2025) have relaxed the positive diagonal-upper zero triangular restriction to a generalized structure, where the first series i loaded positively by factor k (with loading $\lambda_{ik} > 0$) should be higher-order ranked than the first series i' loaded by factor $k-1$, i.e. $i' < i$ for $\lambda_{i',k-1}$. Nevertheless, as for traditional restrictions this generalization is not order-invariant, either.

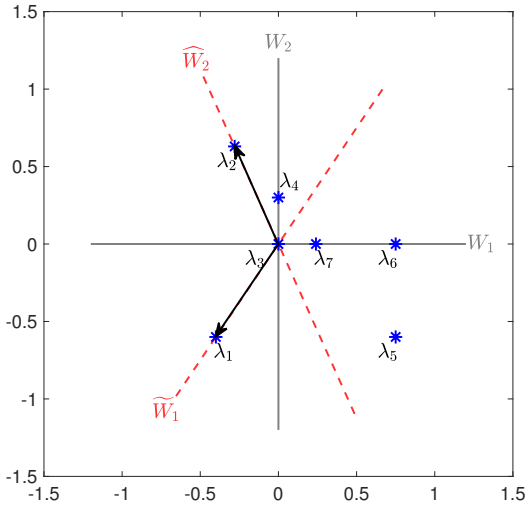
In a high-dimensional data set, it may generally be difficult to identify the leading, factor-founding series, the more so if the number of factors is unknown a priori. Taking non-invariance into consideration, Aßmann et al. (2016), Chan et al. (2018) and Kaufmann and Schumacher (2019) derived order-invariant estimation approaches, avoiding pre-imposing restrictions and obtaining rotational identification by processing the posterior output. We follow this approach in the present paper.

Rather than choosing a specific variable ordering and pre-imposing rotational restrictions, we will induce zero or near-zero factor loadings by specifying a shrinkage prior and estimate the strong or non-zero loadings using data information. An estimated sparse structure for Λ will most probably be destroyed by any rotation H and therefore, rotational identification will be evaluated after estimation rather than imposed before estimation. Rotational identification will essentially involve trivial rotations $H = P_{\pm\rho}$[2], which do not destroy the estimated sparse structure. These rotations re-order factor positions, i.e. the columns of

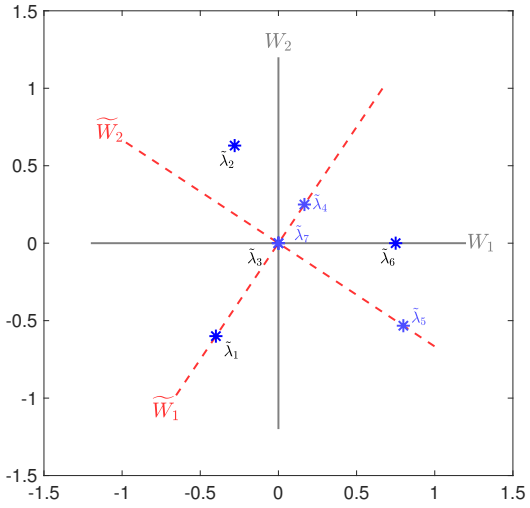
[2] $P_{\pm\rho}$ corresponds to a matrix with potential sign switch of some unit elements in I_K (i.e. substituting -1 for 1) combined with a permutation $\rho = (\rho_1, \ldots, \rho_K)$ of its rows.

Figure 4: Identification restrictions. The axis labels indicate the positive half of the axes. First line: $\lambda_{kk} = 1$, $\lambda_{ik} = 0$, $i < k$. Induces a rotated and (negatively) correlated factor basis, with basis vectors of different lengths (black arrows). Second line: $\lambda_{kk} > 0$, $\lambda_{ik} = 0$, $i < k$, with shrinkage of small coefficients.

$$\Lambda_{\{\widetilde{W}_1,\widetilde{W}_2\}} = \begin{bmatrix} 1.00 & 0 \\ 0 & 1.00 \\ 0 & 0 \\ -0.20 & 0.29 \\ 0.73 & -1.64 \\ -1.13 & -1.07 \\ -0.36 & 0.34 \end{bmatrix}$$



$$\tilde{\Lambda}_{\{\widetilde{W}_1,\widetilde{W}_2\}} = \begin{bmatrix} 0.72 & 0 \\ -0.37 & 0.58 \\ 0 & 0 \\ -0.30 & 0 \\ 0 & -0.96 \\ -0.42 & -0.62 \\ 0 & 0 \end{bmatrix}$$



Λ, and identify factor sign without changing the basic, estimated interpretation of factors.

3 Bayesian framework

Bayesian estimation of parametric factor models relies on standard prior specifications and Markov chain Monte Carlo (MCMC) samplers to simulate from the posterior. The choice of prior specifications is motivated by the structure of high-dimensional economic and financial datasets. These datasets often consist of groups of data containing information on the economic stance in various dimensions, such as production, employment, interest rates, stock markets, trade, etc.. To extract interpretable factors, we use prior distributions for factor loadings inducing shrinkage, such as to discriminate between strong and weak factor-series association. We formulate standard conditional conjugate prior distributions for all remaining parameters. We discuss the prior specification for factor loadings in more details in the next subsection, followed by a description of the MCMC sampler.

3.1 Prior specification

In high-dimensional economic data, it is a stylised fact that usually few factors will capture general underlying economic developments such as business cycle or price (inflation) conditions. These pervasive factors will load, across sub-groups, on many series in the dataset. Several additional, "weaker" factors are usually present, too, which are specific to a sub-group of series or a sector of the economy, reflecting the conditions for example in financial (interest rates, exchange rates) or housing markets. The number of series loaded by these factors may be approximately equal across factors, the groups of loaded series may overlap or be distinct across factors.

To reflect these expectations, we use a prior specification that is factor-invariant, which does not induce factor-specific shrinkage increasing in the number of factors. By allowing for series-specific adaptive factor association, we learn from data whether and which factors can be characterized as pervasive or weaker, i.e. group-specific.

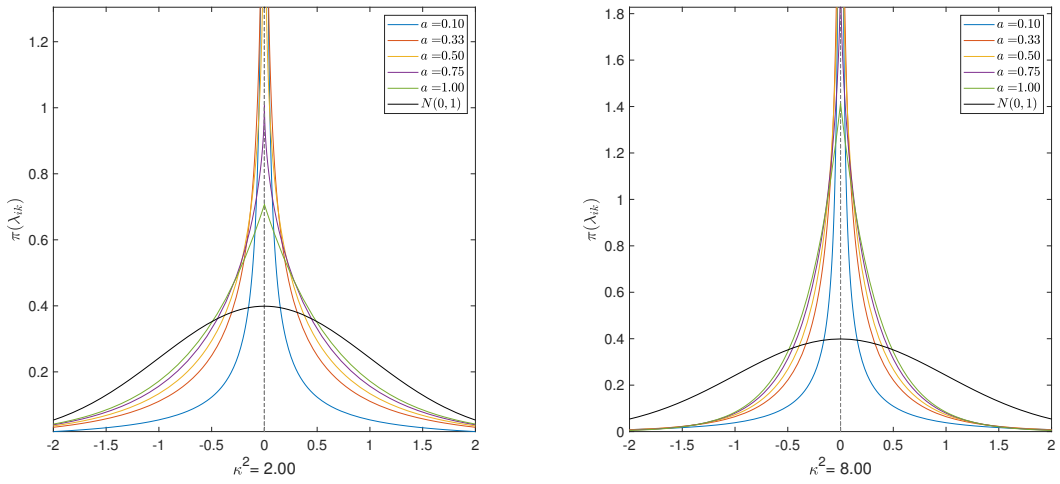
A prior distribution inducing continuous shrinkage is the normal-gamma prior introduced in the context of regression models (Griffin and Brown 2010; Griffin and Brown 2017):

$$\pi\left(\lambda_{ik}|\tau_{ik}\right) = N\left(0, \tau_{ik}\right), \ \pi\left(\tau_{ik}\right) = G(a, a\kappa^2/2), \tag{6}$$

where κ^2 determines global shrinkage and a determines the shape.[3] Shrinkage is locally adaptive as there is a series- and factor-specific posterior update for the prior variance

[3]The specification of the distribution follows Bitto and Frühwirth-Schnatter (2019)

Figure 5: Marginal density $\pi(\lambda_{ik})$ compared with $N(0,1)$



(a) (b)

τ_{ik}. The marginal density of λ_{ik} is available:

$$\pi(\lambda_{ik}) = \frac{\left(\sqrt{a\kappa^2}\right)^{a+1/2}}{\sqrt{\pi}2^{a-1/2}\Gamma(a)}|\lambda_{ik}|^{a-1/2}\mathcal{K}_{a-1/2}\left(\sqrt{a\kappa^2}|\lambda_{ik}|\right),$$

with $\mathcal{K}$ the modified Bessel function of the second kind. Figure 5 plots the density for various settings of κ^2 and a, along with the normal standard distribution. The plots reflect that overall shrinkage increases with κ^2 and a decreasing a renders the distribution increasingly peaked at zero. For $a < 0.5$, the density has a pole at zero. Kaufmann and Strachan (2025) use this prior in a model to extract drifting factors, where factors are independent of each other.

A second kind of prior distribution builds on a point mass-normal mixture distribution (West 2003; Carvalho et al. 2008)

$$\begin{aligned} \pi(\lambda_{ik}|\beta_{ik},\tau_{ik}) &= (1-\beta_{ik})\,\delta_0 + \beta_{ik}N(0,\tau_k),\ \tau_k = IG(g_\tau, G_\tau), \\ \pi(\beta_{ik}|\rho_k) &= (1-\rho_k)\,\delta_0 + \rho_k B(a_0 b_0, a_0(1-b_0)), \\ \pi(\rho_k) &= B(r_0 s_0, r_0(1-s_0)), \end{aligned} \tag{7}$$

where $IG(\cdot)$ and $B(\cdot)$ denote the inverse gamma and beta distribution, respectively. Compared with the normal-gamma prior (6), the point-mass normal prior is hard-thresholding in that it induces exact zeros into the factor loading matrix. We can interpret it as a limiting case of continuous spike and slab prior distributions, introduced for variable selection

in large-dimensional regression models (George and McCulloch 1997). The factor-specific base rate for non-zero factor loadings is represented by ρ_k and governs factor-specific global shrinkage. The hierarchical layer for the probability of a non-zero loading, β_{ik}, introduces a first local-adaptive, series-specific shrinkage. It follows itself a point mass-slab (beta) distribution. Finally, τ_k introduces a second adaptive element, which leads to a factor-specific design of the slab distribution. We can thus learn from the data and adapt the series-specific probability of a non-zero loading and the prior variance of the slab accordingly. When integrating out β_{ik} and τ_k, factor loadings follow a point mass-scaled t distribution with $2g_\tau$ degrees of freedom, $\pi\left(\lambda_{ik}|\rho_k\right) \sim (1-\rho_k b)\delta_0 + \rho_k b\ t_{2g_\tau}\left(0, G_\tau/g_\tau\right)$. The distribution has fatter tails than a normal of equal scale, including a spike at zero.

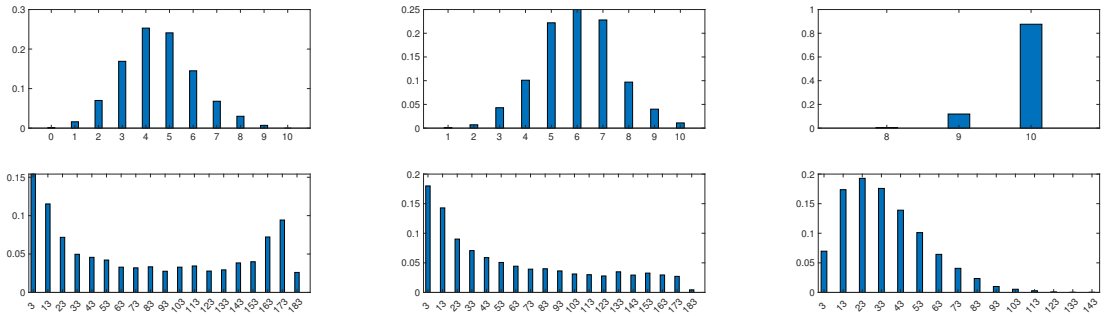
The expected prior probability of a non-zero loading conditional on ρ is $E(\beta_{ik}|\rho_k) = \rho_k b_0$. In a matrix of dimension $N \times K_{\max}$, the number of non-zero factor loadings N_k in each column follows a binomial distribution, $N_k|\rho_k \sim Bin(N, \rho_k b_0)$. By integrating out ρ_k, we can determine the probability of a factor loading on more than two units, $P(N_k > 2) = \int_0^1 P(Bin(N, \rho_k b_0) > 2)B(r_0 s_0, r_0(1-s_0))d\rho$.[4] Finally, over $K_{\max}$ factors, the number of factors loading on more than two units will also follow a binomial distribution $Bin(K_{\max}, P(N_k > 2))$. Figure 6 illustrates the influence of s_0 and r_0 on these distributions by a prior predictive simulation. We simulate 1,000 factor loading matrices of dimension $N \times K_{\max}$ (200×10) out of the hierarchical prior, holding $b_0 = 0.9$ and $a_0 = 0.1$ fixed. The numbers were chosen to correspond to the data dimension we face and settings we use in the empirical illustration. In each panel, the top graph shows $P(K|N_k > 2, K_{\max})$. The bottom graph shows $P(N_k)$ for $N_k > 2$. The graphs lead to following conclusions.

Global shrinkage is determined by s_0, reflecting the expected number of units loaded by a factor a priori. A lower s_0 decreases $P(N_k > 2)$ and leftward shift in the distribution of the number of factors loading on more than two units, i.e. shrinks the number of factors potentially estimated, i.e. $K \leq K_{\max}$. Sparsity within and across factors is determined by r_0. Low values, $r_0 \leq 1$, increase the variance of the base rate across factors. This induces a heterogenous sparse structure, such that factors will load either on many or few units while some will load on a moderate number of units (Panels (a), (b), (d) and (e)). The larger r_0, the more uniformly factors load on units. At the two extremes, $s_0 = 0.2$ and $r_0 = 0.1$ (Panel (a)) induce a very sparse and heterogeneously sparse factor loading matrix, while $s_0 = 0.8$ and $r_0 = 10$ (Panel (f)) induce a relatively dense factor loading matrix. Under this specification, fewer local or group-specific factors may be extracted.

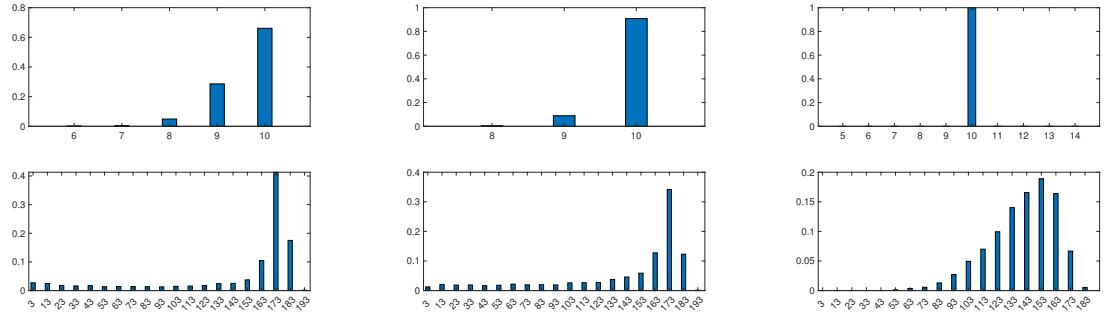
These two prior distributions allow us to design in a flexible way the amount and the structure of sparsity we find appropriate in modelling a specific dataset. There may be

[4]Requiring more than two non-zero loadings on each factor is based on (Anderson and Rubin 1956, Theorem 5.6), which states that at least three loadings must be non-zero to identify one factor. Simple counting shows that a minimum of six distinct covariances is required to identify six unknown parameters of a one-factor model.

Figure 6: Hierarchical sparse prior distribution. Prior predictive simulation of factor loading matrices for $N = 200$, $K_{\max} = 10$, 1,000 repetitions. We set $b_0 = 0.9$ and $a_0 = 0.1$. The evaluations are based on counts of the number of non-zero loadings N_k. The top graph in each panel shows $P(K|N_k > 2, K_{\max})$. The bottom graph shows $P(N_k)$



(a) $s_0 = 0.2$, $r_0 = 0.5$ (b) $s_0 = 0.2$, $r_0 = 1.0$ (c) $s_0 = 0.2$, $r_0 = 10$



(d) $s_0 = 0.8$, $r_0 = 0.5$ (e) $s_0 = 0.8$, $r_0 = 1.0$ (f) $s_0 = 0.8$, $r_0 = 10$

extensions of interest, when data become huge-dimensional or dynamic modelling more sophisticated. An unknown amount of shrinkage in the normal-gamma prior may be learned by specifying hyperprior distributions on κ^2 and a. Bitto and Frühwirth-Schnatter (2019) show that this may be useful for variance and variable selection in large-dimensional sparse dynamic models. In huge-dimensional settings, it may be of interest to learn the number of relevant factors by introducing additional shrinkage along the factor dimension. Those prior distributions are not factor-invariant, though. Bhattacharya and Dunson (2011) introduced the multiplicative gamma process shrinkage prior:

$$\begin{aligned}\pi\left(\lambda_{ik}|\tau_{ik},\kappa_k\right) &= N\left(0,\tau_{ik}\kappa_k\right),\ \tau_{ik}\sim IG(\nu/2,\nu/2),\ \kappa_k^{-1}=\prod_{j=1}^{k}\rho_j,\\ &\rho_1\sim G(a_1,1),\rho_j\sim G(a_2,1),\ j\geq 2,\end{aligned}$$

where κ_k is a global, factor-specific shrinkage parameter. Under the restriction $a_2 > 1$, the κ_ks are stochastically decreasing and induce shrinkage increasing in k. As in the normal-gamma prior (6), τ_{ik} allows for series-specific local-adaptive shrinkage. Legramanti et al. (2020) put a cumulative shrinkage process prior on τ_k

$$\begin{aligned}&\pi\left(\lambda_{ik}|\tau_k\right)=N\left(0,\tau_k\right),\\ &\pi\left(\tau_k|\beta_k\right)=\left(1-\beta_k\right)IG(g_\tau,G_\tau)+\beta_k\delta_{\tau_\infty},\\ &\beta_k=\sum_{l=1}^{k}\rho_l,\ \rho_l=\nu_l\prod_{j=1}^{l-1}\left(1-\nu_j\right),\ \nu_j\sim B(1,\alpha),\end{aligned}$$

where global shrinkage along the factor dimension is stronger with decreasing $\alpha > 0$. Setting τ_∞ to a small, but positive number induces a continuous shrinkage prior on λ_{ik}.

For the remaining parameters we assume standard prior distributions, multivariate normal for factor and normal for idiosyncratic dynamics, respectively, $\pi\left(\text{vec}\left(\Phi\right)\right) = N\left(0,\text{P}_0\right)$, $\pi\left(\psi_i\right) = N\left(0,\text{Q}_0\right)$, $\forall i$, and inverse gamma for idiosyncratic error variances, $\pi\left(\sigma_i^2\right) = IG(g_s,G_{S_i})$, $\forall i$.

The process (2) induces a joint normal prior for factors, $\pi\left(\boldsymbol{f}|\Phi\right) = N\left(0,\text{F}_0\right)$, where $\text{F}_0 = \boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}_f\boldsymbol{\Phi}^{-1\prime}$ is block-diagonal and derived from the implicit, compact representation, see (9) in the following sub-section.

3.2 Compact representation

To expose posterior derivations efficiently, we introduce further notation and write the model in compact form. Collect all data in vector $\boldsymbol{y} = (y_1',\ldots,y_T')'$ and stack factors (including initial conditions) and errors by observation in, respectively, $\boldsymbol{f} = (f_0',\ldots f_T')'$,

$\boldsymbol{\varepsilon} = (\varepsilon_1', \ldots, \varepsilon_t')'$, with similarly defined $\boldsymbol{\eta}$ and $\boldsymbol{\epsilon}$. Model (1)-(3) can then be written as:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{\Lambda}\boldsymbol{\iota}_{\setminus f_0}\boldsymbol{f} + \boldsymbol{\varepsilon}, & (8)\\ \boldsymbol{\Phi}\boldsymbol{f} &= \boldsymbol{\eta},\ \boldsymbol{\eta} \sim N\left(0, \boldsymbol{\Sigma}_f\right), & (9)\\ \boldsymbol{\Psi}\boldsymbol{\varepsilon} &= \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N\left(0, I_T \otimes \Sigma\right), & (10) \end{aligned}$$

where the selection matrix $\boldsymbol{\iota}_{\setminus f_0}$ extracts $(f_1', \ldots, f_T')'$ from $\boldsymbol{f}$ and $\boldsymbol{\Lambda}$ is block-diagonal $\boldsymbol{\Lambda} = I_T \otimes \Lambda$. The autoregressive processes in (9) and (10) are represented in implicit form, with banded matrices

$$\boldsymbol{\Phi} = \begin{bmatrix} I_K & 0 & \ldots & \\ -\Phi & I_K & 0 & \ldots \\ & \ddots & \ddots & \\ & \ldots & -\Phi & I_K \end{bmatrix}, \ \boldsymbol{\Psi} = \begin{bmatrix} I_N & 0 & \ldots & \\ -\Psi & I_N & 0 & \ldots \\ & \ddots & \ddots & \\ & \ldots & -\Psi & I_N \end{bmatrix}. \quad (11)$$

The factor covariance $\boldsymbol{\Sigma}_f = \begin{bmatrix} \Sigma_{f_0} & 0 \\ 0 & I_{KT} \end{bmatrix}$ is block-diagonal, with Σ_{f_0} representing the marginal covariance of initial conditions. Usually, we set it equal to the unconditional variance of the factor process, $E(f_t f_t')$, determined from $\text{vec}\left(\Sigma_f\right) = \left(I_{K^2} - \Phi \otimes \Phi\right)^{-1} \text{vec}\left(I_K\right)$.

Although the dimension of the system is very large, as $\boldsymbol{y}$ and $\boldsymbol{f}$ are of length NT and KT, respectively, the banded nature of system matrices $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ defines a sparse system, which can be exploited and improves computational efficiency considerably.

3.3 Markov chain Monte Carlo sampling

We group the parameters that are relevant for the posterior of factors in $\theta = \{\Lambda, \Phi, \Psi, \Sigma\}$. The parameters governing shrinkage are collected in $\vartheta^\bullet$, with elements determined by the specific shrinkage prior specification for factor loadings. Under the normal-gamma prior $\vartheta^\bullet$ corresponds to $\vartheta^{\text{ng}} = \{\tau^{\text{ng}}\}$, $\tau^{\text{ng}} = \{\tau_{ik} | i = 1, \ldots, N; k = 1, \ldots, K\}$. Under the hierarchical sparse prior, $\vartheta^\bullet$ is $\vartheta^{\text{hs}} = \left\{\tau^{\text{hs}}, \beta, \rho\right\}$, with $\tau^{\text{hs}} = \{\tau_k | k = 1, \ldots, K\}$, $\beta = \{\beta_{ik} | i = 1, \ldots, N; k = 1, \ldots, K\}$ and $\rho = \{\rho_k | k = 1, \ldots K\}$.

The complete data likelihood factorizes

$$L\left(\boldsymbol{y}|\boldsymbol{f}, \theta\right) = \prod_{t=1}^{T} f\left(y_t | f_t, \theta\right), \ f\left(y_t | f_t, \theta\right) = \prod_{i=1}^{N} n\left(\tilde{y}_{it} | \lambda_i \tilde{f}_{it}, \sigma_i\right), \quad (12)$$

where $\tilde{y}_{it} = y_{it} - \psi_i y_{i,t-1}$, $\tilde{f}_{it} = f_t - \psi_i f_{t-1}$ represent series-specific filtered values and $n(v|\text{m}, \text{s})$ indicates the normal density with mean m and variance s evaluated at v.

To obtain a sample from the posterior distribution we sample iteratively from full conditional distributions:

S.1. Sample latent factors conditional on data and parameters, $\pi\left(\boldsymbol{f}|\boldsymbol{y},\theta\right)$.

S.2. Sample parameters conditional on data and factors:

(a) $\pi\left(\Lambda|\boldsymbol{y},\boldsymbol{f},\Psi,\Sigma,\vartheta^{\bullet}\right)$, $\pi\left(\vartheta^{\bullet}|\Lambda\right)$.

(b) $\pi\left(\theta_{\setminus\Lambda}|\boldsymbol{y},\boldsymbol{f},\Lambda\right)$.

S.3. Terminate the iteration by a random permutation $\varrho=(\varrho_1,\ldots,\varrho_K)$ of factor position and sign, permute accordingly factor-specific parameters.

For example, when $K=5$ a permutation of factor position $\rho=(2,1,5,4,3)$, along with a sign-switch of the second and fourth factor is represented by following permutation matrix

$$P_{\pm\rho}=\begin{bmatrix}0&-1&0&0&0\\1&0&0&0&0\\0&0&0&0&1\\0&0&0&-1&0\\0&0&1&0&0\end{bmatrix}\;\rightarrow\;P_{\pm\rho}f_t,\;\Lambda P'_{\pm\rho},\;P_{\pm\rho}\Phi P'_{\pm\rho}.$$

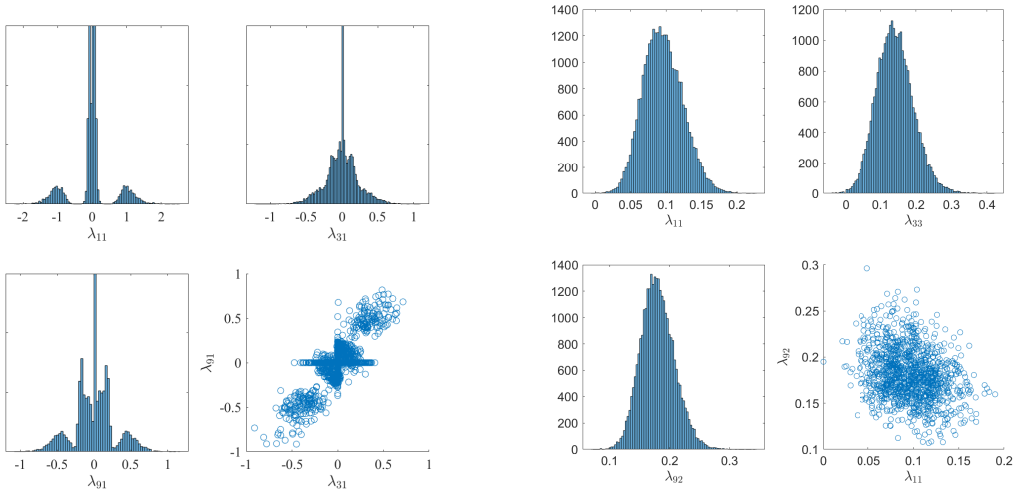
The efficient sampler proposed by Chan and Jeliazkov (2009) is use to obtain a draw of factors (Step **S**.1). In Step **S**.2a, the normal-gamma prior shrinks the loadings of series that are only weakly associated with a factor towards zero, whereas the hierarchical sparse prior induces a sparse posterior. The posterior distributions of remaining parameters $\theta_{\setminus\Lambda}$ in Step **S**.2b are standard. All posterior distributions are derived in Appendix A. Random permutation in Step **S**.3 enforces label and sign switching, which allows us to explore the unconstrained posterior distribution (Frühwirth-Schnatter 2001). Besides improving mixing, it renders the posterior distribution of non-zero loadings multimodal while the one of zero loadings will remain centered at zero. In models of smaller factor dimension, multimodality may help to determine identifying restrictions on the factor loadings.

4 Post-processing for identification

Due to permuting randomly sign and factor position at the end of each sampling iteration, the posterior output will be multimodal with potentially $2^K K!$ modes for factors and factor-specific parameters. Traditional approaches to sort out the draws of a permutation sampler impose inequality restrictions on a single or a group of parameters. In factor modelling such restrictions are usually imposed on factor loadings.

However, in a high-dimensional model of larger factor dimension, with $N\gg K$, it is not obvious to determine which loadings of a specific or a group of series would be appropriate to identify factor positions and sign. Figure 7, Panel (a), plots the MCMC output of

Figure 7: FRED-QD, $K = 7$ unobserved factors: Factor loadings, histograms and scatter plot of (a) unsorted and (b) sorted draws



(a) (b)

the FAVAR estimate for the FRED-QD dataset presented in Subsection 6, including 7 unobserved factors. None of the histograms of selected factor loadings, $(\lambda_{11}, \lambda_{31}, \lambda_{91})$ is suited to identify 7 modes, nor is a combined output across series (see, the bottom left scatter plot in Panel (a)).[5]

As by inducing sparsity in the factor loading matrix we expect to identify specific factor-series associations and factors reflecting distinct patterns, we take an alternative approach. We rely on *factor* rather than loading draws to sort out the posterior output in an automated way. Factor draws sampled from the same posterior distribution will be highly correlated. We exploit this pattern to group or cluster the MCMC factor draws. The means or medians of grouped factor draws identify each a *factor representative*, and provide the basis for re-ordering each draw of factors and factor-specific parameters. Formally, the algorithm proceeds as follows.

Collect all MK factor draws in $\boldsymbol{f}^{MK} = \left\{ f_k^{(m)} | k = 1, \ldots, K; m = 1, \ldots, M \right\}$, with $f_k^{(m)} = \{f_{k1}, \ldots, f_{kT}\}$.

P.1. To determine *factor representatives* f_g, $g = 1, \ldots, G$, pursue either of the following approaches:

[5]Note that because of randomly permuting factor position and signs at the end of each MCMC iteration, the histograms of λ_{ik}, across all k, look similar.

(a) Cluster factor draws based on their correlation (or on k-medoids clustering), determine the mean or median of each group, $\hat{\mathrm{f}}_g$, as estimate of the factor representative, (Kaufmann and Schumacher, 2017, 2019).

(b) Set up a mixture model with $G = K$ components and mixture indicator $z_k^{(m)} = \{1, \ldots, K\}$ to associate each factor draw $f_k^{(m)}$ with one of K components (Kaufmann and Pape 2026):

$$\tilde{f}_k^{(m)} | z_k^{(m)} = g \sim N(\mathrm{f}_g, \mathrm{F}_g), \; g = 1, \ldots, K,$$

where $\tilde{f}_k^{(m)} = \text{sign}(corr(f_k^{(m)}, \mathrm{f}_g)) f_k^{(m)}$ represents the factor draw $f_k^{(m)}$ sign-adjusted to correlate positively with f_g and the mean factor path $\mathrm{f}_g = \{\mathrm{f}_{gt} | t = 1, \ldots, T\}$ of component g is interpreted as *factor representative*.

Formulate a standard mixture prior distribution, $P(z_k^{(m)} = g) = \eta_g$, $\pi\left(\theta^K\right) = \pi\left(\eta_1, \ldots, \eta_K\right) \prod_{g=1}^K \pi\left(\mathrm{f}_g\right) \pi\left(\mathrm{F}_g\right)$, for $\theta^K = \{\eta_g, \mathrm{f}_g, \mathrm{F}_g | g = 1, \ldots, K\}$, and update $\pi\left(\theta^K | \boldsymbol{f}^{MK}\right)$, to obtain posterior inference on the mixture indicator, $\pi\left(z_k^{(m)} | \boldsymbol{f}^{MK}, \theta^K\right)$, and factor representatives $\pi\left(\mathrm{f}_g | \boldsymbol{f}^{MK}\right)$, $\hat{\mathrm{f}}_g = E\left(\mathrm{f}_g | \boldsymbol{f}^{MK}\right)$.

P.2. Remove factor- and sign-switching. Conditional on $\hat{\mathrm{f}}_g$ or $z_k^{(m)}$, post-process the posterior output and retain only those draws (m) corresponding to a unique permutation $\rho = (\rho_1, \ldots, \rho_K)$ of $\{1, \ldots, K\}$, i.e. $\rho_i \neq \rho_j \; \forall i \neq j$. Either:

(a) Re-order each draw (m) according to maximum absolute correlation with the factor representative estimate, $\hat{\mathrm{f}}_g$,

$$f_1^{(m)}, \ldots, f_K^{(m)} := \left\{ f_{\rho_k}^{(m)} \middle| |corr(f_{\rho_k}^{(m)}, \hat{\mathrm{f}}_k)| = \max_{g=1,\ldots,K} |corr(f_{\rho_g}^{(m)}, \hat{\mathrm{f}}_k)|, \; k = 1, \ldots, K \right\}.$$

(b) Re-order each draw (m) in ascending order of mixture indicators

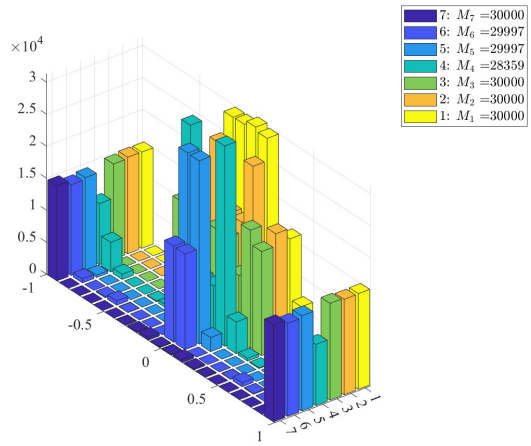
$$f_1^{(m)}, \ldots, f_K^{(m)} := \left\{ f_{z_k^{(m)}}^{(m)} | k = 1, \ldots, K, \; z_1^{(m)} < \cdots < z_K^{(m)} \right\}.$$

Adjust the sign of draws negatively correlated with the factor representative, $f_k^{(m)} := \text{sign}\left(corr(f_k^{(m)}, \hat{\mathrm{f}}_k)\right) f_k^{(m)}$, $k = 1, \ldots, K$.

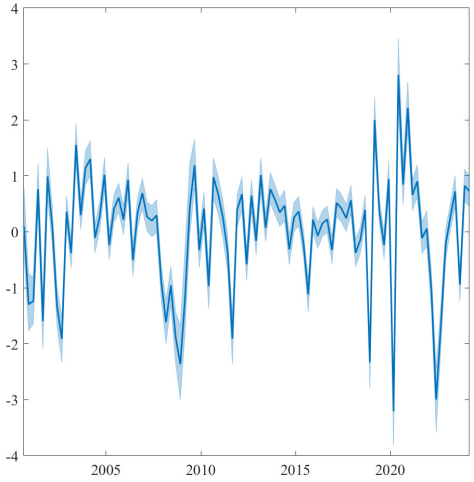
Permute and sign-adjust factor loadings and factor-specific parameters accordingly.

P.3. Determine overall factor order and sign. Based on posterior statistics such as the 95% HPDI, re-order factors in decreasing order of e.g. the number of *exclusive* or *non-zero* loadings; sign-adjust factors so that most factor-specific loadings are positive.

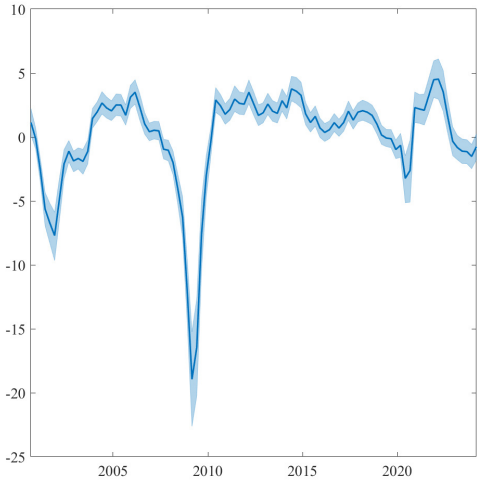
Figure 8: FRED-QD, $K = 7$ unobserved factors: Unsorted factor draws. (a) Histogram of $corr(f^{(1)}, f^{(m)})$, $m = 1, \ldots, M^{(g)}$, at sorting iteration g; the legend gives the number of draws grouped per iteration, M_g, to determine a factor representative ($M^{(1)} = MK$, $M^{(g)} = M^{(g-1))} - M_{g-1}$). (b)–(d) Mean ($f_g$) and 95% HPDI of the grouped, sign-adjusted draws, $g \in \{1, 2, 6\}$, labelled by the interpretation assigned to the factor (see also Figure 1).



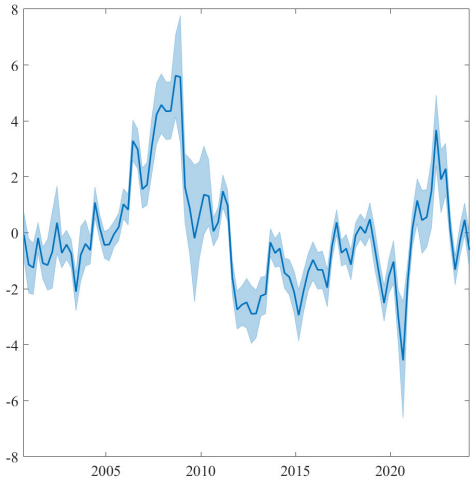
(a)



(b) Stock markets

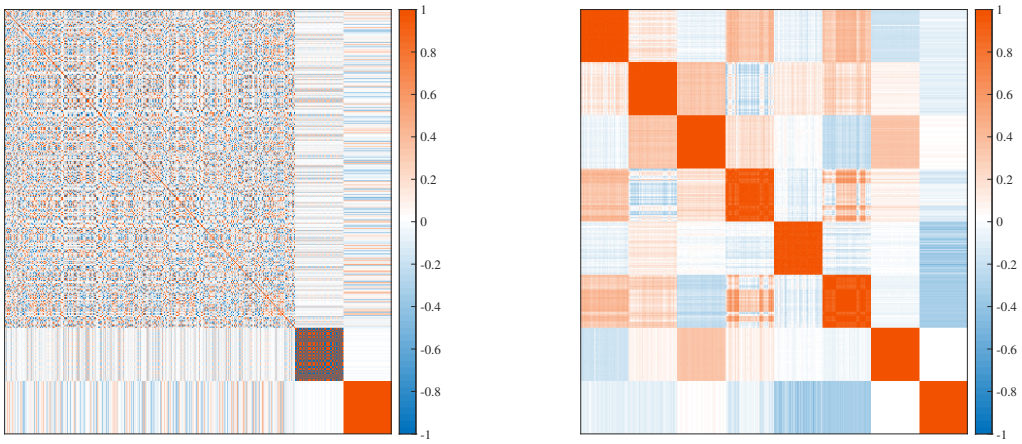


(c) Employment/IP



(d) Housing (sign-reversed)

Figure 9: FRED-QD, $K = 7$ unobserved factors and the federal funds rate as observed factor: Unsorted (left) and sorted (right) factor draws, scatter plot of correlations across selected factor draws.



Remark 1. Step **P**.1 potentially addresses the issue of factor duplication or redundancy when the number of estimated factors, K', overfits the number of underlying factors, $K < K'$. In the first case, a mixture model with a prior specification $\pi(\eta)$ allowing for empty components $G > K$, $G = K'$, can reveal that a mixture with a lower number than K' components underlies the posterior output, i.e. that fewer than K' factors underly the data (Rousseau and Mengersen 2011). Redundancy would be reflected in factors not loading on any series, as data would be uninformative about overfitting factors and their loadings shrink toward zero. Both cases have been illustrated with simulation studies in Kaufmann and Schumacher (2019) and Kaufmann and Pape (2026).

Remark 2. If K corresponds to the number of underlying factors, the posterior output could also be post-processed using the weighted orthogonal Procrustes (WOP) procedure applied in Aßmann et al. (2016), which minimizes the distance between draws by orthogonal rotation. The approach should conserve the sparse factor loading structure, as rotations should essentially reflect permutations of factor position and sign, $P_{\pm\rho}$. See Eckert et al. (2025) for a recent application.

Figure 8 illustrates the identification of factor representatives based on highly correlated draws (Step **P**.1a). Panel (a) displays the histogram of correlations across factor draws remaining at clustering iteration g, $M^{(g)}$, with $M^{(1)} = MK$. At each iteration, we group and extract highly correlated factor draws (of absolute correlation larger than 0.8). The legend indicates the number of draws extracted, M_g. Each histogram is plotted for re-

maining draws $M^{(g)} = M^{(g-1)} - M_{g-1}$. We observe that virtually all 30,000 retained posterior draws of 7 factors ($M^{(1)} = 210,000$) can be assigned to seven distinct groups. For illustration, Panel (b) to (d) plot the mean (as factor representative) and the 95% HPDI for sign-adjusted, clustered factor draws at iteration $g \in \{1, 2, 6\}$. The panels are labelled by the interpretation attached to each factor based on the group of series they mainly load onto (see also Figure 8.

Figure 7, Panel (b) illustrates that the post-processing procedure is successful in identifying factor position and sign. All displayed histograms and the scatter plot reflect unimodal posterior distributions shifted away from zero. Finally, Figure 9 displays a scatter plot across a sample of factor draws. The left panel shows correlations across unsorted draws, where the right-bottom red square refers to observed Factor 8, the federal funds rate, which is not permuted with other factors. Factor 7 corresponds to the Covid factor, which is only sign-permuted at each iteration. The right scatter plot confirms that post-processed factor draws are well separated and identify distinct factor patterns.

5 Generalizations

By inducing sparsity in the factor loading matrix, we can expect to extract distinct and interpretable factors, identified through their strong series-factor associations. This allows us to explore various extensions. In the following, we drop the restriction of independent factor innovations and estimate a full correlation matrix. An estimate of a fully unrestricted, reduced-form state system can then be used for further structural posterior evaluations (Subsection 5.1). It can also be used for scenario-based forecasting, to analyze how alternative economic and financial scenarios affect forecasts of GDP growth or inflation. After describing the extended sampler, we discuss the extension to a FAVAR system in Subsection 5.2. Subsection 5.3 outlines the adjustment of the MCMC sampler to obtain an estimate including a Covid factor loading only from the start of 2020. The proposal yields an estimate in the spirit of Stock and Watson (2025). Finally, other specifications such as extracting drifting factors, modelling time-varying drifts in factors, and suggestions for further research conclude this section.

5.1 Full factor correlation matrix

By inducing sparsity in the factor loading matrix, we eventually obtain a posterior inference on well-identified factors. This allows relaxing conditional independence across factors in the state equation (2):

$$f_t \;=\; \Phi f_{t-1} + \eta_t,\ \eta_t \sim N\left(0, \Omega\right), \tag{13}$$

with $\Omega = \begin{bmatrix} 1 & & * \\ & \ddots & \\ * & & 1 \end{bmatrix}$ a full correlation matrix, i.e. with ones on the diagonal. After estimation, this reduced-form factor model makes the system amenable to a range of established procedures for structural identification, such as short-run, long-run, sign, and maximum forecast error variance identification restrictions.

The parameter space is expanded by Ω, $\theta = \{\Lambda, \Phi, \Omega, \Psi, \Sigma\}$, inducing a generalized prior distribution for factors, which can be derived as outlined in Subsection 3.2: $\pi(\boldsymbol{f}|\Phi, \Omega) = N(0, \mathrm{F}_0)$ with $\mathrm{F}_0 = \boldsymbol{\Phi}^{-1}\boldsymbol{\Omega}\boldsymbol{\Phi}^{-1\prime}$ block-diagonal,

$$\boldsymbol{\Omega} = \begin{bmatrix} \Sigma_f & 0 \\ 0 & I_T \otimes \Omega \end{bmatrix}, \ \Sigma_f \text{ implied by } \operatorname{vec}(\Sigma_f) = (I_{K^2} - \Phi \otimes \Phi)^{-1} \operatorname{vec}(\Omega). \tag{14}$$

It is straightforward to derive the posterior distribution of factors conditional on the generalized prior, and Step **S**.1 of the MCMC sampler remains unchanged. When sampling parameters, Step **S**.2b includes an additional step, sampling from $\pi(\Omega|\boldsymbol{f}, \Phi)$. Formulating a prior distribution for a correlation matrix, where off-diagonals are restricted on the interval $(-1, 1)$ and diagonals are fixed to 1, is not obvious, neither so the derivation of the posterior distribution. Rather than relying on a Metropolis-Hastings step (Zhang et al. 2006; Liu 2008), we apply parameter expansion along the lines of Conti et al. (2014) to facilitate posterior sampling. The idea is to expand the state equation by a factor-specific variance, introduced as a working parameter for each factor. The expansion induces an inverse wishart prior for the unscaled covariance matrix of the factor innovations, which is updated in the standard way to obtain a draw of the full covariance matrix. By rescaling the covariance matrix to a correlation matrix, and the factors and factor-specific parameters accordingly, we return to the original scaled system. Formally:

S′.2. Sample parameters conditional on data and factors.

(b) $\pi\left(\theta_{\setminus\Lambda}|\boldsymbol{y}, \boldsymbol{f}, \Lambda\right)$. Apply parameter expansion to draw from $\pi(\Omega|\boldsymbol{f}, \Phi)$:

i. Draw a variance v_k for each factor as a working parameter from a hierarchical inverse gamma distribution (Huang and Wand 2013)

$$v_k|\Omega, s_k \sim IG\left(\frac{\nu + K - 1}{2}, \frac{s_k\omega_k^-}{2}\right), \ s_k \sim G\left(\frac{1}{2}, \frac{1}{2\nu C_k^2}\right), \tag{15}$$

where ω_k^- is the kth diagonal element of Ω^{-1}. Collect the v_k and s_k in diagonal matrices V and S, respectively. Expand to the unscaled system:

$$\hat{\Omega} = V^{\frac{1}{2}}\Omega V^{\frac{1}{2}} \quad \text{with} \quad \begin{matrix} \hat{f}_t = V^{\frac{1}{2}} f_t, \ \hat{\lambda} = \lambda V^{-\frac{1}{2}}, \\ \hat{\Phi} = V^{\frac{1}{2}}\Phi V^{-\frac{1}{2}} \end{matrix}.$$

ii. Update the prior distribution $\pi\left(\hat{\Omega}|\Omega, S\right) \sim IW(\nu + K - 1, S)$ with the likelihood $L\left(\hat{\boldsymbol{f}}|\hat{\Phi}, \hat{\Omega}\right) = \sum_{t=2}^{T} n\left(\hat{f}_t|\hat{\Phi}\hat{f}_{t-1}, \hat{\Omega}\right)$ and draw from the posterior:

$$\pi\left(\hat{\Omega}|S, \hat{\boldsymbol{f}}, \hat{\Phi}\right) \sim IW\left(\nu + K - 1 + (T-1), S + \sum_{t=2}^{T} \hat{\eta}_t \hat{\eta}_t'\right)$$

with $\hat{\eta}_t = \hat{f}_t - \hat{\Phi}\hat{f}_{t-1}$.

iii. Update deterministically each factor variance k by setting it equal to the kth diagonal element of $\hat{\Omega}$, $v_k = \hat{\omega}_k$. Return to the scaled system:

$$\Omega = V^{-\frac{1}{2}}\hat{\Omega}V^{-\frac{1}{2}} \quad \text{with} \quad \begin{array}{l} f_t = V^{-\frac{1}{2}}\hat{f}_t,\ \lambda = \hat{\lambda}V^{\frac{1}{2}}, \\ \Phi = V^{-\frac{1}{2}}\hat{\Phi}V^{\frac{1}{2}} \end{array}.$$

Parameter expansion takes into consideration that a covariance matrix can be decomposed into two parts, a correlation matrix Ω and a matrix V that contains the variances on its diagonal, $\hat{\Omega} = V^{\frac{1}{2}}\Omega V^{\frac{1}{2}}$. Assuming a hierarchical inverse Wishart prior distribution $\hat{\Omega}|S \sim IW(\nu + K - 1, S)$, the joint distribution of V and S can be factored as $p(V, S|\Omega) = p(V|S, \Omega)p(S)$, and through a change of variables we can show that each diagonal element in V, v_k, follows the inverse gamma distribution in (15). Note that the values for s_k may also be fixed. The hierarchical prior increases efficiency by introducing additional sampling variation. In addition, the hierarchical prior allows us specifying a scale-adjusted prior for $\hat{\Omega}$, as C_k^2 is related to the diagonal elements of $\hat{\Omega}$. This is useful when analyzing factor augmented VAR settings, see the next subsection.

Note that the parameter ν in (15) determines the marginal density for correlation $\hat{\omega}_{kj}$, $k \neq j$:

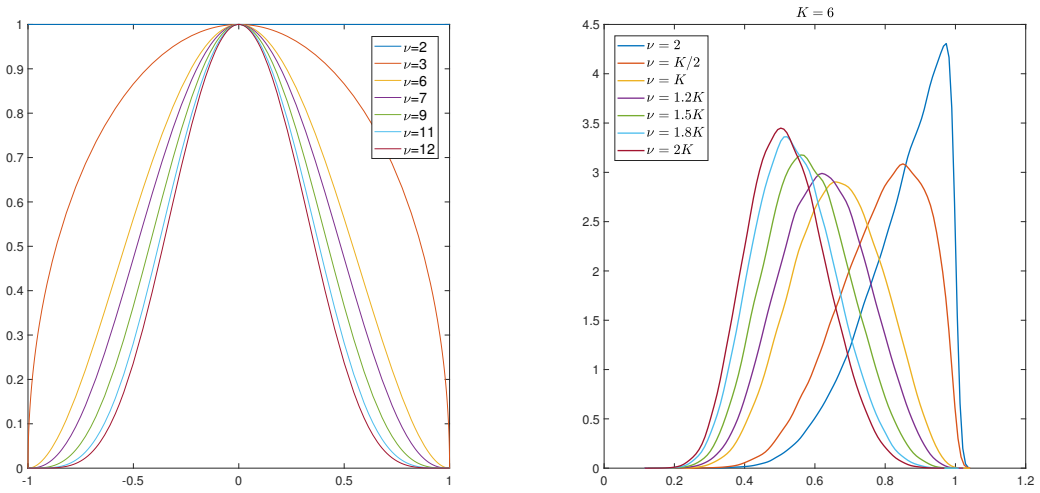
$$\pi\left(\hat{\omega}_{kj}\right) \propto (1 - \hat{\omega}_{kj}^2)^{\nu/2-1}, -1 < \hat{\omega}_{kj} < 1.$$

Setting $\nu = 2$, the marginal density for each correlation is uniform over $(-1, 1)$, while the density is increasingly concentrated at zero with increasing ν, see Panel (a) in Figure 10. We may want to control correlation across factor innovations. Increasing ν for example in proportion to K, shifts the distribution of maximum correlation towards lower values, see Panel (b) in Figure 10 for $K = 6$.

5.2 Factor augmented VAR

We can extend the modelling framework to the analysis of factor augmented VARs (Bernanke et al. 2005; Beyeler and Kaufmann 2021), which have become popular for analyzing transmission processes in many areas, see e.g. Boivin et al. (2009), Ellis et al.

Figure 10: Marginal prior on correlation. Panel (a): Marginal density. Panel (b): Marginal density of maximum correlation when $K = 6$.



(a) (b)

(2014), Stock and Watson (2016), Korobilis and Schröder (2025), among many others. In this setup, the vector $y_t = (y_t^{x\prime}, y_t^{o\prime})'$ is of dimension $N + m$. The data contains a set of m variables, y_t^o, assumed to be measured without error or perfectly reflecting a driving force of remaining data y_t^x. As before, some $K \ll N$ underlying, not directly observable economic factors may characterize common dynamics of y_t^x. These factors may be relevant determinants for the additional m observed series y_t^o, and therefore, we may be interested in analyzing observed and latent factors jointly. The FAVAR representation for $(f_t' \; y_t^{o\prime})'$ writes

$$\begin{aligned} y_t &= \begin{bmatrix} \Lambda^f & \Lambda^o \\ 0 & I_m \end{bmatrix} \begin{bmatrix} f_t \\ y_t^o \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}, \\ \begin{bmatrix} f_t \\ y_t^o \end{bmatrix} &= \Phi \begin{bmatrix} f_{t-1} \\ y_{t-1}^o \end{bmatrix} + \begin{bmatrix} \eta_t^f \\ \eta_t^o \end{bmatrix} \quad \eta_t \text{ i.i.d. } N(0, \Omega), \end{aligned} \tag{16}$$

with dynamics for the N idiosyncratic terms ε_t remaining unchanged to (3). The matrices Λ^f and Λ^o are factor loading matrices of dimension $N \times K$ and $N \times m$, respectively, and I_m reflects the interpretation that y_t^o enter the VAR as *observed factors*. The model is very general, as it allows observed factors y_t^o to load on the other variables y_t^x. For example, including the policy rate as observed factor, a general (sparse) Λ^o allows a policy shock to affect fast-moving variables on impact such as financial market variables, while slow-moving variables, with a zero loading of the observed factor would react with a lag, when

loading unobserved factors react with a lag to the policy shock. Ω is again a full matrix:

$$\Omega = \begin{bmatrix} 1 & & & * \\ & \ddots & & \\ * & & 1 & * \\ * & & * & \omega^o \end{bmatrix}.$$

So, ω_{kj}, $k \neq j$, $k, j \leq K$ reflect correlations, while for $k, j > K$ they reflect covariations.

While the setup remains practically unchanged, the inclusion of observed variables in the state equation induces a prior and posterior distribution of unobserved factors which conditions on these variables. In addition, given that Ω consists of restricted and unrestricted elements, we slightly adjust the parameter expansion step. The MCMC sampler needs the following adjustments.

S″.1. Sample latent factors from $\pi\left(\boldsymbol{f}|\boldsymbol{y}, \boldsymbol{y}^o, \theta\right)$, conditional on observed variables y_t^o:

$$\begin{aligned} y_t^x - \Lambda^o y_t^o = \bar{y}_t &= \Lambda^f f_t + \varepsilon_t, & (17) \\ f_t &= \Phi^f f_{t-1} + \Phi^{fo} y_{t-1}^o + \eta_t^f,\ \eta_t^f \sim N\left(0, \Omega^f\right), & (18) \end{aligned}$$

where the system matrices are obtained by partitioning Φ and Ω accordingly, $\Phi = \begin{bmatrix} \Phi^f & \Phi^{fo} \\ \Phi^{fo} & \Phi^o \end{bmatrix}$ and $\Omega = \begin{bmatrix} \Omega^f & \Omega^{fo} \\ \Omega^{of} & \Omega^o \end{bmatrix}$. Equation (18) implies the following prior distribution for factors:

$$\pi\left(\boldsymbol{f}|\boldsymbol{y}^o, \Phi, \Omega^f\right) = N\left(\boldsymbol{\Phi}^{f^{-1}} \boldsymbol{\Phi}^{fo} \boldsymbol{y}^o, \boldsymbol{\Phi}^{f^{-1}} \boldsymbol{\Omega}^f \boldsymbol{\Phi}^{f^{-1\prime}}\right), \quad (19)$$

with $\boldsymbol{y}^o = (y_0^{o\prime}, \ldots y_{T-1}^{o\prime})'$, where initial conditions may be set equal to first observed values, $y_0^{o\prime} = y_1^{o\prime}$. The matrix $\boldsymbol{\Phi}^f$ has the same banded structure as in (11), $\boldsymbol{\Omega}^f$ is build as in (14) and $\boldsymbol{\Phi}^{fo}$ is of appropriate dimension with matrices Φ^{fo} on its diagonal.

The moments of the posterior distributions are then derived by combining the prior (19) with the likelihood implied by (17),

$$L\left(\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{y}^o, \theta\right) = \prod_{t=1}^{T} \prod_{i=1}^{N} n\left(\tilde{\tilde{y}}_{it}|\lambda_i^f \tilde{f}_{it}, \sigma_i\right).$$

S″.2. (b) $\pi\left(\theta_{\setminus\Lambda}|\boldsymbol{y}, \boldsymbol{y}^o, \boldsymbol{f}, \Lambda\right)$. Adjusts slightly parameter expansion to draw from $\pi\left(\Omega|\boldsymbol{f}, \boldsymbol{y}^o, \Phi\right)$:

i. Sample the $K + m$ diagonal elements for S from the gamma distribution given in (15), setting $C_k^2 = 1$ for $k \leq K$ and $C_k^2 = \hat{\omega}_k^2$ for $k = K+1, \ldots, m$,

where $\hat{\omega}_k^2$ can be set to the variance of residuals from a fitted univariate autoregression for each variable in y_t^o. Sample K elements v_k from

$$v_k|\Omega, s_k \sim IG\left(\frac{\nu + (K+m) - 1}{2}, \frac{s_k \omega_k^-}{2}\right).$$

Collect the v_k and s_k in diagonal matrices V^f and S, respectively. Expand to the unscaled system with $V = \begin{bmatrix} V^f & 0 \\ 0 & I_m \end{bmatrix}$, see Step $\mathbf{S}'$.2(b)i.

ii. Draw from the posterior

$$\pi\left(\hat{\Omega}|S, \hat{\boldsymbol{f}}, \boldsymbol{y}^o, \hat{\Phi}\right) \sim IW\left(\nu + (K+m) - 1 + (T-1), S + \sum_{t=2}^{T} \hat{\eta}_t^f \hat{\eta}_t^{f\prime}\right)$$

with $\hat{\eta}_t^f = f_t - \Phi^f f_{t-1} - \Phi^{fo} y_{t-1}^o$.

iii. Update deterministically V^f (see Step $\mathbf{S}'$.2(b)iii) and return to the scaled system.

5.3 Modelling the Covid period

The period extending over the Covid pandemic has been challenging for econometric modelling, in particular at the outbreak of the pandemic, when it was unclear whether the effects would be long-lasting. It became apparent, though, that the disruption would be transitory, showing up as extreme, short-lived volatility in economic time series. Various approaches have been proposed to capture this unusual pattern. In VAR modelling, Lenza and Primiceri (2022) propose to dummy out those extreme values, when we assume that underlying dynamics remain the same. However, uncertainty may be underestimated when disregarding those data. In factor modelling, we could explicitly take into account outliers in each series, as proposed by Antolin-Diaz et al. (2024). However, in a large dataset, extreme volatile observations may show up synchronously in many data series. Therefore, the volatile period may be captured by an underlying factor as well. Stock and Watson (2025) follow this approach and introduce a Covid factor, which is active from March 2020 onwards. They assume that the same underlying pre-Covid factors and factor dynamics remain, respectively, relevant and the same throughout the Covid period. In the following we outline the extension that includes a Covid factor in the parametric setup and the adjustments in the sampling steps to obtain posterior inference in the spirit of Stock and Watson (2025). We will use the approach to estimate the FAVAR for the FRED-QD dataset in Section 6.

The state equation includes a Covid factor f_t^c, which is active only from the start of the pandemic in the first quarter of 2020. The active period is determined by a dummy

indicator $D_t^c = 1$, for $t \geq 2020Q1$, and $D_t^c = 0$ otherwise. The loadings vector of that factor is denoted by λ^c, and factor dynamics described in the transition equation are of varying dimension:

$$y_t = \begin{bmatrix} \Lambda^f & \Lambda^o & \lambda^c \\ 0 & I_m & 0 \end{bmatrix} \begin{bmatrix} f_t \\ y_t^o \\ f_t^c D_t^c \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}. \tag{20}$$

Collect *pre-Covid* factors in $f_t^* = (f_t', f_t^{o\prime})'$ and write the transition equation compactly

$$\begin{bmatrix} f_t^* \\ f_t^c D_t^c \end{bmatrix} = \begin{bmatrix} \Phi^* & \Phi^{*c} D_{t-1}^c \\ \phi^{c*} D_{t-1}^c & \phi^c D_{t-1}^c \end{bmatrix} \begin{bmatrix} f_{t-1}^* \\ f_{t-1}^c D_t^c \end{bmatrix} + \begin{bmatrix} \eta_t^* \\ \eta_t^c D_t^c \end{bmatrix}, \tag{21}$$

with η_t i.i.d. $N(0, \Omega_t)$:

$$\Omega_t = \begin{bmatrix} \Omega^* & \omega^{*c\prime} D_t^c \\ \omega^{*c} D_t^c & D_t^c \end{bmatrix}. \tag{22}$$

Conditional on y_t^o, taking into account the time-varying dimensions of system matrices in (20)-(22) it is relatively straightforward to re-design the sampler along the lines described in ($\mathbf{S}''.1$) to obtain a joint posterior draw of unobserved and the Covid factors.[6] Parameters are sampled in two steps, deriving in the first one the marginal posterior of loadings and dynamics based on pre-Covid unobserved factors only. Conditional on these pre-Covid effects and dynamics, we derive the additional contribution of the Covid factor:

$\mathbf{S}'''.2$. Sample parameters conditional on data and factors:

(a) $\pi(\Lambda | \boldsymbol{y}, \boldsymbol{f}, \boldsymbol{y}^o, \Psi, \Sigma, \vartheta^\bullet)$. Sample from

i. $\pi\left(\Lambda^f, \Lambda^o | \boldsymbol{y}, \boldsymbol{f}^*, \Psi, \Sigma, \vartheta^\bullet\right)$ over the pre-Covid period.

ii. $\pi\left(\lambda^c | \boldsymbol{y}, \boldsymbol{f}^*, \boldsymbol{f}^c, \Lambda^f, \Lambda^o, \Psi, \Sigma, \vartheta^\bullet\right)$ over the Covid period. To derive the posterior distribution, we condition on pre-Covid factor effects, i.e. base the update on the regression relationship

$$\left(y_t - \Lambda^f f_t - \Lambda^o y_t^o\right) D_t^c = \lambda^c f_t^c D_t^c + \varepsilon_t D_t^c.$$

The posterior distributions of loadings and hyperparameters $\vartheta^\bullet$ are derived analogously, as outlined in Appendix A.2.

(b) $\pi\left(\theta_{\setminus\Lambda} | \boldsymbol{y}, \boldsymbol{f}^*, \boldsymbol{f}^c, \Lambda\right)$.

i. $\pi(\Phi | \boldsymbol{f}^*, \boldsymbol{f}^c, \Omega)$ is updated in three steps:

- $\pi(\Phi^* | \boldsymbol{f}^*, \Omega^*)$ updated over the pre-Covid period, for $t = 1 - D_t^c$.

[6] Practical implementation can work with the full system, i.e. unobserved and Covid factors stacked over the full sample. Setting the prior precision for the Covid factor to a very large number shrinks the Covid factor to zero in the pre-Covid period. Set those observations to zero before sampling parameters.

- $\pi\left(\Phi^{*c}|\boldsymbol{f}^*, \boldsymbol{f}^c, \Phi^*, \Omega^*\right)$ over the Covid period, where we condition on pre-Covid dynamics to estimate the marginal contribution of the Covid factor:

$$f_t^* - \Phi^* f_{t-1}^* \;\; = \;\; \Phi^{*c} f_{t-1}^c + \eta_t^*, \text{ for } t|D_{t-1}^c = 1.$$

- $\pi\left(\phi^{c*}, \phi^c|\boldsymbol{f}^*, \boldsymbol{f}^c\right)$ over the Covid period, based on:

$$f_t^c = \phi^{c*} f_{t-1}^* + \phi^c f_{t-1}^c + \eta_t^c, \text{ for } t|D_{t-1}^c = 1.$$

ii. The posterior update of correlations governing Ω_t is factored into a *marginal* update of Ω^* and a *conditional* update of ω^{*c}, based on a structural representation inspired by Chan (2022).

- Block-diagonalize the current estimate of the full correlation matrix Ω_T with respect to the Covid factor errors,

$$\begin{bmatrix} I & 0 \\ a & 1 \end{bmatrix} \Omega_T \begin{bmatrix} I & a' \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \Omega^* & 0 \\ 0 & \tilde{\omega}^c \end{bmatrix} = \tilde{\Omega}_T, \tag{23}$$

  which allows us to represent the residuals of system (21) in a block-diagonal structural form:

$$\begin{bmatrix} I & 0 \\ a & 1 \end{bmatrix} \begin{bmatrix} \eta_t^* \\ \eta_t^c \end{bmatrix} = \begin{bmatrix} \eta_t^* \\ \tilde{\eta}_t^c \end{bmatrix}, \quad \begin{bmatrix} \eta_t^* \\ \tilde{\eta}_t^c \end{bmatrix} \sim N\left(0, \tilde{\Omega}_T\right), \; t|D_{t-1}^c = 1, \tag{24}$$

  with $\eta_t^* = f_t^* - \Phi^* f_{t-1}^* - \Phi^{*c} f_{t-1}^c$ and $\eta_t^c = f_t^c - \phi^{c*} f_{t-1}^* - \phi^c f_{t-1}^c$.
- Sample expansion parameters $\{v_k, s_k\}$ for all $K+m+1$ factors as described in $\mathbf{S}''$.2(b)i, where ω_k^- represents the inverse of the diagonal kth element of the block-diagonalized matrix $\tilde{\Omega}_T$. Collect $K+m$ elements related to f_t^* in $S^* = \text{diag}(s_1, \ldots, s_{K+m})$ and specify V^* similar to V in $\mathbf{S}''$.2(b)i and $V = \text{diag}(V^*, v_{K+m+1}^{\frac{1}{2}})$. Expand to the unscaled system, to obtain $\hat{\eta}_t^* = V^{*\frac{1}{2}} \eta_t^*$ and $\hat{\eta}_t^c = v_{K+m+1}^{\frac{1}{2}} \eta_t^c$
- Sample the covariance matrix of the pre-Covid factor errors using the residuals over the whole period,

$$\hat{\Omega}^*|\hat{\eta}^* \sim IW\left(\nu + K + m + 2 + (T-1), \; S^* + \sum_{t=2}^{T} \hat{\eta}_t^* \hat{\eta}_t^{*\prime}\right).$$

- Use the structural representation in (24) to update the (co-)variance(s) of the Covid factor errors. Specify conjugate prior distributions (Chan 2022)

$$\hat{\omega}^c \sim IG\left(\frac{\nu + K + m + 2}{2}, \frac{s_c}{2}\right), \; -a|\hat{\omega}^c \sim N\left(0, \hat{\omega}^c S^{*-1}\right).$$

Collect the residuals in $\hat{H}^* = (\hat{\eta}_t^*)'_{t|D_{t-1}^c=1}$ and $\hat{h}^c = (\hat{\eta}_t^c)_{t|D_{t-1}^c=1}$, such that $\hat{h}^c = -\hat{H}^* a' + \hat{\tilde{\eta}}^c$, and denote by T^c the number of residuals in $\hat{h}^c$. The posterior distributions are

$$\begin{aligned}\hat{\omega}^c|\hat{h}^c, \hat{H}^* &\sim IG\left(\frac{\nu + K + m + 2 + T^c}{2}, \frac{s_c + \hat{h}^{c\prime}\hat{h}^c - b_1' B_1^{-1} b_1}{2}\right), \\ -a'|\hat{\omega}^c, \hat{h}^c, \hat{H}^* &\sim N\left(b_1, \hat{\omega}^c B_1\right),\end{aligned}$$

with $B_1 = \left(\hat{H}^{*\prime}\hat{H}^* + S^*\right)^{-1}$ and $b_1 = B_1 \hat{H}^{*\prime}\hat{h}^c$.

- Recover $\hat{\Omega}_T$:

$$\hat{\Omega}_T = \begin{bmatrix} I & 0 \\ -a & 1 \end{bmatrix} \begin{bmatrix} \hat{\Omega}^* & 0 \\ 0 & \hat{\omega}^c \end{bmatrix} \begin{bmatrix} I & -a' \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \hat{\Omega}^* & -\hat{\Omega}^* a' \\ -a\hat{\Omega}^* & a\hat{\Omega}^* a' + \hat{\omega}^c \end{bmatrix}.$$

Update deterministically V, with the first K and the last diagonal elements equal to the corresponding diagonal elements of $\hat{\Omega}$, and return to the scaled system, see $\mathbf{S}'$.2(b)iii.

5.4 Trending factors and further extensions

So far, common factors were extracted from mean-adjusted, non-trending data, as such they reflected common dynamics without extracting the common trending behaviour of data. Kaufmann and Strachan (2025) develop a factor model including common and idiosyncratic deterministic and stochastic processes for trending data. The basic specification models independent factors and allows for a drift in factors and idiosyncratic components, i.e. Equations (2) and (3) are extended to

$$\begin{aligned} f_t &= \mu + \Phi f_{k,t-1} + \eta_t, \eta_t \sim i.i.d.\ N\left(0, I_K\right), \ \Phi \text{ diagonal}, \\ \varepsilon_t &= \delta + \Psi \varepsilon_{t-1} + \epsilon_t, \ \epsilon_t \sim i.i.d.\ N\left(0, \Sigma\right). \end{aligned} \tag{25}$$

The decomposition of series-specific drifts into a factor-driven and an idiosyncratic component is based on a statistical model, allowing for heterogeneous factor-driven drifts across series. Factor-series association is induced by specifying a sparse normal-gamma prior for factor loadings. Factor drifts are data-determined by a weighted average over series-specific drifts, where the weights correspond to factor loadings. Hence, those series most strongly loaded by factors determine the factor drift. The estimation approach can be applied to extract time-varying factor trends as well:

$$\begin{aligned} f_t &= \mu_t + \Phi f_{k,t-1} + \eta_t, \eta_t \sim i.i.d.\ N\left(0, I_K\right), \ \Phi \text{ diagonal}, \\ \mu_t &= \mu_{t-1} + \xi_t, \ \xi_t \sim N\left(0, \Sigma_\xi\right), \ \Sigma_\xi \text{ diagonal}. \end{aligned} \tag{26}$$

In specifications (25) and (26) factors trend independently and factor and idiosyncratic trends affect series independently, which a priori rules out cointegration across series and across factors. An encompassing approach, starting from a specification for series in levels to evaluate restrictions induced by potential cointegration relationships is the subject of ongoing research.

A further extension to consider is relaxing the independence across idiosyncratic components. In a high-dimensional and heterogeneous dataset, inducing sparsity in the factor loading matrix usually leads to the extraction of the main, pervasive factors, and may also identify some weaker, less pervasive factors, i.e. factors relevant for a specific sub-group of data. This local correlation structure could also be captured by allowing for weak correlation across idiosyncratic components:

$$\varepsilon_t \quad = \quad \Psi\varepsilon_{t-1} + \epsilon_t,\ \epsilon_t \sim i.i.d.\ N\left(0, \Sigma\right),\ \Psi \text{ full},\ \Sigma \text{ full},$$

where the amount of correlation would be controlled by specifying a shrinkage prior distribution for off-diagonal elements in Ψ and a hierarchical inverse wishart prior distribution for Σ (Huang and Wand 2013), see also (15).

6 Illustration

For illustration, we estimate a FAVAR model for the FRED-QD dataset (McCracken and Ng 2021), as analyzed in Beyeler and Kaufmann (2021). The data are transformed as indicated in Beyeler and Kaufmann (2021), with adaptations to log levels for exchange rates and some stock market indices. The model is estimated for the period 2000, first quarter, to 2024, first quarter, a period including two exceptionally stressful periods, the financial crisis and the Covid-19 pandemic. We estimate a basic specification for $K = 7$ unobserved factors, including a Covid factor as outlined in Subsection 5.3 and the federal funds rate (FFR) as observed factor ($m = 1$). Seven factors were suggested by Stock and Watson (2005) for structural analysis and also by McCracken and Ng (2021). Likewise, eight interpretable (rotated) factors were identified in Bai and Ng (2013) or Despois and Doz (2023). The estimation period includes part of the Great Moderation and the prolonged period of low interest rates. All factors except the stock market factor are also extracted when the data sample starts in 1984Q1, the beginning of the Great Moderation period. The stock market separates from the housing factor in the shorter sample, which captures the diverging evolution of housing and stock markets shortly before the pandemic.[7]

[7]These and further comparisons between factors extracted for different sample periods are available upon request. This illustrates that sparse factor modelling is exposed to the same issue as traditional dynamic factor modelling, although our comparisons suggest that the number of factors extracted may not be inflated as much as in the principal components approach (Breitung and Eickmeier 2011).

Table 1: FRED-QD: Prior specifications, for $K + m = 8$, including a Covid factor. The prior for λ_{ij} is specified for the sparse specification (7). The observed factor y^o is the federal funds rate ($m = 1$).

Parameters	Settings
$\lambda_{ik}\|\beta_{ik}, \tau_{ik}$, $\beta_{ik}\|\rho_k$	$g_\tau = 2$, $G_\tau = 0.0025$, $a_0 = 0.1$, $b_0 = 0.9$, $r_0 = 1$, $s_0 = 0.2$.
$\Phi_1, \ldots, \Phi_p$	$p = 2$, $\phi_{kj,l} \sim N\left(0, \mathrm{P}_{0,kjl}\right)$, $\mathrm{P}_{0,kjl} = c_{kj}\varphi/l^2$, $\varphi = 0.25$, $c_{kj} = 0.25$, $k \neq j$, and $c_{kk} = 1$ otherwise, $l = 1, \ldots, p$.
$\hat{\Omega}\|\Omega, S$	$\nu = 2(K + m)$, $C_k^2 = 1$, $k \leq K$ and $C_k^2 = \hat{\omega}_k^2$, $k = K + 1$, $\hat{\omega}_k^2$ equals the residual variance of a fitted AR(2) to y^o.
$\psi_{i1}, \ldots, \psi_{iq}$, σ_i	$q = 2$, $\psi_{il} \sim N\left(0, \mathrm{Q}_{0,il}\right)$, $\mathrm{Q}_{0,il} = \varphi/l^2$, $\varphi = 0.25$, $l = 1, \ldots, q$, $g_s = 2$, $G_{S_i} = \hat{\sigma}_i^2$ $\hat{\sigma}_i^2$ equals the residual variance of a fitted AR(2) to each y_i.

We include two lags in idiosyncratic and factor dynamics, $q = p = 2$ and leave the factor innovation matrix Ω unrestricted.[8]The posterior distributions of higher-order dynamic lags turn out to be centered at zero. The specification of prior distributions is displayed in Table 1. We specify rather a loose prior for the base rate ρ_k of non-zero factor loadings, as we expect data to be informative about the posterior. MCMC sampling proceeds along the lines exposed in Subsection 5.3. Posterior evaluations are based on 30,000 MCMC draws, after discarding 15,000 initial draws to remove dependence on initial conditions. To process the posterior output, we rely on correlations across draws to identify factor representatives, applying Steps **P**.1a and **P**.2a described in Subsection 4.

Factor-series association is plotted in Figure 1, based on non-zero loadings identified as the mean across factor loading draws for which $\beta_{ik}^{(m)} > 0.5$ and associated with a posterior median of $\beta_{ik}|\cdot$ larger than 0.5, $q_{0.5}(\beta_{ik}|\cdot) > 0.5$:

$$\delta_{q_{0.5}(\beta_{ik}|\cdot)>0.5} \sum_m \lambda_{ik}^{(m)} \delta_{\beta_{ik}^{(m)}>0.5} / \sum_m \delta_{\beta_{ik}^{(m)}>0.5},$$

where $\delta_u = 1$ if u is fulfilled. Figure 11 plots the mean and a 95% HPDI interval of identified factors. The Covid factor loads on many series and captures the extreme, transitory volatility at the beginning of the pandemic. The loadings of the other factors suggest the same interpretation as before the pandemic. The patterns extracted during the Covid period also match narrative evidence. For example, we observe the rebound in production and employment after the sharp drop in the first half of 2020. Likewise, the price factor captures the mounting inflationary pressure from 2021 onwards.

Table 2 displays the posterior mean of correlations across factor innovations, with a 68%

[8]We include two lags for the idiosyncratic components to capture the dynamics of a series if it is not loaded by any factor.

Figure 11: FRED-QD: Mean factors, with 95% HPDI.

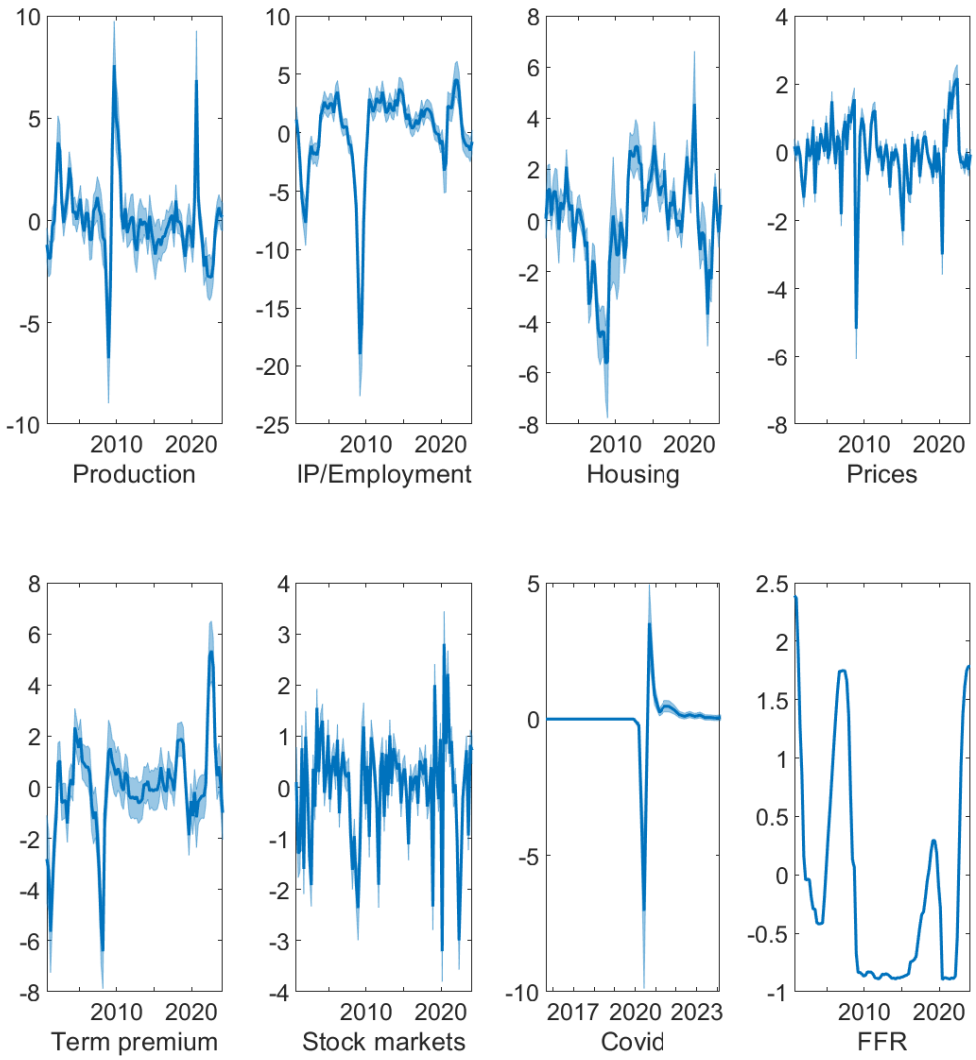


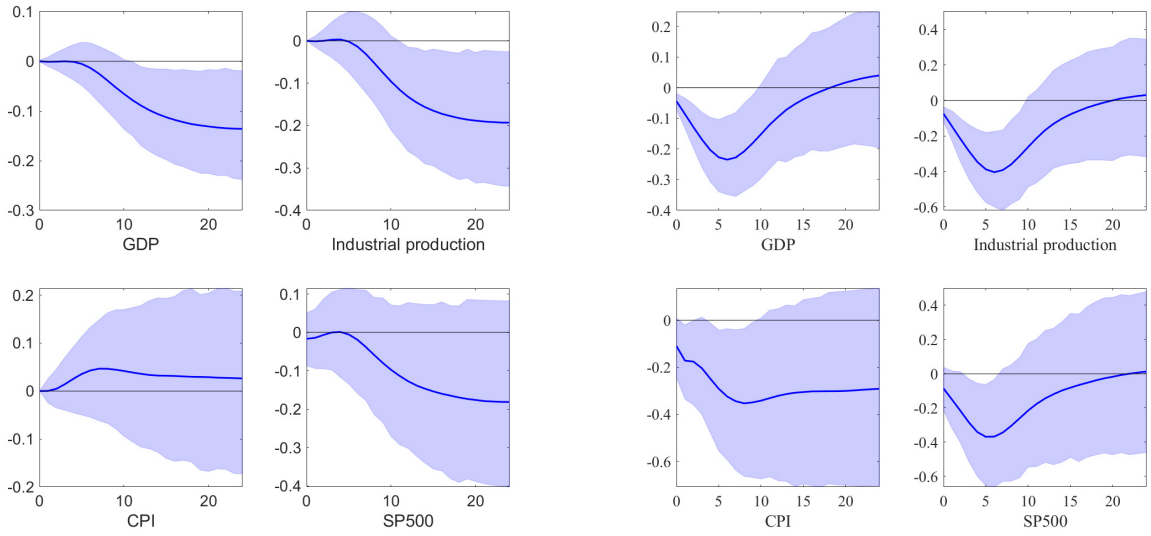
Table 2: FRED-QD: Correlations across factor innovations. For better comparison, the (co-)variance(s) of the federal funds rate (FFR) are standardized. Posterior mean, $E(\Omega|\boldsymbol{y},\cdot)$, and 68% HPDI.

Production	1.0							
Employment/IP	**0.43**	1.0						
	(0.33,0.55)							
Housing	**0.13**	0.08	1.0					
	(0.01,0.27)	(-0.09,0.25)						
Prices	**0.21**	**0.36**	**-0.27**	1.0				
	(0.10,0.34)	(0.27,0.48)	(-0.39,-0.16)					
Term premium	0.11	-0.02	**-0.32**	0.08	1.0			
	(-0.02,0.23)	(-0.14,0.10)	(-0.44,-0.21)	(-0.04,0.19)				
Stock markets	**0.25**	0.08	**0.15**	-0.04	-0.04	1.0		
	(0.15,0.37)	(-0.04,0.21)	(0.04,0.27)	(-0.16,0.07)	(-0.15,0.08)			
Covid	-0.10	0.06	-0.09	0.16	-0.08	**-0.14**	1.0	
	(-0.31,0.10)	(-0.13,0.28)	(-0.27,0.07)	(-0.01,0.34)	(-0.25,0.08)	(-0.30,-0.00)		
FFR	**0.21**	**0.27**	-0.07	**0.41**	0.09	-0.01	**0.22**	1.0
	(0.09,0.34)	(0.15,0.38)	(-0.19,0.05)	(0.31,0.53)	(-0.03,0.21)	(-0.12,0.11)	(0.02,0.40)	

HPDI reported in brackets. Although quite a few are shifted away from zero, mean correlations remain moderate. Interestingly, the Covid and stock market innovations are negatively correlated, which also matches narrative evidence.

Figures 12 and 13 illustrate further posterior structural evaluations. Figure 12 presents cumulated impulse responses of selected series to factor shocks, based on pre-Covid factor dynamics, i.e. excluding the Covid factor. Panel (a) identifies structural shocks by Cholesky decomposition of Ω^*, ordering factor innovations as follows: Employment/IP, production, prices, FFR , term premium, housing, stock markets. Whereas a shock in the FFR has a permanent negative mean effect on GDP and industrial production, we observe that the effect on the consumer price index (CPI) and the S&P500 index remains insignificant over the whole horizon (based on the 68% highest posterior density interval). Since the FFR fell to historically low levels over the sample period, it ceased temporarily to be the main policy instrument. The Federal Reserve deployed unconventional monetary policy measures, essentially transmitted to the real economy by affecting long term spreads. Panel (b) shows impulse responses to a shock identified by maximizing the forecast error variance of the spread between the 10-year government bond yield and the 3-month treasury bill rate over a horizon of six quarters (maxFEV 10y-3m). Impulse responses seem more consistent with reactions to a monetary policy shock. GDP and industrial production show a negative reaction before returning to their initial level. The responses of CPI and the S&P500 index are also more plausible, although uncertainty remains substantial. The mean response of CPI remains negative over the whole horizon, while the mean negative response of the S&P500 index is transitory.

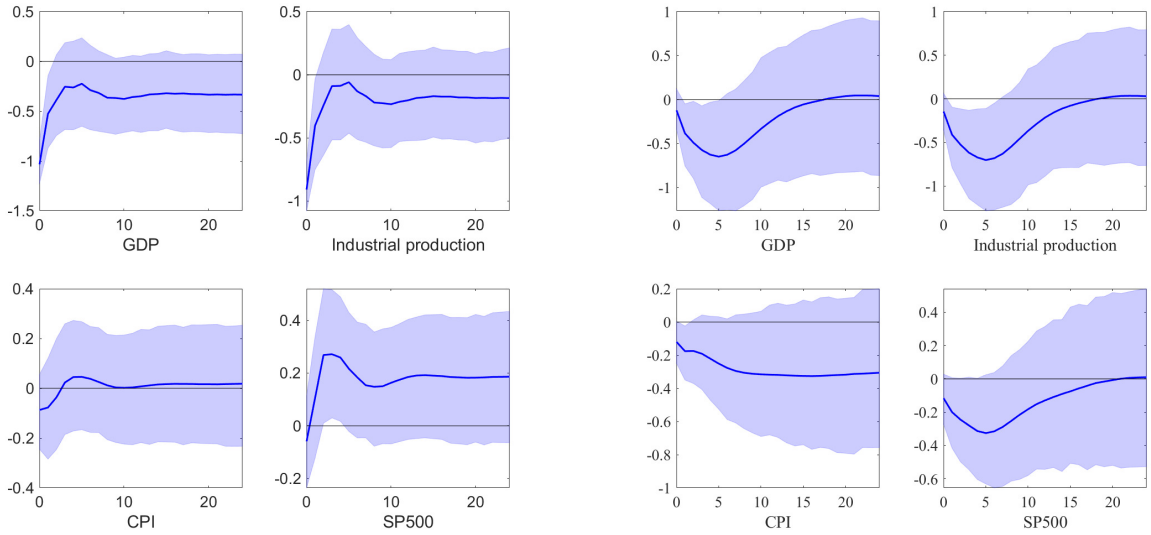
Figure 12: FRED-QD: Mean cumulated responses with a 68% HPDI; derived for dynamics excluding the Covid period. (a) Impulse responses to a shock in the federal funds rate identified by Cholesky decomposition; Factor cholesky ordering: Employment/IP, production, prices, federal funds rate, term premium, housing, stock markets. (b) Impulse responses to a shock maximizing the forecast error variance of the spread between the 10-year government bond yield and the 3-month treasury bill rate over a horizon of six quarters.



(a) FFR (b) maxFEV 10y-3m

Figure 13 plots cumulated impulse responses based on factor dynamics during the Covid period, i.e. including the Covid factor. Panel (a) plots impulse responses to a negative Covid shock, identified by ranking the Covid factor first and the remaining factors as indicated before. The persistent negative mean response of GDP suggests that the Covid shock acted like a negative productivity shock. The Covid shock had no effect on prices, and the positive mean response of the S&P500 index reflects the surge of technology stocks at the onset of the pandemic. The impulse responses to a shock identified by maxFEV 10y-3m remain consistent with pre-pandemic responses. The wider HPD intervals reflect greater uncertainty, owing to the small number of observations to estimate dynamics related to the Covid factor.

Figure 13: FRED-QD: Mean cumulated responses with a 68% HPDI; derived for dynamics including the Covid period. (a) Impulse responses to a negative shock in the Covid factor; Factor cholesky ordering: Covid, Employment/IP, production, prices, federal funds rate, term premium, housing, stock markets. (b) Impulse responses to a shock maximizing the forecast error variance of the spread between the 10-year government bond yield and the 3-month treasury bill rate over a horizon of six quarters.



(a) Covid

(b) maxFEV 10y-3m

7 Conclusion

Sparse factor modelling in econometrics is useful when factor identification, i.e. the extraction of interpretable factors is of interest. Sparsity is induced into the factor loading matrix by specifying a prior distribution which shrinks or sets to zero weak factor-series association. The normal-gamma and a hierarchical point mass-normal prior distributions are discussed in more details.

An illustrative discussion of rotational identification motivates the proposed order-invariant estimation approach. Different ordering of variables ranked first in the dataset may imply very different factor interpretations when rotational identification restrictions are imposed before estimation. In the worst case, an underlying sparse structure may be distorted. Therefore, we suggest to estimate factor-series association in a data-driven way, i.e. to estimate the model without pre-specifying rotational identification restrictions. Factor identification, including factor position and sign, is obtained by post-processing the MCMC output.

Extracted factors capture very distinct dynamic features governing various groups of variables. Therefore, inducing sparsity in the factor loading matrix allows relaxing the restrictions of independence usually imposed on factor innovations. We discuss extensions to estimate a full correlation matrix and add observed factors to the factor state equation (FAVAR model). A further extension includes a Covid factor, which loads only during the pandemic period starting from the 2020 onwards. The reduced-form estimate can then be used for further structural evaluations or scenario-based forecasting.

An application to the FRED-QD dataset uses the obervation sample spanning the period 2000, first quarter, to 2024, first quarter. The Covid factor loads on many series and captures the extreme, transitory volatility at the beginning of the pandemic. The loadings of the other factors suggest the same interpretation as before the pandemic. We discuss cumulated impulse responses of selected variables based on pre-pandemic dynamics and pandemic dynamics including the Covid factor. Since the Federal Reserve deployed unconventional monetary policy measures, essentially transmitted to the real economy by affecting long term spreads, we identify a policy shock by maximizing the forecast error variance of the spread between the 10-year government bond yield and the 3-month treasury bill rate over a horizon of six quarters. Impulse responses seem consistent with reactions to a monetary policy shock, although uncertainty remains considerable. A (negative) Covid shock acts like a negative productivity shock on GDP and has a positive, persistent mean effect on the S&P500 index.

Several open questions offer interesting avenues for further research on sparse factor modelling. An encompassing model, starting from a specification for series in levels and deriving representations for trend-adjusted series, would allow us to analyze the implications

of cointegration across series and across factors. Further relaxations, such as allowing for weak correlations across idiosyncratic innovations, could re-allocate some of the common correlation currently captured by weaker, group-specific factors to the idiosyncratic components. The parametric sparse factor model would then come closer to dynamic approximate factor models.

References

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics 18*, 338–357.

Anderson, T. and H. Rubin (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*, pp. 111–150. Berkeley, California: University of California Press.

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2024). Advances in nowcasting economic activity: The role of heterogeneous dynamics and fat tails. *Journal of Econometrics 238*(2), 105634.

Aßmann, C., J. Boysen-Hogrefe, and M. Pape (2016). Bayesian analysis of static and dynamic factor models: An ex-post approach towards the rotation problem. *Journal of Econometrics 192*, 190–206.

Aßmann, C., S. Kaufmann, and M. Pape (2025). Invited discussion: Sparse bayesian factor analysis when the number of factors is unknown (with discussion). *Bayesian Analysis 20*, 274–282.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of Econometrics 176*, 18–29.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics 120*, 387–422.

Beyeler, S. and S. Kaufmann (2021). Reduced-form factor augmented VAR - Exploiting sparsity to include meaningful factors. *Journal of Applied Econometrics 36*, 989–1012.

Bhattacharya, A. and D. B. Dunson (2011). Sparse Bayesian infinite factor models. *Biometrika 98*, 291–306.

Bitto, A. and S. Frühwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics 210*, 75–97.

Boivin, J., M. P. Giannoni, and I. Mihov (2009). Sticky prices and monetary policy: Evidence from disaggregated us data. *The American Economic Review 99*, 350–384.

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

Breitung, J. and S. Eickmeier (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics 163*(1), 71–84.

Carvalho, C. M., J. Chang, J. E. Lucas, J. R. Nevins, W. Quanli, and M. West (2008). High-dimensional sparse factor modeling: Applications in gene expression genomics. *Journal of the American Statistical Association 103*, 1438–1456.

Chan, J., R. Leon-Gonzalez, and R. W. Strachan (2018). Invariant inference and efficient computation in the static factor model. *Journal of the American Statistical Association 113*, 819–828.

Chan, J. C. (2022). Asymmetric conjugate priors for large Bayesian VARs. *Quantitative Economics 13*, 1145–1169.

Chan, J. C. and I. Jeliazkov (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation 1*, 101–120.

Conti, G., S. Frühwirth-Schnatter, J. J. Heckman, and R. Piatek (2014). Bayesian exploratory factor analysis. *Journal of Econometrics 183*, 31–57.

Despois, T. and C. Doz (2023). Identifying and interpreting the factors in factor models via sparsity: Different approaches. *Journal of Applied Econometrics 38*, 533–555.

Eckert, F., P. Kronenberg, H. Mikosch, and S. Neuwirth (2025). Tracking economic activity with alternative high-frequency data. *Journal of Applied Econometrics 40*, 270–290.

Ellis, C., H. Mumtaz, and P. Zabczyk (2014). What lies beneath? A time-varying favar model for the UK transmission mechanism. *The Economic Journal 124*, 668–699.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic factor model: Identification and estimation. *Review of Economics & Statistics 82*, 540–554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2001). Coincident and leading indicators for the euro area. *The Economic Journal 111*, C62–C85.

Freyaldenhoven, S. (2022). Factor models with local factors — Determining the number of relevant factors. *Journal of Econometrics 229*, 80–102.

Frühwirth-Schnatter, S. (2001). MCMC estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*, 194–209.

Frühwirth-Schnatter, S., D. Hosszejni, and H. F. Lopes (2025). Sparse Bayesian factor analysis when the number of factors is unknown. *Bayesian Analysis 20*, 213–344.

George, E. I. and R. E. McCulloch (1997). Approaches for Bayesian variable selection. *Bayesian Statistics 5*, 609–620.

Geweke, J. and G. Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies 9*, 557–587.

Griffin, J. and P. Brown (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis 5*, 171–188.

Griffin, J. and P. Brown (2017). Hierarchical shrinkage priors for regression models. *Bayesian Analysis 12*, 135–159.

Huang, A. and M. P. Wand (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis 8*, 439–452.

Jöreskog, K. (1969, June). A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika 34*(2), 183–202.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika 36*(2), 109–133.

Kaufmann, S. and M. Pape (2026). Bayesian (non-)unique sparse factor modelling. *Econometrics and Statistics 2nd Special Issue on Bayesian Methods in Statistics and Econometrics: Structured dependence and robust inference at scale* forthcoming.

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

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

Kaufmann, S. and R. Strachan (2025). Dynamic factor models with common (drifting) stochastic trends. Study Center Gerzensee.

Kim, C.-J. and C. R. Nelson (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime-switching. *Review of Economics & Statistics 80*, 188–201.

Korobilis, D. and M. Schröder (2025). Monitoring multi-country macroeconomic risk: A quantile factor-augmented vector autoregressive (QFAVAR) approach. *Journal of Econometrics 249*, 105730.

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra (2009). Circos: An information aesthetic for comparative genomics. *Genome Research 19*, 1639–1645.

Lawley, D. and A. Maxwell (1971). *Factor Analysis as a Statistical Method* (Second edition ed.). London: Butterworth & Co (Publishers) Ltd.

Legramanti, S., D. Durante, and D. B. Dunson (2020). Bayesian cumulative shrinkage for infinite factorizations. *Biometrika 107*, 745–752.

Lenza, M. and G. E. Primiceri (2022). How to estimate a VAR after March 2020. *Journal of Applied Econometrics 37*(4), 688–699.

Liu, X. (2008). Parameter expansion for sampling a correlation matrix: An efficient GPX-RPMH algorithm. *Journal of Statistical Computation and Simulation 78*, 1065–1076.

Lopes, H. F. and M. West (2004). Bayesian model assessment in factor analysis. *Statistica Sinica 14*, 41–67.

McCracken, M. W. and S. Ng (2021). FRED-QD: A quarterly database for macroeconomic research. *Federal Reserve Bank of St. Louis REVIEW First Quarter*, 1–44.

Millsap, R. E. (2001). When Trivial Constraints Are Not Trivial: The Choice of Uniqueness Constraints in Confirmatory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal 8*(1), 1–17.

Rousseau, J. and K. Mengersen (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(5), 689–710.

Ročková, V. and E. I. George (2016). Fast Bayesian factor analysis via automatic rotations to sparsity. *Journal of the American Statistical Association 111*, 1602–1622.

Stock, J. and M. Watson (1989). New indexes of coincident and leading economic indicators. In O. Blanchard and S. Fischer (Eds.), *NBER Macroeconomics Annual*, pp. 351–409. Cambridge, Mass.: MIT Press.

Stock, J. and M. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*, 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*, 147–162.

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. Working Paper 11467, NBER.

Stock, J. H. and M. W. Watson (2016). Chapter 8 - dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2, pp. 415–525. Elsevier B.V.

Stock, J. H. and M. W. Watson (2025). Recovering from COVID. *Brookings Papers on Economic Activity 56*(1), 297–346.

West, M. (2003). Bayesian factor regression models in the "large p, small n" paradigm. *Bayesian Statistics 7*, 723–732.

Zhang, X., W. J. Boscardin, and T. R. Belin (2006). Sampling correlation matrices in Bayesian models with correlated latent variables. *Journal of Computational and Graphical Statistics 15*, 880–896.

A Posterior distributions

A.1 The factors $\boldsymbol{f}$, $\pi(\boldsymbol{f}|\boldsymbol{y},\theta)$

To derive the posterior distribution of factor, we use the compact representation (8)-(10) in Subsection 3.2. We display the derivation for general autoregressive settings p and q for, respectively, factor and idiosyncratic dynamics. Condense series-specifically:

$$\begin{aligned}\Psi(L)y_t = \tilde{y}_t &= \Lambda f_t - \Lambda \odot (\psi_{\cdot 1} \otimes \mathbf{1}_{1\times K}) f_{t-1} - \cdots - \Lambda \odot (\psi_{\cdot q} \otimes \mathbf{1}_{1\times K}) f_{t-q} + \epsilon_t, \\ &\quad \epsilon_t \sim N(0,\Sigma), \\ f_t &= \Phi_1 f_{t-1} + \cdots + \Phi_p f_{t-p} + \eta_t, \quad \eta_t \sim N(0, I_k),\end{aligned}$$

where $\psi_{\cdot j}$ stacks all lag j-specific coefficients ψ_{ij}, $i = 1,\ldots,N$, $\odot$ and $\otimes$ represent the Hadamard and the Kronecker product, respectively. The row vector $\mathbf{1}_{1\times K}$ contains as elements K ones. We stack the observations into the compact matrix representation:

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{\Lambda}}\boldsymbol{f} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, I_{T-q} \otimes \Sigma), \tag{27}$$

$$\boldsymbol{\Phi}\boldsymbol{f} = \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim N(0, \boldsymbol{\Sigma}_f), \tag{28}$$

where $\tilde{\boldsymbol{y}} = (\tilde{y}_1', \ldots, \tilde{y}_T')'$ contains all observed data, $\boldsymbol{f} = \left(f'_{q+1-\max(p,q)}, \ldots, f'_{q+1}, \ldots, f'_T\right)'$ stacks all unobserved factors, including initial states. The matrices $\tilde{\boldsymbol{\Lambda}}$ and $\boldsymbol{\Phi}$ are respectively of dimension $(T-q)N \times (T+d)K$ and square $(T+d)K$, with $d = (p-q)I\{p > q\}$. Typically, these matrices are sparse and banded around the main diagonal:

$$\tilde{\boldsymbol{\Lambda}} = \left[\begin{array}{c|cccccc} \mathbf{0}_{(T-q)N\times dK} & -\Lambda \odot (\psi_{\cdot q} \otimes \mathbf{1}_{1\times K}) & \ldots & \Lambda & 0 \ldots & & 0 \\ & & \ddots & \ddots & & \ddots & \vdots \\ & 0 \ldots & & 0 & & -\Lambda \odot (\psi_{\cdot q} \otimes \mathbf{1}_{1\times K}) & \ldots \quad \Lambda \end{array}\right],$$

$$\boldsymbol{\Phi} = \left[\begin{array}{cccccccc} I_p \otimes I_K & & 0 & \ldots & & & & \\ \hline -\Phi_p & \ldots & -\Phi_1 & I_K & 0 & \ldots & & \\ & & & & \ddots & & & \\ & \ldots & 0 & -\Phi_p & \ldots & -\Phi_1 & I_K & \end{array}\right],$$

$$\boldsymbol{\Sigma}_f = \left[\begin{array}{ccc} I_p \otimes \Sigma_f & 0 & \ldots \\ 0 & & \\ \vdots & & I_{T+d-p} \otimes I_K \end{array}\right],$$

where Σ_f represents the variance of the initial states (see below).

We use the sampler proposed in Chan and Jeliazkov (2009) to sample $\boldsymbol{f}$ in one sweep. Given the representation in (27)-(28), the complete data likelihood has a normal density

$$f\left(\tilde{\boldsymbol{y}}|\boldsymbol{f},\theta\right)\sim N\left(\tilde{\boldsymbol{\Lambda}}\boldsymbol{f},I_{T-q}\otimes\Sigma\right). \tag{29}$$

For the unobserved states, from (28) we obtain the following prior:

$$\begin{aligned}\boldsymbol{f}|\theta &\sim N\left(0,\mathrm{F}_0\right), \\ &\quad \mathrm{F}_0^{-1}=\boldsymbol{\Phi}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\Phi}.\end{aligned} \tag{30}$$

If in $\boldsymbol{\Sigma}_f$, the variance of the initial states, Σ_f, is not chosen to be diffuse, we may set it equal to the implied unconditional factor variance. From the companion form of a VAR(p) process, $\mathbf{F}_t=\tilde{\boldsymbol{\Phi}}\mathbf{F}_{t-1}+\boldsymbol{\eta}_t$, $\boldsymbol{\eta}_t\sim N\left(0,\left[\begin{array}{c}I_K\quad 0_{K\times(p-1)K}\\ 0_{(p-1)K\times pK}\end{array}\right]\right)$, with $\tilde{\boldsymbol{\Phi}}=\left[\begin{array}{c}\Phi_1,\ldots,\Phi_p\\ I_{K(p-1)},0_{K(p-1)\times K}\end{array}\right]$, we obtain $E(\mathbf{F}_t\mathbf{F}_t')=\tilde{\boldsymbol{\Phi}}E(\mathbf{F}_{t-1}\mathbf{F}_{t-1}')\tilde{\boldsymbol{\Phi}}'+\Sigma_{\boldsymbol{\eta}}$ and $\Sigma_{\mathbf{F}}=\tilde{\boldsymbol{\Phi}}\Sigma_{\mathbf{F}}\tilde{\boldsymbol{\Phi}}'+\Sigma_{\boldsymbol{\eta}}$. The vec operator yields

$$\mathrm{vec}(\Sigma_{\mathbf{F}})=\left[\mathbf{I}_{(Kp)^2}-(\tilde{\boldsymbol{\Phi}}\otimes\tilde{\boldsymbol{\Phi}})\right]^{-1}\times\mathrm{vec}\left(\Sigma_{\boldsymbol{\eta}}\right),$$

from which we can retrieve the corresponding values for Σ_f.

Combining the prior with the likelihood, the posterior of the factors is:

$$\begin{aligned}\boldsymbol{f}|\tilde{\boldsymbol{y}},\theta &\sim N\left(\mathrm{f},\mathrm{F}\right), \\ \mathrm{F}^{-1} &= \mathrm{F}_0^{-1}+\tilde{\boldsymbol{\Lambda}}'\left(I_{T-q}\otimes\Sigma^{-1}\right)\tilde{\boldsymbol{\Lambda}}, \\ \mathrm{f} &= \mathrm{F}\left(\tilde{\boldsymbol{\Lambda}}'\left(I_{T-q}\otimes\Sigma^{-1}\right)\tilde{\boldsymbol{y}}\right).\end{aligned}$$

To avoid the full inversion of F^{-1} we take the Cholesky decomposition, $\mathrm{F}^{-1}=L'L$, then $\mathrm{F}=L^{-1}L^{-1'}$. We obtain a draw $\boldsymbol{f}$ by setting $\boldsymbol{f}=\mathrm{f}+L^{-1}\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a $(T+d)k$ vector of independent draws from the standard normal distribution.

A.2 Factor loadings and hyperparameters, $\pi\left(\Lambda|\boldsymbol{y},\boldsymbol{f},\theta_{\backslash\Lambda},\vartheta^{\bullet}\right)$, $\pi\left(\vartheta^{\bullet}|\Lambda\right)$

The normal-gamma prior distribution (6) on Λ leads to the following posterior update:

1. $\pi\left(\Lambda|\boldsymbol{y},\boldsymbol{f},\theta_{\backslash\Lambda},\tau^{ng}\right)$: Conditional on factors, factor loadings can be updated independently across series, $\left(\lambda_{i1},\ldots,\lambda_{iK}\right)'|\cdot\sim N\left(\mathrm{l}_i,\mathrm{L}_i\right)$, $i=1,\ldots,N$, with

$$\mathrm{L}_i=\left(\sigma_i^{-1}\tilde{f}_i'\tilde{f}_i+\mathrm{diag}\left(\tau_{i1}^{-1},\ldots,\tau_{iK}^{-1}\right)\right)^{-1},\ \mathrm{l}_i=\mathrm{L}_i\left(\sigma_i^{-1}\tilde{f}_i'\tilde{y}_i\right),$$

where $\tilde{y}_i=\left(\tilde{y}_{i1},\ldots,\tilde{y}_{iT}\right)'$ collects filtered observations and $\tilde{f}_i=\left(\tilde{f}_{i1},\ldots,\tilde{f}_{iT}\right)'$ is the matrix of series-specifically filtered factors.

2. $\pi\left(\tau^{ng}|\Lambda\right)$: Conditional on Λ, the posterior distribution of τ_{ik} takes the form of a generalized inverse gaussian,

$$\tau_{ik}|\cdot \sim GIG\left(a-1/2, a\kappa^2, \lambda_{ik}^2\right).$$

Under the hierarchical prior (7), the posterior distribution of Λ will be sparse, too. We simulate from following distributions:

1. $\pi\left(\Lambda|\boldsymbol{y}, \boldsymbol{f}, \theta_{\setminus\Lambda}, \vartheta^{hs}\right)$: Conditional on factors, we update Λ factor-wise. The hierarchical prior implies a common base rate of a non-zero factor loading of $E\left(\beta_{ik}\right) = \rho_k b_0$ across variables. Marginalizing over β_{ik}, the marginal prior becomes

$$\pi\left(\lambda_{ik}|\rho_k\right) \sim (1-\rho_k b_0)\delta_0(\lambda_{ik}) + \rho_k b_0 N\left(0, \tau_k\right).$$

For each factor k, transform the variables to

$$\tilde{y}_{it}^{(k)} = \tilde{y}_{it} - \sum_{l=1, l\neq k}^{K} \lambda_{il}\tilde{f}_{i,lt} = \lambda_{ik}\tilde{f}_{i,kt} + \epsilon_{it},$$

which isolates the effect of factor k on variable i. Combine the marginal prior with data information to sample independently across i from

$$\begin{aligned}\pi\left(\lambda_{ik}|\cdot\right) &= \prod_{t=1}^{T} \pi(\tilde{y}_{it}^{(k)}|\cdot)\left\{(1-\rho_k b_0)\delta_0(\lambda_{ik}) + \rho_k b_0 N\left(0, \tau_k\right)\right\}, \\ &= P\left(\lambda_{ik}=0|\cdot\right)\delta_0(\lambda_{ik}) + P\left(\lambda_{ik}\neq 0|\cdot\right) N\left(m_{ik}, M_{ik}\right),\end{aligned}$$

with observation density $\pi(\tilde{y}_{it}^{(k)}|\cdot) = N\left(\lambda_{ik}\tilde{f}_{i,kt}, \sigma_i\right)$ and where

$$M_{ik} = \left(\sigma_i^{-1}\sum_{t=1}^{T}\tilde{f}_{i,kt}^2 + \tau_k^{-1}\right)^{-1}, \; m_{ik} = M_{ik}\left(\sigma_i^{-1}\sum_{t=1}^{T}\tilde{f}_{i,kt}\tilde{y}_{it}^{(k)}\right).$$

The posterior odds of a non-zero factor loading in (31) are given by:

$$\frac{P\left(\lambda_{ik}\neq 0|\cdot\right)}{P\left(\lambda_{ik}=0|\cdot\right)} = \frac{\pi\left(\lambda_{ik}\right)|_{\lambda_{ik}=0}}{\pi\left(\lambda_{ik}|\cdot\right)|_{\lambda_{ik}=0}}\frac{\rho_k b_0}{1-\rho_k b_0} = \frac{N\left(0; 0, \tau_k\right)}{N\left(0; m_{ik}, M_{ik}\right)}\frac{\rho_k b_0}{1-\rho_k b_0}.$$

2. $\pi\left(\vartheta^{hs}|\Lambda\right)$: Conditional on λ we update hyperparameters:

(a) $\pi\left(\beta_{ik}|\lambda_{ik}, \rho_k\right)$: If $\lambda_{ik} = 0$

$$\begin{aligned}&\pi(\beta_{ik}|\lambda_{ik}=0,\cdot) \propto (1-\beta_{ik})\left[(1-\rho_k)\delta_0(\beta_{ik}) + \rho_k B\left(a_0 b_0, a_0(1-b_0)\right)\right], \\ &P(\beta_{ik}=0|\lambda_{ik}=0,\cdot) \propto (1-\rho_k), \quad P(\beta_{ik}\neq 0|\lambda_{ik}=0,\cdot) \propto (1-b_0)\rho_k.\end{aligned}$$

That is, with posterior odds $(1-b_0)\rho_k/(1-\rho_k)$ we sample from $B\left(a_0b_0, a_0(1-b_0)+1\right)$ and otherwise set β_{ik} equal to zero. Conditional on $\lambda_{ik} \neq 0$ we obtain

$$\pi(\beta_{ik}|\lambda_{ik} \neq 0, \cdot) \propto \beta_{ik} N\left(\lambda_{ik}; 0, \tau_k\right)\left[(1-\rho_k)\delta_0(\beta_{ik}) + \rho_k B\left(a_0b_0, a_0(1-b_0)\right)\right],$$
$$P(\beta_{ik} = 0|\lambda_{ik} \neq 0, \cdot) = 0, \quad P(\beta_{ik} \neq 0|\lambda_{ik} \neq 0, \cdot) = 1.$$

In this case we sample β_{ik} from $B\left(a_0b_0 + 1, a_0(1-b_0)\right)$.

(b) $\pi\left(\tau_k, \rho_k|\Lambda\right)$: Independently across factors, the posterior distributions take the form of inverse gamma, $\pi\left(\tau_k|\cdot\right) \sim IG\left(g_k, G_k\right)$, and beta distributions, $\pi\left(\rho_k|\cdot\right) \sim B\left(r_{1k}, r_{2k}\right)$, with

$$g_k = g_\tau + \frac{1}{2}\sum_{i=1}^{N} I\left\{\lambda_{ik} \neq 0\right\}, \quad G_k = G_\tau + \frac{1}{2}\sum_{i=1}^{N} {\lambda_{ik}}^2,$$

$$r_{1k} = r_0 s_0 + S_k, \quad r_{2k} = r_0(1-s_0) + N - S_k, \text{ and } S_k = \sum_{i=1}^{N} I\left\{\beta_{ik} \neq 0\right\},$$

and $I\left\{\cdot\right\}$ is the indicator function.

A.3 Remaining parameters, $\pi\left(\theta_{\backslash\Lambda}|\boldsymbol{y}, \boldsymbol{f}\right)$

Remaining parameters have standard posterior distributions:

1. Assuming a normal prior distribution for $\text{vec}\left([\Phi_1, \ldots, \Phi_p]\right) \sim N(\mathrm{p}_0, \mathrm{P}_0)$, the posterior is normal, $\Phi_1, \ldots, \Phi_p|\boldsymbol{f} \sim N\left(\mathrm{p}, \mathrm{P}\right)$ with

$$\mathrm{P} = \left(\left(\boldsymbol{f}^{-\prime}\boldsymbol{f}^{-} \otimes \Omega^{-1}\right) + \mathrm{P}_0^{-1}\right)^{-1}, \ \mathrm{p} = \mathrm{P}\left(\left(\boldsymbol{f}^{-\prime} \otimes \Omega^{-1}\right)\boldsymbol{f} + \mathrm{P}_0^{-1}\mathrm{p}_0\right),$$

where $\boldsymbol{f}^{-}$ is the matrix of lagged factors, $\boldsymbol{f}^{-} = \begin{bmatrix} f'_{-p+1} & \cdots & f'_0 \\ \vdots & & \vdots \\ f'_{T-1} & \cdots & f'_{T-p} \end{bmatrix}$, and $\Omega = I_K$ in the basic model (2).

2. Assuming a normal prior distribution for series-specific idiosyncratic dynamics, $\psi_{i1}, \ldots, \psi_{iq} \sim N(\mathrm{q}_0, \mathrm{Q}_0)$, leads to a normal posterior update $N(\mathrm{q}, \mathrm{Q})$, with

$$\mathrm{Q} = \left(\sigma_i^{-1}\hat{\varepsilon}_i^{-\prime}\hat{\varepsilon}_i^{-} + \mathrm{Q}_0^{-1}\right), \ \mathrm{q} = \mathrm{Q}\left(\sigma_i^{-1}\hat{\varepsilon}_i^{-\prime}\hat{\varepsilon}_i + \mathrm{Q}_0^{-1}\mathrm{q}_0\right),$$

where $\hat{\varepsilon}_i$ stacks residuals $\hat{\varepsilon}_{it} = y_{it} - \lambda_i f_t$ for $t = q+1, \ldots, T$ and $\hat{\varepsilon}_i^{-}$ stacks lagged values $(\hat{\varepsilon}_{i,t-1}, \ldots, \hat{\varepsilon}_{i,t-q})$, for $t = q, \ldots, T-q$.

Finally, a inverse gamma prior distribution for individual $\sigma_i \sim IG(g_s, G_s)$ leads to an inverse gamma posterior, updated individually across series, $\sigma_i|\cdot = IG(g_i, G_i)$, with $g_i = g_s + 0.5(T-q)$, $G_i = G_s + 0.5\sum_{t=q+1}^{T}\hat{\epsilon}_{it}^2$, $\hat{\epsilon}_{it} = \hat{\varepsilon}_{it} - \psi_{i1}\hat{\varepsilon}_{i,t-1} - \cdots - \psi_{iq}\hat{\varepsilon}_{i,t-q}$.