Dynamic factor models with common (drifting) stochastic trends

Sylvia Kaufmann, Rodney W. Strachan

Working Paper 24.02

Dynamic factor models with common (drifting) stochastic trends

Sylvia Kaufmann[*], Rodney W. Strachan[†]

January 24, 2024

Abstract

It is common to transform data to stationarity, such as by differencing and demeaning, before estimating factor models in macroeconomics. Imposing these transformations, however, limit opportunities to learn about trending behaviour. Trends and deterministic processes can play a central role in the behaviour of macroeconomic processes and so it is important to be able to characterise these features of the data. In this paper, we develop a model of common and idiosyncratic deterministic and stochastic processes in a factor model. We work with the unidentified model. A judicious choice of parameter expansion and post-processing ensures the model avoids a non-invariant specification such that the inference is data driven and the computation is efficient.

1 Motivation

In high-dimensional multivariate dynamic modelling, it is common practice to transform the data prior to analysis by differencing and then demeaning the differenced data. The data are sometimes also then standardized, but our focus is on the detrending transformation. There are good reasons for employing these transformations as, for example, they often improve computation or simply limit the scope of the study to reduce complexity by excluding consideration of the deterministic processes. There are, however, many important questions in economics that are informed by the behaviour of the deterministic

[*]Study Center Gerzensee and University of Basel
[†]University of Queensland, Brisbane

processes, such as the existence and form of balanced growth paths or steady state deterministic growth paths. It is, therefore, important in such cases to model these processes. This paper proposes a specification for a factor model that permits analysis of both the stochastic and the deterministic terms.

There are several approaches to decomposing vector series into common and idiosyncratic, stochastic and deterministic processes. Nonstationary univariate linear processes may be decomposed into nonstationary (permanent) and stationary (transitory) components using approaches developed, for example, in Beveridge and Nelson (1981). In a vector process, if there are common stochastic trends that capture all of the long-run, stochastic, non-stationary drifting component and the idiosyncratic components follow stationary processes then the processes may cointegrate (Engle and Granger 1987; Stock and Watson 1988; Sims et al. 1990). In the study of cointegrating systems, the multivariate Beveridge-Nelson decomposition identifies the cointegrating space and, as such, the permanent common and transitory idiosyncratic components of the system. This decomposition can be obtained from the vector error correction model (VECM) of the cointegrating VAR. Using the VECM, we may also extract the deterministic processes of the common components and the idiosyncratic deterministic processes for each variable in the system (Johansen 1995). Existing factor models do not fit within these decompositions, and VECMs may not be an appropriate choice to analyze high-dimensional vector series.

A factor model is a natural candidate to analyze high-dimensional data sets (Forni et al. 2000; Stock and Watson 2002a; Stock and Watson 2002b). As mentioned, most applications analyze non-trending and standardized data, where (drifting) trends are removed by mean-adjusting first differences of the (log) level series. A factor model fitted to first-differenced, mean-adjusted series implies that common factor and idiosyncratic shocks both have a permanent level effect, although permanent effects of idiosyncratic shocks sometimes are ruled out by assumption (such as a noninvertible MA process). Given the usual ex-ante data transformation, extracted factors will capture joint dynamics without extracting a potential common drift. Hence, extracted factors represent the stochastic part of the trend component of the level series – the sum of shocks – without incorporating the deterministic component – the drift.

Interest often lies in the relative contributions of common and idiosyncratic processes to total variation of the variables in the system. Such a decomposition of the stochastic processes can prove highly informative on a range of economic questions. Similarly, the deterministic terms may be decomposed into both common and idiosyncratic components and empirical analysis of these terms can inform us on a number of important issues. Therefore, it will be of interest to extract factors that include both the stochastic as well as the deterministic drift driving each series.

In the present paper, we specify a model to analyze time series displaying common as well as idiosyncratic deterministic drifts in levels. We extract factors that follow a unit root process with potential drift to capture the common drifting trend component. We do not consider cointegration to limit the scope to decomposition of the deterministic and stochastic trends into common and idiosyncratic terms, i.e. the idiosyncratic term in the factor model is not restricted to be over-differenced. As the factor-driven drift component may not account for all of a series-specific drift, the idiosyncratic non-stationary component potentially may also display a drift. In addition, we extend the model to take into account that the series-specific factor-driven drift may (randomly) deviate from the extracted factor drift. This allows us to capture common trends in series which behave as cointegrated including a trend.

We contribute to the literature on factor modelling by explicitly modelling series-specific deterministic terms. Given that we do not pre-impose overidentifying restrictions, posterior inference on the factor space, factor identification and interpretation is essentially data driven and self-organizing. The specification of the factor model raises identification issues, as the series-specific drift has to be decomposed into the common and idiosyncratic components. Moreover, without setting factor-identifying or rotational restrictions prior to posterior inference leaves so-called factor founders, i.e. factor-determining series, undetermined and renders factor-specific drifts undetermined, too. Factor interpretation or the factor space is determined by inducing shrinkage into the factor loading matrix (Griffin and Brown 2017), and rotational identification is obtained by post-processing the draws from the posterior distribution (Chan *et al.* 2018, Kaufmann and Schumacher 2019). To identify factor drifts, we derive the non-centered representation of the factor model (Papaspiliopoulos et al. 2007). We implement a parameter expansion procedure (Liu and Wu 1999) to sample factor loadings and obtain draws of factor drifts corresponding to averages across series-specific drifts weighted by factor loadings (i.e. factor strength). Posterior inference and post-processing allows us to characterize the common factors and decompose the level series into a common factor-driven drifting and an idiosyncratic, in most series non-trending, trend component.

In related work, non-stationary factors usually are extracted after de-meaning and standardizing the data. Most closely related to the present paper is Stock and Watson (1989), who extract a coincident indicator from four time series, transformed to growth rates by taking first differences of the logarithmic levels and mean-adjusted to remove the drift prior to estimation. The drift (δ) of the coincident indicator is backed out after estimation based on the Kalman gain. The deviation of the series's unconditional (average) growth rate from the factor growth rate represents the idiosyncratic deterministic component, $D_i = \overline{\Delta Y}_i - \delta$. Kim and Nelson (1998) use the approach to back out the drift of a coincident economic indicator with Markov switching growth rate. The decomposition implies a non-stationary process for the idiosyncratic component in the level series. In a more recent paper, Müller et al. (2022) model the low-frequency component of a large panel of GDP growth series. A local linear trend extracts the global common drift, while additional hierarchical levels capture group-specific, highly persistent (low-frequency) drift components. To render a factor model consistent with predictions of theoretical asset pricing models, Favero and Melone (2020) propose a factor error correction model, where the deviation of variables' levels from implied factor levels induces a mean-reverting reaction in returns. The cointegration assumption implies a stationary series-specific idiosyncratic component.

Only few papers published between Stock and Watson (1989) and Müller et al. (2022), Favero and Melone (2020), such as Bai and Ng (2004) and Barigozzi et al. (2016), explicitly account for series-specific deterministic terms. However, given that there is no interest in decomposing these terms into a factor-driven and an idiosyncratic component, the terms cancel out when the factor model is analyzed and estimated for variables transformed to non-trending, mean-adjusted series. Corona et al. (2020) analyze a non-stationary factor model, without including a drift. The extracted factor displays a drift, however, which in principle violates the symmetric distributional assumption of factor shocks. This is also the case in Bai (2004), who additionally assumes stationary idiosyncratic components in the level series. This implies cointegration between non-stationary factors and observed variables, as shown in Banerjee et al. (2017).

The next section introduces the specification of the factor model, discusses identification issues, and introduces an extension interpretable as a factor-driven drift model with random effects. Section 3 outlines the prior specifications and the Markov chain Monte Carlo (MCMC) sampler. We discuss parameter expansion in more detail and how to obtain identification of factor drifts. In Section 5, we estimate the model using datasets from various studies (Bai 2004; Eden et al. 2021; Müller et al. 2022) to illustrate the procedure. Section 6 concludes.

2 Model specification and identification issues

2.1 Econometric specification

We specify a factor model for the standardized growth rate, y_{it}, of series $i = 1, \ldots N$ in period $t = 1, \ldots, T$. The model will include both common factor (μ_k) and idiosyncratic (μ_i) drifts terms. The specification for the growth rate y_{it} is:

$$
\begin{aligned}
y_{it} &= \sum_{k=1}^{K} \lambda_{ik} f_{kt} + \varepsilon_{it} && (1) \\
f_{kt} &= \mu_k + \phi_{k1} f_{k,t-1} + \cdots + \phi_{kp} f_{k,t-p} + \nu_{kt}, \nu_{kt} \sim N(0,1) && (2) \\
\varepsilon_{it} &= \mu_i + \psi_{i1} \varepsilon_{i,t-1} + \cdots + \psi_{iq} \varepsilon_{i,t-q} + \epsilon_{it},\ \epsilon_{it} \sim N(0, \sigma_i) && (3)
\end{aligned}
$$

where we can see that the full deterministic process is $\sum_{k=1}^{K} \lambda_{ik}\mu_k + \mu_i$. As mentioned, the term μ_i captures the series-specific or idiosyncratic drift component of y_{it} that is unexplained by the common drifts μ_k, $k = 1, \ldots, K$. We assume independent factor dynamics, and for scaling purposes, the factor-specific shocks are assumed to be independent standard normal; $\nu_{kt} \sim N(0, 1)$. The idiosyncratic components ε_{it} evolve independently. Although the factors evolve independently, each series may be loaded by any factor as the factor loadings are unrestricted.

Stack all series at time t into vectors, such that $y_t = (y_{1,t}, \ldots, y_{N,t})'$, $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{N,t})'$ and $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{N,t})'$ are $(N \times 1)$ vectors. Next, define $(K \times 1)$ vectors $f_t = (f_{1,t}, \ldots, f_{K,t})'$ and $\nu_t = (\nu_{1,t}, \ldots, \nu_{K,t})'$. We gather the K common drifts μ_k into the $(K \times 1)$ vector μ and the N idiosynchratic drifts μ_i into the $(N \times 1)$ vector μ_ι. The model may then be written more compactly as:

$$
\begin{aligned}
y_t &= \lambda f_t + \varepsilon_t & (4)\\
\Phi(L) f_t &= \mu + \nu_t, \nu_t \sim N(0, I_K) & (5)\\
\Psi(L)\varepsilon_t &= \mu_\iota + \epsilon_t,\ \epsilon_t \sim N(0, \Sigma) & (6)
\end{aligned}
$$

where I_K is the identity matrix of dimension K and $\Phi(L) = I_K - \Phi_1 L - \cdots - \Phi_p L^p$ and $\Psi(L) = I_N - \Psi_1 L - \cdots - \Psi_q L^q$ are diagonal processes, $\Phi_j = \text{diag}(\phi_{1j}, \ldots, \phi_{Kj})$, $\Psi_j = \text{diag}(\psi_{1j}, \ldots, \psi_{Nj})$, and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_N)$. As usual in time series notation, we introduce the expressions $\Phi(1) = I_K - \sum_{j=1}^{p} \Phi_j$ and $\Psi(1) = I_N - \sum_{j=1}^{q} \Psi_j$.

We estimate the model without imposing rotational identification restrictions. In particular, none of the usual restrictions are imposed on the $(N \times K)$ matrix of factor loadings; $\lambda = \{\lambda_{ik}\}$. For example, a common identification scheme would impose an upper zero triangular structure on the matrix λ, such that $\lambda_{ik} = 0$ for $k > i$, $i = 1, \ldots, K$ and an identity covariance matrix for ν_t. This structure would define the first factor as the only one influencing the first series, y_{1t}, the first and second factors the only ones influencing the second series, y_{2t}, and so on. Alternatively, a leading identity matrix, such that $\lambda_{ii} = 1$, $\lambda_{ik} = 0$, $i \neq k$, $i, k = 1, \ldots, K$, would identify the leading series as distinct factors (this would also permit an unrestricted and unknown covariance matrix for ν_t but still identify the factors). Obviously, this would need a careful choice of the leading variables. Given that this may be a difficult choice in a large, maybe heterogeneous dataset, we estimate factor loadings without pre-determining these factor founders. This strategy leaves the factor-specific growth rate, μ_k, undetermined, too. We will retrieve μ_k as an average of series-specific growth rates weighted by factor loadings, i.e. factor strength. To solve the identification issue, in the next section we re-parameterize the specification into a non-centered representation (Papaspiliopoulos et al. 2007), and scale factor drifts in a parameter expansion step (Liu and Wu 1999).

2.2 Re-parametrization and extended specification

2.2.1 Non-centered series-specific trending

Define $\mu^* = (I_K - \sum_{j=1}^{p} \Phi_j)^{-1}\mu = \Phi(1)^{-1}\mu$ and $\mu_\iota^* = (1 - \sum_{j=1}^{q} \Psi_j)^{-1}\mu_\iota = \Psi(1)^{-1}\mu_\iota$ the unconditional growth rates of, respectively, factors and idiosyncratic components to re-parameterize model (1)-(3):

$$
\begin{aligned}
y_t &= \lambda\mu^* + \lambda(f_t - \mu^*) + \mu_\iota^* + (\varepsilon_t - \mu_\iota^*) \\
&= \lambda\mu^* + \mu_\iota^* + \lambda(f_t - \mu^*) + (\varepsilon_t - \mu_\iota^*) \qquad (7) \\
(f_t - \mu^*) &= \Phi_1(f_{t-1} - \mu^*) + \cdots + \Phi_p(f_{t-p} - \mu^*) + \nu_t \qquad (8) \\
(\varepsilon_t - \mu_\iota^*) &= \Psi_1(\varepsilon_{t-1} - \mu_\iota^*) + \cdots + \Psi_q(\varepsilon_{t-q} - \mu_\iota^*) + \epsilon_t \qquad (9)
\end{aligned}
$$

Representation (7) reveals that factor-specific unconditional (conditional) growth rates μ^* (μ) and λ are not identifiable from the data without imposing a scale identifying restriction. We will induce scaling using parameter expansion and interweaving into specification (7) to retrieve an update of the common growth rates as an average of series-specific growth rates weighted by factor loadings λ.

2.2.2 A random effects model for series-specific factor growth

In the non-centered specification (7)-(9), the kth factor-driven series-specific drift, $\lambda_{ik}\mu_k^*$, is exactly proportional to the factor drift. In other words, conditional on λ, the specification defines a pooled model for series-specific drifts. The factor-driven series-specific drift has – up to proportionality – exactly the same drift as factors. We extend the specification to capture a feature that we may (and in fact do) observe in data. Usually, series trend with the factor(s), at a rate that randomly differs from the factor drifts, however.

To capture this additional feature, we may extend the non-centered *pooled* specification (7)-(9) to a *random effects* factor growth components model:

$$
\begin{aligned}
y_{it} &= \lambda_i\,(\mu^* + u_i^*) + \lambda_i(f_t - \mu^*) + \mu_{\iota,i}^* + (\varepsilon_{it} - \mu_{\iota,i}^*) \\
&= \lambda_i\,(\mu^* + u_i^*) + \mu_{\iota,i}^* + \lambda_i(f_t - \mu^*) + (\varepsilon_{it} - \mu_{\iota,i}^*) \qquad (10) \\
(f_t - \mu^*) &= \Phi_1(f_{t-1} - \mu^*) + \cdots + \Phi_p(f_{t-p} - \mu^*) + \nu_t \qquad (11) \\
(\varepsilon_{it} - \mu_{\iota,i}^*) &= \psi_{i1}(\varepsilon_{i,t-1} - \mu_{\iota,i}^*) + \cdots + \psi_{iq}(\varepsilon_{i,t-q} - \mu_{\iota,i}^*) + \epsilon_{it} \qquad (12)
\end{aligned}
$$

with $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{iK})$ and $u_i^*|U_0 \sim N(0, U_0)$ a $K \times 1$ vector of random effects, and where $\mu_{\iota,i}^*$ and ε_{it} correspond to, respectively, $\mu_{\iota,i}^*$ and ε_{it} in Equation (7), purged of $\lambda_i u_i^*$ (see below). The diagonal variance matrix U_0 indicates the degree of heterogeneity across series' factor-driven drifts. Rather than fixing U_0, by specifying a hierarchical prior and drawing proper posterior inference, we will learn from the data the degree of heterogeneity across series-specific factor-driven drifts.

In (7), a potential a series-specific random factor-drift effect, $\lambda_i u_i^*$, would be captured by the idiosyncratic component $\mu_{\iota,i}^* = \tilde{\mu}_{\iota,i}^* + \lambda_i u_i^*$. Allocating the random effect to the factor component leads to:

$$\begin{aligned}
y_{it} &= \lambda_i \mu^* + \lambda_i (f_t - \mu^*) + \mu_{\iota,1}^* + (\varepsilon_t - \mu_{\iota,i}^*) \\
&= \lambda_i \mu^* + \lambda_i (f_t - \mu^*) + (\tilde{\mu}_{\iota,i}^* + \lambda_i u_i^*) + (\varepsilon_t - (\tilde{\mu}_{\iota,i}^* + \lambda_i u_i^*)) \\
&= \lambda_i (\mu^* + u_i^*) + \lambda_i (f_t - \mu^*) + \tilde{\mu}_{\iota,i}^* + (\overbrace{\varepsilon_t - \lambda_i u_i^*}^{\tilde{\varepsilon}_{it}} - \tilde{\mu}_{\iota,i}^*) \\
(\tilde{\varepsilon}_{it} - \tilde{\mu}_{\iota,i}^*) &= \psi_{i1} (\tilde{\varepsilon}_{i,t-1} - \tilde{\mu}_{\iota,i}^*) + \cdots + \psi_{iq} (\tilde{\varepsilon}_{i,t-q} - \tilde{\mu}_{\iota,i}^*) + \epsilon_{it}
\end{aligned}$$

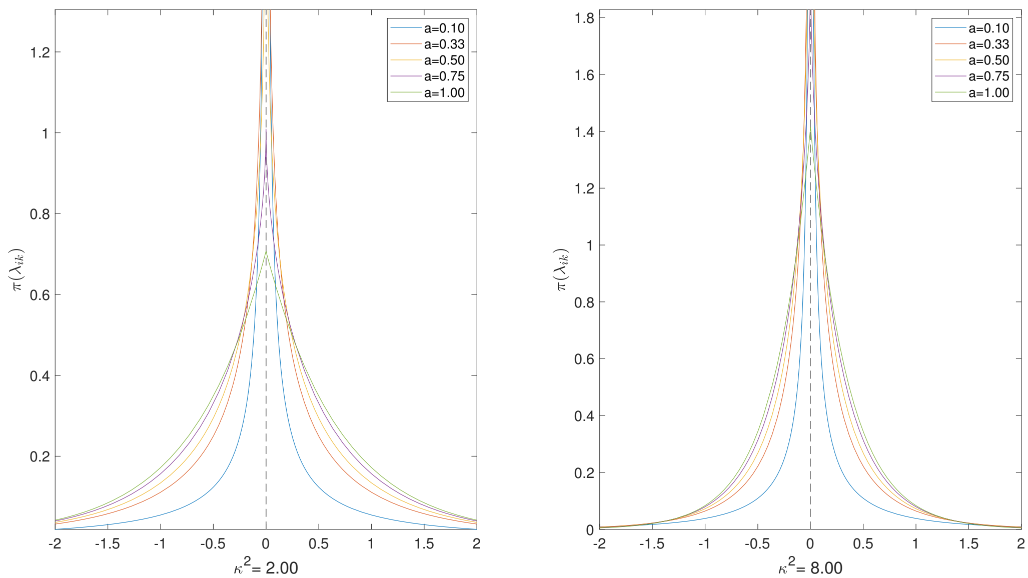
In Equations (10) and (12), we have re-defined $\mu_{\iota,i}^* := \tilde{\mu}_{\iota,i}^* = \mu_{\iota,i}^* - \lambda_i u_i^*$ and $\varepsilon_{it} := \tilde{\varepsilon}_{it} = \varepsilon_{it} - \lambda_i u_i^*$, to keep notation unchanged.

3 Bayesian Markov chain Monte Carlo inference

The specification of independent factor dynamics spans an independent factor basis. As factor founders remain undetermined a priori, we use a shrinkage prior on factor loadings which induces a data-driven, self-organizing association of series to factors. Groups of series with similar drifts and dynamics will strongly be loaded by the same factor, and ultimately determine the interpretation of factors. In the following, we use a global-local normal-gamma shrinkage prior (Griffin and Brown 2017), while other priors inducing shrinkage such as the spike-and-slab (George and McCulloch 1997) or exact zeros (point mass-normal mixture, West 2003) have also proven useful to infer a sparse, interpretable structure of factor loadings. The normal-gamma prior is useful here, as induced distributions in the parameter expansion step can be derived straightforwardly, while these may be more involved to derive under mixture shrinkage distributions.

We introduce additional hierarchical levels for factor and idiosyncratic drifts. A normal-gamma shrinkage prior specification on idiosyncratic drifts reflects our prior expectation to capture the observed drift in a series from factor drifts when the series displays a drift that is common to a group of series. As drifts may be very different across groups of series, specifying a factor-independent prior variance for factor drifts may necessitate an overly diffuse prior specification. Hence, by specifying a normal-inverse gamma prior for factor drifts we learn from data about the appropriate factor-specific prior drift variance.

Figure 1: Marginal density of λ_{ik} under a hierarchical normal-gamma prior.



3.1 Prior specifications

Inducing shrinkage into an otherwise unrestricted factor loading matrix will ultimately associate series of similar drift and dynamics to the same factor, and will allow us to potentially determine the interpretation of factors. Here, we implement a hierarchical normal-gamma shrinkage prior on factor loadings (Griffin and Brown 2017), inducing factor-series specific local shrinkage:[1]

$$\pi\left(\lambda_{ik}|\tau_{ik}\right) \;=\; N\left(0, \tau_{ik}\right), \quad \pi(\tau_{ik}) = G\left(a, a\kappa^2/2\right) \tag{13}$$

where the gamma hyperprior specification on τ_{ik} induces a global-local shrinkage on factor loadings. Overall shrinkage is determined by κ^2, the larger κ^2 the stronger the shrinkage. For $a < .5$, the marginal distribution of λ_{ik} has a pole at zero. Figure 1 illustrates the marginal density of λ_{ik} induced by various parameter specifications.

For the remaining parameters, we specify standard, independent distributions.

- Normal (hierarchical) distributions for parameters governing the common factor

[1] Additional hierarchical layers may be introduced when the dataset is high-dimensional or the factor loading matrix expected to be very sparse (Cadonna et al. 2020).

processes, $\mu = \{\mu_k | k = 1, \ldots, K\}$ and $\mathbf{\Phi} = \{\phi_{kj} | k = 1, \ldots, K; j = 1, \ldots, p\}$:

$$\begin{aligned}
\pi\left(\mu, \mathbf{\Phi} | M_0\right) &= \prod_{k=1}^{K} \pi\left(\mu_k | M_{k0}\right) \pi\left(\phi_{k1}, \ldots, \phi_{kp}\right) \\
&\quad \pi\left(\mu_k | M_{k0}\right) = N(\mathrm{m}_0, M_{k0}), \pi\left(\phi_{k1}, \ldots, \phi_{kp}\right) = N(\mathrm{p}_0, \mathrm{P}_0) \\
&\quad \pi\left(M_{k0}\right) = IG\left(\mathrm{s}_M, \mathrm{S}_M\right)
\end{aligned}$$

with p_0 and P_0 of dimension $p \times 1$ and $p \times p$, respectively. The prior variance of drifts may be fixed at $M_{k0} = \mathrm{S}$. However, by specifying an inverse gamma prior, $IG\left(\mathrm{s}_M, \mathrm{S}_M\right)$, we may learn from data about the prior variance. By using a hierarchical specification, we avoid specifying an overly diffuse prior on factor drifts, which may be necessary in the presence of large differences in observed drifts across groups of series.

- A hierarchical normal distribution for random effects $\boldsymbol{u}^* = [u_1^*, \ldots, u_N^*]'$, $u_i^* \sim N(0, U_0)$, U_0 diagonal with elements $U_{k0} \sim IG(\upsilon_0, \mathcal{U}_0)$, to learn from data the extent of heterogeneity across series-specific factor-driven drifts.

- Normal (hierarchical) distributions for parameters governing the idiosyncratic processes, $\mu_\iota = \{\mu_{\iota,i} | i = 1, \ldots, N\}$, $\mathbf{\Psi} = \{\psi_{ij} | i = 1, \ldots, N; j = 1, \ldots, q\}$ and Σ:

$$\begin{aligned}
\pi\left(\mu_\iota, \mathbf{\Psi} | M_0^i\right) &= \prod_{i=1}^{N} \pi\left(\mu_{\iota,i} | M_{i0}\right) \pi\left(\psi_{i1}, \ldots, \psi_{iN}\right) \\
&\quad \pi\left(\mu_{\iota,i} | M_{i0}\right) = N(\mathrm{m}_{i0}, M_{i0}),\ \pi\left(\psi_{i1}, \ldots, \psi_{iN}\right) = N(\mathrm{q}_0, \mathrm{Q}_0) \\
&\quad \pi\left(M_{i0}\right) = G\left(a_{\mu_i}, a_{\mu_i} \kappa_{\mu_i}^2 / 2\right) \\
\pi\left(\Sigma\right) &= \prod_{i=1}^{N} \pi\left(\sigma_i\right) \\
&\quad \pi\left(\sigma_i\right) = IG(\mathrm{s}_0, \mathrm{S}_0)
\end{aligned}$$

with q_0 and Q_0 of dimension $q \times 1$ and $q \times q$, respectively. By specifying a gamma prior on M_{i0}, we induce shrinkage in $\mu_{\iota,i}$. This reflects our preference for a non-trending idiosyncratic component, that is, we prefer to capture the observed series-specific drift by some factor drifts, if the observed drift is similar to the drift of a group of series.

Specification (2) induces a normal prior for the factors, $\boldsymbol{f} = (f'_{-p+1}, \ldots, f'_1, \ldots, f'_T)'$, $\pi\left(\boldsymbol{f} | \mu, \mathbf{\Phi}\right) = N\left(\mathrm{f}_0, \mathrm{F}_0\right)$, see Appendix A.1 for implied moments f_0 and F_0.

3.2 The likelihood

In analogy to factors, we collect all observations in vector $\boldsymbol{y} = \{y_{it}|i = 1, \ldots, N, t = 1, \ldots T\}$, or $\boldsymbol{y}_i = \{y_{it}|t = 1, \ldots N\}$. Parameters are collected in $\theta = \{\lambda, \boldsymbol{\tau}, \mu, M_0, \mu_\iota, M_0^i, \boldsymbol{\Psi}, \boldsymbol{\Phi}, \Sigma\}$, $\boldsymbol{\tau} = \{\tau_{ik}|i = 1, \ldots, N; k = 1, \ldots, K\}$, $M_0 = \{M_{k0}|k = 1, \ldots, K\}$, $M_0^i = \{M_{i0}|i = 1, \ldots, N\}$.

Conditional on factors $\boldsymbol{f}$, the likelihood factorizes:

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

where $\tilde{y}_{it} = \psi_i(L)y_{it}$, $\tilde{f}_{it} = \psi_i(L)\left(f_t + u_i^*\right)$ represent series-specific filtered values and $n(\mathrm{v}|\mathrm{m}, \mathrm{s})$ indicates the normal density with mean m and variance s evaluated at v.

3.3 Posterior sampler

To infer the posterior we combine the prior (Section 3.1) with the data likelihood, and draw iteratively from

S.1. $\pi\left(\boldsymbol{f}|\boldsymbol{y}, \boldsymbol{u}^*, \theta\right)$: We apply the sampler as detailed in Appendix A.1, where series are filtered and purged from the idiosyncratic and the random factor growth effects $\tilde{y}_t = \Psi(L)y_t - \mu_\iota - \tilde{u}$, with $\tilde{u} = \Psi(1)\left(\lambda_1 u_1^*, \ldots, \lambda_N u_N^*\right)'$.

S.2. $\pi\left(\lambda, \boldsymbol{\tau}|\boldsymbol{y}, \boldsymbol{f}, \boldsymbol{u}^*, \theta\right) = \pi\left(\boldsymbol{\tau}|\lambda\right)\pi\left(\lambda|\boldsymbol{y}, \boldsymbol{f}, \boldsymbol{u}^*, \theta\right)$: We apply parameter expansion to obtain a joint draw of λ and series' unconditional growth rates, see the next subsection for a detailed derivation. Conditional on λ, we obtain an update on μ^*, which scales factor drifts, and an update on $\boldsymbol{\tau}$:

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

where GIG indicates the generalized inverse Gaussian distribution.

S.3. $\pi\left(u^*|\boldsymbol{y}, \boldsymbol{f}, \theta\right)$: Based on the centered specification, derive posterior moments for series-specific random effects $\pi\left(u_i^*|\boldsymbol{y},\boldsymbol{f},\theta\right)=N\left(\mathrm{u}_i,\mathrm{U}_i\right)$; and update $U_{k0}|\boldsymbol{u}^*$:

$$\begin{aligned}
\tilde{y}_{it} &= \psi_i(L)\left(y_{it}-\lambda_i f_t\right)=\psi_i(1)\lambda_i u_i^*+\mu_{\iota,i}+\epsilon_{it} \\
&\quad \text{with } u_i^*\sim N(0,U_0),\ \mu_{\iota,i}\sim N\left(\mathrm{m}_{i0},M_{i0}\right),\ \epsilon_{it}\sim N\left(0,\sigma_i\right) \\
&\quad \mathrm{U}_i=\left(T\lambda_i'\psi_i(1)\left(M_{i0}+\sigma_i\right)^{-1}\psi_i(1)\lambda_i+U_0^{-1}\right)^{-1} \\
&\quad \mathrm{u}_i=\mathrm{U}_i\left(\lambda_i'\psi_i(1)\left(M_{i0}+\sigma_i\right)^{-1}\sum_t(\tilde{y}_{it}-\mathrm{m}_{i0})\right) \\
U_{k0}|\boldsymbol{u}^* &\sim IG\left(\upsilon_0+.5N,\mathcal{U}_0+.5\sum_{i=1}^N u_{ik}^{*2}\right)
\end{aligned}$$

S.4. $\pi\left(\mu_\iota,\boldsymbol{\Psi}|\boldsymbol{y},\boldsymbol{f},\boldsymbol{u}^*,\lambda,\Sigma,M_0^i\right)$, and $\pi\left(\Sigma|\boldsymbol{y},\boldsymbol{f},\lambda,\mu_\iota,\boldsymbol{\Psi}\right)$ independently across equations: For $i=1,\ldots,N$, based on

$$\begin{aligned}
y_{it}-\lambda_i\left(f_t+u_i^*\right)=\varepsilon_{it} &= \mu_{\iota,i}+\psi_{i1}\varepsilon_{i,t-1}+\cdots+\psi_{iq}\varepsilon_{i,t-q}+\epsilon_{it},\ \epsilon_{it}\sim N(0,\sigma_i) \\
&= X_{it}'(\mu_{\iota,i},\psi_{i1},\ldots,\psi_{iq})'+\epsilon_{it},\ X_{it}'=(1,\varepsilon_{it-1},\ldots,\varepsilon_{i,t-q})
\end{aligned}$$

and given prior specifications, the posterior is joint normal:

$$(\mu_{\iota,i},\psi_{i1},\ldots,\psi_{iq})'|\boldsymbol{y}_i,\boldsymbol{f},\lambda_i,\sigma_i,M_{i0}\sim N(\mathrm{q}_i,\mathrm{Q}_i)$$

$$\mathrm{Q}_i=\left(\sum_{t=q+1}^T X_{it}X_{it}'/\sigma_i+\left[\begin{array}{cc} M_{i0}^{-1} & 0 \\ 0 & \mathrm{Q}_0^{-1}\end{array}\right]\right)^{-1},\ \mathrm{q}_i=\mathrm{Q}_i\left(\sum_{t=q+1}^T X_{it}\varepsilon_{it}/\sigma_i\right)$$

Given a draw $(\mu_{\iota,i},\psi_{i1},\ldots,\psi_{iq})$, we update σ_i and draw from:

$$\begin{aligned}
&\sigma_i|\boldsymbol{y}_i,\boldsymbol{f},\lambda_i,\mu_{\iota,i},\psi_{i1},\ldots,\psi_{iq}\sim \\
&\qquad IG\left(\mathrm{s}_0+.5(T-q),\mathrm{S}_0+.5\sum_{t=q+1}^T\left(\varepsilon_{it}-X_{it}'(\mu_{\iota,i},\psi_{i1},\ldots,\psi_{iq})'\right)^2\right)
\end{aligned}$$

Under a hierarchical prior for $\mu_{\iota,i}$, we update M_{i0}

$$\pi\left(M_{i0}|\mu_{\iota,i}\right)=GIG\left(a_{\mu_i}-1/2,a_{\mu_i}\kappa_{\mu_i}^2,(\mu_{\iota,i}-\mathrm{m}_{i0})^2\right)$$

S.5. $\pi\left(\mu,\boldsymbol{\Phi}|\boldsymbol{f},M_0\right)$ independently across factor-specific equations: For $k=1,\ldots,K$, based on

$$\begin{aligned}
f_{kt} &= \mu_k+\phi_{k1}f_{k,t-1}+\cdots+\phi_{kp}f_{k,t-p}+\nu_{kt},\ \nu_{kt}\sim N(0,1) \\
&= X_{kt}'(\mu_k,\phi_{k1},\ldots,\phi_{kp})'+\nu_{kt},\ X_{kt}'=(1,f_{k,t-1},\ldots,f_{k,t-p})
\end{aligned}$$

and given prior specifications, the posterior is joint normal:

$$(\mu_k, \phi_{k1}, \ldots, \phi_{kp})'|\boldsymbol{f}, M_{k0} \sim N(\mathrm{p}_k, \mathrm{P}_k)$$
$$\mathrm{P}_k = \left(\sum_{t=p+1}^{T} X_{kt}X'_{kt} + \begin{bmatrix} M_{k0}^{-1} & 0 \\ 0 & \mathrm{P}_0^{-1} \end{bmatrix} \right)^{-1}, \ \mathrm{p}_k = \mathrm{P}_k \left(\sum_{t=p+1}^{T} X_{kt} f_{kt} \right)$$

Under a hierarchical prior for μ_k, we update M_{k0}

$$\pi\left(M_{k0}|\mu_k\right) = IG\left(\mathrm{s}_M + .5, \mathrm{S}_M + .5(\mu_k - \mathrm{m}_0)^2\right)$$

S.6. Terminate the iteration by a random permutation $\varrho = (\varrho_1, \ldots, \varrho_K)$ of factor position and sign, permute accordingly factor-specific parameters. Given that factor ordering and sign remain unidentified while sampling, random permutation enforces label and sign switching and allows us to explore the unconstrained posterior distribution (Frühwirth-Schnatter 2001).

3.4 Parameter expansion

We use parameter expansion to sample λ in step **S**.2, based on re-parametrization (10)-(12). Observation equation (10) shows that λ is not identifiable from data, unless we pin down factor unconditional (conditional) drifts. Gathering terms and filtering series-specifically:

$$y_{it} = \lambda_i \left(\mu^* + u_i^*\right) + \mu^*_{\iota,i} + \lambda_i(f_t - \mu^*) + (\varepsilon_{it} - \mu^*_{\iota,i}) \tag{14}$$
$$\psi_i(L)y_{it} = \psi_i(1)\left(\sum_k \lambda_{ik}(\mu_k^* + u_{ik}^*) + \mu^*_{\iota,i}\right) + \sum_k \lambda_{ik}\psi_i(L)\left(f_{kt} - \mu_k^*\right) + \epsilon_{it} \tag{15}$$
$$\tilde{y}_t = c + \Lambda \tilde{f}_t^* + \epsilon_t \tag{16}$$

where $\tilde{y}_t = (\psi_1(L)y_{1t}, \ldots, \psi_N(L)y_{Nt})'$ and $\tilde{f}_t^* = ((\psi_1(L)(f_t - \mu^*)', \ldots, (\psi_N(L)(f_t - \mu^*))')'$ stack series-specific filtered series, c stacks filtered series-specific unconditional growth rates or drifts $c_i = \psi_i(1)\left(\sum_k \lambda_{ik}(\mu_k^* + u_{ik}^*) + \mu^*_{\iota,i}\right)$. The matrix Λ is block-diagonal, each row i containing row vector i of λ.

Parameter expansion allows joint sampling of λ and c independently across series. Conditional on λ, we obtain an update of the unconditional factor drift μ^* as a weighted average over series-specific drifts, with weights determined by factor loadings. We proceed in three steps:

PX.1. Equation (5) implies unconditional moments for factors, $f_t|\mu, \mathbf{\Phi} \sim N(\mu^*, \Sigma_{\nu^*})$, with $\mu^* = \Phi(1)^{-1}\mu$ and Σ_{ν^*} implied by

$$\text{vec}(\mathbf{\Sigma}_{\nu^*}) = \left(I_{(Kp)^2} - \left(\tilde{\mathbf{\Phi}} \otimes \tilde{\mathbf{\Phi}}\right)\right)^{-1} \text{vec}\left(\left[\begin{array}{c} I_K \quad 0_{K\times K(p-1)} \\ 0_{K(p-1)\times Kp} \end{array}\right]\right),$$

$$\tilde{\mathbf{\Phi}} = \left[\begin{array}{c} \Phi_1, \ldots, \Phi_p \\ I_{K(p-1)}, 0_{K(p-1)\times K} \end{array}\right]$$

Draw values μ^* from

$$\begin{aligned} &\mu^*|\boldsymbol{f}, \mathbf{\Phi}, M_0 \sim N\left(\mathrm{m}^*, \mathrm{M}^*\right) \\ &\mathrm{M}^* = \left(T\Sigma_{\nu^*}^{-1} + \mathrm{M}_0^{*-1}\right)^{-1},\ \mathrm{m}^* = \mathrm{M}^*\left(\sum_t \Sigma_{\nu^*}^{-1} f_t + \mathrm{M}_0^{*-1}\mathrm{m}_0^*\right) \end{aligned}$$

with implied prior parameters

$$\pi\left(\mu^*|\mathbf{\Phi}, M_0\right) = N\left(\mathrm{m}_0^*, \mathrm{M}_0^*\right),\ \mathrm{m}_0^* = \Phi(1)^{-1}\mathrm{m}_0,\ \mathrm{M}_0^* = \Phi(1)^{-1}M_0\Phi(1)^{-1\prime}$$

Draw u_i^* from $N\left(0, U_0\right)$.

PX.2. Mean-adjust $\boldsymbol{f}$, $f_t^* = f_t - \mu^*$ and update c and λ based on (16), independently across series.

Conditional on working parameters μ^*, $\boldsymbol{u}^*$, and $\Psi(L)$, $\boldsymbol{\tau}$, the implied prior on c_i is normal, $\pi\left(c_i|\mu^*, \boldsymbol{u}^*, \psi_i(1), \tau_i, M_{i0}\right) = N(\mathrm{c}_{i0}, \mathrm{C}_{i0})$, with $\mathrm{c}_{i0} = \psi_i(1)\mathrm{m}_{i0}^* = \mathrm{m}_{i0}$, $\mathrm{C}_{i0} = \psi_i(1)^2\left(\sum_{k=1}^K \tau_{ik}\left(\mu_k^* + u_{ik}^*\right)^2 + \mathrm{M}_{i0}^*\right)$, with implied prior moments $\mathrm{m}_{i0}^* = \psi_i(1)^{-1}\mathrm{m}_{i0}$, $\mathrm{M}_{i0}^* = \psi_i(1)^{-2}M_{i0}$.

Combined with observations:

$$\begin{aligned} \tilde{y}_{it} &= c_i + \lambda_i \tilde{f}_t^{*(i)} + \epsilon_{it},\ t = 1, \ldots, T \\ \tilde{\boldsymbol{y}}_i &= \underbrace{\left[\begin{array}{cc} 1 & \tilde{f}_1^{*(i)\prime} \\ \vdots & \\ 1 & \tilde{f}_T^{*(i)\prime} \end{array}\right]}_{\mathbf{X}_i} \left[\begin{array}{c} c_i \\ \lambda_i' \end{array}\right] + \boldsymbol{\epsilon}_i, \qquad \tilde{f}_t^{*(i)} = \psi_i(L) f_t^* \end{aligned}$$

we update (c_i, λ_i) and draw from a normal posterior

$$\begin{aligned} &\pi\left((c_i, \lambda_i)'|\tilde{\boldsymbol{y}}_i, \mathbf{X}_i, \sigma_i\right) = N(\mathrm{l}_i, \mathrm{L}_i) \\ &\mathrm{L}_i = \left(\mathbf{X}_i'\mathbf{X}_i/\sigma_i + \mathrm{L}_{i0}^{-1}\right)^{-1},\ \mathrm{l}_i = \mathrm{L}_i\left(\mathbf{X}_i'\tilde{\boldsymbol{y}}_i/\sigma_i + \mathrm{L}_{i0}^{-1}\mathrm{l}_{i0}\right) \end{aligned}$$

where prior moments are defined accordingly:

$$\mathrm{L}_{i0} = \mathrm{diag}(\mathrm{C}_{i0}, \tau_{i1}, \ldots, \tau_{iK}),\ \mathrm{l}_{i0} = (\mathrm{c}_{i0}, 0_{1\times K})'$$

PX.3. Conditional on λ, and c, update working parameters μ^* based on a random effects model for series-specific unconditional growth rates.[2]

Based on the regression model:
$\psi_i(1)^{-1}c_i = (\lambda_i(\mu^* + u_i^*)) + \mu_{\iota,i}^*$
with $\mu^* \sim N(\mathrm{m}^*, \mathrm{M}^*)$, $u_i^* \sim N(0, U_0)$, $\mu_{\iota,i}^* \sim N(\mathrm{m}_{i0}^*, \mathrm{M}_{i0}^*)$
Derive the marginal model:
$\psi_i(1)^{-1}c_i = \lambda_i\mu^* + e_i,\ e_i \sim N(\mathrm{m}_{i0}^*, \lambda_i \mathrm{U}_0 \lambda_i' + \mathrm{M}_{i0}^*)$
Collect terms:
$\Psi(1)^{-1}c = \lambda\mu^* + e,\ e \sim N\left(\mathrm{m}_0^{i*}, \mathrm{E}\right)$, E diagonal with elements $\lambda_i U_0 \lambda_i' + \mathrm{M}_{i0}^*$

We obtain the following updating distribution:

$$\mu^*|e, \lambda = N\left(\mathrm{m}^+, \mathrm{M}^+\right)$$
$$\mathrm{M}^+ = \left(\lambda'\mathrm{E}^{-1}\lambda + \mathrm{M}^{*-1}\right)^{-1},\ \mathrm{m}^+ = \mathrm{M}^+\left(\lambda'\mathrm{E}^{-1}(\Psi(1)^{-1}c - \mathrm{m}_0^{i*}) + \mathrm{M}^{*-1}\mathrm{m}^*\right)$$

Add μ^* to f_t^* to update $f_t = f_t^* + \mu^*$.

4 Posterior identification and interpretation

4.1 Post-processing

The sampler produces M draws from the posterior distribution. As the sampler enforces switching of factor position and sign at the end of each iteration, the output does not allow us to make posterior inference on factors and factor-related parameters. We proceed as described in Kaufmann and Schumacher (2019), and also use the same settings for grouping draws (β^{corr}) and determine the relevant groups (β^{draws}). However, here we use draws of *factors* stacked with respective factor-specific draws of N loadings to determine *factor representatives*. We proceed as follows:

P.1. Collect all KM stacked *factor* and *factor loading* draws,

$$\bar{\boldsymbol{f}}^{(l)} = \left(f_{k1}^{(m)}, \ldots, f_{kT}^{(m)}, \lambda_{1k}^{(m)}, \ldots, \lambda_{Nk}^{(m)}\right)'$$
$$l = 1, \ldots, KM;\ l = k + K(m-1),\ k = 1, \ldots, K,\ m = 1, \ldots, M$$

[2] For completeness, we keep notation m_{i0}^* in the following. Note that usually the prior specification will imply $\mathrm{m}_{i0}^* = 0$.

and group those into a group $\mathcal{J}_g$ for which absolute correlation is larger than $\beta^{corr} = .8$

$$\mathcal{J}_g = \left\{ \bar{\boldsymbol{f}}^{(j)} | j = 1, \dots, N_g; |corr(\bar{\boldsymbol{f}}^{(j_1)}, \bar{\boldsymbol{f}}^{(j_2)})| > .8, \ \forall j_1, j_2 \in \{1, \dots, N_g\} \right\}$$

Retain those groups for posterior inference which contain at least $\beta^{draws} M$ draws, $\beta^{draws} = .8$.

Sign-identify the group and define the group-specific mean or median a *factor representative*, $\bar{f}_g$, $g = 1, \dots, G$, $G \leq K$.

P.2. If $G = K$, re-order each draw (m) according to maximum absolute correlation with factor representatives, and retain only draws with a unique assignment. Define a permutation $\rho = (\rho_1, \dots, \rho_K) \in \{1, \dots, K\}$, $\rho_1 \neq \dots \neq \rho_K$ such that:

$$\bar{\boldsymbol{f}}_1^{(m)}, \dots, \bar{\boldsymbol{f}}_K^{(m)} := \rho\left(\bar{\boldsymbol{f}}^{(m)}\right) = \left\{ \bar{\boldsymbol{f}}_{\rho_k}^{(m)} \middle| |corr(\bar{\boldsymbol{f}}_{\rho_k}^{(m)}, \bar{f}_k)| = \max_{j=1,\dots,K} |corr(\bar{\boldsymbol{f}}_{\rho_j}^{(m)}, \bar{f}_k)| \right\}$$

Adjust the sign of draws negatively correlated with the factor representative, $\bar{\boldsymbol{f}}_k^{(m)} = sign\left(corr(\bar{\boldsymbol{f}}_k^{(m)}, \bar{f}_k)\right) corr(\bar{\boldsymbol{f}}_k^{(m)}, \bar{f}_k)$, $k = 1, \dots, K$.

If $G < K$, we re-order draws such that the factors $k = 1, \dots, G$ are the ones highest correlated with factor representatives. Only those are retained for posterior inference. See also the remark below. Overall, this step removes label- and sign-switching.

P.3. Based on selected posterior statistics such as the median or mean of factor loadings, or *significant* factor loadings determined by means of the 95% highest posterior density interval, we may re-determine factor position and the orientation of the factor basis. We may re-order factors according to the number of *exclusive* or *significant* loadings of a factor. We also may require most loadings of a specific factor to be positive, sign-adjust draws of factors, factor loadings and factor-specific parameters accordingly.

Remark to P.1: Highly correlated draws of factors and factor loadings represent draws from the same posterior distribution. If all fitted K factors are relevant to capture data dynamics, we should be able to identify $G = K$ groups of M draws each. Nevertheless, the number of draws to identify a group is set to less than M, $\beta^{draws} M$, to account for the fact that some posterior distributions may overlap, and hence some draws may not be clearly assigned to a group. If a model overfits the number of factors, $G < K$, overfitted factors and factor loadings, $\bar{\boldsymbol{f}}^{(l^*)}$, will be drawn from the prior distribution as there is no information in the data for updating. These draws will only be loosely (cross-)correlated and not assigned to nor forming a relevant group, i.e. either $|corr(\bar{\boldsymbol{f}}^{(l^*)}, \bar{\boldsymbol{f}}^{(j_1)})| < .8$, for any $j_1 \neq l^*$, or $N_{g^*} \leq \beta^{draws} M$ for $\mathcal{J}_{g^*}$ collecting draws of over-fitting factors. On the other hand, if a model underfits the relevant number of factors, potentially $G > K$ and groups should collect maybe even less than $\beta^{draws}M$ posterior draws each. Ultimately, this step determines the number of factors. This suggests it is reasonable to start out with a model potentially overfitting the number of factors and reducing it, until $G = K$ factor representatives are extracted from posterior draws.

4.2 Trend components

The sorted MCMC output lends to posterior evaluation like interpretation of factor means or determining factor founders based on large, non-zero factor loadings. Given y_t is expressed in growth rates, for some applications it may be of interest to derive factor-driven (F_t) or idiosyncratic (E_{it}) trend components of the level series Y_t using the Beveridge-Nelson decomposition (Beveridge and Nelson 1981).

$$\begin{aligned} F_t &= \Phi(1)^{-1}\left(\mu t + \sum_{t=1}^{t} \nu_t\right) \\ E_{it} &= \psi_i(1)^{-1}\left(\mu_{\iota,i} t + \sum_{t=1}^{t} \epsilon_{it}\right) \end{aligned}$$

To adjust to the volatility of series, multiply with series-specific standard deviations, $s = (s_1, \ldots, s_N)$, $s_i = \sqrt{1/T \sum_t (y_t - \bar{y})^2}$, $\bar{y} = 1/T \sum_t y_t$. For specification (1) we obtain the series-specific decomposition:

$$\begin{aligned} s \odot y_t &= s \odot \lambda f_t + s \odot \varepsilon_t \\ f_t &= \mu + \Phi_1 f_{t-1} + \cdots + \Phi_p f_{t-p} + \nu_t \\ \varepsilon_t &= \mu^i + \Psi_1 \varepsilon_{t-1} + \cdots + \Psi_q \varepsilon_{t-q} + \epsilon_t \\ Y_{it} &= \underbrace{s_i \lambda_i \Phi(1)^{-1}\left(\mu t + \sum_{j=1}^{t} \nu_t\right)}_{\tau_{it}^f\text{: factor trend}} + \underbrace{s_i \psi_i(1)^{-1}\left(\mu_{\iota,i} t + \sum_{j=1}^{t} \epsilon_{it}\right)}_{\tau_{it}^s\text{: idiosyncratic trend}} + Y_{i0} + \text{cycle}_{it} \end{aligned}$$

For the model with random effects (10), the decomposition is:

$$Y_{it} = \underbrace{s_i \lambda_i \left((u_i^* + \Phi(1)^{-1}\mu)t + \Phi(1)^{-1} \sum_{j=1}^{t} \nu_t\right)}_{\tau_{it}^f\text{: factor trend}} + \underbrace{s_i \psi_i(1)^{-1}\left(\mu_{\iota,i} t + \sum_{j=1}^{t} \epsilon_{it}\right)}_{\tau_{it}^s\text{: idiosyncratic trend}} + Y_{i0} + \text{cycle}_{it}$$

These decompositions may be useful to derive a specific aggregate from disaggregate components. Note that we may also derive factor-specific trend components in a similar way. When factors are correlated or factor dynamics more general, upon identification structural shock-specific trend components may be compiled.

5 Illustrations

5.1 Data and specifications

For illustration, this section estimates the model for three sets of yearly data, which display considerable persistence and a trending behaviour. The series for each of the three data sets are plotted in Figure 2 in, respectively, panels (a), (b) and (c). We first revisit the data used in Bai (2004)[3]. Panel (a) plots the sixty yearly series from Bai (2004), spanning 1948 to 2000, of U.S. sectoral employment (Full-time equivalent workers). The data are ordered by the level in 2000, and we observe that while most of the series display an upward trend, some of them are trending negatively. Next, we fit the model to a balanced panel of $N = 111$ yearly real GPD per capita series, running from 1960 to 2019 (Panel (b)).[4] The data corresponds to a sub-sample of data used in Müller, Stock, and Watson (2022), who extract a series-specific low-frequency trend factor from an unbalanced sample covering the years 1900 to 2017. Figure 2 (b) plots the series in log levels, ordered in levels as of 2019. With few exceptions, countries in 2019 ended up with a higher per-capita real GDP level than in 1960.[5] Finally, we will estimate the model for a challenging dataset, plotted in Panel (c) of Figure 2. The data includes relative capital goods prices grouped into two broad classes: ICT and Non-ICT prices. Subcategories consist of Equipment and Structures, Residential (R), Non-residential (N) and Consumer durables (C). The dataset has been used in Eden and Gaggl (2019), who show that capital composition matters to disentangle the sources of the decreasing labor income share in the United States.[6] The left panel plots the logarithm of all $N = 108$ yearly series spanning the period 1982 to 2018, while the right panel zooms in by excluding the series with extraordinary initial (relative) price levels. Obviously, there is a lot of heterogeneity across data trends, where some series display an overall decrease in prices of 99% while others reach an increase in prices of more than 200%.

To determine the number of factors in each data set, we proceed as discussed in Subsection 4.1. We capture factor dynamics by setting $p = 1$, and present results for $q = 0$.[7] As series do not trend in exact proportion, we allow for heterogeneity across series-specific

[3] They are available at https://www.bea.gov/data/employment/employment-by-industry, and we concatenate data from Table 6.5B and 6.5C.

[4] Data available at https://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt100.

[5] The countries with a lower real GDP per capita level were, ranking from largest to weakest decrease, the Democratic Republic of the Congo, the Central African Republic, Venezuela, Madagascar, Haiti, the Niger, the Gambia, and Guinea-Bissau, of which the first six rank within the poorest ten countries as of 2019.

[6] We thank Maya Eden and Paul Gaggl for providing us with the data set.

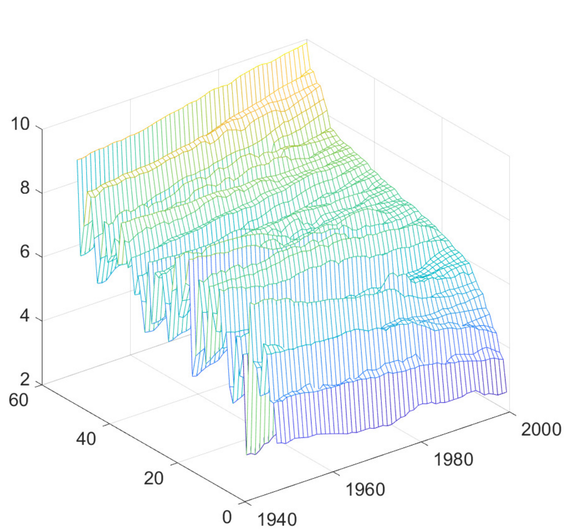
[7] Qualitatively, results do not change substantially when idiosyncratic dynamics are set to $q = 1$. The idiosyncratic component gets more weight for some series, though. This leads to a much better fit of data.

factor-driven trends by estimating a random effects specification for factor-driven drifts, see Subsection 2.2.2.

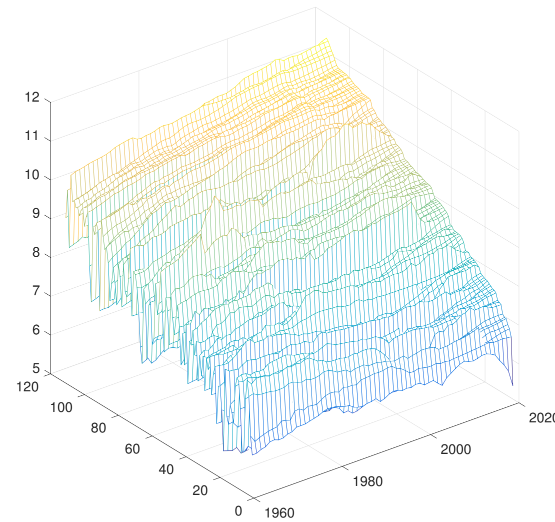
We set equal hyperparameters across applications. For factor processes we set $\mathrm{m}_0 = 0$, $\mathrm{s}_M = 3$ and $\mathrm{S}_M = .18$, which induces a mean and variance for M_{k0} of, respectively, .09 and .01, or a 95% highest density interval (HDI) of $(0.02, 0.22)$; $p_0 = 0$ and $P_0 = \mathrm{c}I_K$ diagonal with $\mathrm{c} = .16$. The prior for random effects is specified with $\upsilon_0 = 3$, $\mathcal{U}_0 = .5$, which induces a mean and variance for prior heterogeneity U_{k0} of, respectively, .25 and .0625, or a 95% HDI of $(.04, .62)$. For idiosyncratic drifts, we specify $\mathrm{m}_{i0} = 0$, $a_{\mu_i} = 0.75$ and $\kappa^2_{\mu_i} = 8$. This induces a mean and variance of, respectively, .25 and .08, or a 95% HDI of $(0, .83)$, for M_{i0}. The inverse gamma prior on σ_i is rather diffuse, specified with $\mathrm{s}_0 = 2$ and $\mathrm{S}_0 = 1$, inducing a prior mean of 1 and no second moment.

To obtain posterior inference, we run a first chain of 11,000 iterations, then obtain five new starting values by applying five random orthogonal rotations to the factor loading matrix and factor-specific parameters. With these, we run five parallel chains of 11,000 iterations. We discard the 6,000 initial values from each chain, which leaves us with 30,000 iterations for posterior evaluation.

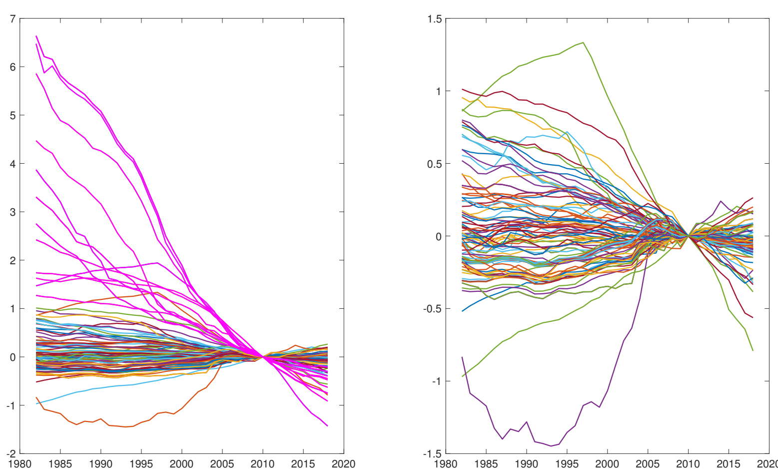
Figure 2: Data. (a) Log of sectoral employment ordered by the level in 2000; (b) Log of real per capita GDP ordered by the level in 2019; (c) Log prices of capital goods (left), zoom-in, excluding series with initial level above 1.2 (right).



(a)



(b)



(c)

5.2 Sectoral employment

For sectoral employment data, we identify $K = 3$ factors that adequately capture common data trends. Figure 3, Panel (a), plots the posterior median of factor trend components with a 50% highest posterior density interval (HPDI). Factor 2 compares well to the linear combination of factors rotated towards total employment in Bai (2004) (ibid. Figure 6). The correlation of total employment growth with our extracted factor trend growth amounts to roughly .90. In addition, Factor 1 accomodates a structural break in the level of some series, while since 1980 Factor 3 captures a decreasing employment trend in some sectors. Panel (b) of Figure 3 illustrates the amount of heterogeneity estimated across series-specific factor-driven trends. The boxplots suggest that heterogeneity is particulary present for Factor 3.

Figure 4, Panel (a), displays a heatmap of mean factor loadings. Inducing shrinkage identifies well those factors loading most strongly on each series. Panel (b) visualizes factor-series association for loadings shifted away from zero (based on a 95% HPDI) by a circos plot (Krzywinski et al. 2009).[8] Employment in the manufacturing, mining and agricultural sectors are mainly driven by Factors 2 and 3. On the other hand, Factor 1 loads mainly on employment in the services, finance, trade and transportation sectors.

Finally, Figure 5 plots the posterior median trend components along with the data series for employment in selected sectors. As example, Factor 1 (2, 3) is the strongest loading on employment in the banking sector (Printing and publishing, Metal mining), whereas Factor 1 (2) loads negatively (positively) on Stone, clay, and glass products. Note however, that loadings can be interpreted in absolute values, as a sign-switch of negative loadings can be adjusted by re-scaling the series-specific random error. This is a further advantage of a factor-driven drift with random error model.[9] Overall, the factor-driven trend component (red) follows quite closely the data series. When we add the idiosyncratic trend components (yellow), we obtain the overall series-specific trend component (purple). The purple line virtually matches observed data (blue), which confirms a very good fit of observed data.[10]

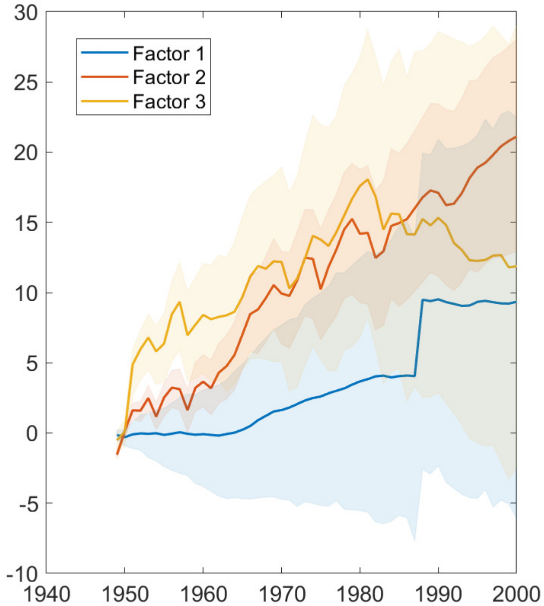
[8] The size of the factor blocks corresponds to the number of non-zero factor loadings, while the thickness of the chord end is proportional to the importance of the common component (the sum over the absolute posterior mean factor loadings).

[9] Negative loadings may be interpreted in absolute values, when we adjust the series-specific error term:

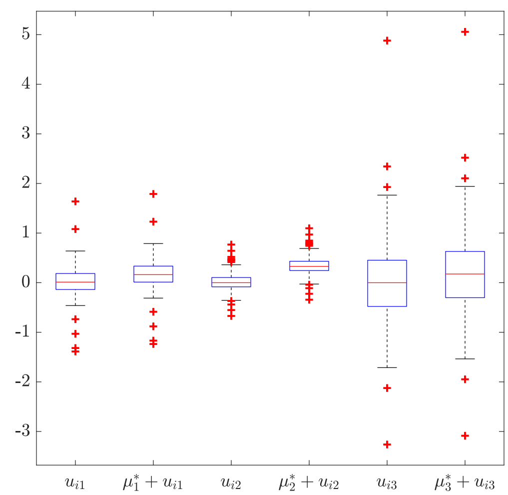
$$\begin{aligned}
&\tilde{\lambda}_{ik} = -\lambda_{ik} \\
&\lambda_{ik}(\mu_k^* + u_{ik}) = \tilde{\lambda}_{ik}(-\mu_k^* - u_{ik}) = \tilde{\lambda}_{ik}(\mu_k^* + \tilde{u}_{ik}) \\
&\text{where } \tilde{u}_{ik} \text{ is determined by} \\
&-\mu_k^* - u_{ik} = \mu_k^* + \tilde{u}_{ik} \Rightarrow \tilde{u}_{ik} = -2\mu_k^* - u_{ik}
\end{aligned}$$

[10] We plot the series-specific cyclical components in Figure B.12, Panel (a) of Appendix B. They display

Figure 3: (a) Posterior median factor trend components with 50% HPDI interval; (b) Boxplot of mean series-specific random effects $\boldsymbol{u}^*$ (u_{ik}) and series-specific factor-driven unconditional growth rates $\boldsymbol{\mu}^* + \boldsymbol{u}^*$ ($\mu_k^* + u_{ik}$)



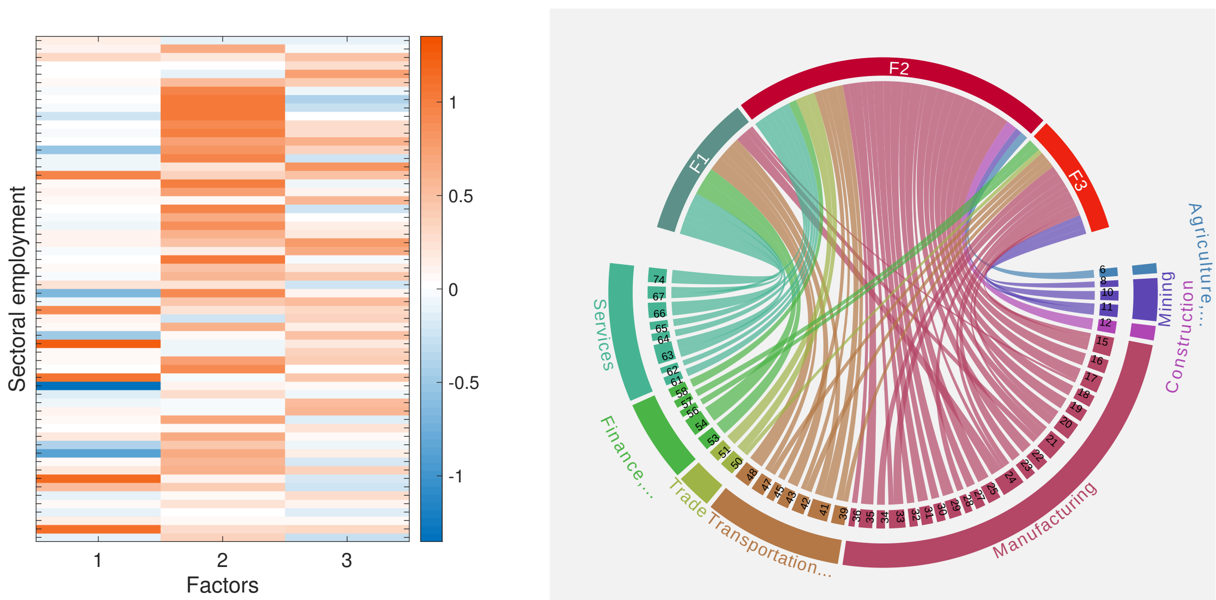
(a) Factor trend



(b) Random effects

no remaining trending behavior.

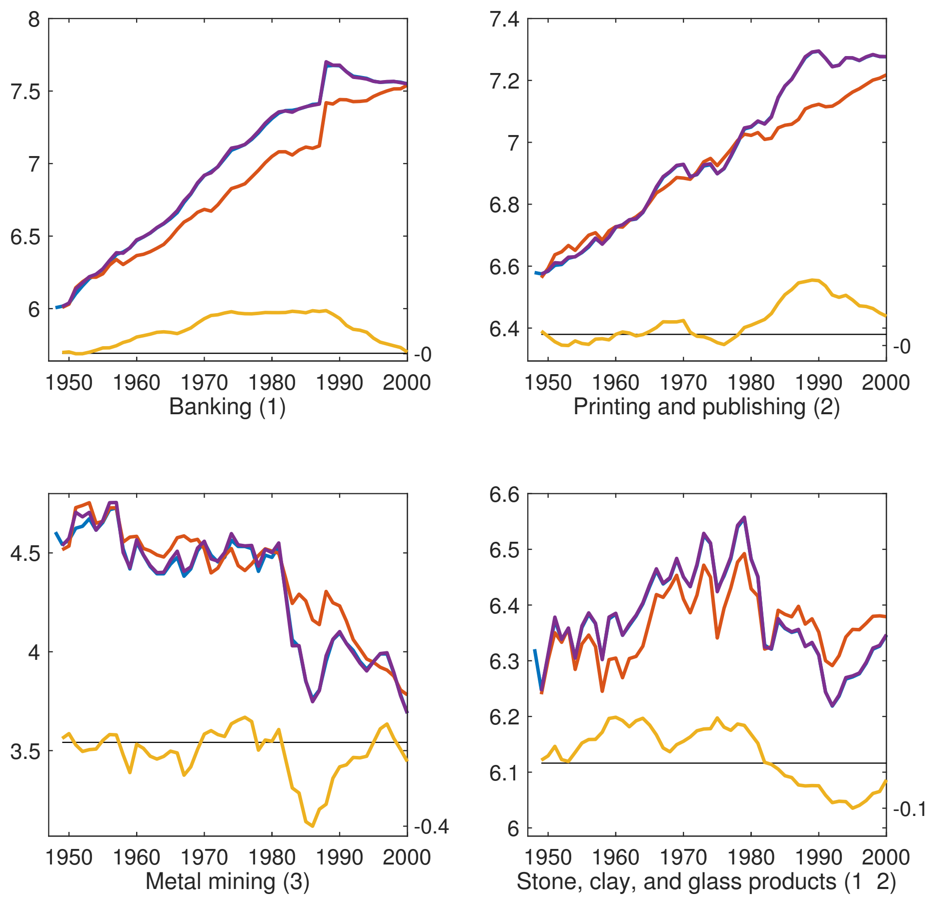
Figure 4: Sectoral employment: Panel (a): Heatmap of posterior mean factor loadings; Panel (b): Series-factor association based on non-zero loadings according to a 95% HPDI.



(a) Mean loadings

(b) Circos for non-zero loadings (95% HPDI)

Figure 5: Selected sectors: Data (blue) and posterior median stochastic trend component (Factor-driven in red, idiosyncratic in yellow, total in purple). In parentheses the factor(s) most strongly determining the series (According to loadings shifted away from zero on the basis of a 95% HPDI).



5.3 GDP per capita

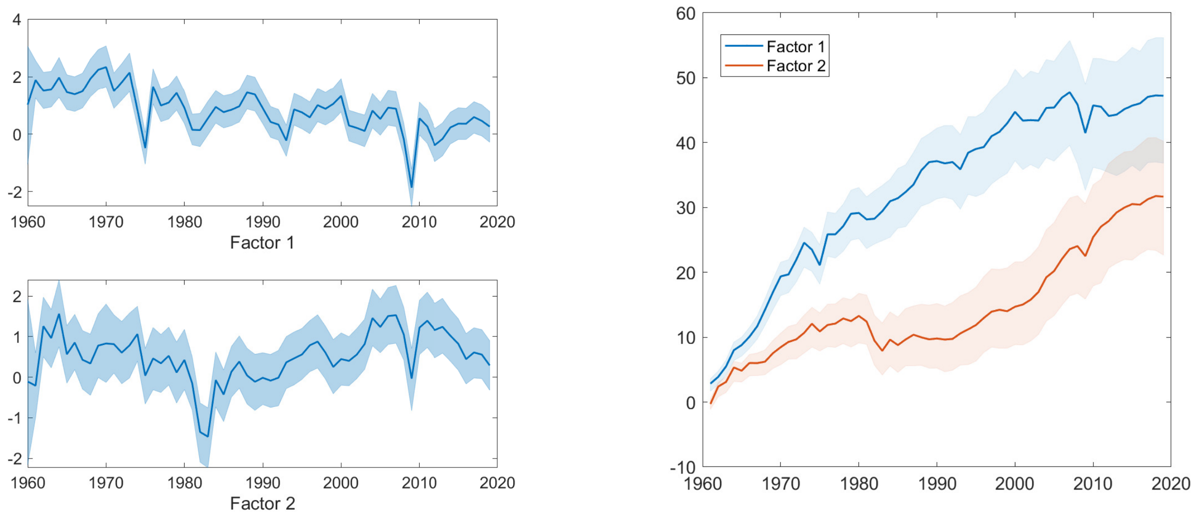
For real GDP per capita series, we extract two factors, see Figure 6. The left panel shows that the first factor reflects well the recession in the 1970s and the great financial crisis in 2008/2009, as well as the milder episodes at the beginning of the 1980s, 1990s and 2000s. Over the long run, this factor also reflects an overall slowdown in growth. The second factor displays a significant drop during the recession of the 1980s, after which growth recovered strongly. This contrasts developments after the financial crisis, as growth steadily slowed down until 2019. Panel (b) shows how this translates into factor trend components. The converging trend of Factor 2 came to a halt after 2010. Since then, both trend components evolve according to σ-convergence.[11]

The different factor growth paths suggest that the first factor should load mainly on developed countries while the second factor should affect mainly developing or emerging market countries. Figure 7 visualizes mean factor loadings (Left panels) and mean factor loadings shifted away from zero according to a 95% HPDI (Right panels). Indeed, "non-zero" loadings reveal that Factor 1 mainly drives developed countries, while Factor 2 mainly loads on developing countries. As examples, we plot the trend decomposition for various countries in Figure 8. As shown in Figure 7, the United States are mainly loaded by Factor 1. In Panel (a) of Figure 8 we see that the factor-driven median trend component follows closely the level series, whereas the idiosyncratic median trend component captures a minor share of the data trend. China is not strongly loaded by either factor. Nevertheless, the factor-driven component trends upwards (red). However, from 1980 onwards, the factor component is not strong enough to capture the accelerating growth in the level series, and an increasing share is taken up by the idiosyncratic trend component (yellow). Venezuela and the Gambia are two countries which end up with a lower real per capita GDP level than in 1960. In both countries, the idiosyncratic trend component captures the bulk of level dynamics, as their economic developments depart strongly from those captured by extracted factors. Overall, it is again the case that the sum of factor-driven and idiosyncratic trend components matches well observed data series.[12]

[11] See also Canova (2004).

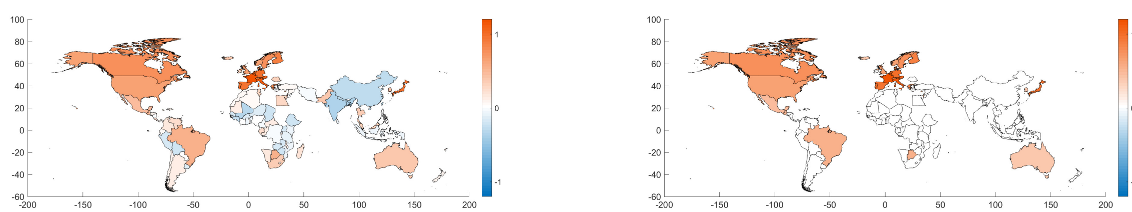
[12] The series-specific cyclical components are plotted in Figure B.12 of Appendix B.

Figure 6: Real per capita GDP. Extracted factors.

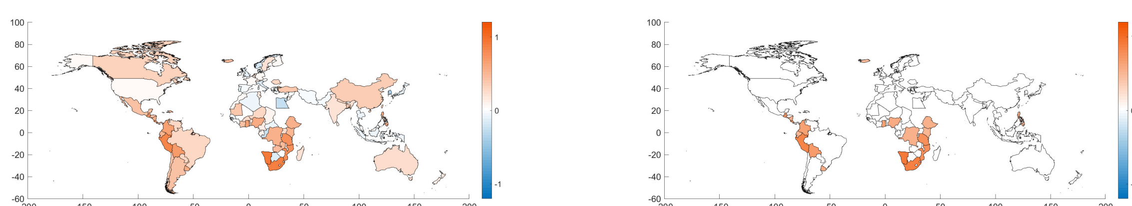


(a) Mean factors (95% HPDI) (b) Median factor trend components (50% HPDI)

Figure 7: Real per capita GDP: Posterior mean factor loadings.

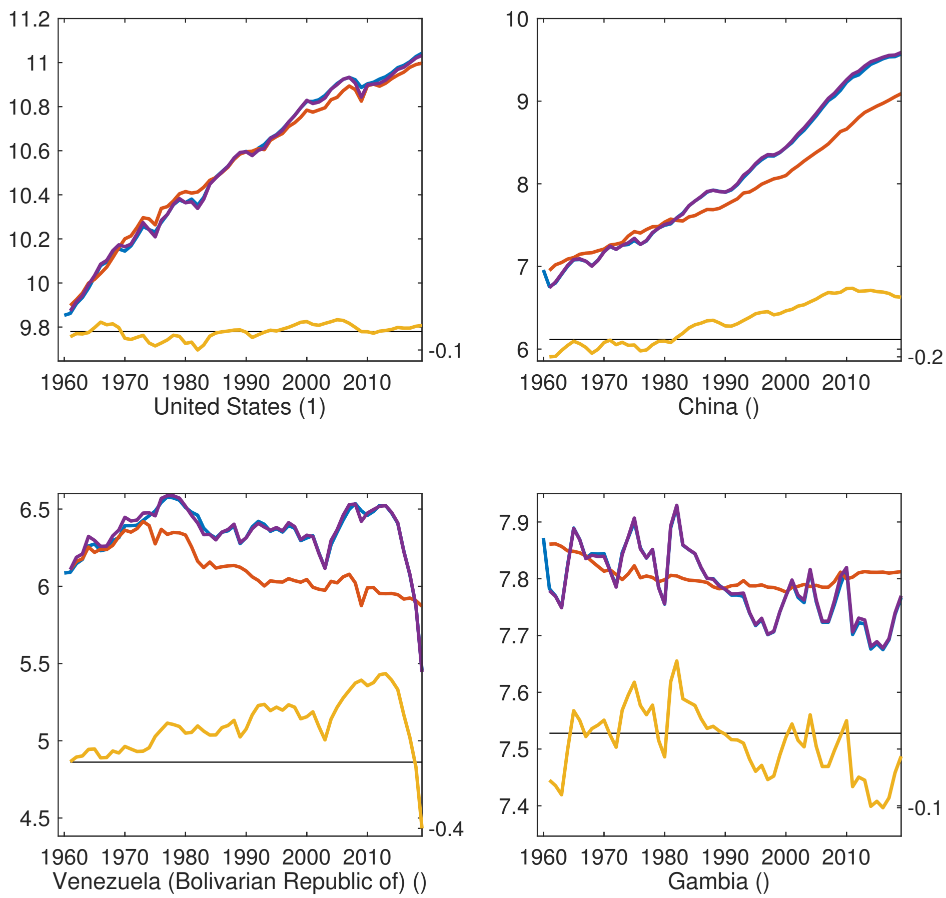


(a) Factor 1 (b) Factor 1, non-zero (95% HPDI)



(c) Factor 2 (d) Factor 2, non-zero (95% HPDI)

Figure 8: Selected countries: 100 times the log level of the original series (blue) and posterior median stochastic trend component (Factor-driven in red, idiosyncratic in yellow, total in purple). In parentheses the factor most strongly determining the series (According to loadings shifted away from zero on the basis of a 95% HPDI).

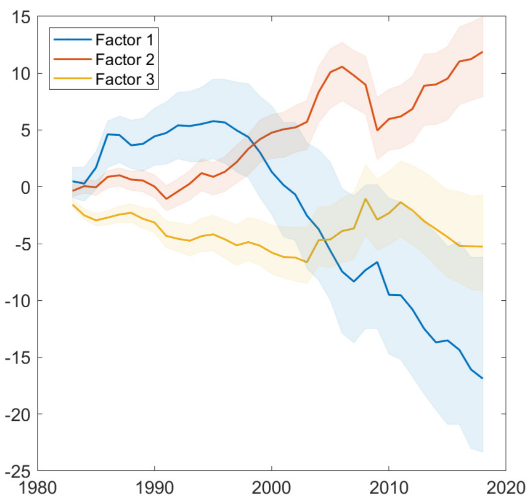


5.4 Relative capital goods prices

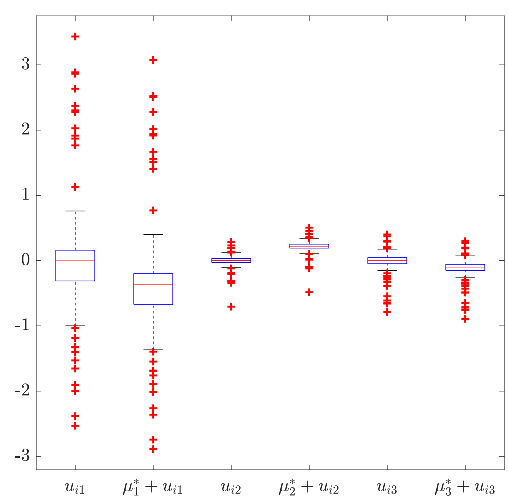
Despite the large heterogeneity across observed series trends, as few as $K = 3$ factors capture well data dynamics in levels. Figure 9, Panel (a) plots the posterior median along with a 50% HPDI. The factors capture well the different trending patterns of series. Factor 1 is the one that captures the strong and steady decline in relative prices observed for some series. Heterogeneity across series-specific factor-driven drifts is largest for this factor (Panel (b)). The circos graph in Figure 10, Panel (a), associates Factor 1 mainly with ICT goods and equipment for consumer durables. Factor 2 loads mainly on structures, non-residential and residential as well, and displays an overall increase in prices (Figure 9, Panel (a)). The factor captures a strong increase in prices during the build-up of the house price bubble from 2004 to 2008, and the subsequent collapse in 2008/2009 coincides with the great financial crisis. The third factor captures an overall mild decrease in prices, showing a transitory increase during the build-up of the house price bubble and three more years (2004 to 2009). The circos graph reveals that Factor 3 loads on non-residential equipment and some, mainly non-residential, structure series. The different blocks of series loaded by respective factors are also reflected in the heatmap of posterior mean factor loadings (Figure 10, Panel (b)).

Finally, Figure 11 plots selected series along with there median trend components. The factor-driven trend component (red) captures quite well the series-specific trending behavior. In some cases, the idiosyncratic trend component (yellow) takes up a persistent level-adjusting share, see the top panels in the figure. In other cases, depicted in the bottom panels, the idiosyncratic component adjusts for transitory deviations of the factor-driven trend component from the level of observed data. Overall, the total trend component (purple) again provides a good fit to the data.

Figure 9: Capital goods prices. Panel (a): Posterior median factor trend components with 50% HPDI interval; Panel (b): Boxplot of mean series-specific random effects $\boldsymbol{u}^*$ (u_{ik}) and series-specific factor-driven unconditional growth rates $\boldsymbol{\mu}^* + \boldsymbol{u}^*$ ($\mu_k^* + u_{ik}$)

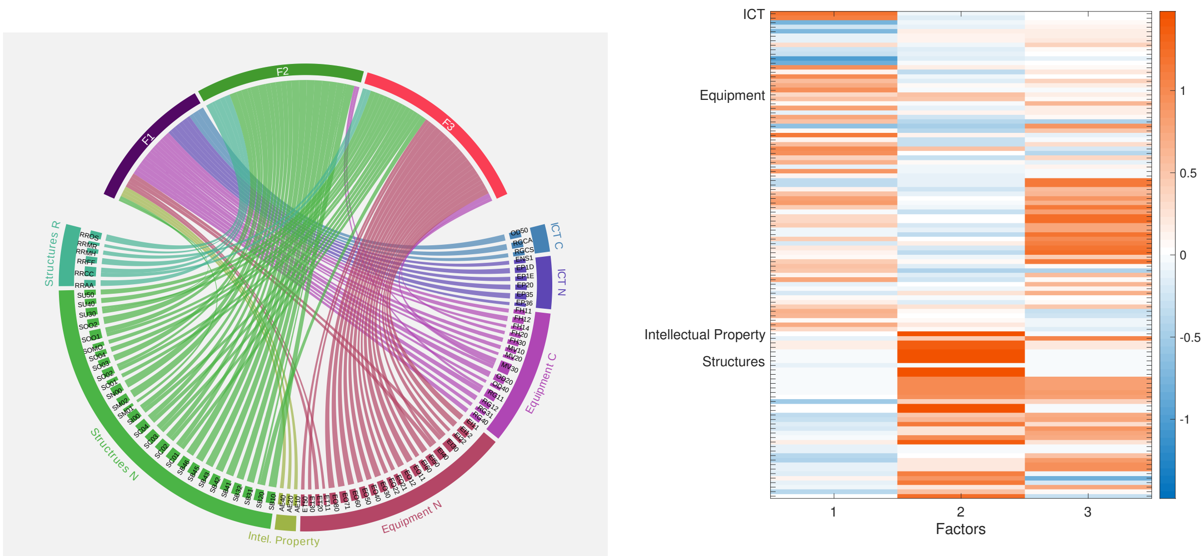


(a)



(b)

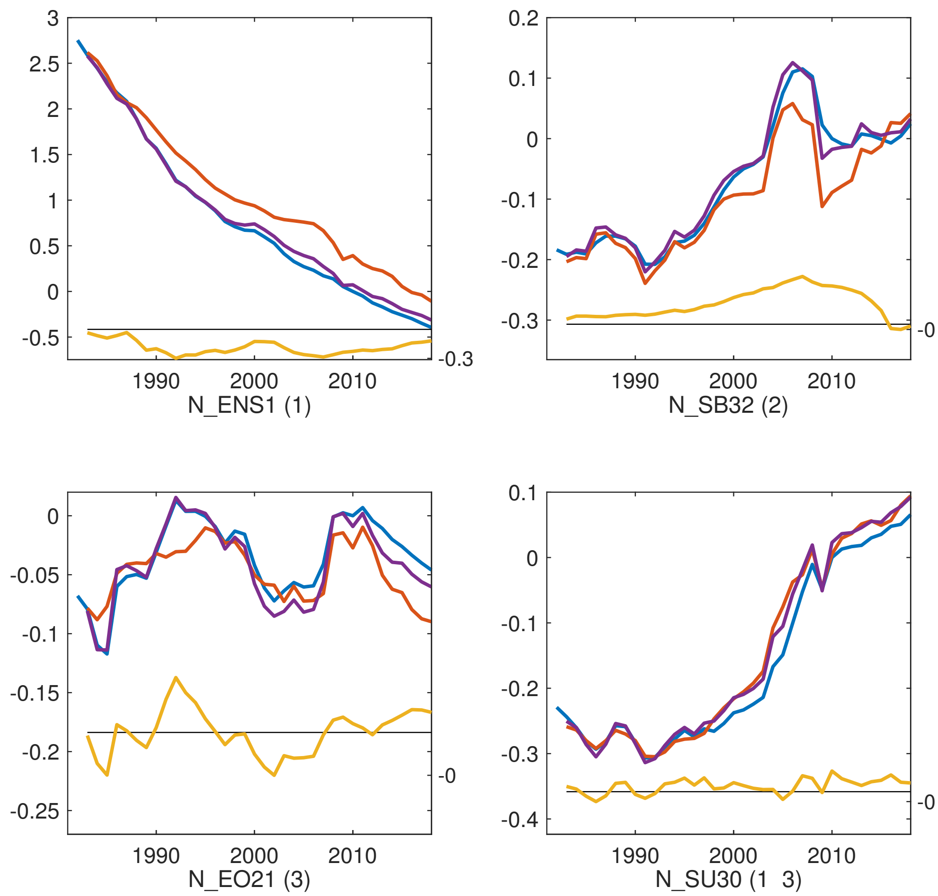
Figure 10: Capital goods prices. Panel (a): Series-factor association based on loadings shifted away from zero according to a 95% HPDI; Panel (b): Heatmap of posterior mean factor loadings.



(a) Circos for non-zero loadings (95% HPDI)

(b) Mean loadings

Figure 11: Capital goods prices. Selected series: 100 times the log level of the original series (blue) and posterior median stochastic trend component (Factor-driven in red, idiosyncratic in yellow, total in purple). In parentheses the factor(s) most strongly (significantly) determining the series.



6 Conclusion

Much work to date has considered the dynamics and common stochastic trends in multivariate systems. However, there are sound economic arguments for studying the deterministic trends that drive many series, an area of investigation that is precluded by approaches that difference and de-mean the data prior to analysis. This paper presents an approach to extract common and idiosynchratic deterministic components from many series using a dynamic factor model and so addresses a little studied area in the literature.

The specification of the factor model allows us to extract and identify multiple drifting factors from high-dimensional data series displaying a common long-run deterministic drift. A general specification captures data features in a flexible manner. We do not restrict the idiosyncratic component to be stationary, as series, though trending in common, may not be cointegrated. As factor-specific drifts may not account for all series-specific drift, we allow for a series-specific factor-driven drift, modelled as a factor drift with series-specific random effects. Remaining idiosyncratic drift is captured by a potential drift in the idiosyncratic component.

In applying this approach to sectoral employment data, we are able to obtain new results. We clearly visualize heterogeneity of factor drifts with considerable dispersion of the effect of the factor capturing declining employment since 1980. In a study of GDP growth rates of many countries, we are able to distinguish between trends in developed countries and those in developing countries. Finally, in application to relative capital goods prices, factors driving groupings of goods were identified along with the degree of heterogeneity in each group. In each application, we find evidence for a new characterisation for the factors and both the stochastic and deterministic processes driving the data. The precision of the results is achieved through a judicious approach to shrinkage and use of parameter expansions to avoid identifying restrictions that could distort inference.

Improved characterisation of factors driving economic processes leads to better understanding of a range of economic questions. Although we do not consider it in this paper, it is likely that the more precise estimation of factors within a flexible range of processes will improve forecasts and structural analysis.

References

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics 122*, 137–183.

Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica 72*, 1127–1177.

Banerjee, A., M. Marcellino, and I. Masten (2017). Structural FECM: Cointegration in large-scale structural FAVAR models. *Journal of Applied Econometrics 32*, 1069–1086.

Barigozzi, M., M. Lippi, and M. Luciani (2016). Non-stationary dynamic factor models for large datasets. Working Paper 2016-024, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington D.C.

Beveridge, S. and C. R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics 7*, 151–174.

Cadonna, A., S. Frühwirth-Schnatter, and P. Knaus (2020). Triple the Gamma — A unifying shrinkage prior for variance and variable selection in sparse state space and TVP models. *Econometrics 8*.

Canova, F. (2004). Testing for convergence clubs in income per-capita: A predictive density approach. *International Economic Review 45*, 49–77.

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.

Corona, F., P. Poncela, and E. Ruiz (2020). Estimating non-stationary common factors: Implications for risk sharing. *Computational Economics 55*, 37–60.

Eden, M. and P. Gaggl (2019). Capital composition and the declining labor share. Working Paper 7996, CESifo.

Eden, M., P. Gaggl, and S. Kaufmann (2021). Defining the right aggregate (working title). mimeo.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica 55*, 251–276.

Favero, C. A. and A. Melone (2020). Asset pricing vs asset expected returning in factor-portfolio models. Working Paper 14417, CEPR.

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.

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

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

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

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

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.

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.

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.

Liu, J. S. and Y. N. Wu (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association 448*, 1264–1274.

Müller, U. K., J. H. Stock, and M. W. Watson (2022). An econometric model of international growth dynamics for long-horizon forecasting. *The Review of Economics and Statistics 104*, 857–876.

Papaspiliopoulos, O., G. O. Roberts, and M. Sköld (2007). A general framework for the parametrization of hierarchical models. *Statistical Science 22*, 59–73.

Sims, C. A., J. H. Stock, and M. W. Watson (1990). Inference in linear time series models with some unit roots. *Econometrica 58*, 113–144.

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 (1988). Variable trends in economic time series. *Journal of Economic Perspectives 2*, 147–174.

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

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

A Posterior distributions

A.1 The factors $\boldsymbol{f}$

To derive the posterior, we condense the right-hand specification of (4) in Section 2 and cast it into the state space representation:

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

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

$$\tilde{\boldsymbol{y}} = \boldsymbol{\Lambda f} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, I_{T-q} \otimes \Sigma) \tag{17}$$

$$\boldsymbol{\Phi f} = \boldsymbol{\mu} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim N(0, \mathbf{S}) \tag{18}$$

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 $\boldsymbol{\Lambda}$ 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:

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

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{1}_{p\times 1} \otimes \Phi(1)^{-1}\mu \\ \mathbf{1}_{(T+d-p)\times 1} \otimes \mu \end{bmatrix}, \quad \boldsymbol{\Phi} = \left[\begin{array}{cccccccc}
 & I_p \otimes I_K & & & 0 & \ldots & & \\
\hline
-\Phi_p & \ldots & -\Phi_1 & I_K & 0 & \ldots & & \\
 & & & & \ddots & & & \\
 & \ldots & 0 & -\Phi_p & \ldots & -\Phi_1 & I_K &
\end{array}\right],$$

$$\mathbf{S} = \begin{bmatrix} I_p \otimes \Sigma_\nu^0 & 0 & \ldots \\ 0 & & \\ \vdots & I_{T+d-p} \otimes I_K & \end{bmatrix}$$

where Σ_ν^0 represents the variance of the initial states (see below).

We adapt the sampler proposed in Chan and Jeliazkov (2009) to sample $\boldsymbol{f}$ in one sweep. Given the representation in (17)-(18), the complete data likelihood has a normal density

$$f(\tilde{\boldsymbol{y}}|\boldsymbol{f}, \theta) \sim N(\boldsymbol{\Lambda f}, I_{T-q} \otimes \Sigma) \tag{19}$$

For the unobserved states, from (18) we obtain the following prior:

$$\begin{aligned} \boldsymbol{f}|\theta \quad &\sim \quad N\left(\mathrm{f}_0, \mathrm{F}_0\right) \\ & \quad\quad \mathrm{F}_0^{-1} = \boldsymbol{\Phi}'\mathbf{S}^{-1}\boldsymbol{\Phi} \\ & \quad\quad \mathrm{f}_0 = \boldsymbol{\Phi}^{-1}\boldsymbol{\mu} \end{aligned} \tag{20}$$

If in $\mathbf{S}$, the variance of the initial states, Σ_ν^0, 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, $\mathbf{F}_t = \tilde{\boldsymbol{\Phi}}\mathbf{F}_{t-1} + \boldsymbol{\nu}_t$, $\boldsymbol{\nu}_t \sim N\left(0, \begin{bmatrix} I_K & 0_{K\times(p-1)K} \\ \multicolumn{2}{c}{0_{(p-1)K\times pK}} \end{bmatrix}\right)$, with $\tilde{\boldsymbol{\Phi}} = \begin{bmatrix} \Phi_1, \ldots, \Phi_p \\ I_{K(p-1)}, 0_{K(p-1)\times K} \end{bmatrix}$, we obtain $E(\mathbf{F}_t\mathbf{F}_t') = \tilde{\boldsymbol{\Phi}}E(\mathbf{F}_{t-1}\mathbf{F}_{t-1}')\tilde{\boldsymbol{\Phi}}' + \Sigma_{\boldsymbol{\nu}}$ and $\Sigma_{\mathbf{F}} = \tilde{\boldsymbol{\Phi}}\Sigma_{\mathbf{F}}\tilde{\boldsymbol{\Phi}}' + \Sigma_{\boldsymbol{\nu}}$. The vec operator yields

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

from which we can retrieve the corresponding values for Σ_ν^0.

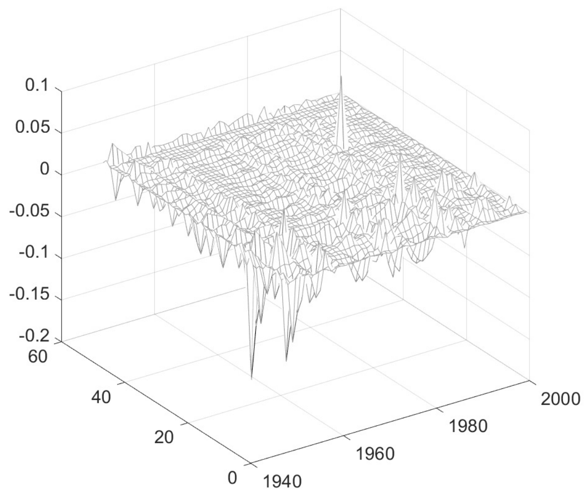
Combining the prior with the likelihood, the posterior of the factors is:

$$\begin{aligned} \boldsymbol{f}|\tilde{\boldsymbol{y}}, \theta \quad &\sim \quad N\left(\mathrm{f}, \mathrm{F}\right) \\ \mathrm{F}^{-1} \quad &= \quad \mathrm{F}_0^{-1} + \boldsymbol{\Lambda}'\left(I_{T-q} \otimes \Sigma^{-1}\right)\boldsymbol{\Lambda} \\ \mathrm{f} \quad &= \quad \mathrm{F}\left(\boldsymbol{\Lambda}'\left(I_{T-q} \otimes \Sigma^{-1}\right)\tilde{\boldsymbol{y}} + \mathrm{F}_0^{-1}\mathrm{f}_0\right) \end{aligned} \tag{21}$$

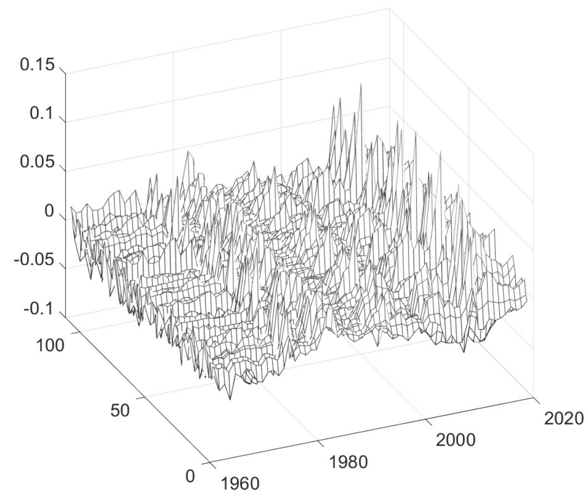
To avoid the full inversion of F^{-1} we take the Cholesky decomposition, $\mathrm{F}^{-1} = L'L$, then $\mathrm{F} = L^{-1}L^{-1\prime}$. We obtain a draw $\boldsymbol{f}$ by setting $\boldsymbol{f} = \mathrm{f} + L^{-1}\boldsymbol{\nu}$, where $\boldsymbol{\nu}$ is a $(T+d)k$ vector of independent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

B Additional figures

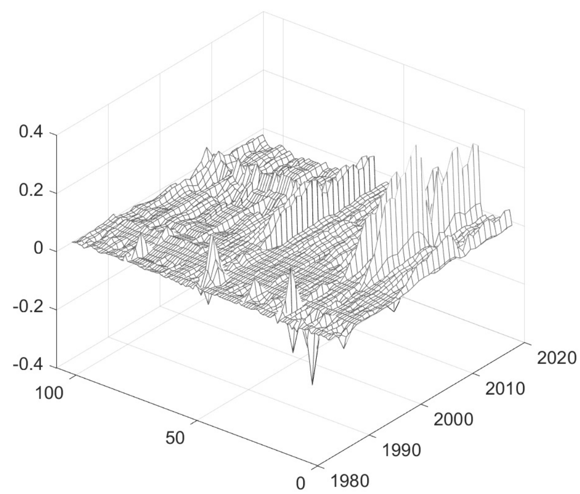
Figure B.12: Mean series-specific cyclical components (Difference between the series and the total trend component)



(a) Sectoral employment

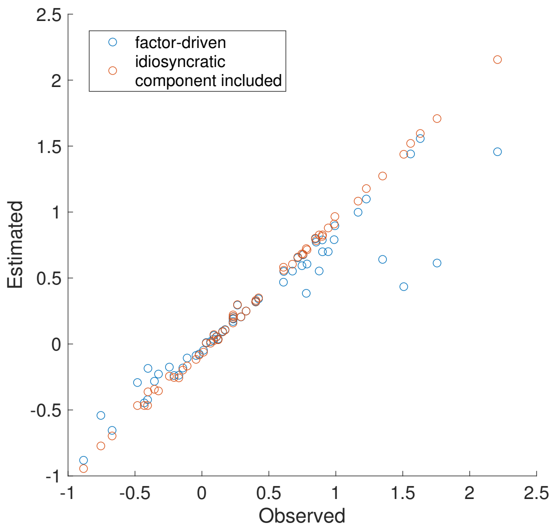


(b) Real per capita GDP

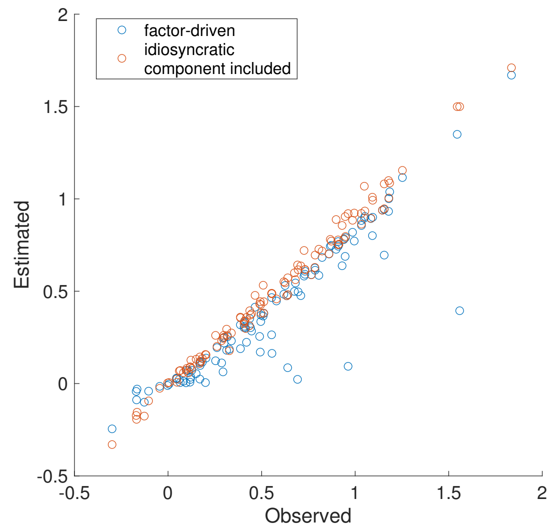


(c) Capital goods prices

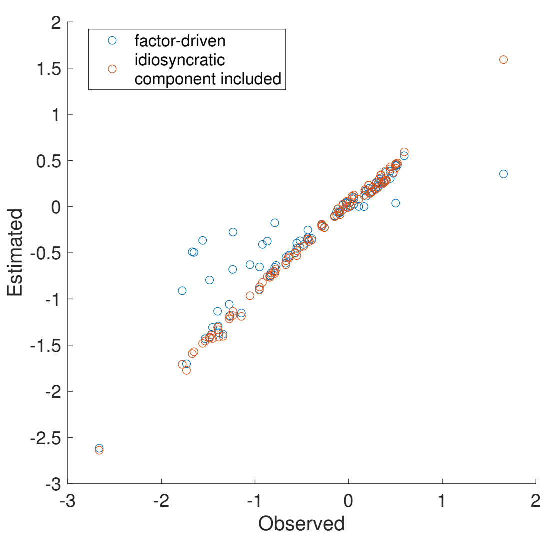
Figure B.13: Unconditional series-specific growth rate against unconditional, respectively, factor-driven and overall trend-driven mean growth rate.



(a) Sectoral employment



(b) GDP per capita



(c) Capital goods prices