# A Multilevel Factor Model for Economic Activity with Observation-Driven Dynamic Factors

Mariia Artemova, Francisco Blasques, Siem Jan Koopman

*Vrije Universiteit Amsterdam and Tinbergen Institute, the Netherlands*

April 6, 2023

**Abstract**

We analyze the role of industrial and non-industrial production sectors in the US economy by adopting a novel multilevel factor model. The proposed model is suitable for high-dimensional panels of economic time series and allows for interdependence structures across multiple sectors. The estimation procedure is based on a multistep least squares method which is simple and fast in its implementation. By analyzing the shock propagation process throughout the network of interconnections, we corroborate some of the key findings about the role of industrial production in the US economy, quantify the importance of propagation effects and shed new light on dynamic sectoral linkages.

*Keywords:* Dynamic factor model, Interconnectedness, Output growth.
*JEL classification codes:* C22, C32, C38, C51.

## 1 Introduction

In the last few decades, the world has witnessed a significant relocation of manufacturing jobs, with production migrating across countries and resulting in the emergence of a global supply chain structure of unprecedented size and complexity. As a result, many western countries, like the United States (US) have seen a decline in manufacturing jobs. In the US, the weight of industrial production (IP) sector has diminished from 25% to 18% of gross domestic product (GDP) in the period 1977-2011. However, recent evidence from Foerster et al. (2011) and Andreou et al. (2019) suggests that the IP sector still plays a key role in aggregate economic activity and GDP fluctuations in the United States. Particularly, Andreou et al. (2019) decompose sectoral dynamics across industrial production and non-industrial production (non-IP) sectors into three unobservable factors, (i) a common factor which explains variations in both IP and non-IP sectors; (ii) a group-specific factor which is exclusive to IP sectors, and (iii) a group-specific factor which is exclusive to non-IP sectors. The authors find that the common factor explains about 90% of the variability in the aggregate growth index of IP sectors and that the IP group-specific factor has very little additional explanatory power during the period 1977-2011. This means effectively that a single common factor can be interpreted as an IP factor. So while the role of manufacturing itself is declining, the IP sectors as a whole are still very much relevant.

Foerster et al. (2011) make further use of a structural factor analysis to conclude that within the IP sector, nearly all the variability of quarterly growth rates is associated with common factors. They note that month-to-month and quarter-to-quarter variations in the IP index are puzzlingly large as apparently the variability across IP sectors does not "average out". Foerster et al. (2011) point out that there are competing explanations for this puzzling observation, including (i) that IP fluctuations may be driven by common shocks affecting IP sectors as a whole, (ii) that sector-specific shocks affecting large IP sectors may be responsible for high aggregate IP variability, or (iii) that input-output linkages may help shocks to propagate across IP sectors, meaning that sector-specific shocks will not average out.

In this paper, we revisit the role of the IP and non-IP sectors in the US economy, as well as the linkages between the sectors. However, we approach the data using a novel dynamic multilevel factor model which captures temporal dynamics in the time-series data. In contrast to the previous studies, this allows us to (i) analyze sectoral shocks as uncorrelated temporal 'innovations', (ii) study the dynamic propagation of these shocks across sectors,

(iii) forecast common and group-specific factors at any horizon, and (iv) analyze sectoral linkages in a one-step-ahead predictability sense. In the empirical analysis, we have found that propagation effects are important sources of the variability of the aggregate indices. A key novel finding is that non-IP sectors are closely connected to both the IP and non-IP groups once the dynamic cumulative effects are taken into account.

Factor models are popular and important tools for dimensionality reduction in large data sets which, in economics and finance, often consist of panels of time series variables. Specifically, each individual time series variable in the panel is modelled as a linear combination of the factors, with the factor weights referred to as loadings. The interpretations of the factors are implicitly given by the loadings but are ambiguous for larger panels. To facilitate the interpretation of the factors and investigate group-specific dynamic co-movements, the factor model can be made subject to more structural dependencies. For example, when variables are related to geographical entities (regions, countries), it can be expected that neighboring regions may share common (dynamic) features. Similarly, when variables are related to different industries, the time series variables may be subject to shocks common to these industries. Such panel data structures arise naturally in many applications in economics and finance. We can relate our model to a large body of literature on hierarchical and/or multilevel factor models, which are also referred to as block-factor models. The block-factor models are considered by Kose et al. (2003), Crucini et al. (2011) and Choi et al. (2018) for extracting international business cycles, by Wang (2008) for analyzing co-movements between variables in the real and financial sectors, by Diebold et al. (2008) and Bai & Wang (2015) in their studies of international bond yields co-movements, among many others. A different class of hierarchical dynamic factor models is explored for macroeconomic forecasting by Moench et al. (2013).

Our modelling framework distinguishes itself from the existing models in two important directions. First, we model the factors as observation-driven time-varying parameters where the concept of observation-driven is discussed in general terms by Cox et al. (1981). This enables us to develop an estimation method that is straightforward in comparison to other approaches. For example, when the factors are parameter-driven and specified as dynamic stochastic (autoregressive) processes, more involved Bayesian estimation procedures may need to be adopted; see, e.g. Kose et al. (2003), Diebold et al. (2008), Moench et al. (2013) and Bai & Wang (2015). In contrast, our estimation procedure relies on basic least squares. The interest in fast and simple estimation procedure for the multilevel factor model has also been highlighted in Breitung & Eickmeier (2014).

Second, our dynamic model specification allows us to obtain forecast and impulse response functions which are formulated straightforwardly and require no additional derivations and computations. In the studies of Wang (2008), Breitung & Eickmeier (2014), Choi et al. (2018) and Andreou et al. (2019), a static version of the multilevel factor model is adopted with estimation relying on principal component analysis and canonical correlation analysis. While the estimation procedure for their static version of the model also relies on basic methods, it is more challenging to obtain forecasts and impulse response functions which can clearly be a shortcoming in empirical studies.

We propose a new model formulation where the dynamic factors are *driven* by the weighted linear combination of the cross-sectional data. In the spirit of Westerlund & Urbain (2015) and Karabiyik & Westerlund (2021), we allow the factors to be *driven* by the cross-sectional averages of the data which leads to a straightforward interpretation of the factors. For example, when considering a two-level model, the common factor represents the conditional (or predicted) mean of the group averages, while the group-specific factor is the conditional mean of the group in deviations from the common factor. We establish the stationarity and ergodicity of the limit dynamic factors and show that the filter of the true factors is invertible. We also give conditions for the consistency and asymptotic normality of the multistep least squares estimator. In a set of Monte Carlo experiments, we demonstrate that the estimators behave well in finite samples and that the model is approximating the dynamics of several data generating processes accurately.

Using the proposed model, we analyze the co-movements within and between different US economy sectors and their interconnections. We confirm the finding of Andreou et al. (2019) that industrial production is a dominant factor in the US economy, however, we approach this question from the quarter-ahead predictive sense. We also find that non-IP sectors are less related to the common factor than the IP sectors. However, there is also empirical evidence of sectoral interconnectedness, especially for the sectors with input-output relations (Long & Plosser, 1987). Our model gives us additional insight on how economic shocks propagate and dissipate through the network of connections across sectors and across time. Particularly, based on the dynamic impulse response analysis, we find that there are immediate effects of shocks between IP sectors. However, cumulative effects are more pronounced when non-IP sectors are involved. Therefore, even though the non-IP sectors are less related to the common factor, non-IP sectors are tightly linked both to IP and non-IP sectors. Overall, we find that shock propagation plays an important role in explaining the variation in the aggregate indices and that the role of the propagation effects is larger for the non-IP sectors than for the IP ones.

The outline of the paper is as follows. In Section 2, we introduce our multilevel factor model framework and describe a multistep estimation procedure. We establish stochastic properties of the model as well as consistency and asymptotic normality of the estimators. Section 3 summarizes the results of the Monte Carlo experiments. In Section 4, we present and explore the results of our empirical study for US economic activity. Section 5 concludes. The Appendix contains the proofs of the main theoretical results. The Supplementary Appendix (SA) contains information about the dataset, additional empirical and Monte Carlo results, details on forecasts, impulse response functions and group connectedness measures as well as further technical lemmas and proofs.

# 2 The Model

## 2.1 Dynamic observation-driven multilevel factors

Let $\mathbf{y}_t$ be an $N$-variate random variable which is observed for time periods $t = 1, \ldots, T$. Suppose the elements of the vector $\mathbf{y}_t$ can be categorized into $S$ groups with $N_s$ variables in each group, such that $N = \sum_{s=1}^{S} N_s$. We let each variable be related to a common factor and a group-specific factor. The *observation-equation* of our dynamic factor model is then given by

$$\mathbf{y}_t^s = \mathbf{\Lambda}_s^c f_t + \mathbf{\Lambda}_s^g g_t^s + \boldsymbol{\varepsilon}_t^s, \quad \text{for } t = 1, \ldots, T \text{ and } s = 1, \ldots, S,$$

where $\mathbf{y}_t^s = (y_{1,t}^s, \ldots, y_{N_s,t}^s)^\top$ is the $N_s \times 1$ dimensional vector corresponding to variables in group $s$, $f_t$ and $g_t^s$ are the unobserved common and group-specific factors, respectively, with $\mathbf{\Lambda}_s^c$ and $\mathbf{\Lambda}_s^g$ being the corresponding $N_s \times 1$ vectors of loadings, respectively, and $\boldsymbol{\varepsilon}_t^s = (\boldsymbol{\varepsilon}_{1,t}^s, \ldots, \boldsymbol{\varepsilon}_{N_s,t}^s)^\top$ is the mean-zero vector of identically distributed shocks, possibly subject to idiosyncratic serial dependence. Depending on the application at hand, the factors $f_t$ and $g_t^s$ can also be interpreted as global and region-specific factors, respectively.

In matrix notation, the model can be written as

$$\mathbf{y}_t = \mathbf{\Lambda}^c f_t + \mathbf{\Lambda}^g \mathbf{g_t} + \boldsymbol{\varepsilon}_t, \quad \text{for} \quad t = 1, \ldots, T, \tag{1}$$

where $\mathbf{y}_t = \left(\mathbf{y}_t^{1\top}, \ldots, \mathbf{y}_t^{S\top}\right)^\top$ is an $N \times 1$ dimensional vector of cross-sectional variables which collects all variables in all groups, $\mathbf{g}_t$ is an $S \times 1$ vector containing all group-specific factors, that is $\mathbf{g}_t = \left(g_t^1, \ldots, g_t^S\right)^\top$, $\mathbf{\Lambda}^c = \left(\mathbf{\Lambda}_1^{c\top}, \ldots, \mathbf{\Lambda}_S^{c\top}\right)^\top$ is an $N \times 1$ vector of unknown loadings associated with the common factor $f_t$, $\mathbf{\Lambda}^g = \text{block-diag}(\mathbf{\Lambda}_1^g, \ldots, \mathbf{\Lambda}_S^g)$ is an $N \times S$ block-diagonal matrix of unknown coefficients for the group-specific factors $\boldsymbol{g}_t$, and $\boldsymbol{\varepsilon}_t =$

$\left(\boldsymbol{\varepsilon}_t^{1\top}, \ldots, \boldsymbol{\varepsilon}_t^{S\top}\right)^\top$ is an $N$-variate stationary mean-zero sequence with covariance matrix $\mathbb{C}\text{ov}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}$.

The dynamic factors are modeled as observation-driven processes. The common factor $f_t$ takes into account between-group co-movements and is driven by all cross-sectional variables,

$$f_{t+1} = \alpha + \beta \left( \frac{1}{S} \sum_{s=1}^{S} \left( \frac{1}{N_s} \sum_{i=1}^{N_s} y_{i,t}^s \right) - f_t \right) + \gamma f_t. \tag{2}$$

Intuitively, when $\beta \geq 0$ and $|\gamma| < 1$, $f_t$ tracks the time-varying "group conditional expectation". We filter the "group conditional expectation", rather than the conditional expectation of $\mathbf{y}_t$, to account for possibly different number of variables in each group. In contrast, the conditional expectation of $\mathbf{y}_t$ would easily be dominated by the groups with large number of elements.

The group-specific factors take into account the information on the co-movements only from the variables of the corresponding group. Naturally, these co-movements can be common as well as group-specific. To exclude the impact of the common factor, we model the group-specific factors as the conditional mean of $\mathbf{y}_t^s$ in deviations. Thus, they are driven by the within-group co-movements in deviations from the common factor. Therefore, we have the following updating equation for the group-specific factors,

$$g_{t+1}^s = \alpha_s + \beta_s \left( \frac{1}{N_s} \sum_{i=1}^{N_s} y_{i,t}^s - f_t \frac{1}{N_s} \sum_{i=1}^{N_s} \lambda_{i,s}^c - g_t^s \right) + \gamma_s g_t^s, \tag{3}$$

for $s = 1, \ldots, S$, with $\boldsymbol{\Lambda}_s^c = (\lambda_{1,s}^c, \ldots, \lambda_{N_s,s}^c)^\top$ and $\lambda_{i,s}^c$ being the exposure of variable $i$ in group $s$ to the common factor $f_t$. Hence, if $\beta_s \geq 0$ and $|\gamma_s| < 1$, $g_t^s$ captures a time-varying location of the group $s$ in deviations from the common factor. Since the factors are driven by the averages, we standardize the series beforehand. Hence, we can set $\alpha = \alpha_s = 0$ for

$s = 1, \ldots, S$.

We focus on this formulation with one common and one group-specific factor per each group as it is in line with our application. Besides, in the majority of applications for multilevel factor models there is evidence of the presence of only one common factor affecting all groups; see, for example, Moench et al. (2013), Bai & Wang (2015) and Andreou et al. (2019). On the other hand, further factor levels can be explicitly incorporated into the model by introducing similarly the next level updating equations for them. For instance, variable-specific factors can be added. Furthermore, to better capture the factors' dynamics, more autoregressive terms and more lags of the mean reversion terms could be included into the updating equations (2)–(3).

The proposed model is closely related to the parameter-driven local level model as in Durbin & Koopman (2012) and to the score-driven models of Creal et al. (2013), Creal et al. (2014) when the innovations are Gaussian. We opt for the observation-driven approach since, in contrast to the parameter-driven models, nonlinearities in the updating equations can be easily incorporated. For example, the filter can be made robust to the extreme values by bounding their influence on the update. Moreover, even in the case of standard factor model, full maximum likelihood estimation methods of large-dimensional parameter-driven models using Kalman filter are computationally demanding as discussed, for example, in Engle & Watson (1981) and Jungbacker & Koopman (2015). Our model formulation, as discussed in the next section, enables us to develop fast and simple estimation procedure based on the least squares criterion functions.

## 2.2 Estimation procedure

In this section, we discuss the estimation of the parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}^c, \boldsymbol{\Lambda}^c, \boldsymbol{\theta}^g, (\boldsymbol{\Lambda}^g_1, \ldots, \boldsymbol{\Lambda}^g_S))^\top$, where the vector $\boldsymbol{\theta}^c = (\beta, \gamma)^\top$ contains the parameters of the common factor and $\boldsymbol{\theta}^g = (\boldsymbol{\theta}^g_1, \ldots, \boldsymbol{\theta}^g_S)^\top$ of the group-specific factors with $\boldsymbol{\theta}^g_s = (\beta_s, \gamma_s)^\top$ for $s = 1, \ldots, S$.

In factor models, the loadings and factors are not separately identifiable. However, once the factors are known, their loadings can be estimated. Moreover, in the multilevel factor model, both the common and group-specific factors are unobserved and need to be estimated. We propose a sequential estimation procedure. First, we estimate the parameters $\boldsymbol{\theta}^c$ of the updating equation for the common factor $f_t$, and filter the factor $f_t$ itself. Next, given the common factor estimates, we obtain the loadings $\boldsymbol{\Lambda}^c$, and estimate the residuals of this regression. Finally, using the residuals, we estimate the parameters $\boldsymbol{\theta}^g$ featured in the updating equation for the group-specific factors, filter the group-specific factors $\mathbf{g_t}$, and estimate their respective loadings $\boldsymbol{\Lambda}^g$. The sequential estimation of the factors is in some sense similar to the two-step principal component estimator approach for the multilevel factor models; see, for example, Beck et al. (2009). Below we state in details the steps of the estimation procedure.

**Step I.** Estimate the static parameters $\boldsymbol{\theta}^c$ of the common factor by minimizing the following criterion function,

$$Q_T^{(1)}(\boldsymbol{\theta}^c) = \frac{1}{T}\sum_{t=2}^{T}\left(\frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{N_s}\sum_{i=1}^{N_s} y_{i,t}^s\right) - f_t(\boldsymbol{\theta}^c, \hat{f}_1)\right)^2,$$

where the criterion function depends on the common factor $f_t(\boldsymbol{\theta}^c, \hat{f}_1)$ initialized at $\hat{f}_1$. Given the estimate of $\boldsymbol{\theta}^c$, the common factor itself can be filtered recursively using equation (2).

**Step II.** Given the filtered common factor, estimate the vector of loadings $\boldsymbol{\Lambda}^c$ by minimizing the ordinary least squares loss function $Q_T^{(2)}(\boldsymbol{\Lambda}^c)$. Hence,

$$\hat{\boldsymbol{\Lambda}}_T^c = \left(\sum_{t=2}^{T}\left(\hat{f}_t(\hat{\boldsymbol{\theta}}_T^c, \hat{f}_1)\right)^2\right)^{-1}\sum_{t=2}^{T}\hat{f}_t(\hat{\boldsymbol{\theta}}_T^c, \hat{f}_1)\mathbf{y}_t,$$

which is a standard linear regression where the loadings are coefficients and the estimated common factor is a regressor.

The residuals from the last regression, $\hat{\boldsymbol{\xi}}_t = \mathbf{y}_t - \hat{\boldsymbol{\Lambda}}_T^c \hat{f}_t$, capture the effects that influence individual series but are not common to all of them. This includes not only the individual effects but also the group-specific ones. Hence, the residuals contain information about the group-specific co-movements as well as about the variable-specific ones. We use the residuals from Step II to estimate the group-specific factors and loadings on them. We again do it in steps as factors and loadings are not separately identifiable without restrictions. Steps III and IV are similar to Steps I and II in the sense that we first evaluate the group-specific factors and then turn to the estimation of the loadings on them.

**Step III.** Estimate the parameters of the group-specific factors $\boldsymbol{\theta}^g$ using the following least squares criterion function:

$$Q_T^{(3)}(\boldsymbol{\theta}^g) = \frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{S}\left(\frac{1}{N_s}\left(\sum_{i=1}^{N_s}\hat{\xi}_{i,t}^s\right) - g_t^s(\boldsymbol{\theta}^g, \hat{g}_1^s)\right)^2,$$

where $\hat{\boldsymbol{\xi}}_t = ((\hat{\boldsymbol{\xi}}_t^1)^\top, \ldots, (\hat{\boldsymbol{\xi}}_t^S)^\top)^\top$ and $\hat{g}_1^s$ is a filter initialization of the group-specific factors. Given the estimated factor parameters $\boldsymbol{\theta}^g$, the group-specific factors can be filtered out using updating equation (3).

**Step IV.** The loadings on the group-specific factors are estimated using again OLS by minimizing the sum of the squared residuals $Q_T^{(4)}(\boldsymbol{\Lambda}^g)$ given the estimated factors.

The sequential procedure allows us not to impose restrictions on the factors or loadings explicitly. Alternatively, the first two steps could be combined in one by imposing constraint $\frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s}\lambda_{i,s}^c = 1$, and steps III and IV could be combined by imposing $\frac{1}{N_s}\sum_{i=1}^{N_s}\lambda_{i,s}^g = 1$, $s = 1, \ldots, S$, where $\lambda_{i,s}^g$ is the loading of variable $i$ in sector $s$ on its group-specific factor $g_t^s$. However, when the cross-sectional dimension is large the joint estimation can be computationally demanding. Therefore, the sequential procedure is still preferable.

## 2.3 Forecasts, impulse response functions and network analysis

Besides the simple and convenient step-by-step estimation of the parameters, the model in Section 2.1 also offers immediate access to forecasts and impulse response functions (IRFs). In particular, the simple forward-iteration of the filtering equations allows us to produce $h$-step-ahead forecasts of the unobserved dynamic factors $f_{T+h}$ and $\mathbf{g}_{T+h}$, as well as of the data $\mathbf{y}_{T+h}$, for any $h = 1, 2, \ldots$. The IRFs can further be used to analyze the impact of unit-specific shocks over the cross-section and over time, giving us an overview of group interconnectedness and network structures (Diebold et al., 2008). Section C in the SA provides the details on how forecasts and impulse response functions are defined, and how they can be obtained in practice. Additionally, it discusses different measures of interconnectedness which can be calculated to obtain insights into the network structure of the data.

## 2.4 Stochastic properties of the observation-driven filters

In this section, we analyze the properties of our model as a filter for unobserved factors. Naturally, the observation-driven filters are initialized at some values $\hat{f}_1 \in \mathbb{R}$ and $\hat{\boldsymbol{g}}_1 \in \mathbb{R}^S$ at time $t = 1$, take observations $\{\mathbf{y}_t\}_{t\in\mathbb{N}}$ as given and update the filtered dynamic factors according to equations (2) and (3), providing us with filtered sequences $\{\hat{f}_t(\boldsymbol{\theta}, \hat{f}_1)\}_{t\in\mathbb{N}}$ and $\{\hat{\mathbf{g}}_t(\boldsymbol{\theta}, \hat{\mathbf{g}}_1)\}_{t\in\mathbb{N}}$. To shorten the notation, we sometimes suppress the dependence of the filtered sequences on the initializations $\hat{f}_1$ and $\hat{\boldsymbol{g}}_1$ as well as on $\boldsymbol{\theta}$, and denote the filtered sequences as $\{\hat{f}_t\}_{t\in\mathbb{N}}$ and $\{\hat{\boldsymbol{g}}_t\}_{t\in\mathbb{N}}$.

Propositions 1 and 2 state conditions for the invertibility of the sequences $\{\hat{f}_t(\boldsymbol{\theta}, \hat{f}_1)\}_{t\in\mathbb{N}}$ and $\{\hat{\mathbf{g}}_t(\boldsymbol{\theta}, \hat{\mathbf{g}}_1)\}_{t\in\mathbb{N}}$ as well as the stationarity, ergodicity and bounded moments of the limit sequences $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and $\{\mathbf{g}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$. Since the initial values $\hat{f}_1$ and $\hat{\mathbf{g}}_1$ are almost surely incorrect, filter invertibility plays a crucial role in ensuring that the influence of the initialization vanishes suitably fast as $t \to \infty$. Additionally, establishing suitable stochastic properties for the limit sequences $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and $\{\mathbf{g}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is important for the proof of consistency and asymptotic normality of the least squares estimators, since the stochastic properties of the filtered sequences are directly related to the stochastic properties of the least squares objective functions. For further discussion of the importance of filter invertibility we refer to Straumann & Mikosch (2006) and Wintenberger (2013).

The following assumptions state conditions on the parameter space and the properties of the data for the filters to be invertible.

**Assumption 1.** *$\Theta^c$, $\Theta^g$, $\Theta^{\lambda_c}$, and $\Theta^{\lambda_g}$ are compact parameter spaces. $\Theta := \Theta^c \times \Theta^g \times \Theta^{\lambda_c} \times \Theta^{\lambda_g} \subseteq \mathbb{R}^{2(S+1+N)}$.*

**Assumption 2.** *$\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic sequence.*

**Assumption 3.** *$\mathbb{E}|y^s_{i,t}|^2 < \infty, i = 1, \ldots, N_s$ and $s = 1, \ldots, S$.*

**Proposition 1.** *Let assumptions 1–3 hold. The sequence $\{\hat{f}_t(\boldsymbol{\theta}, \hat{f}_1)\}_{t\in\mathbb{N}}$ initialized at $\hat{f}_1 \in \mathbb{R}$ converges exponentially almost surely (e.a.s.) to a unique strictly stationary and ergodic (SE) sequence $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ uniformly over the parameter space,*

$$\sup_{\boldsymbol{\theta}\in\Theta} |\hat{f}_t(\boldsymbol{\theta}, \hat{f}_1) - f_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0 \quad \text{as } t \to \infty,$$

*if and only if $\Theta$ is such that $|\gamma - \beta| < 1$ $\forall \boldsymbol{\theta} \in \Theta$. Moreover, the filter limit sequence satisfies $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |f_t(\boldsymbol{\theta})|^2 < \infty$.*

**Proposition 2.** *Let the conditions of Proposition 1 hold. The sequence $\{\hat{g}^s_t(\boldsymbol{\theta}, \hat{g}^s_1)\}_{t\in\mathbb{N}}$ initialized at $\hat{g}^s_1 \in \mathbb{R}$ converges e.a.s. to a unique SE sequence $\{g^s_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ uniformly on $\Theta$, $\sup_{\boldsymbol{\theta}\in\Theta} |\hat{g}^s_t(\boldsymbol{\theta}, \hat{g}^s_1) - g^s_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$ as $t \to \infty$, if and only if $\Theta$ is such that $|\gamma_s - \beta_s| < 1$ $\forall \boldsymbol{\theta} \in \Theta$ and $s = 1, \ldots, S$. Additionally, the limit sequences satisfy $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |g^s_t(\boldsymbol{\theta})|^2 < \infty$.*

## 2.5 Asymptotic properties of the estimators

In this section, we formulate conditions for the consistency and asymptotic normality of the least squares estimators detailed in Steps I – IV. For ease of exposition, we use the following notation for the Step I criterion function, $\hat{Q}_T^{(1)}(\boldsymbol{\theta}^c) = \frac{1}{T}\sum_{t=2}^T q^{(1)}(\mathbf{y}_t, \hat{f}_t(\boldsymbol{\theta}^c), \boldsymbol{\theta}^c)$, where $q^{(1)}(\mathbf{y}_t, \hat{f}_t(\boldsymbol{\theta}^c), \boldsymbol{\theta}^c) := \left(\frac{1}{S}\sum_{s=1}^S \left(\frac{1}{N_s}\sum_{i=1}^{N_s} y_{i,t}^s\right) - \hat{f}_t(\boldsymbol{\theta}^c, \hat{f}_1)\right)^2$ and superscript (1) denotes the step of the estimation procedure. We use a similar notation for the criterion functions $\hat{Q}_T^{(2)}(\boldsymbol{\Lambda}^c)$, $\hat{Q}_T^{(3)}(\boldsymbol{\theta}^g)$, and $\hat{Q}_T^{(4)}(\boldsymbol{\Lambda}^g)$.

Theorems 1–3 establish the consistency of the estimators for each step of the estimation procedure described in Section 2.2 as well as the consistency of the plug-in filters $\hat{f}_t(\hat{\boldsymbol{\theta}}_T^c, \hat{f}_1)$ and $\hat{\mathbf{g}}_t(\hat{\boldsymbol{\theta}}_T^g, \hat{\mathbf{g}}_1)$. To proceed with the proofs of consistency of the loadings and group-specific factors we first establish that the common factor itself converges. The next theorem reveals that the filter invertibility ensures both strong consistency of the estimator of the common factor static parameters as well as the convergence of the plug-in estimator.

**Theorem 1** (Consistency: Step I)**.** *Let assumptions 1–3 hold. Then the least squares estimator $\hat{\boldsymbol{\theta}}_T^c(\hat{f}_1)$ is strongly consistent for $\boldsymbol{\theta}_0^c$ for any initialization $\hat{f}_1 \in \mathbb{R}$,*

$$\hat{\boldsymbol{\theta}}_T^c(\hat{f}_1) \xrightarrow{a.s.} \boldsymbol{\theta}_0^c \quad \text{as } T \to \infty.$$

*Furthermore, a plug-in filter $\hat{f}_t(\hat{\boldsymbol{\theta}}_T^c, \hat{f}_1)$ converges almost surely (a.s.),*

$$|\hat{f}_t(\hat{\boldsymbol{\theta}}_T^c, \hat{f}_1) - f_t(\boldsymbol{\theta}_0^c)| \xrightarrow{a.s.} 0 \ \text{ as } \ t, T \to \infty,$$

*where $\boldsymbol{\theta}_0^c \in \Theta^c$ is a minimizer of a limit criterion function $Q_\infty^{(1)}(\boldsymbol{\theta}^c)$.*

The next theorem establishes consistency of the estimator of the loadings. The estimates are obtained using OLS, hence the conditions for the consistency are overall the same as for the regular least squares estimator. For the consistency of the least squares estimator we usually need to impose some assumptions on the regressors such as stationarity and ergodicity. However, in our case, they can only hold for the limit sequence and not for the filter itself. Moreover, the common factor is evaluated at $\hat{\boldsymbol{\theta}}_T^c$ and not at $\boldsymbol{\theta}_0^c$.

**Theorem 2** (Consistency: Step II)**.** *Let assumptions 1–3 hold. Then the least squares estimator* $\hat{\boldsymbol{\Lambda}}_T^c$ *is strongly consistent for* $\boldsymbol{\Lambda}_0^c$*,* $\hat{\boldsymbol{\Lambda}}_T^c \xrightarrow{a.s.} \boldsymbol{\Lambda}_0^c \quad \text{as } T \to \infty$*, where* $\boldsymbol{\Lambda}_0^c \in \Theta^{\lambda_c}$ *is a minimizer of a limit criterion function* $Q_\infty^{(2)}(\boldsymbol{\Lambda}^c)$*.*

Theorem 3 covers the consistency of the Steps III and IV estimators. Finally, Theorem 4 establishes the asymptotic normality of the stepwise least squares estimator.

**Theorem 3** (Consistency: Steps III and IV)**.** *Let assumptions 1-3 hold. Then the estimator* $\hat{\boldsymbol{\theta}}_T^g(\hat{g}_1)$ *is strongly consistent for* $\boldsymbol{\theta}_0^g$ *for any initialization* $\hat{\mathbf{g}}_1 \in \mathbb{R}^S$*,*

$$\hat{\boldsymbol{\theta}}_T^g(\hat{\mathbf{g}}_1) \xrightarrow{a.s.} \boldsymbol{\theta}_0^g \quad as \quad T \to \infty,$$

*where* $\boldsymbol{\theta}_0^g \in \Theta^g$ *is a minimizer of a limit criterion function* $Q_\infty^{(3)}(\boldsymbol{\theta}^g)$*.*

*Furthermore, the plug-in filter* $\hat{\mathbf{g}}_t(\hat{\boldsymbol{\theta}}_T^g, \hat{\mathbf{g}}_1)$ *converges a.s.,* $\|\hat{\mathbf{g}}_t(\hat{\boldsymbol{\theta}}_T^g, \hat{\mathbf{g}}_1) - g_t(\boldsymbol{\theta}_0^g)\| \xrightarrow{a.s.} 0$ *as* $t, T \to \infty$*. The least squares estimator* $\hat{\boldsymbol{\Lambda}}_T^g$ *is strongly consistent for* $\boldsymbol{\Lambda}_0^g$*, i.e.* $\hat{\boldsymbol{\Lambda}}_T^g \xrightarrow{a.s.} \boldsymbol{\Lambda}_0^g$ *as* $T \to \infty$*, where* $\boldsymbol{\Lambda}_0^g \in \Theta^{\lambda_g}$ *is a minimizer of a limit criterion function* $Q_\infty^{(4)}(\boldsymbol{\Lambda}^g)$*.*

**Assumption 4.** $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ *is near epoch dependent (NED) of size* $-1$ *on a strongly mixing process of size* $-r/(r-1)$ *for some* $r > 2$*.*

**Assumption 3.a.** $\mathbb{E}\|\mathbf{y}_t\|^{2r} < \infty$ *with* $r$ *defined in Assumption 4.*

**Assumption 5.** $\boldsymbol{\theta}_0^c \in \text{int}(\Theta^c)$*,* $\boldsymbol{\Lambda}_0^c \in \text{int}(\Theta^{\lambda_c})$*,* $\boldsymbol{\theta}_0^g \in \text{int}(\Theta^g)$*,* $\boldsymbol{\Lambda}_0^g \in \text{int}(\Theta^{\lambda_g})$*.*

**Theorem 4** (Asymptotic Normality)**.** *Let assumptions 1, 2, 3.a and 4–5 hold. Then for every* $\hat{f}_1 \in \mathbb{R}$ *and* $\hat{\boldsymbol{g}}_1 \in \mathbb{R}^s$ *the four-step least squares estimator* $\hat{\boldsymbol{\theta}}_T$ *satisfies*

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{B}(\boldsymbol{\theta}_0)\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)) \quad as\ T \to \infty,$$

*where* $\boldsymbol{B}(\boldsymbol{\theta}_0) := \lim_{T\to\infty} \frac{1}{T}\mathbb{V}ar\left(\sum_{t=2}^{T} \nabla_{\boldsymbol{\theta}} q_t(\boldsymbol{\theta}_0)\right)$, *with*
$\nabla_{\boldsymbol{\theta}} q_t(\boldsymbol{\theta}_0) := \left(\nabla^{\top}_{\boldsymbol{\theta}^c} q_t^{(1)}(\boldsymbol{\theta}_0), \nabla^{\top}_{\mathrm{vec}\,\boldsymbol{\Lambda}^c} q_t^{(2)}(\boldsymbol{\theta}_0), \nabla^{\top}_{\boldsymbol{\theta}^g} q_t^{(3)}(\boldsymbol{\theta}_0), \nabla^{\top}_{(\boldsymbol{\Lambda}_1^g,\ldots,\boldsymbol{\Lambda}_S^g)} q_t^{(4)}(\boldsymbol{\theta}_0)\right)^{\top}$ *and*

$$\boldsymbol{A}(\boldsymbol{\theta}_0) := \begin{bmatrix} \boldsymbol{A}_1(\boldsymbol{\theta}_0) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{A}_{21}(\boldsymbol{\theta}_0) & \boldsymbol{A}_2(\boldsymbol{\theta}_0) & \mathbf{0} & \mathbf{0} \\ \boldsymbol{A}_{31}(\boldsymbol{\theta}_0) & \boldsymbol{A}_{32}(\boldsymbol{\theta}_0) & \boldsymbol{A}_3(\boldsymbol{\theta}_0) & \mathbf{0} \\ \boldsymbol{A}_{41}(\boldsymbol{\theta}_0) & \boldsymbol{A}_{42}(\boldsymbol{\theta}_0) & \boldsymbol{A}_{43}(\boldsymbol{\theta}_0) & \boldsymbol{A}_4(\boldsymbol{\theta}_0) \end{bmatrix}.$$

As expected the asymptotic variance of the Step I estimator has a standard form, while the variance of the Step II estimator is affected by the previous estimation step, the variance of the Step III estimator is affected by the two previous steps, and so on. This is implied by a lower triangular structure of matrix $\boldsymbol{A}(\boldsymbol{\theta})$. The exact expressions for the elements of the matrices $\boldsymbol{A}(\boldsymbol{\theta})$ and $\boldsymbol{B}(\boldsymbol{\theta})$ are provided in SA Section D.1.

As noted earlier the common factor captures the "group conditional expectation" of the data, while the group-specific factors describe the conditional means of $\mathbf{y}_t^s$, in deviations from the common factor. From the observation equation (1) we notice, however, that $\frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s} y_{i,t}^s = f_t + \frac{1}{S}\sum_{s=1}^{S} g_t^s + \tilde{\varepsilon}_t$, where $\tilde{\varepsilon}_t = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s}\varepsilon_{i,t}^s$. Hence, the forecast for the "group conditional expectation" is based on the common factor as well as on the averages of the group-specific factors. Intuitively, this means that if in the current period many groups experience downturn then in the next period we expect the downturn in the common factor as well. From the perspective of "factors disentangling", this means that the group-specific factors can introduce a 'bias' on the filtered common factor. The following remark highlights however that this bias vanishes for large $S$, or when the group factors represent a martingale difference sequence.

**Remark.** *Let either (i)* $\mathbb{E}\left[\frac{1}{S}\sum_{s=1}^{S} g_t^s|\mathcal{F}_{t-1}\right] = 0$*; or (ii)* $\{g_t^s\}_{t\in\mathbb{Z}}$ *is a martingale difference sequence for all* $s = 1,\ldots,S$*; or (iii)* $S\to\infty$*. Then,* $\hat{f}_t$ *is an unbiased filter of the common* *factor* $f_t$*, i.e.* $\mathbb{E}\Big[\hat{f}_t(\boldsymbol{\theta}_0^c)|\mathcal{F}_{t-1}\Big] = \mathbb{E}\Big[f_t(\boldsymbol{\theta}_0^c)|\mathcal{F}_{t-1}\Big]$.

# 3 Monte Carlo study

To investigate the performance of the proposed estimation procedure for the multilevel factor model, we perform a series of Monte Carlo experiments. The primary data generating process (DGP) under consideration is:

$$\begin{aligned}
\mathbf{y}_t &= \tilde{\boldsymbol{\Lambda}}^c \tilde{f}_t + \tilde{\boldsymbol{\Lambda}}^g \tilde{\mathbf{g}}_t + \tilde{\boldsymbol{\varepsilon}}_t, && \tilde{\boldsymbol{\varepsilon}}_t \overset{i.i.d.}{\sim} N(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}), \quad t = 1, \ldots, T, \\
\tilde{f}_{t+1} &= \kappa \tilde{f}_t + \xi_{t+1}, && \xi_t \overset{i.i.d.}{\sim} 0.5N(0,1), \\
\tilde{\mathbf{g}}_{t+1} &= \boldsymbol{\psi}^\top \tilde{\mathbf{g}}_t + \boldsymbol{\eta}_{t+1}, && \boldsymbol{\eta}_t \overset{i.i.d.}{\sim} 0.5N(\mathbf{0}_S, \mathbf{I}_S).
\end{aligned} \tag{4}$$

We use the following parameter values in the simulations: $\kappa = 0.9, \psi_s \sim U([0.75, 0.9])$, $\tilde{\lambda}^c_{i,s} \sim U([0,1])$ and $\tilde{\lambda}^g_{i,s} \sim U([0,1])$ for $s = 1, \ldots, S$. Furthermore, to interpret the factors as conditional means, we normalize the loadings such that $\frac{1}{S}\sum_{s=1}^S \frac{1}{N_s}\sum_{i=1}^{N_s} \tilde{\lambda}^c_{i,s} = 1$ and $\frac{1}{N_s}\sum_{j=1}^{N_s} \tilde{\lambda}^g_{j,s} = 1$ for each group $s$ where $\tilde{\boldsymbol{\Lambda}}^g_s := (\tilde{\lambda}^g_{1,s}, \ldots, \tilde{\lambda}^g_{N_s,s})$.

The aim of this study is to investigate the finite sample properties of the static and time-varying parameters as well as the fit of the data. The number of Monte Carlo simulations is set to 1000. In the primary setup, we put $N_s = 10, S = 10$ and $T = 300$. For each simulation, we generate observations for 500 time periods and discard those for the first 200 time periods to reduce the influence of the starting point. In some experiments we additionally consider other DGPs, different sample sizes $T$ and different numbers of groups $S$. When this is the case, it is explicitly stated in the text.

## 3.1 Static parameters

First, we analyze the finite sample properties of the static parameter estimates. This includes the factor parameters in $\boldsymbol{\theta}^c$ and $\boldsymbol{\theta}^g$ in model equations (2) and (3) as well as the loadings $\boldsymbol{\Lambda}^c$ and $\boldsymbol{\Lambda}^g$ in model equation (1). We consider three different sample sizes $T = \{300, 600, 1200\}$. The kernel density plots of the estimated parameters for different sample sizes are presented in Figure 1. Since we simulate data from the parameter-driven model (4) and we estimate the parameters from the observation-driven model discussed in Section 2, all the parameters are reparameterized. Therefore, to obtain the reparameterized values we simulate large time series of length $T = 1,000,000$ for $\mathbf{y}_t$ and then estimate the model (1)–(3) using the estimation procedure discussed in Section 2.2. We then check that for this length of the time series the estimates are invariant to the seed used in the simulation and do not change with an increase in the sample size. The results in Figure 1 confirm that the estimated parameters are concentrated around the reparameterized parameters and collapse towards them as the sample size increases.

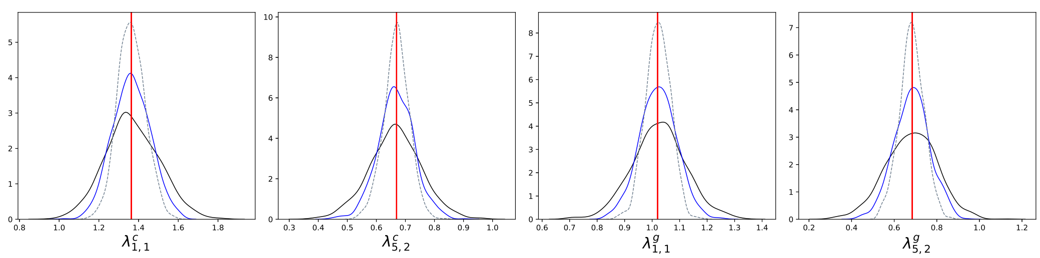



Figure 1: **Kernel density plots of the estimated parameters.** The results are based on 1000 Monte Carlo simulations for 3 sample sizes and with $S = 10$ and $N_s$=10. Since the dimension of the parameter vector is large, not all results are shown, but the results for other parameters are similar. The vertical red line indicates the reparameterized value of the parameter obtained using the large sample.

## 3.2 Filtering

Next, we analyze how well our model and estimation procedure can forecast the group conditional mean, which is $\frac{1}{N_s}\sum_{i=1}^{N_s} \lambda^c_{i,s}\tilde{f}_t + \tilde{g}^s_t$. For this purpose, we consider the DGP as stated above as well as a range of other patterns for the common and group-specific factors.

The patterns for the factors are summarized in Table E.10 in the SA. The patterns include smooth dynamics such as an AR(1) process or a sine curve, but also abrupt changes such as breaks and ramps.

We simulate time series of size $T = 300$ and estimate observation-driven model presented in Section 2. In Figure 2 we demonstrate that our model captures well the dynamics of the group conditional mean given different dynamic specifications for $f_t$ and $\mathbf{g}_t$. In most cases, the true value of the group mean (red dashed line) is within the 95% confidence bounds. Hence, the model as able to adapt to an abrupt single break but also to changes caused by multiple breaks. The additional results of the filtering of the common and group-specific factors are presented in SA Section E.

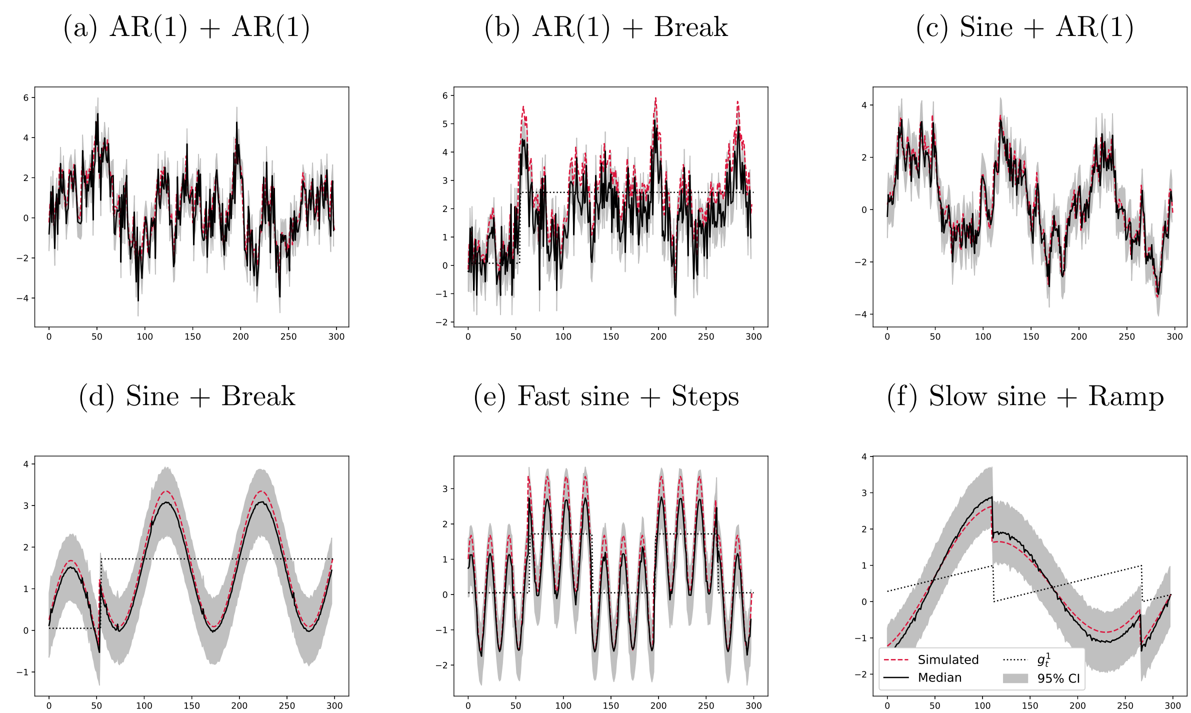



Figure 2: **One-step ahead forecast of the group 1 conditional mean.** The results are based on 1000 Monte Carlo simulations for different patterns for $f_t$ and $\mathbf{g}_t$ presented in Table E.10. Red dashed line denotes the true value of the group mean, solid black line represents the median value of the group mean computed over 1000 simulations and the gray shaded area indicates the 95% confidence bounds. For illustrative purposes, in some cases we also present the dynamics of the group-specific factor using the black dotted line to highlight the influence of the break, steps, or ramp.

### 3.3 Out-of-sample forecast performance

Finally, we evaluate out-of-sample forecast performance of the proposed model. We again use the primary DGP (4) to simulate the data. We conduct one, two, and three-step-ahead out-of-sample forecasts $M = 50$ times using a rolling-window of size $T$. In particular, we simulate time series of size $T + M$, estimate parameters using $\mathbf{y}_t, \ldots, \mathbf{y}_{T+t}$, then produce one, two, or three-step ahead forecasts for each of the series and repeat this procedure $M = 50$ times. In these experiments, we consider a small sample size $T = 100$ as well as a large sample size $T = 300$.

For each of the forecasts we compute an $N \times 1$ vector of mean squared errors, $MSE = \frac{1}{M}\sum_{m=1}^{M}(\mathbf{y}_{T+m} - \hat{\mathbf{y}}_{T+m})^2$, and of mean absolute errors, $MAE = \frac{1}{M}\sum_{m=1}^{M}|\mathbf{y}_{T+m} - \hat{\mathbf{y}}_{T+m}|$, where $\hat{\mathbf{y}}_{T+m}$ is either one, two, or three-step-ahead forecast. We compare results of our model to the results of the static factor model with one, two and three factors included which we indicate as PC1, PC2 and PC3, respectively. The static factor models are estimated using a two-step estimation procedure and the forecasts for them are produced iteratively. In Table 1, we present the average of the MSE and MAE ratios where the average is taken across the simulations and cross-sections. We find that according to both MSE and MAE ratios our model outperforms all static factor model specifications for one- and two-step-ahead forecasts and it performs equally well as the static factor models in the case of the three-step-ahead forecasts.

## 4 IP and Non-IP sectors in US Economic Activity

Our empirical study investigates the importance of Industrial Production (IP) and non-IP sectors in US economic activity as well as the linkages between the two sectors. IP and non-IP sectors constitute the Gross Domestic Product (GDP) index which is the most important measure of economic activity. Since many sectors are interconnected, they can be subject

| | | *Obs-driv/PC1* | | *PC2/PC1* | | *PC3/PC1* | |
|---|---|---|---|---|---|---|---|
| | | MSE | MAE | MSE | MAE | MSE | MAE |
| | $h = 1$ | 0.863 | 0.732 | 0.978 | 0.951 | 0.961 | 0.915 |
| $T = 100$ | $h = 2$ | 0.937 | 0.871 | 0.985 | 0.966 | 0.973 | 0.941 |
| | $h = 3$ | 0.980 | 0.959 | 0.990 | 0.977 | 0.981 | 0.959 |
| | $h = 1$ | 0.880 | 0.758 | 0.978 | 0.950 | 0.959 | 0.911 |
| $T = 300$ | $h = 2$ | 0.949 | 0.893 | 0.985 | 0.965 | 0.972 | 0.939 |
| | $h = 3$ | 0.993 | 0.982 | 0.989 | 0.975 | 0.981 | 0.957 |

Table 1: **Average out-of-sample MSE and MAE ratios for the observation driven, PC2, and PC3 models relative to the PC1 model.** The averages are taken across the simulations and cross-section. PC$k$ denotes a factor model with $k$ factors, $h$ is the forecast horizon and $T$ the rolling window size.

to shock spillovers as well as to common economic shocks. These shocks can potentially explain the large variability of the aggregate IP index that was documented by Foerster et al. (2011). We adopt our proposed multilevel factor model to analyze the co-movements within and between different US industries as well as their interconnectedness.

## 4.1 Data description

The dataset consists of quarterly standardized growth rates between 1977Q1 and 2011Q4 of IP and non-IP sectors in the US. In particular, we have the growth rates of 87 IP sectors and 42 non-IP sectors. The IP data is provided by the Board of Governors of the Federal Reserve System (FED) and the seasonally adjusted quarterly time series data for each of the sectors is retrieved from the FRED database[1]. The IP data is disaggregated up to the four-digit level in the North American Industry Classification System (NAICS) for the year 2002. The non-IP data is provided by Andreou et al. (2019).

We aim to identify both common and group-specific factors. To accommodate the within-group co-movements, we distribute sectors into groups. The groups are defined according to the NAICS and correspond to the two and three-digit aggregation levels of the year 2002 for the non-IP and IP sectors, respectively. In this way, we obtain 24 groups

[1]https://fred.stlouisfed.org

for the IP data and 11 groups for the non-IP data; see the SA for more information.

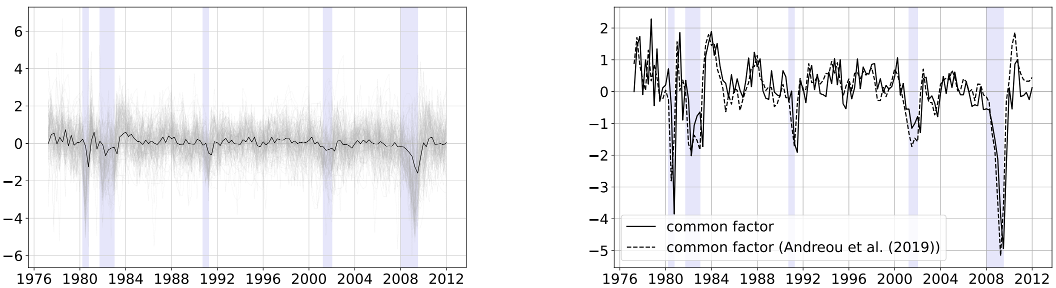
For the period under consideration, the variables are published at different time frequencies. The IP sector data is released on a monthly basis, while non-IP sector data is published on an annual basis. We follow Andreou et al. (2019) by considering yearly non-IP sector data and quarterly IP sector data. Hence, we have 35 years of observations for the low-frequency (non-IP) variables and 140 quarters of observations for the high-frequency (IP) variables. To obtain a fully quarterly dataset, we disaggregate the annual (non-IP) data into quarterly series using a mix of the Al-Osh (1989) and Silva & Cardoso (2001) methods; see the SA for more information.

## 4.2 Common factor estimates

Figure 3a presents the observed data and the filtered common factor. This estimated factor appears to correctly capture the "group conditional expectation" of the data and the variation in this factor is closely related to the economic cycles. In particular, strong downturns of the common factor are well aligned with the US recessions. Specifically, large falls are evident during the early 1980s recession, early 1990s recession, September 9/11 attacks, and perhaps most profoundly, during the great recession. Therefore, we conclude that the common factor indeed summarizes the aggregate shocks in the economy accurately.

We also provide a comparison with the common factor reported in Andreou et al. (2019) and which is similar to our filtered dynamic common factor (Figure 3b). Both factor estimates are based on a similar dataset and group-factor model. However, the estimates of Andreou et al. (2019) are based on principal components (using all data) while our estimates are coming from an observation-driven "real-time" filter (only using past data). Overall, we find that the correlation between the two sets of factor estimates is high (approximately, 72% contemporaneous and 90% with one lag/forward).

Furthermore, we measure the importance of the common factor to the sectors by the




(a) Non-IP and IP time series (grey lines) and the filtered common factor (black line).

(b) The common factor from our model and the one from Andreou et al. (2019).

Figure 3: **The estimated common factor.** Light purple shaded areas correspond to the recession periods as established by the National Bureau of Economic Research (`https://www.nber.org/cycles.html`). In panel (b), the estimated common factor from our model is standardized to facilitate comparison with Andreou et al. (2019).

coefficient of determination $R^2$ from the regressions of the sectoral growth rates on our filtered common factor. In Table 2 we present the rankings based on this $R^2$ for the IP and non-IP sectors. We find that the $R^2$ for the IP sectors is higher than for the non-IP sectors. Hence, the common factor explains more variability in the IP sectors data. This is in line with the findings in Andreou et al. (2019) where the authors find that the common factor is more related to IP data. We also find that among the non-IP sectors the common factor plays the most important role in explaining the variability of the Administration and support services, Construction, and Wholesale trade sectors. Among the top ten ranked IP sectors, most of them belong to the Fabricated metal products (FMP), Furniture (Furnit), and Machinery (Mach) groups.

Overall, We find that the fit ($R^2$) is low for many sectors when regressed only on the common factor; hence, a few sectors contain all information about the dynamics of the common factor. Interestingly, the sectors for which the common factor provides a good fit do not necessarily have the largest weight in the IP or GDP indices.

The close similarity of our results compared to those obtained by Andreou et al. (2019) is interesting since we use (partially) different datasets, different models, different methods for factor estimation, and also different ways of treating mixed frequencies and of defining

| IP sectors | Group | $R^2$(%) | non-IP sectors | Group | $R^2$(%) |
|---|---|---|---|---|---|
| Com. and serv. ind. machin. & other gen. purpose machin. | Mach | 53.263 | Administr. and support servic. | PBS | 44.475 |
| Forging and stamping | FMP | 46.361 | Construction | Constr | 41.145 |
| Metalworking machinery | Mach | 45.712 | Wholesale trade | WT | 34.627 |
| Coating, engraving, heat treating | FMP | 43.953 | Accommodation | AER | 32.936 |
| Other fabricated metal prod. | FMP | 43.234 | Miscel. prof., scientif.&tech. servic. | PBS | 27.82 |
| Machine shops, turned product | FMP | 41.67 | Other transport.&support activ. | TW | 26.563 |
| Architectural and structur. metals | FMP | 39.7 | Gov. enterprises (Federal) | Gov | 25.241 |
| Household and instit. furniture | Furnit | 37.982 | Retail trade | RT | 24.125 |
| Other miscell. manufact. | Miscel | 33.817 | Rail transportation | TW | 22.767 |
| Office and other furniture | Furnit | 33.767 | Warehousing and storage | TW | 22.475 |

Table 2: **Regression results of the sectoral growth rates on the common factor.** We demonstrate the top ten ranked IP and non-IP sectors together with the group name according to the $R^2$ of the regression of the sectoral growth rates on the estimated common factor. The group name abbreviations are outlined in Tables A.1 and A.2 in the SA.

sector groups [2]. However, in contrast, our model and methods can provide forecasts and impulse response functions.

## 4.3 Granularity and network analysis

Next, we examine group interconnectedness in terms of the shock spillovers. Given that our results are coming from a dynamic model, we can examine the contribution of the contemporaneous shocks as well as of the propagation effects on the industrial production aggregate index. Furthermore, we study how these shocks propagate and dissipate through the network of connections by means of the network representation based on impulse response functions (IRFs).

A high IP index variability may arise because of (i) common contemporaneous shocks affecting all the sectors, (ii) large sectors having large non-averaging out effects, and (iii) propagation of the shocks; see Foerster et al. (2011). The covariance terms of the sectoral growth rates contain information on these three effects while the 'uncorrelated' innovations summarize the contemporaneous effects (i) and (ii) excluding the propagation effects. Therefore, for making an effort to explain the large variability in the IP index, we pro-

[2] We did a robustness check where we split the data into 2 groups as in Andreou et al. (2019). The results are rather similar and are available upon request.

vide an insight into the importance of contemporaneous common shocks by analyzing the covariance matrix of the residuals. In particular, we compare the standard deviations of the aggregate indexes by taking account of the covariances and by not taking account of these (Table 3). Moreover, to quantify the importance of the propagation effects, we consider both the aggregated indexes computed as the weighted averages of the sectoral (IP or non-IP) growth rates as well as the weighted average of the residuals. The results of this analysis show that both the contemporaneous shocks and propagation effects are important sources of the IP index variation, together accounting for approximately 65% of the variation. While propagation is less important, it still accounts for a non-negligible part of the variation (around 24%). The aggregate non-IP index has substantially less variability while it has contemporaneous and propagation effects jointly accounting for 50% of the total variability and propagation effects accounting for around 32%. We may conclude that the effects of the contemporaneous shocks are stronger for IP than for non-IP sectors.

| | I. 'Uncorrelated innovations' | | II. Sectoral growth rates | |
|---|---|---|---|---|
| | IP sectors | non-IP sectors | IP sectors | non-IP sectors |
| With sectoral covariation | 4.25 | 0.26 | 5.58 | 0.38 |
| Without sectoral covariation | 1.77 | 0.14 | 1.97 | 0.19 |

Table 3: **Standard deviations with and without the sectoral covariance terms.** Similar to Foerster et al. (2011), the entries of the rows labeled 'with sectoral covariation' are the sample standard deviations of $\sum_{i=1}^{N} w_{i,t} u_{i,t}^{G}$ (I), where $u_{i,t}^{G}$ denote 'uncorrelated' innovations adjusted for the original standard deviations of the sectoral growth rates, and $\sum_{i=1}^{N} w_{i,t} y_{i,t}^{G}$ (II), where $y_{i,t}^{G}$ are non-standardized sectoral growth rates. Superscript $G = \{\text{IP}, \text{non-IP}\}$ highlights that the sum is taken over the series that either belong to IP or non-IP group. The rows labeled 'without sectoral covariance' refer to the standard deviation computed without taking into account the sectoral covariance terms. As a robustness check we also considered equal weights, the results are similar and not included.

To analyze the importance of the large sectors in explaining large variability of the aggregate indices, in Figure 4, we report the top ten sectors with the largest standard deviation of the destandardized 'uncorrelated' innovations, namely potentially influential shock-transmitters. These results demonstrate that among the large shock transmitters only 2-3 of them have a large weight in the aggregate index which indicates that the large sectors are unlikely to be the dominant source of the large variations in the aggregate IP index.

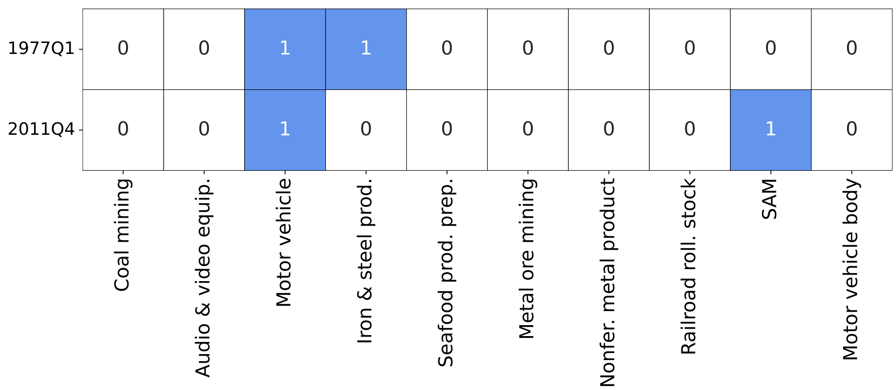
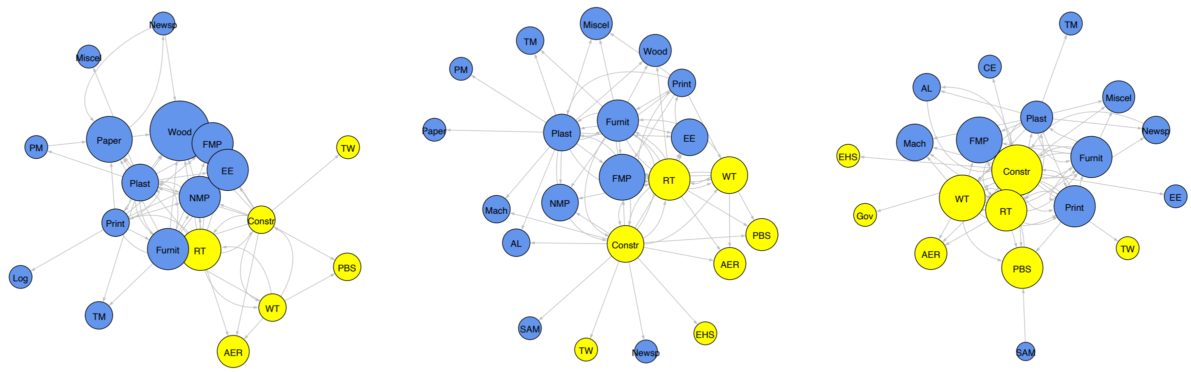



Figure 4: **Sectors with the largest standard deviations of the residuals.** The figure presents the top ten IP sectors with the largest standard deviation of the destandardized 'uncorrelated' innovations. The blue cells correspond to the sectors that are also in the top ten according to the weight of the aggregate industrial production index in year 1997Q1 and 2011Q4.

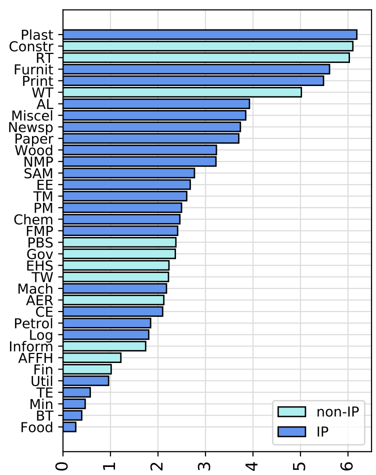
Moreover, given that our model is dynamic, it allows tracking shocks propagation through the networks of connections using the approach of Diebold & Yılmaz (2014). However, we adapt this approach to account for group rather than sector interconnectedness by considering the averages of the generalized impulse responses of Koop et al. (1996) within each group. Moreover, we examine how the shocks propagate and accumulate through the network by focusing on the cumulative impulse responses with different time periods after the shock occurrence. The networks summarize information on the effect of the shock from one group to another with the strength of the connection being determined by the strength of the (cumulative) responses. To quantify this effect we look at the (average) group impulse responses due to the shock in one of the groups. Specifically, by considering the shocks to one group (all the sectors within the group) we quantify the responses of all other groups. This provides us with the pairwise connectedness measures between the shock-transmitter group and all other groups. We assume that the size of the shock in the group-transmitter is proportional to the number of the sectors in the group as well as to the size of the residuals of the series within this group. Then we repeat the analysis by considering another group as a shock-transmitter and quantifying the responses of all other groups. The network adjacency matrix is obtained once all the groups have been considered as potential shock transmitters.



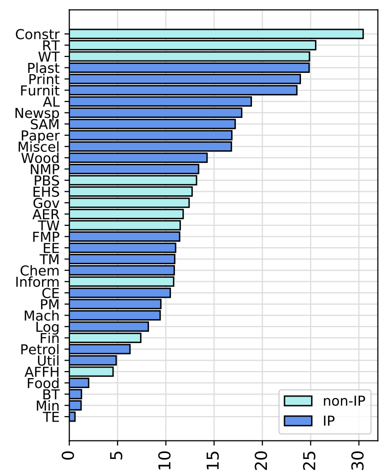

(a) $h = 0$ (b) $h = 1$ (c) $h = 4$

Figure 5: **Networks based on the group pairwise directional connectedness for different periods after the shock.** The nodes correspond to the different groups of economic activity. The color of the node indicates whether the group is IP (blue) or non-IP (yellow). The size of the node is based on the (rescaled) degree centrality. The group name abbreviations are explained in Tables A.1 and A.2 in the SA. The nodes with weak connections, are not shown in the plot.

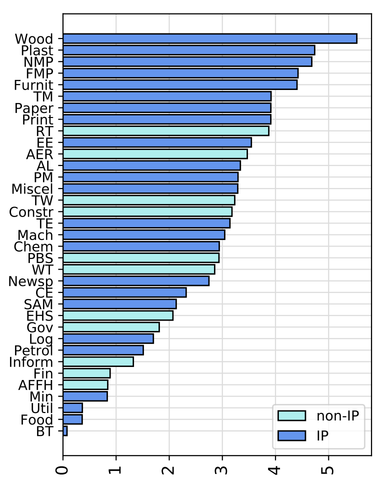
The networks based on the cumulative group average impulse responses are presented in Figure 5. We consider different horizons after the shock occurrence: contemporaneous response ($h = 0$), 1 quarter ($h = 1$) and one year ($h = 4$) after the shock. For illustrative purposes, we only present the pairs that have the strongest connection (defined as the 95th percentile of all average pairwise connections, in absolute terms). The networks reveal several interesting results. First, at $h = 0$ there are many links between IP groups, a few from non-IP to IP groups and between non-IP groups, and no links from IP to non-IP groups. Furthermore, we find that the edges between the non-IP groups at $h = 0$ are preserved at $h = 4$, although many new edges between non-IP groups appear as the horizon increases. In contrast, the number of links between IP groups decreases as the horizon $h$ increases, while the number of links between IP and non-IP groups substantially increases. Therefore, we observe immediate effects of the shocks between IP groups which can potentially be explained by the presence of the input-output relations between the sectors. The cumulative effects are more pronounced between IP and non-IP sectors and between non-IP sectors. The latter indicates that when non-IP groups are involved it takes time for the shocks to accumulate. It confirms our earlier finding that the role of the propagation effects is larger for non-IP than for IP sectors. Further, the results show that one year after the shock ($h = 4$) the "hub" groups become more evident. In particular, Construction (Constr) and Wholesale trade (WT) are involved in most of the edges which implies that these groups have larger cumulative effects on the others.



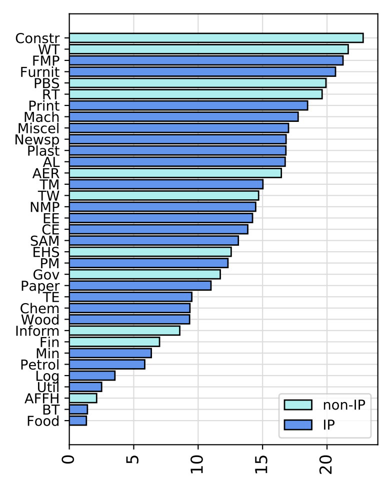

(a) out-degree centrality $h = 0$




(b) out-degree centrality $h = 4$




(c) in-degree centrality $h = 0$




(d) in-degree centrality $h = 4$

Figure 6: **Ranking based on the absolute value of the in- and out-degree centrality measures at $h = 0$ and $h = 4$.** Details on the centrality measures can be found in the SA C. The group name abbreviations are explained in Tables A.1 and A.2 in the SA.

As the network representation only shows the strongest connections, in Figure 6 we present a ranking based on the in- and out-degree centrality measures for the different periods after the shock occurrence. This allows us to identify the most central groups based on how the node responds to the shock in other groups (the in-degree centrality) and how other groups respond to the shock in the node (the out-degree centrality). For the bottom ten ranked groups, we find that at $h = 0$ and $h = 4$ the composition remains almost unchanged according to both in- and out-degree centrality. Hence, several groups are constantly isolated in the network since they do not strongly respond to the shocks in other groups and do not transmit shocks to the neighbors much. For the top ten ranked groups based on the out-degree centrality, the composition remains almost unchanged when the horizon increases. Consequently, there are few groups that transmit pronouncedly shocks to others and the effect accumulates over time. Among the top ten ranked groups based on the in-degree centrality the number of non-IP groups increases substantially when the horizon increases. Hence, while the effect on the IP groups is immediate, for the shocks to non-IP groups it takes time to accumulate. We again find that Construction, Retail and Wholesale trade sectors are the most central according to both in- and out-degree centrality measures. Among the IP groups, Plastics, Furniture and Printing groups have large in- and out-degree centrality measures both at $h = 0$ and $h = 4$. In Figure 5 it is shown that these groups are tightly linked (in both directions) to Construction, Wholesale and Retail trade groups which transmit large shocks, leading to high in-degree centrality but also high out-degree centrality.

# 5 Conclusion

We have introduced a new parsimonious multilevel factor model with observation-driven factor dynamics. The model accounts for different types of factors such as common and group-specific factors, which can be relevant in many applications. The method is easy to apply in practice since our proposed estimation procedure is simple and fast. Moreover, the dynamic model generates forecasts and impulse response functions in a standard fashion. This dynamic feature can provide insight on how economic shocks propagate and dissipate through the network of connections across groups/sectors and across time. We further have established theoretical stochastic properties of the filters and asymptotic properties of the estimators. In Monte Carlo experiments, we have established that our estimators behave well in finite samples. In the empirical study, we have studied the role of the IP and non-IP industries in the US economy. The results confirm the importance of the IP sectors in explaining the variability of aggregate shocks. A key novel finding is that the non-IP groups/sectors are more tightly linked to both the IP and non-IP groups when the cumulative effects over time are accounted for. Our proposed dynamic factor model is able to measure these propagation effects in an effective way and without relying on heavy computational methods.

# References


Al-Osh, M. (1989). A dynamic linear model approach for disaggregating time series data. *Journal of Forecasting*, *8*(2), 85–96.

Andreou, E., Gagliardini, P., Ghysels, E., & Rubin, M. (2019). Inference in group factor models with an application to mixed-frequency data. *Econometrica*, *87*(4), 1267–1305.

Bai, J., & Wang, P. (2015). Identification and Bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics*, *33*(2), 221–240.

Beck, G. W., Hubrich, K., & Marcellino, M. (2009). Regional inflation dynamics within and across euro area countries and a comparison with the United States. *Economic Policy*, *24*(57), 142–184.

Blasques, F., Francq, C., & Laurent, S. (2023). Quasi score-driven models. *Journal of Econometrics*, *234*(1), 251–275.

Blasques, F., van Brummelen, J., Koopman, S. J., & Lucas, A. (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics*, *227*(2), 325–346.

Breitung, J., & Eickmeier, S. (2014). Analyzing business and financial cycles using multi-level factor models. *Advances in Econometrics*, *35*, 177–214.

Choi, I., Kim, D., Kim, Y. J., & Kwark, N.-S. (2018). A multilevel factor model: Identification, asymptotic theory and applications. *Journal of Applied Econometrics*, *33*(3), 355–377.

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., ... Lauritzen, S. L. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, *8*(2), 93–115.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

Creal, D., Schwaab, B., Koopman, S. J., & Lucas, A. (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *Review of Economics and Statistics*, *96*(5), 898–915.

Crucini, M. J., Kose, M. A., & Otrok, C. (2011). What are the driving forces of international business cycles? *Review of Economic Dynamics*, *14*(1), 156–175.

Diebold, F. X., Li, C., & Yue, V. Z. (2008). Global yield curve dynamics and interactions: a dynamic Nelson–Siegel approach. *Journal of Econometrics*, *146*(2), 351–363.

Diebold, F. X., & Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, *182*(1), 119–134.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford university press.

Engle, R., & Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, *76*(376), 774–781.

Foerster, A. T., Sarte, P.-D. G., & Watson, M. W. (2011). Sectoral versus aggregate shocks: A structural factor analysis of industrial production. *Journal of Political Economy*, *119*(1), 1–38.

Jungbacker, B., & Koopman, S. (2015). Likelihood-based dynamic factor analysis for measurement and forecasting. *Econometrics Journal*, *18*, C1–C21.

Karabiyik, H., & Westerlund, J. (2021). Forecasting using cross-section average–augmented time series regressions. *The Econometrics Journal*, *24*(2), 315–333.

Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, *74*(1), 119–147.

Kose, M. A., Otrok, C., & Whiteman, C. H. (2003). International business cycles: World, region, and country-specific factors. *The American Economic Review*, *93*(4), 1216–1239.

Long, J. B., & Plosser, C. I. (1987). Sectoral vs. aggregate shocks in the business cycle. *The American Economic Review*, *77*(2), 333–336.

Moench, E., Ng, S., & Potter, S. (2013). Dynamic hierarchical factor models. *Review of Economics and Statistics*, *95*(5), 1811–1817.

Newey, K., & McFadden, D. (1994). Large sample estimation and hypothesis. *Handbook of Econometrics, IV, Edited by R.F. Engle and D.L. McFadden*, 2112–2245.

Pötscher, B. M., & Prucha, I. (1997). *Dynamic nonlinear econometric models: Asymptotic theory.* Springer Science & Business Media.

Silva, J. S., & Cardoso, F. (2001). The Chow-Lin method using dynamic models. *Economic Modelling*, *18*(2), 269–280.

Straumann, D., & Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics*, *34*(5), 2449–2495.

Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.

Wang, P. (2008). Large dimensional factor models with a multi-level factor structure: Identification, estimation and inference. Working paper, Hong Kong University of Science and Technology, Hong Kong.

Westerlund, J., & Urbain, J.-P. (2015). Cross-sectional averages versus principal components. *Journal of Econometrics*, *185*(2), 372–377.

White, H. (1996). *Estimation, inference and specification analysis* (No. 22). Cambridge university press.

Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH (1,1) model. *Scandinavian Journal of Statistics*, *40*(4), 846–867.

# Appendix

This appendix contains the proofs of the theorems stated in Section 2.5. Other theoretical details can be found in the SA Section D. We adopt the common notation for the norms.

Particularly, we use a Euclidean norm for vectors, that is for any vector $\mathbf{x}$, $\|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}}$, and a Frobenius norm for matrices, that is for any matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\| = \sqrt{\text{trace}(\boldsymbol{A}'\boldsymbol{A})}$.

The vector $\boldsymbol{\theta}$ collects all the unknown parameters. To shorten further notation, we introduce $\boldsymbol{\theta} := (\boldsymbol{\theta}^{(1)^\top}, \boldsymbol{\theta}^{(2)^\top}, \boldsymbol{\theta}^{(3)^\top}, \boldsymbol{\theta}^{(4)^\top})^\top$ where superscript $i$ corresponds to the step of the estimation procedure. For example, $\boldsymbol{\theta}^{(1)} \equiv \boldsymbol{\theta}^c$.

*Proof of Theorem 1:* The existence and measurability of the estimator $\hat{\boldsymbol{\theta}}_T^{(1)}$ follow straightforwardly from Theorem 2.11 in White (1996) since the criterion function is continuous on all arguments and the parameter space $\Theta^c$ is compact.

To prove consistency it is sufficient to verify the uniform convergence of the criterion function to the limit criterion function and the identifiable uniqueness of the minimizer $\boldsymbol{\theta}_0^{(1)}$ of the limit criterion function (White, 1996, Theorem 3.4). For notational convenience, we define $Q_\infty^{(1)}(\boldsymbol{\theta}^{(1)}) := \mathbb{E}\left[q^{(1)}\left(\mathbf{y}_t, f_t(\boldsymbol{\theta}^{(1)}), \boldsymbol{\theta}^{(1)}\right)\right]$, $Q_T^{(1)}(\boldsymbol{\theta}^{(1)}) := \frac{1}{T}\sum_{t=2}^T q^{(1)}\left(\mathbf{y}_t, f_t(\boldsymbol{\theta}^{(1)}), \boldsymbol{\theta}^{(1)}\right)$, and $\hat{Q}_T^{(1)}(\boldsymbol{\theta}^{(1)}) := \frac{1}{T}\sum_{t=2}^T q^{(1)}\left(\mathbf{y}_t, \hat{f}_t(\boldsymbol{\theta}^{(1)}, \hat{f}_1), \boldsymbol{\theta}^{(1)}\right)$. To show the uniform convergence of the criterion function we proceed in a similar manner as in Blasques et al. (2022). By the triangle inequality we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}^c}\left|\hat{Q}_T^{(1)}(\boldsymbol{\theta}) - Q_\infty^{(1)}(\boldsymbol{\theta})\right| \le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}^c}\left|\hat{Q}_T^{(1)}(\boldsymbol{\theta}) - Q_T^{(1)}(\boldsymbol{\theta})\right| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}^c}\left|Q_T^{(1)}(\boldsymbol{\theta}) - Q_\infty^{(1)}(\boldsymbol{\theta})\right|. \quad (5)$$

The strong uniform convergence of the second term in (5) follows by Lemma TA.1. For the first term on the right hand side, by the triangle inequality we have

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}^c}\left|\hat{Q}_T^{(1)}(\boldsymbol{\theta}) - Q_T^{(1)}(\boldsymbol{\theta})\right| &\le \frac{1}{T}\sum_{t=2}^T \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}^c}|(\phi(\mathbf{y}_t) - \hat{f}_t(\boldsymbol{\theta}))^2 - (\phi(\mathbf{y}_t) - f_t(\boldsymbol{\theta}))^2| \\ &\le \frac{1}{T}\sum_{t=2}^T \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}^c}|\hat{f}_t^2(\boldsymbol{\theta}) - f_t^2(\boldsymbol{\theta})| + \frac{2}{T}\sum_{t=2}^T |\phi(\mathbf{y}_t)| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}^c}|\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})|, \end{aligned} \quad (6)$$

where $\phi(\mathbf{y}_t) := \frac{1}{S}\sum_{s=1}^S \frac{1}{N_s}\sum_{i=1}^{N_s} y_{i,t}^s$. By Proposition 1 $\sup_{\boldsymbol{\theta}\in\Theta^c}|\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$ as $t \to \infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|f_t(\boldsymbol{\theta})|^2 < \infty$. Hence, by Lemma 2.1 in Straumann & Mikosch (2006) and Corollary TA.15 in Blasques et al. (2022), the expression (6) converges to 0 a.s. since sequence $\{|\phi(\mathbf{y}_t)|\}_{t\in\mathbb{Z}}$ is SE (implied by Assumption 2) with $\mathbb{E}\log^+|\phi(\mathbf{y}_t)| < \infty$ (implied by Assumption 3). Hence, the uniform convergence of the criterion function in (5) holds.

Now we turn to the identifiable uniqueness of $\boldsymbol{\theta}_0^{(1)} \in \Theta$. For a nonlinear least squares criterion function the minimum is achieved at the conditional mean. Therefore, the uniqueness condition is that $f_t(\boldsymbol{\theta}^{(1)}) \neq f_t(\boldsymbol{\theta}_0^{(1)})$ for all $\boldsymbol{\theta}^{(1)} \neq \boldsymbol{\theta}_0^{(1)}$ on a set of non-zero probability (Newey & McFadden (1994)). To prove uniqueness we proceed by contradiction. We assume that there is $\boldsymbol{\theta}^{(1)} \neq \boldsymbol{\theta}_0^{(1)}$ such that $f_t(\boldsymbol{\theta}_0^{(1)}) = f_t(\boldsymbol{\theta}^{(1)})$ on a set of non-zero probability. This implies that for every $t$,

$$(\beta - \beta_0)\phi(\mathbf{y}_t) = (\gamma_0 - \beta_0 - (\gamma - \beta))f_t(\boldsymbol{\theta}_0^{(1)}). \tag{7}$$

We notice that the left hand side in (7) is $\mathcal{F}_t$ measurable, while the right hand side is $\mathcal{F}_{t-1}$ measurable. For this equality to hold we should have $\beta = \beta_0$ and $\gamma = \gamma_0$ which leads to the contradiction. Given the uniqueness of $\boldsymbol{\theta}_0^{(1)}$, identifiable uniqueness immediately follows from the continuity of the limit criterion function and compactness of the set $\Theta$; see Pötscher & Prucha (1997). Hence, $\hat{\boldsymbol{\theta}}_T^{(1)} \xrightarrow{a.s.} \boldsymbol{\theta}_0^{(1)}$ as $T \to \infty$. The convergence of the plug-in estimator follows by Lemma TA.2 thus completing the proof. ■

*Proof of Theorem 2:* We denote the critetion function evaluated at the filtered time-varying parameter as $\hat{Q}_T^{(2)}(\boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(1)}) := \frac{1}{T}\sum_{t=2}^T q^{(2)}\left(\mathbf{y}_t, \hat{f}_t(\boldsymbol{\theta}^{(1)}), \boldsymbol{\theta}^{(2)}\right)$, the criterion function evaluated at the limit time-varying parameter as $Q_T^{(2)}(\boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(1)}) := \frac{1}{T}\sum_{t=2}^T q^{(2)}\left(\mathbf{y}_t, f_t(\boldsymbol{\theta}^{(1)}), \boldsymbol{\theta}^{(2)}\right)$, and the limit criterion function as $Q_\infty^{(2)}(\boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(1)}) := \mathbb{E}\left[q^{(2)}(\mathbf{y}_t, f_t(\boldsymbol{\theta}^{(1)}), \boldsymbol{\theta}^{(2)})\right]$, where superscript (2) refers to the step of the estimation procedure.

The existence and measurability of the estimator $\hat{\boldsymbol{\theta}}_T^{(2)}$ follows from Theorem 2.15 for two-stage estimators in White (1996) since $\Theta^{\lambda_c}$ is compact and criterion function is continuous.

We highlight that the criterion function $\hat{Q}_T^{(2)}(\boldsymbol{\theta}^{(2)}, \hat{\boldsymbol{\theta}}_T^{(1)})$ depends on the filtered time-varying parameter that was initialized at some value $\hat{f}_1$ at time $t = 1$ and evaluated at the first stage parameter estimate $\hat{\boldsymbol{\theta}}_T^{(1)}$. By the triangle inequality

$$\sup_{\boldsymbol{\theta}\in\Theta^{\lambda_c}} \left|\hat{Q}_T^{(2)}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_T^{(1)}\right) - Q_\infty^{(2)}\left(\boldsymbol{\theta}, \boldsymbol{\theta}_0^{(1)}\right)\right| \leq \sup_{\boldsymbol{\theta}\in\Theta^{\lambda_c}} \left|\hat{Q}_T^{(2)}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_T^{(1)}\right) - Q_T^{(2)}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_T^{(1)}\right)\right| + \sup_{\boldsymbol{\theta}\in\Theta^{\lambda_c}} \left|Q_T^{(2)}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_T^{(1)}\right) - Q_\infty^{(2)}\left(\boldsymbol{\theta}, \boldsymbol{\theta}_0^{(1)}\right)\right|. \quad (8)$$

The first term in the expression above goes to zero almost surely by the same argument as in the proof of Theorem 1 since the filter is uniformly invertible. The uniform convergence of the criterion function then follows since the second term in (8) converges to zero almost surely by Lemma TA.3.

Identifiable uniqueness condition is satisfied, since $\boldsymbol{\Lambda}^c f_t(\boldsymbol{\theta}_0^c)$ is a conditional mean and $\boldsymbol{\Lambda}^c f_t(\boldsymbol{\theta}_0^c) \neq \tilde{\boldsymbol{\Lambda}}^c f_t(\boldsymbol{\theta}_0^c)$ for $\boldsymbol{\Lambda}^c \neq \tilde{\boldsymbol{\Lambda}}^c$ and every $t$ (Newey & McFadden (1994)), the criterion function is continuous and the parameter space $\Theta^{\lambda_c}$ is compact. ∎

*Proof of Theorem 4:* In the proof of asymptotic normality of the four-stage estimator, we rely on the theorems in White (1996) for two-stage estimators and theory developed in Blasques et al. (2023) for establishing the asymptotic normality of the estimators that are based on the filtered time-varying parameters.

For convenience, we denote as $\hat{\boldsymbol{Q}}_T(\boldsymbol{\theta})$ a 4-dimensional vector with the criterion functions of each step of the estimation procedure as the elements of the vector, i.e. $\hat{\boldsymbol{Q}}_T(\boldsymbol{\theta}) = (\hat{Q}_T^{(1)}(\boldsymbol{\theta}^{(1)}), \hat{Q}_T^{(2)}(\boldsymbol{\theta}^{(2)}), \hat{Q}_T^{(3)}(\boldsymbol{\theta}^{(3)}), \hat{Q}_T^{(4)}(\boldsymbol{\theta}^{(4)}))^\top$, where the 'hat' highlights that the criterion functions depend on the filtered time-varying parameters $\hat{f}_t$ and $\hat{\boldsymbol{g}}_t$. The 'blocks' of the estimator $\hat{\boldsymbol{\theta}}_T$ are obtained sequentially in four steps by minimizing corresponding criterion functions given the filtered time-varying parameters $\hat{f}_t$ and $\hat{\boldsymbol{g}}_t$, i.e. $\hat{\boldsymbol{\theta}}_T = (\hat{\boldsymbol{\theta}}_T^{(1)}, \hat{\boldsymbol{\theta}}_T^{(2)}, \hat{\boldsymbol{\theta}}_T^{(3)}, \hat{\boldsymbol{\theta}}_T^{(4)})^\top = \arg\min_{\boldsymbol{\theta}\in\Theta} \hat{\boldsymbol{Q}}_T(\boldsymbol{\theta})$.

First, we derive the asymptotic distribution of the estimator $\tilde{\boldsymbol{\theta}}_T$ which minimizes the criterion function $\boldsymbol{Q}_T(\boldsymbol{\theta})$ evaluated at the limit sequences $f_t$ and $\boldsymbol{g}_t$. Then, we show that

$\hat{\boldsymbol{\theta}}_T$ has the same asymptotic distribution as $\tilde{\boldsymbol{\theta}}_T$.

The proof of the asymptotic normality of $\tilde{\boldsymbol{\theta}}_T$ is based on Theorem 6.10 in White (1996) which we generalize to the 4-step estimator. By the mean value theorem, we have $\nabla_{\boldsymbol{\theta}} \boldsymbol{Q}_T(\tilde{\boldsymbol{\theta}}_T) = \nabla_{\boldsymbol{\theta}} \boldsymbol{Q}_T(\boldsymbol{\theta}_0) + \boldsymbol{A}_T(\boldsymbol{\theta}_T^{\star}) \left(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right)$, where, with some abuse of notation, $\boldsymbol{\theta}_T^{\star}$ lies (row-wise) between $\boldsymbol{\theta}_0$ and $\tilde{\boldsymbol{\theta}}_T$ and $\boldsymbol{A}_T(\boldsymbol{\theta}) := \frac{1}{T}\boldsymbol{A}_t(\boldsymbol{\theta})$ with $\boldsymbol{A}_t$ as defined in (D.37). Since $\tilde{\boldsymbol{\theta}}_T$ is an $m$-estimator and assuming that $\boldsymbol{A}_T(\boldsymbol{\theta}_T^{\star})$ is invertible, we obtain $\sqrt{T}\left(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) = -\left(\boldsymbol{A}_T(\boldsymbol{\theta}_T^{\star})\right)^{-1} \sqrt{T} \nabla_{\boldsymbol{\theta}} \boldsymbol{Q}_T(\boldsymbol{\theta}_0)$. Therefore, the asymptotic normality of $\tilde{\boldsymbol{\theta}}_T$ follows if (a.) $\nabla_{\boldsymbol{\theta}} \boldsymbol{Q}_T(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{B}(\boldsymbol{\theta}_0))$ as $T \to \infty$; (b.) $\boldsymbol{A}_T(\boldsymbol{\theta}_T^{\star}) \xrightarrow{P} \boldsymbol{A}(\boldsymbol{\theta}_0)$ as $T \to \infty$; (c.) $\boldsymbol{A}(\boldsymbol{\theta}_0)$ is nonsingular.

Condition (a.) holds since it is implied by Lemma TA.7. The non-singularity of the limit, condition (c.), follows by the uniqueness of the minimum $\boldsymbol{\theta}_0$ which is established in the proofs of Theorems 1–3. Theorems 1–3 and Theorem 18.10 (vi) in Van der Vaart (2000) imply $\tilde{\boldsymbol{\theta}}_T^{\star} \xrightarrow{P} \boldsymbol{\theta}_0$ as $T \to \infty$ thus ensuring condition (b.), see Lemma TA.9. By Slutsky lemma we obtain $\sqrt{T}\left(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{B}(\boldsymbol{\theta}_0)\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0))$ as $T \to \infty$.

If $\sqrt{T}\left\|\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T\right\| \xrightarrow{P} 0$ as $T \to \infty$, then by Theorem 18.10(iv) in Van der Vaart (2000), $\hat{\boldsymbol{\theta}}_T$ has the same asymptotic distribution as $\tilde{\boldsymbol{\theta}}_T$. By the mean value theorem

$$\nabla_{\boldsymbol{\theta}} \boldsymbol{Q}_T(\tilde{\boldsymbol{\theta}}_T) = \nabla_{\boldsymbol{\theta}} \boldsymbol{Q}_T(\hat{\boldsymbol{\theta}}_T) + \boldsymbol{A}_T(\boldsymbol{\theta}_T^{\star}) \left(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T\right), \tag{9}$$

where $\boldsymbol{\theta}_T^{\star}$ lies, with an abuse of notation, (row-wise) between $\hat{\boldsymbol{\theta}}_T$ and $\tilde{\boldsymbol{\theta}}_T$. Rearranging the terms in (9) and exploiting the fact that $\nabla_{\boldsymbol{\theta}} \boldsymbol{Q}_T(\tilde{\boldsymbol{\theta}}_T) = \nabla_{\boldsymbol{\theta}} \hat{\boldsymbol{Q}}_T(\hat{\boldsymbol{\theta}}_T) = 0$, we obtain $\sqrt{T}\left(\nabla \hat{\boldsymbol{Q}}_T(\hat{\boldsymbol{\theta}}_T) - \nabla \boldsymbol{Q}_T(\hat{\boldsymbol{\theta}}_T)\right) = \boldsymbol{A}_T(\boldsymbol{\theta}_T^{\star})\sqrt{T}\left(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T\right)$.

By Lemma TA.8, the left hand side converges to $\mathbf{0}$ almost surely. Lemma TA.9 together with $\boldsymbol{\theta}_T^{\star} \xrightarrow{P} \boldsymbol{\theta}_0$ imply that $\boldsymbol{A}_T(\boldsymbol{\theta}_T^{\star}) \xrightarrow{P} \boldsymbol{A}(\boldsymbol{\theta}_0)$ as $T \to \infty$. Therefore, we have $\sqrt{T}\left\|\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T\right\| \xrightarrow{P} 0$ as $T \to \infty$, which completes the proof. ■

# Acknowledgements

F. Blasques is thankful to the Dutch Science Foundation (NWO) for financial support (VI.Vidi.195.099).

# SUPPLEMENTARY APPENDIX

A Multilevel Factor Model

with Observation-Driven Dynamics

Mariia Artemova, Francisco Blasques, Siem Jan Koopman

# A Data

| Group | Group abbreviation | $N$ | NAICS | Group | Group abbreviation | $N$ | NAICS |
|---|---|---|---|---|---|---|---|
| 1. Agriculture, forestry, fishing, and hunting | AFFH | 2 | 11 | 7. Finance, insurance, real estate, rental, and leasing | Fin | 6 | 52-53 |
| 2. Construction | Constr | 1 | 23 | 8. Professional and business services | PBS | 6 | 54-56 |
| 3. Wholesale trade | WT | 1 | 42 | 9. Educational services, health care, and social assistance | EHS | 4 | 61-62 |
| 4. Retail trade | RT | 1 | 44-45 | 10. Arts, entertainment, recreation, accommodation, and food services | AER | 5 | 71-72, 81 |
| 5. Transportation and warehousing | TW | 8 | 48-49 | 11. Government | Gov | 4 | 92 |
| 6. Information | Inform | 4 | 51 | | | | |
| Total: 11 groups, 42 series | | | | | | | |

Table A.1: **Non-Industrial Production groups and number ($N$) of series in each group.** The table presents the names of the groups together with their abbreviations as well as the number ($N$) of series in each group. The column NAICS code corresponds to the NAICS code that was used to define the groups. The abbreviation stated in the table is further used in the network analysis.

| Group | Group abbreviation | $N$ | NAICS | Group | Group abbreviation | $N$ | NAICS |
|---|---|---|---|---|---|---|---|
| 12. Logging | Log | 1 | 1133 | 24. Chemicals | Chem | 6 | 325 |
| 13. Mining | Min | 4 | 211, 212 | 25. Plastics and rubber products | Plast | 2 | 326 |
| 14. Support activities for mining | SAM | 1 | 213 | 26. Nonmetallic mineral product | NMP | 5 | 327 |
| 15. Utilities | Util | 2 | 221 | 27. Primary metals | PM | 4 | 331 |
| 16. Food | Food | 9 | 311 | 28. Fabricated metal product | FMP | 9 | 332 |
| 17. Beverage and tobacco | BT | 2 | 312 | 29. Machinery | Mach | 6 | 333 |
| 18. Textile mills and textile product mills | TM | 5 | 313, 314 | 30. Computer and electronic product | CE | 6 | 334 |
| 19. Apparel, leather and allied products | AL | 2 | 315, 316 | 31. Electrical equipment, appliance, and component | EE | 4 | 335 |
| 20. Wood products | Wood | 3 | 321 | 32. Transportation equipment | TE | 7 | 336 |
| 21. Paper | Paper | 2 | 322 | 33. Furniture and related product | Furnit | 2 | 337 |
| 22.Printing and related support activ. | Print | 1 | 323 | 34. Miscellaneous | Miscel | 2 | 339 |
| 23. Petroleum and coal prod. | Petrol | 1 | 324 | 35. Newspaper, periodical, book, and directory publishers | Newsp | 1 | 5111 |
| Total: 24 groups, 87 series | | | | | | | |

Table A.2: **Industrial Production groups and number ($N$) of series in each group.** Most of the groups were defined based on the three-digit NAICS level except Logging and Newspapers groups for which only four-digit level is available according to the Board of Governors of the Federal Reserve System. Several groups were further united roughly according to their appearance in the input-output table. Particularly, we unite the following groups: Mining = Oil and gas extraction + Mining (excluding Oil and gas extraction); Utilities=Electric power generation, transmission and distribution+Natural gas distribution; Textile mills+Textile product mills; Apparel+Leather and allied product. For further explanations, we refer to Table A.1 in Supplementary Appendix.

| Group abbreviation | Sector | NAICS code | Weight |
|---|---|---|---|
| Log | Logging | 1133 | 0.24 |
| Min | Oil and gas extraction | 211 | 6.5 |
| Min | Coal mining | 2121 | 1.06 |
| Min | Metal ore mining | 2122 | 0.4 |
| Min | Nonmetallic mineral mining and quarrying | 2123 | 0.65 |
| SAM | Support activities for mining | 213 | 1.21 |
| Util | Electric power generation, transmission, and distribution | 2211 | 8.06 |
| Util | Natural gas distribution | 2212 | 1.61 |
| Food | Animal food | 3111 | 0.43 |
| Food | Grain and oilseed milling | 3112 | 0.8 |
| Food | Sugar and confectionery product | 3113 | 0.53 |
| Food | Fruit and vegetable preserving and specialty food | 3114 | 1.03 |
| Food | Dairy product | 3115 | 0.83 |
| Food | Animal slaughtering and processing | 3116 | 1.34 |
| Food | Seafood product preparation and packaging | 3117 | 0.14 |
| Food | Bakeries and tortilla | 3118 | 1.21 |
| Food | Other food | 3119 | 1.21 |
| BT | Beverage | 3121 | 1.34 |
| BT | Tobacco | 3122 | 1.13 |
| TM | Fiber, yarn, and thread mills | 3131 | 0.19 |
| TM | Fabric mills | 3132 | 0.58 |
| TM | Textile and fabric finishing and fabric coating mills | 3133 | 0.26 |
| TM | Textile furnishings mills | 3141 | 0.33 |
| TM | Other textile product mills | 3149 | 0.21 |
| AL | Apparel | 315 | 1.47 |
| AL | Leather and allied product | 316 | 0.26 |
| Wood | Sawmills and wood preservation | 3211 | 0.39 |
| Wood | Veneer, plywood, and engineered wood product | 3212 | 0.31 |
| Wood | Other wood product | 3219 | 0.68 |
| Paper | Pulp, paper, and paperboard mills | 3221 | 1.63 |
| Paper | Converted paper product | 3222 | 1.43 |
| Print | Printing and related support activities | 323 | 2.28 |
| Petrol | Petroleum and coal products | 3241 | 2.11 |
| Chem | Basic chemical | 3251 | 2.34 |
| Chem | Resin, synthetic rubber, and synthetic fibers | 3252 | 1.1 |
| Chem | Pesticide, fertilizer, and other agricultural chemical | 3253 | 0.48 |
| Chem | Pharmaceutical and medicine | 3254 | 2.89 |
| Chem | Paints, coating, and adhesive | 3255 | 0.53 |

Table A.3: **Industrial Production sectors** (I/II).

| Group abbreviation | Sector | NAICS code | Weight |
|---|---|---|---|
| Chem | Soap, cleaning compound, and toilet preparation | 3256 | 2.35 |
| Plast | Plastics product | 3261 | 2.39 |
| Plast | Rubber product | 3262 | 0.76 |
| NMP | Clay product and refactory | 3271 | 0.24 |
| NMP | Glass and glass product | 3272 | 0.58 |
| NMP | Cement and concrete product | 3273 | 0.84 |
| NMP | Lime and gypsum product | 3274 | 0.11 |
| NMP | Other nonmetallic mineral product | 3279 | 0.37 |
| PM | Iron and steel products | 3311 | 1.47 |
| PM | Alumina and aluminum production and processing | 3313 | 0.46 |
| PM | Nonferrous Metal (Except Aluminum) Production and Processing | 3314 | 0.47 |
| PM | Foundries | 3315 | 0.7 |
| FMP | Forging and stamping | 3321 | 0.49 |
| FMP | Cutlery and handtool | 3322 | 0.32 |
| FMP | Architectural and structural metals | 3323 | 1.14 |
| FMP | Boiler, Tank, and Shipping Containers | 3324 | 0.53 |
| FMP | Hardware | 3325 | 0.26 |
| FMP | Spring and wire product | 3326 | 0.19 |
| FMP | Machine shops, turned product, and screw, nut, and bolt | 3327 | 1.07 |
| FMP | Coating, engraving, heat treating, and allied activities | 3328 | 0.42 |
| FMP | Other fabricated metal product | 3329 | 1.28 |
| Mach | Agriculture, construction, and mining machinery | 3331 | 1.17 |
| Mach | Industrial machinery | 3332 | 0.69 |
| Mach | Commercial and service industry machinery and other general purpose machinery | 3333 | 2.09 |
| Mach | Ventilation, heating, air-conditioning, and commercial refrigeration equipment | 3334 | 0.68 |
| Mach | Metalworking machinery | 3335 | 0.77 |
| Mach | Engine, turbine, and power transmission equipment | 3336 | 0.71 |
| CE | Computer and peripheral equipment | 3341 | 1.51 |
| CE | Communications equipment | 3342 | 1.39 |
| CE | Audio and video equipment | 3343 | 0.15 |
| CE | Semiconductor and other electronic component | 3344 | 2.53 |
| CE | Navigational, measuring, electromedical, and control instruments | 3345 | 2.56 |
| CE | Magnetic and Optical Media | 3346 | 0.18 |
| EE | Electric lighting equipment | 3351 | 0.32 |
| EE | Household appliances | 3352 | 0.46 |
| EE | Electrical equipment | 3353 | 0.83 |
| EE | Other electrical equipment and component | 3359 | 0.82 |
| TE | Motor vehicle | 3361 | 2.49 |
| TE | Motor vehicle body and trailer | 3362 | 0.39 |
| TE | Motor vehicle parts | 3363 | 2.92 |
| TE | Aerospace product and parts | 3364 | 3.21 |
| TE | Railroad rolling stock | 3365 | 0.19 |
| TE | Ship and boat building | 3366 | 0.51 |
| TE | Other transportation equipment | 3369 | 0.18 |
| Furnit | Household and institutional furniture and kitchen cabinet | 3371 | 0.85 |
| Furnit | Office and other furniture | 3372,9 | 0.63 |
| Miscel | Medical equipment and supplies | 3391 | 1.32 |
| Miscel | Other Miscellaneous Manufacturing | 339 | 1.31 |
| Newsp | Newspaper, Periodical, Book, and Directory Publishers | 5111 | 3.55 |

Table A.4: **Industrial Production sectors** (II/II).

As mentioned in the main text, to obtain a fully quarterly dataset, we disaggregate the annual (non-IP) data into quarterly series using a mix of the Al-Osh (1989) and Silva & Cardoso (2001) methods. Here, we provide the details on the disaggregation methodology.

Overall, the methods for temporal disaggregation can be divided into two groups: methods that use some high-frequency related series, indicator variables, and smoothing approaches that do not rely on any indicator variable. The indicator variables exploit the fact that economic time series tend to co-move together. For example, due to economic events affecting them in the same way. Hence, temporal disaggregation methods with indicator variables estimate the unobserved sub-period values of the target series such that the short-term dynamics of the indicator variable is preserved and the temporal additivity constraint is satisfied.

Whenever an indicator variable is available we apply a dynamic Chow-Lin regression method as proposed in Silva & Cardoso (2001). In total, we find indicators for 21 out of 42 non-IP sectors. The list of the used indicators is presented in Table A.5. The choice of the indicator variables is based on a correlation between the target series and aggregated indicator, and the explanatory power of the indicator in the regression used for the disaggregation. The indicator variables are also chosen such that they are related to the target variable from the economic perspective. Overall, in many cases, we use growth rates of employed people in the corresponding group as an indicator variable; for the trade and transportation groups, we often use the growth rates of the real imports.

Since it is not always possible to find a good indicator for the disaggregation, we also use the method proposed in Al-Osh (1989), which is based on a linear dynamic model and ARMA representation of the unknown series corresponding to the disaggregated target variable. The measurement equation comes from the temporal additivity constraint, while the transition equation is based on ARMA representation, which is cast to a state space form. To choose the order of the ARMA model we follow the procedure proposed by Al-Osh

(1989). For the majority of the series we use AR(1) representation.

| Sector | Indicator (growth rates) | $\rho(\%)$ | $\bar{R}^2(\%)$ |
|---|---|---|---|
| Farms | Gross value added | 82.45 | 67 |
| Construction | All Employees | 83.6 | 69 |
| Wholesale trade | Imports | 52.37 | 25.2 |
| Retail trade | Imports<br>Real PCE | 70<br>73 | 58 |
| Air transportation | Average Weekly Hours<br>Exports<br>Real PCE: Transport. services | 51.96<br>46.1<br>46.4 | 42 |
| Rail transportation | Exports<br>Imports | 54.7<br>50.2 | 33.5 |
| Truck transportation | Average Weekly Hours<br>Employment/Population (Men) | 32.4<br>80 | 67 |
| Other transport.&support activ. | Imports<br>Real PCE: Transport. serv.<br>Employm./Popul. (Men) | 64.77<br>40.5<br>69.1 | 57.8 |
| Warehousing and storage | Imports<br>Production and Nonsuperv. Employees | 71.67<br>66.23 | 53.2 |
| Legal services | Real PCE: Services | 62.47 | 37.2 |
| Computer systems design and related services | All Employees | 66.84 | 43 |
| Miscel. profes., scientific, and technical services | Employment/Population | 63.25 | 38.1 |
| Administrative and support services | All Employees<br>Employment/Population | 75.4<br>76.4 | 60 |
| Ambulatory health care services | Real PCE: Household Consumption Expenditures: Health Care | 54.65 | 27.7 |
| Performing arts, spectator sports, museums | All Employees | 52.8 | 25.7 |
| Accommodation | Employment/Population (Men) | 78.9 | 61 |
| Food services and drinking places | All Employees | 65.5 | 41.2 |
| Other services, except government | All Employees | 71.98 | 50.4 |
| General government (FEDERAL) | All Employees | 69.27 | 46.4 |
| Government enterprises (FEDERAL) | Population<br>Employment/Population | 46.1<br>52.7 | 44.4 |
| General government (States and Local) | All Employees | 65.5 | 41.1 |

Table A.5: **Indicators used in the dynamic Chow-Lin regression for disaggregation of the non-IP sectoral growth rates.** All indicator variables are seasonally adjusted and were collected using the FRED economic database. The series all employees, average weekly hours, and production and nonsupervisory employees correspond to the particular group. For example, the series All Employees used for the disaggregation of the Administrative and support services sector series is All Employees in Professional and business services group.

# B Additional empirical results

## B.1 Factors and loadings

This section contains additional empirical results for the estimated factors and loadings. Particularly, in Figure B.1, we demonstrate the estimated loadings for all the IP and non-IP sectors. We find that most of the loadings on the common factor are positive. Hence, most of the sectors respond procyclically to the aggregate shocks. The largest negative loadings correspond, for example, to the General government (federal) (Gov group) and Federal reserve banks, credit intermediation, and related activities sectors (Fin group). Intuitively, we expect these sectors to respond counter-cyclically, because, for example, during crises they reflect the adoption of stimulus packages. Nevertheless, we find that the $R^2$ for all the sectoral series with the negative loadings on the common factor are low, hence aggregate shocks do not play important role in explaining the variability in them.

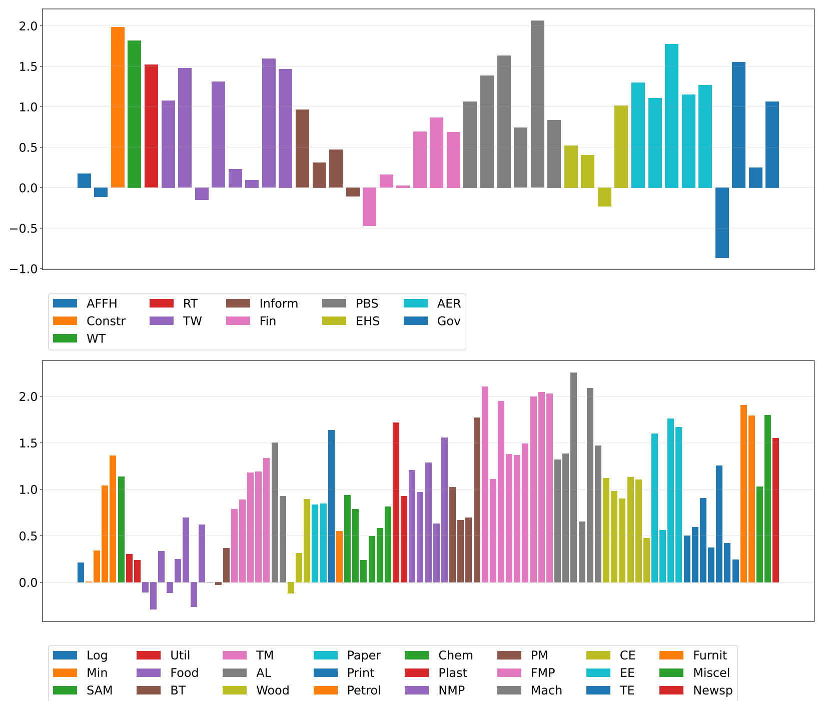



Figure B.1: **Estimated loadings on the common factor: non-IP sectors (top) and IP sectors (bottom).** The color of the bar indicates the group that the series belong to. The groups' name abbreviations are provided in Tables A.1 and A.2.

Once the group-specific factors and respective loadings have been estimated, we can investigate the importance of the group-specific factors in explaining the variability of the sectoral data. Specifically, to rank the sectors based on the importance of the group-specific factor, we compute the increments in $R^2$ (Table B.6). The increments are defined as a difference between the $R^2$ calculated after Step IV of the estimation procedure, hence when regressed on both the common and corresponding group-specific factors, and the $R^2$ when regressed only on the common factor, thus $R^2$ after Step II.

| IP sectors | $\Delta R^2(\%)$ | non-IP sectors | $\Delta R^2(\%)$ |
|---|---|---|---|
| ***Ten sectors with largest increment in $R^2$*** | | | |
| Support activities for mining | 29.878 | Hospitals & nursing & residen. care facil. | 36.066 |
| Apparel | 15.456 | Forestry, fishing, & related activ. | 29.162 |
| Leather & allied prod. | 14.769 | Wholesale trade | 28.845 |
| Medical equipment & supplies | 12.941 | General gov. (States & Local) | 27.822 |
| Textile & fabric finishing and coating mills | 10.374 | Motion picture & sound record. industr. | 27.341 |
| Newspaper, period., book & direct. publishers | 10.032 | Water transportation | 26.933 |
| Computer & peripheral equipm. | 9.97 | Rent. & leasing services | 26.634 |
| Basic chemical | 9.549 | Construction | 24.868 |
| Metalworking machinery | 9.470 | Management of companies & enterprises | 22.076 |
| Printing & related support activ. | 8.658 | Miscel. prof., scientific & technical services | 21.714 |
| ***Ten sectors with smallest increment in $R^2$*** | | | |
| Oil & gas extraction | 0.055 | Insurance carriers and related activities | 2.538 |
| Converted paper product | 0.051 | Rail transportation | 2.470 |
| Nonmetal. mineral mining & quarr. | 0.043 | Food services & drinking places | 2.166 |
| Ventilat.,heating,air-condition. | 0.041 | Warehousing & storage | 2.060 |
| Pharmaceut. and medicine | 0.015 | Other services, except gov. | 1.675 |
| Animal slaught. and processing | 0.015 | Other transportation & support activ. | 1.635 |
| Metal ore | 0.005 | Accommodation | 1.291 |
| Railroad rolling stock | 0.001 | Truck transportation | 1.002 |
| Ship & boat building | 0.001 | Air transportation | 0.096 |
| Coal mining | 0.000 | Funds, trusts, & other financial vehi. | 0.075 |

Table B.6: **Regression of the sectoral growth rates on the common and group-specific factors.** We demonstrate the top and bottom ten ranked IP and non-IP-sectors according to the increments in the $R^2$. The increment is computed as a difference between the $R^2$ of the regression of the sectoral growth rate on the common and corresponding group-specific factors and the $R^2$ of the regression of the sectoral growth rate on only the common factor.

We find that the increments are larger for the non-IP sectors, hence the group-specific factors explain more variation in the non-IP sectors. Therefore, the common factor is more related to IP sectors while the group-specific factor dynamics are more important for non-IP sectors. Overall, the largest increments in the $R^2$ appear for Hospitals and nursing and residential care facilities, Support activities for mining, Forestry, fishing and related activities, and Wholesale trade sectors. We note that many of the top ten ranked non-IP sectors had low $R^2$ when regressed only on the common factor. Therefore, these sectors are mostly related to the group-specific factors. The exceptions are Construction, Wholesale trade and Miscellaneous professional, scientific and technical service sectors, which have both high $R^2$ when regressed only on the common factor and large increments, hence both factors are important for these sectors.

In Figure B.2, we also demonstrate that most of the loadings on the group-specific factors are positive, especially for the non-IP sectors, therefore the majority of the sectors co-move within the group. Moreover, the loadings are larger for the non-IP sectors than for the IP sectors. We also find that the non-IP group specific factors have on average larger variation than the common factor while for the IP sectors we observe the opposite (Table B.7).

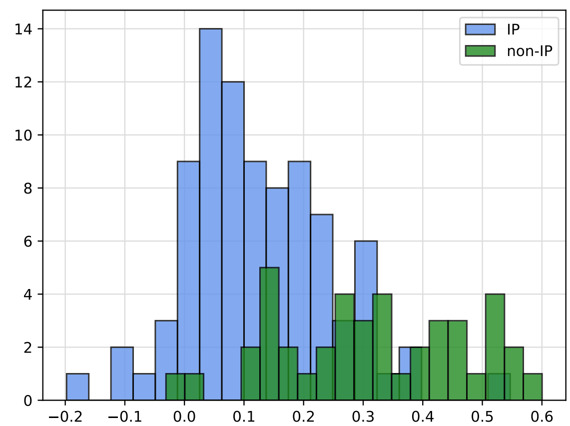



Figure B.2: **Histogram of the estimated loadings on the group-specific factor.** For comparison of the loadings between the different groups, we report the rescaled loadings where the scale is equal to the standard deviation of the corresponding group-specific factor.

| | Common factor | IP factors | non-IP factors |
|---|---|---|---|
| All | 0.104 | 0.044 | 0.155 |
| Top 3 | - | 0.195 | 0.259 |
| Top 5 | - | 0.146 | 0.227 |

Table B.7: **Variance of the factors.** We report the sample variance of the common factor, the average of the sample variances of the group-specific factors that correspond to either IP sectors or non-IP sectors. We also report the average of the variances among the top 3 and top 5 sectors with the largest variance among the IP and non-IP sectors.

## B.2 Model specification

Before proceeding with the granularity and network analysis we examine the residuals obtained after Step IV of the estimation procedure to make sure that they correspond to proper uncorrelated 'innovations' in the time-series sense. From the standard diagnostics we find the traces of significant autocorrelation in the residuals after the extraction of the dynamic factors. This means that further idiosyncratic dynamics is present in the data. To account for this we additionally model the residuals as AR(1) processes. The coefficients of the estimated AR(1) model are shown in Figure B.3. We find that coefficients for the non-IP sectors are typically larger than for the IP sectors. Once we fit the AR(1) models on the residuals we test again for the presence of the autocorrelation. The results of the multivariate autocorrelation test (Lütkepohl, 2005) are presented in Table B.8. We find no evidence of autocorrelation up to lag 9. Therefore, we model residuals using AR(1) model. Additionally, we test whether the residuals are Gaussian since our (G)IRF analysis is based on this assumption. However, we reject the hypothesis of the normally distributed innovations. Therefore, we stress that we should interpret our network empirical results with cautious.

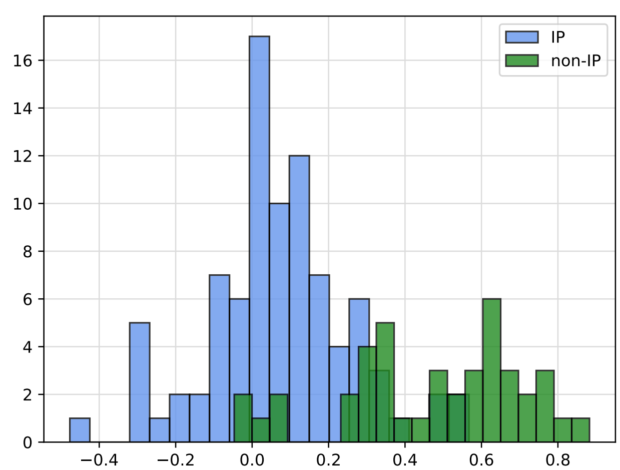



Figure B.3: **Histogram of the estimated coefficients of the AR(1) models.** The univariate AR(1) models were estimated using the residuals after Step IV.

| N of lags | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $p$-value | 0.325 | 0.277 | 0.279 | 0.260 | 0.260 | 0.262 | 0.273 | 0.265 | 0.273 |

Table B.8: **Results of the multivariate autocorellation test.**

# C Forecasts, IRFs and connectedness

One-step-ahead forecasts ($h = 1$) directly follow from the updating equations since the factors are updated based on a mean reversion term and an autoregressive part: the update only uses current values of $f_T$, $\mathbf{g}_T$ and $\mathbf{y}_T$. Forecasts at longer horizons can be obtained by using predicted values in a recursive way. For example, $h$-step-ahead forecast for $\hat{\mathbf{y}}_{T+h}$ can be produced as follows:

$$\hat{\mathbf{y}}_{T+h} = \mathbb{E}[\mathbf{y}_{T+h}|\mathcal{F}_T] = \mathbf{\Lambda}^c\mathbb{E}[f_{T+h}|\mathcal{F}_T] + \mathbf{\Lambda}^g\mathbb{E}[\mathbf{g}_{T+h}|\mathcal{F}_T],$$

where $\mathbb{E}[f_{T+h}|\mathcal{F}_T]$ and $\mathbb{E}[\mathbf{g}_{T+h}|\mathcal{F}_T]$ are $h$-step-ahead predictions for the common and group-specific factors, respectively. Intuitively, using this model we can assess which part of the increase or decrease in $y_{i,t+h}$ is expected to be attributed to the common factor and/or to the group-specific factor.

In policy analysis it is also important to produce Impulse Response Functions (IRFs) that show how variables respond to the impulse in another variable (or group of variables). For example, if we are interested in the IRF of each variable due to a shock in one of the groups $s$, then numerically we set the size of the shock $\boldsymbol{\varepsilon}_t^s = \mathbf{1}_{N_s}/N_s$ (or proportionally to the standard deviations) and all other elements of the vector of innovations to zero. Additionally, we set $f_t = 0$ and $\mathbf{g}_t = \mathbf{0}$. Then we iteratively obtain impulse response function as a function of $h$ by generating $\mathbf{y}_{t+h}$ according to the system equations (1)–(3). Analytically, this procedure can be summarized as follows:

$$\frac{\partial \mathbf{y}_{t+h}}{\partial \boldsymbol{\varepsilon}_t^s} = \mathbb{E}[\mathbf{y}_{t+h}|\mathcal{F}_t, \boldsymbol{\varepsilon}_t^s] - \mathbb{E}[\mathbf{y}_{t+h}|\mathcal{F}_t],$$

where $\mathcal{F}_t$ denotes the information set up to time $t$ that contains observations before the shock occurrence.

When covariance matrix $\boldsymbol{\Sigma}$ of innovations is not diagonal, generalized impulse response (GIRF) analysis proposed by Koop et al. (1996) can be considered instead. Particularly, in our application we assume that in group $s$ an exogenous shock occurs with the size of the shock $\boldsymbol{\varepsilon}_t^s = \boldsymbol{\Sigma}_s^{1/2}\mathbf{e}_s$ where $\mathbf{e}_s$ is an $N_s \times 1$ vector such that $\mathbf{e}_s = (1/N_s, \ldots, 1/N_s)^\top$. When innovations are Gaussian contemporaneous responses in other variables can be obtained as follows:

$$\frac{\partial y_{i,t}^s}{\partial \boldsymbol{\varepsilon}_t^s} = \frac{\partial \varepsilon_{i,t}^s}{\partial \boldsymbol{\varepsilon}_t^s} = \boldsymbol{\Sigma}_{i,s}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\varepsilon}_t^s,$$

where $\boldsymbol{\Sigma}_{i,s}$ is an $N_s \times 1$ vector that contains covariance terms between sector $i$ and sectors in group $s$. For further horizons GIRFs are obtained according to system equations (1)–(3) as discussed above. The square root of the group $s$ covariance matrix $\boldsymbol{\Sigma}_s$ can be obtained using, for example, Cholesky or Schur decomposition.

In our application, we measure group interconnectedness using the generalized impulse response functions. Specifically, we use approach proposed in Diebold & Yılmaz (2014), but instead of variance decompositions we focus on (G)IRFs. We also modify the approach to account for the group rather than individual interconnectedness. Below we give the details about our group connectedness measures.

Assume that at time $t = 0$ a shock occurs in each of the innovations of variables in group $s$ and we are interested in how other groups respond to this shock. For this, we can compute (generalized) impulse responses of all variables due to a shock arising in one of the other groups as discussed above. We denote period $t$ response of variable $k$ due to a shock in group $s$ as $IRF_{k\leftarrow s,t}$ for $k = 1, \ldots, N$ and $s = 1, \ldots, S$. We are further interested in the connectedness between the groups, hence we denote average pairwise connectedness measures from group $s$ to group $i$ $t$ periods after the shock occurrence as $\theta_{is,t} := \frac{1}{N_i}\sum_{k=1}^{N_i} IRF_{k\leftarrow s,t}$, which is the average impulse responses of group $i$ due to a shock in group $s$. In Table C.9, we schematically illustrate the connectedness matrix, which consists of the average pairwise connectedness measures for each group.

The average pairwise connectedness measure will be uninformative about the strengths of the connections when the responses of the units within a group are of different signs. Therefore, if we are interested in the strengths of the links we denote the average absolute pairwise connectedness measure from group $s$ to $i$ as $\theta_{is,t} := \frac{1}{N_i}\sum_{k=1}^{N_i} |IRF_{k\leftarrow s,t}|$, which is not sensitive to the sign of the response.

| | Group 1 | Group 2 | $\cdots$ | Group $S$ | In-degree |
|---|---|---|---|---|---|
| Group 1 | $\theta_{11,t}$ | $\theta_{12,t}$ | $\cdots$ | $\theta_{1S,t}$ | $\sum_{j=1,j\neq 1}^{S} \theta_{1j,t}$ |
| Group 2 | $\theta_{21,t}$ | $\theta_{22,t}$ | $\cdots$ | $\theta_{2S,t}$ | $\sum_{j=1,j\neq 2}^{S} \theta_{2j,t}$ |
| | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | |
| Group $S$ | $\theta_{S1,t}$ | $\theta_{S2,t}$ | $\cdots$ | $\theta_{SS,t}$ | $\sum_{j=1,j\neq S}^{S} \theta_{Sj,t}$ |
| Out-degree | $\sum_{i=1,i\neq 1}^{S} \theta_{i1,t}$ | $\sum_{i=1,i\neq 2}^{S} \theta_{i2,t}$ | $\cdots$ | $\sum_{i=1,i\neq S}^{S} \theta_{iS,t}$ | |

Table C.9: **Connectedness table.** $\theta_{ij,t} := \frac{1}{N_i}\sum_{k=1}^{N_i} IRF_{k\leftarrow j,t}$ is a group average impulse response to a shock in sector $j$. $IRF_{k\leftarrow j,t}$ denotes impulse response in sector $k$ due to a shock arising in group $j$ $t$ periods after the shock occurrence.

To further identify the most central groups in the network we compute their centrality. Specifically, we use pairwise connectedness measures (in- and out-degree centrality measures) as in Diebold & Yılmaz (2014), which take into account the direction of the relations and the strength. In-degree centrality shows how the node responds to its neighbors, while out-degree centrality indicates how big is the response of the neighbors due to the impulse in the node (Table C.9). We use these measures in our application to summarize results from the impulse response analysis.

# D Technical Appendix

To shorten further notation in the proofs of Propositions 1 and 2, we introduce

$$\phi(\mathbf{y}_t) := \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s} y_{i,t}^s,$$

$$\psi(\mathbf{y}_t^s, f_t; \mathbf{\Lambda}_s^c) := \frac{1}{N_s}\sum_{i=1}^{N_s}\left(y_{i,t}^s - \lambda_{i,s}^c f_t\right),$$

then the factors' updating equations take the following form

$$f_{t+1} = \beta\phi(\mathbf{y}_t) + (\gamma - \beta) f_t, \tag{D.1}$$

$$g_{t+1}^s = \beta_s \psi(\mathbf{y}_t^s, f_t; \mathbf{\Lambda}_s^c) + (\gamma_s - \beta_s) g_t^s, \quad s = 1, \ldots, S. \tag{D.2}$$

We also denote the stochastic recurrence equations for the common factor as $f_{t+1} := \Phi(\mathbf{y}_t, f_t; \boldsymbol{\theta})$ and for the group-specific factors $g_t^s := \Psi(\mathbf{y}_t^s, f_t, g_t^s; \boldsymbol{\theta})$, $s = 1, \ldots, S$.

## D.1 Derivatives

### Derivatives of the time-varying parameters

In this section, we provide the analytical expressions for the derivatives of $f_t(\boldsymbol{\theta})$ and $\boldsymbol{g}_t(\boldsymbol{\theta})$ with respect to parameters $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$, $\boldsymbol{\theta}^{(3)}$, $\boldsymbol{\theta}^{(4)}$, where $\boldsymbol{\theta}^{(1)} \equiv \boldsymbol{\theta}^c$, $\boldsymbol{\theta}_s^{(2)} \equiv \mathbf{\Lambda}_s^c$, $\boldsymbol{\theta}_s^{(3)} \equiv \boldsymbol{\theta}_s^g$, $\boldsymbol{\theta}_s^{(4)} \equiv \mathbf{\Lambda}_s^g$ for $s = 1, \ldots, S$. These expressions are further required for the asymptotic covariance matrix of the estimators.

$$f'_{t+1}(\boldsymbol{\theta}) := \frac{\partial f_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(1)}} = \boldsymbol{\zeta}_f(\mathbf{y}_t, f_t(\boldsymbol{\theta})) + (\gamma - \beta) f'_t(\boldsymbol{\theta}), \tag{D.3}$$

$$f''_{t+1}(\boldsymbol{\theta}) := \frac{\partial^2 f_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(1)} \partial \boldsymbol{\theta}^{(1)\top}} = \boldsymbol{C}(f'_t(\boldsymbol{\theta})) + (\gamma - \beta) f''_t(\boldsymbol{\theta}), \tag{D.4}$$

where $\boldsymbol{\zeta}_f(\mathbf{y}_t, f_t) := (\phi(\mathbf{y}_t) - f_t(\boldsymbol{\theta}), f_t(\boldsymbol{\theta}))^\top$ and $\boldsymbol{C}(x(\boldsymbol{\theta})) := \left(x(\boldsymbol{\theta})[-1, 1] + (x(\boldsymbol{\theta})[-1, 1])^T\right)$.

We further introduce the following notation for the derivatives of $\boldsymbol{g}_t(\boldsymbol{\theta})$,

$$\begin{aligned}
\boldsymbol{g}'_{t+1}(\boldsymbol{\theta}) &:= \left( \left( \frac{\partial g^1_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(3)}_1} \right)^\top, \dots, \left( \frac{\partial g^S_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(3)}_S} \right)^\top \right)^\top, \\
\boldsymbol{g}''_{t+1}(\boldsymbol{\theta}) &:= \text{block-diag} \left( \left( \frac{\partial^2 g^1_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(3)}_1 \partial \boldsymbol{\theta}^{(3)\top}_1} \right)^\top, \dots, \left( \frac{\partial^2 g^S_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(3)}_S \partial \boldsymbol{\theta}^{(3)\top}_S} \right)^\top \right)^\top, \\
\frac{\partial \boldsymbol{g}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(2)}} &:= \left( \frac{\partial g^1_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(2)}_1}, \dots, \frac{\partial g^S_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(2)}_S} \right)^\top, \\
\frac{\partial \boldsymbol{g}_t(\boldsymbol{\theta})}{\partial f} &:= \left( \frac{\partial g^1_t(\boldsymbol{\theta})}{\partial f}, \dots, \frac{\partial g^S_t(\boldsymbol{\theta})}{\partial f} \right)^\top.
\end{aligned}$$

For $s = 1, \dots, S$ we have

$$g^{s\,\prime}_{t+1}(\boldsymbol{\theta}) := \frac{\partial g^s_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(3)}_s} = \boldsymbol{\zeta}_g(\mathbf{y}^s_t, f_t(\boldsymbol{\theta}), g^s_t(\boldsymbol{\theta}); \boldsymbol{\Lambda}^c_s) + (\gamma_s - \beta_s) g^{s\prime}_t(\boldsymbol{\theta}), \tag{D.5}$$

$$g^{s\,\prime\prime}_{t+1}(\boldsymbol{\theta}) := \frac{\partial^2 g^s_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(3)}_s \partial \boldsymbol{\theta}^{(3)\top}_s} = \boldsymbol{C}(g^{s\prime}_t(\boldsymbol{\theta})) + (\gamma_s - \beta_s) g^{s\prime\prime}_t(\boldsymbol{\theta}), \tag{D.6}$$

$$\frac{\partial g^s_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(1)}} = \frac{\partial g^s_{t+1}(\boldsymbol{\theta})}{\partial f} f'_{t+1}(\boldsymbol{\theta}),$$

$$\frac{\partial g^s_{t+1}(\boldsymbol{\theta})}{\partial f} = -\beta_s \frac{1}{N_s} \sum_{i=1}^{N_s} \lambda_{i,s} + (\gamma_s - \beta_s) \frac{\partial g^s_t(\boldsymbol{\theta})}{\partial f}, \tag{D.7}$$

$$\begin{aligned}
\frac{\partial g^s_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(2)}_s} &= -\beta_s \frac{1}{N_s} \boldsymbol{\iota}_{N_s} f_t(\boldsymbol{\theta}) + (\gamma_s - \beta_s) \frac{\partial g^s_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(2)}_s}, \\
\frac{\partial^2 g^s_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(3)}_s \boldsymbol{\theta}^{(1)\top}} &= [-1,\ 0]^\top \left( \frac{1}{N_s} \sum_{i=1}^{N_s} \lambda_{i,s} \right) f'_t(\boldsymbol{\theta})^\top + [-1,\ 1]^\top \frac{\partial g^s_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(1)\top}} + (\gamma_s - \beta_s) \frac{\partial^2 g^s_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(3)}_s \boldsymbol{\theta}^{(1)\top}}, \\
\frac{\partial^2 g^s_{t+1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(3)}_s \boldsymbol{\theta}^{(2)\top}_s} &= [-1,\ 0]^\top \frac{1}{N_s} f_t(\boldsymbol{\theta}) \boldsymbol{\iota}^\top_{N_s} + [-1,\ 1]^\top \frac{\partial g^s_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(2)\top}_s} + (\gamma_s - \beta_s) \frac{\partial^2 g^s_t}{\partial \boldsymbol{\theta}^{(3)}_s \boldsymbol{\theta}^{(2)\top}_s},
\end{aligned}$$

where $\boldsymbol{\zeta}_g(\mathbf{y}^s_t, f_t(\boldsymbol{\theta}), g^s_t(\boldsymbol{\theta}); \boldsymbol{\Lambda}^c_s) := (\psi(\mathbf{y}^s_t, f_t(\boldsymbol{\theta}); \boldsymbol{\Lambda}^c_s) - g^s_t(\boldsymbol{\theta}), g^s_t(\boldsymbol{\theta}))^\top$.

In practice, we approximate the derivates recursively, that is

$$\hat{f}'_{t+1}(\boldsymbol{\theta}) = \boldsymbol{\zeta}_f(\mathbf{y}_t, \hat{f}_t(\boldsymbol{\theta})) + (\gamma - \beta) \hat{f}'_t(\boldsymbol{\theta}),$$

where the filtered sequence depends on the initial values $\hat{f}_1$ and $\hat{f}'_1$. Similarly,

$$\hat{f}''_{t+1}(\boldsymbol{\theta}) = \boldsymbol{C}(\hat{f}'_t(\boldsymbol{\theta})) + (\gamma - \beta) \hat{f}''_t(\boldsymbol{\theta}).$$

Similar recursions follow for $\hat{\boldsymbol{g}}_t'(\boldsymbol{\theta})$, $\hat{\boldsymbol{g}}_t''(\boldsymbol{\theta})$, $\frac{\partial \boldsymbol{g}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(2)}}$, $\frac{\partial \boldsymbol{g}_t(\boldsymbol{\theta})}{\partial f}$, $\frac{\partial^2 g_{t+1}^s(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_s^{(3)} \boldsymbol{\theta}^{(1)\top}}$ and $\frac{\partial^2 g_{t+1}^s(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_s^{(3)} \boldsymbol{\theta}_s^{(2)\top}}$.

## Expressions for the derivatives of the criterion functions

Below, we provide the expressions for the derivatives of the stepwise criterion functions with respect to the corresponding parameter of interest. The superscript indexes correspond to the step of the estimation procedure. We also introduce the notation $\boldsymbol{\theta}^{(i:1)} := (\boldsymbol{\theta}^{(1)\top}, \ldots, \boldsymbol{\theta}^{(i)\top})^\top$.

First, we provide the expressions for the first-order derivatives:

$$\nabla_{\boldsymbol{\theta}^{(1)}} q_t^{(1)}(\boldsymbol{\theta}^{(1)}) := \frac{\partial q_t^{(1)}(\boldsymbol{\theta}^{(1)})}{\partial \boldsymbol{\theta}^{(1)}} = \frac{\partial q_t^{(1)}(\boldsymbol{\theta}^{(1)})}{\partial f} f_t'(\boldsymbol{\theta}^{(1)}), \tag{D.8}$$

$$\nabla_{\boldsymbol{\theta}^{(2)}} q_t^{(2)}(\boldsymbol{\theta}^{(2:1)}) := \frac{\partial q_t^{(2)}(\boldsymbol{\theta}^{(2:1)})}{\partial \boldsymbol{\theta}^{(2)}} = -2 f_t(\boldsymbol{\theta}^{(1)})(\mathbf{y}_t - \boldsymbol{\Lambda}^c f_t(\boldsymbol{\theta}^{(1)})), \tag{D.9}$$

$$\nabla_{\boldsymbol{\theta}_s^{(3)}} q_t^{(3)}(\boldsymbol{\theta}^{(3:1)}) := \frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial \boldsymbol{\theta}_s^{(3)}} = \frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial g^s} g_t^{s\prime}(\boldsymbol{\theta}^{(3:1)}),$$

$$\nabla_{\boldsymbol{\theta}_s^{(4)}} q_t^{(4)}(\boldsymbol{\theta}) := \frac{\partial q_t^{(4)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_s^{(4)}} = -2(\mathbf{y}_t^s - \boldsymbol{\Lambda}_s^c f_t(\boldsymbol{\theta}^{(1)}) - \boldsymbol{\Lambda}_s^g g_t^s(\boldsymbol{\theta}^{(3:1)})) g_t^s(\boldsymbol{\theta}^{(3:1)}), \tag{D.10}$$

$$\frac{\partial q_t^{(1)}(\boldsymbol{\theta}^{(1)})}{\partial f} = -2\left(\frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s} y_{i,t}^s - f_t(\boldsymbol{\theta}^{(1)})\right), \tag{D.11}$$

$$\frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial g^s} = -2\left(\frac{1}{N_s}\sum_{i=1}^{N_s} y_{i,t}^s - \frac{1}{N_s}\sum_{i=1}^{N_s} \lambda_{i,s}^c f_t(\boldsymbol{\theta}^{(1)}) - g_t^s(\boldsymbol{\theta}^{(3:1)})\right), \tag{D.12}$$

with $f_t'(\boldsymbol{\theta})$ and $g_t^{s\prime}(\boldsymbol{\theta})$ as defined in equations (D.3) and (D.5), respectively. We further introduce the notation

$$\frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial \boldsymbol{g}} := \left(\frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial g^1}, \ldots, \frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial g^S}\right)^\top,$$

$$\nabla_{\boldsymbol{\theta}^{(3)}} q_t^{(3)}(\boldsymbol{\theta}^{(3:1)}) := \left(\left(\frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial g^1} g_t^{1\prime}(\boldsymbol{\theta}^{(3:1)})\right)^\top, \ldots, \left(\frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial g^S} g_t^{S\prime}(\boldsymbol{\theta}^{(3:1)})\right)^\top\right)^\top,$$

$$\nabla_{\boldsymbol{\theta}^{(4)}} q_t^{(4)}(\boldsymbol{\theta}) := \left(\left(\nabla_{\boldsymbol{\theta}_1^{(4)}} q_t^{(4)}(\boldsymbol{\theta})\right)^\top, \ldots, \left(\nabla_{\boldsymbol{\theta}_S^{(4)}} q_t^{(4)}(\boldsymbol{\theta})\right)^\top\right)^\top.$$

We turn to the second-order derivatives. We have

$$\nabla_{\boldsymbol{\theta}^{(1)}\boldsymbol{\theta}^{(1)}} q_t^{(1)}(\boldsymbol{\theta}^{(1)}) := \frac{\partial^2 q_t^{(1)}(\boldsymbol{\theta}^{(1)})}{\partial \boldsymbol{\theta}^{(1)} \partial \boldsymbol{\theta}^{(1)\top}} = 2 f_t'(\boldsymbol{\theta}^{(1)}) f_t'(\boldsymbol{\theta}^{(1)})^\top - 2(\phi(\mathbf{y}_t) - f_t(\boldsymbol{\theta}^{(1)})) f_t''(\boldsymbol{\theta}^{(1)}), \tag{D.13}$$

$$\nabla_{\boldsymbol{\theta}^{(2)}\boldsymbol{\theta}^{(1)}} q_t^{(2)}(\boldsymbol{\theta}^{(1)}) := \frac{\partial^2 q_t^{(2)}(\boldsymbol{\theta}^{(2:1)})}{\partial \boldsymbol{\theta}^{(2)} \partial \boldsymbol{\theta}^{(1)\top}} = \left(-2\mathbf{y}_t + 4\boldsymbol{\Lambda}^c f_t(\boldsymbol{\theta}^{(1)})\right) f_t'(\boldsymbol{\theta}^{(1)})^\top, \tag{D.14}$$

$$\nabla_{\boldsymbol{\theta}^{(2)}\boldsymbol{\theta}^{(2)}} q_t^{(2)}(\boldsymbol{\theta}^{(2:1)}) := \frac{\partial^2 q_t^{(2)}(\boldsymbol{\theta}^{(2:1)})}{\partial \boldsymbol{\theta}^{(2)} \partial \boldsymbol{\theta}^{(2)\top}} = 2 f_t^2(\boldsymbol{\theta}^{(1)}) \boldsymbol{I}_N. \tag{D.15}$$

We further denote $\nabla_{\boldsymbol{\theta}^{(i)}\boldsymbol{\theta}^{(k)}} q_t^{(i)}(\boldsymbol{\theta}^{(i:1)}) :=$ block-diag $\left(\nabla_{\boldsymbol{\theta}_1^{(i)}\boldsymbol{\theta}_1^{(k)}} q_t^{(i)}(\boldsymbol{\theta}^{(i:1)}), \ldots, \nabla_{\boldsymbol{\theta}_S^{(i)}\boldsymbol{\theta}_S^{(k)}} q_t^{(i)}(\boldsymbol{\theta}^{(i:1)})\right)$ for $i = 3, 4$ and $k \leq i$. Then, for $s = 1, \ldots, S$ we have

$$\nabla_{\boldsymbol{\theta}_s^{(3)}\boldsymbol{\theta}^{(1)}} q_t^{(3)}(\boldsymbol{\theta}^{(3:1)}) := \frac{\partial^2 q_t^{(3)}\left(\boldsymbol{\theta}^{(3:1)}\right)}{\partial \boldsymbol{\theta}_s^{(3)} \partial \boldsymbol{\theta}^{(1)\top}} = 2\left(\frac{1}{N_s}\sum_{i=1}^{N_s} \lambda_{i,s} + \frac{\partial g_t^s(\boldsymbol{\theta}^{(3:1)})}{\partial f}\right) g_t^{s\prime}(\boldsymbol{\theta}^{(3:1)}) f_t^{\prime}(\boldsymbol{\theta}^{(1)})^\top$$

$$- 2(\psi(\mathbf{y}_t^s, f_t(\boldsymbol{\theta}^{(1)}), \boldsymbol{\Lambda}_s^c) - g_t^s(\boldsymbol{\theta}^{(3:1)}))\frac{\partial^2 g_t^s(\boldsymbol{\theta}^{(3:1)})}{\partial \boldsymbol{\theta}_s^{(3)} \partial \boldsymbol{\theta}^{(1)\top}}, \tag{D.16}$$

$$\nabla_{\boldsymbol{\theta}_s^{(3)}\boldsymbol{\theta}_s^{(2)}} q_t^{(3)}(\boldsymbol{\theta}^{(3:1)}) := \frac{\partial^2 q_t^{(3)}\left(\boldsymbol{\theta}^{(3:1)}\right)}{\partial \boldsymbol{\theta}_s^{(3)} \partial \boldsymbol{\theta}_s^{(2)\top}} = 2 g_t^{s\prime}(\boldsymbol{\theta}^{(3:1)}) \left(\frac{1}{N_s} f_t(\boldsymbol{\theta}^{(1)}) \boldsymbol{\iota}_{N_s} + \frac{\partial g_t^s(\boldsymbol{\theta}^{(3:1)})}{\partial \boldsymbol{\theta}_s^{(2)}}\right)^\top$$

$$- 2(\psi(\mathbf{y}_t^s, f_t(\boldsymbol{\theta}^{(1)}), \boldsymbol{\Lambda}_s^c) - g_t^s(\boldsymbol{\theta}^{(3:1)}))\frac{\partial^2 g_t^s(\boldsymbol{\theta}^{(3:1)})}{\partial \boldsymbol{\theta}_s^{(3)} \partial \boldsymbol{\theta}_s^{(2)\top}}, \tag{D.17}$$

$$\nabla_{\boldsymbol{\theta}_s^{(3)}\boldsymbol{\theta}_s^{(3)}} q_t^{(3)}(\boldsymbol{\theta}^{(3:1)}) := \frac{\partial^2 q_t^{(3)}\left(\boldsymbol{\theta}^{(3:1)}\right)}{\partial \boldsymbol{\theta}_s^{(3)} \partial \boldsymbol{\theta}_s^{(3)\top}}$$

$$= 2\left(g_t^{s\prime}(\boldsymbol{\theta}^{(3:1)}) g_t^{s\prime}(\boldsymbol{\theta}^{(3:1)})^\top\right) - 2(\psi(\mathbf{y}_t^s, f_t(\boldsymbol{\theta}^{(1)}), \boldsymbol{\Lambda}_s^c) - g_t^s(\boldsymbol{\theta}^{(3:1)})) g_t^{s\prime\prime}(\boldsymbol{\theta}^{(3:1)}), \tag{D.18}$$

$$\nabla_{\boldsymbol{\theta}_s^{(4)}\boldsymbol{\theta}^{(1)}} q_t^{(4)}(\boldsymbol{\theta}) := \frac{\partial^2 q_t^{(4)}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_s^{(4)} \partial \boldsymbol{\theta}^{(1)\top}}$$

$$= 2\boldsymbol{\Lambda}_s^c (f_t^{\prime}(\boldsymbol{\theta}^{(1)}))^\top g_t^s(\boldsymbol{\theta}^{(3:1)}) - 2(\mathbf{y}_t^s - \boldsymbol{\Lambda}_s^c f_t(\boldsymbol{\theta}^{(1)}) - 2\boldsymbol{\Lambda}_s^g g_t^s(\boldsymbol{\theta}^{(3:1)})) \left(\frac{\partial g_t^s(\boldsymbol{\theta}^{(3:1)})}{\partial \boldsymbol{\theta}^{(1)}}\right)^\top, \tag{D.19}$$

$$\nabla_{\boldsymbol{\theta}_s^{(4)}\boldsymbol{\theta}_s^{(2)}} q_t^{(4)}(\boldsymbol{\theta}^{(3:1)}) := \frac{\partial^2 q_t^{(4)}\left(\boldsymbol{\theta}^{(3:1)}\right)}{\partial \boldsymbol{\theta}_s^{(4)} \partial \boldsymbol{\theta}_s^{(2)\top}}$$

$$= 2 g_t^s(\boldsymbol{\theta}^{(3:1)}) f_t(\boldsymbol{\theta}^{(1)}) \boldsymbol{I}_{N_s} - 2(\mathbf{y}_t^s - \boldsymbol{\Lambda}_s^c f_t(\boldsymbol{\theta}^{(1)}) - 2\boldsymbol{\Lambda}_s^g g_t^s(\boldsymbol{\theta}^{(3:1)})) \frac{\partial g_t^s}{\partial \boldsymbol{\theta}_s^{(2)}}^\top, \tag{D.20}$$

$$\nabla_{\boldsymbol{\theta}_s^{(4)}\boldsymbol{\theta}_s^{(3)}} q_t^{(4)}(\boldsymbol{\theta}^{(3:1)}) := \frac{\partial^2 q_t^{(4)}\left(\boldsymbol{\theta}^{(3:1)}\right)}{\partial \boldsymbol{\theta}_s^{(4)} \partial \boldsymbol{\theta}_s^{(3)\top}}$$

$$= -2(\mathbf{y}_t^s - \boldsymbol{\Lambda}_s^c f_t(\boldsymbol{\theta}^{(1)}) - 2\boldsymbol{\Lambda}_s^g g_t^s(\boldsymbol{\theta}^{(3:1)})) g_t^{s\prime}(\boldsymbol{\theta}^{(3:1)})^\top, \tag{D.21}$$

$$\nabla_{\boldsymbol{\theta}_s^{(4)}\boldsymbol{\theta}_s^{(4)}} q_t^{(4)}(\boldsymbol{\theta}) := \frac{\partial q_t^{(4)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_s^{(4)} \partial \boldsymbol{\theta}_s^{(4)\top}} = 2(g_t^s(\boldsymbol{\theta}^{(3:1)}))^2 \boldsymbol{I}_{N_s}, \tag{D.22}$$

$$\frac{\partial^2 q_t^{(1)}(\boldsymbol{\theta}^{(1)})}{\partial f^2} = 2, \frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial g^{s2}} = 2, \tag{D.23}$$

where $\boldsymbol{I}_N$ denotes an $N \times N$ identity matrix.

## D.2 Proofs and other technical lemmas

*Proof of Proposition 1:* Given the linearity of the updating equation (D.1), iterating it backwards for the filter $\hat{f}_t$ initialized at some value $\hat{f}_1 \in \mathbb{R}$ we have

$$\hat{f}_{t+1}(\boldsymbol{\theta}) = \sum_{i=0}^{t-1} (\gamma - \beta)^i \beta \phi(\mathbf{y}_{t-i}) + (\gamma - \beta)^t \hat{f}_1. \tag{D.24}$$

Proposition 1 states that the effect of the initialization $\hat{f}_1$ asymptotically vanishes as $t \to \infty$ and the filter $\hat{f}_t$ converges to the limit $f_t$ which, if exists, is then defined as,

$$f_{t+1}(\boldsymbol{\theta}) = \sum_{i=0}^{\infty} (\gamma - \beta)^i \beta \phi(\mathbf{y}_{t-i}). \tag{D.25}$$

First, we establish the stochastic properties of the limit sequence. By monotone convergence theorem and triangle inequality, we have

$$\mathbb{E} \sum_{i=0}^{\infty} \left| (\gamma - \beta)^i \beta \phi(\mathbf{y}_{t-i}) \right| \leq \sum_{i=0}^{\infty} |\gamma - \beta|^i |\beta| \mathbb{E} |\phi(\mathbf{y}_{t-i})| < \infty, \tag{D.26}$$

where the inequality follows by Lemma 2.1 in Straumann & Mikosch (2006) since $|\gamma - \beta| < 1$ and the sequence $\{|\mathbf{y_t}|\}_{t \in \mathbb{Z}}$ is SE with $\mathbb{E}\|\mathbf{y}_t\|^2 < \infty$, which is implied by Assumptions 2 and 3. Therefore, the series $\sum_{i=0}^{\infty} \left| (\gamma - \beta)^i \beta \phi(\mathbf{y}_{t-i}) \right|$ is finite almost surely which implies that $\sum_{i=0}^{\infty} (\gamma - \beta)^i \beta \phi(\mathbf{y}_{t-i})$ converges almost surely. From Krengel's theorem it then follows that the limit sequence $\{f_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ is strictly stationary and ergodic since it is a measurable function of $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ which by Assumption 2 is strictly stationary and ergodic sequence. From Pötscher & Prucha (1997, Theorem 6.10) it also follows that $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |f_t(\boldsymbol{\theta})|^2 < \infty$.

From (D.24) and (D.25) we obtain,

$$\sup_{\boldsymbol{\theta} \in \Theta} |\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})| = \sup_{\boldsymbol{\theta} \in \Theta} |(\gamma - \beta)^t (\hat{f}_1 - f_1(\boldsymbol{\theta}))| \leq \sup_{\boldsymbol{\theta} \in \Theta} |\gamma - \beta|^t \sup_{\boldsymbol{\theta} \in \Theta} |\hat{f}_1 - f_1(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0 \text{ as } \ t \to \infty,$$

where the convergence follows by Straumann & Mikosch (2006, Lemma 2.1) since $\sup_{\boldsymbol{\theta} \in \Theta} |\gamma - \beta| < 1$ and $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} \log^+ |f_1(\boldsymbol{\theta})| < \infty$. The former is guaranteed by condition $|\gamma - \beta| < 1$ and compactness of $\Theta$, while the latter is ensured by the existence of the second moment since $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} \log^+ |f_1(\boldsymbol{\theta})| < \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |f_1(\boldsymbol{\theta})|^2 < \infty$.

Conversely, if $|\gamma - \beta| > 1$ then the sum in (D.26) would diverge, while if $|\gamma - \beta| = 1$ it may diverge, hence the condition $|\gamma - \beta| < 1$ is a necessary and sufficient condition.

Finally, we show the uniqueness of the limit sequence $\{f_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ by contradiction. Assume the existence of two SE solutions to (D.1), $\{f_t\}_{t \in \mathbb{Z}}$ and $\{\tilde{f}_t\}_{t \in \mathbb{Z}}$, then for $t = t^*$ such that $f_{t^*} \neq \tilde{f}_{t^*}$ we have,

$$0 < |f_{t^*} - \tilde{f}_{t^*}| = |\gamma - \beta|^i \left| f_{t^*-i} - \tilde{f}_{t^*-i} \right|, \quad \forall i \geq 0.$$

We know that $|\gamma - \beta|^i \xrightarrow{e.a.s.} 0$ as $i \to \infty$ and $|f_{t^*-i} - \tilde{f}_{t^*-i}| = O_P(1)$ as it is strictly stationary, hence $\mathbb{P}\left(f_t = \tilde{f}_t\right) = 1$ and uniqueness follows, which completes the proof. ■

The claim about the moment bound of the limit sequence can be generalized to any $k \geq 2$ if the sequence $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ has sufficient number of moments.

**Corollary TA.1.** *If $\mathbb{E}\|\mathbf{y}_t\|^k < \infty$ with $k \geq 2$ and conditions of Proposition 1 are satisfied, then the limit sequence $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ satisfies $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|f_t(\boldsymbol{\theta})|^k < \infty$.*

*Proof of Proposition 2:* The proof of this proposition is slightly different from the proof of the Proposition 1 since $\hat{\boldsymbol{g}}_t$ depends not on the limit time-varying parameter $f_t$ but rather on the filtered time-varying parameter $\hat{f}_t$. Therefore, we are dealing with the perturbed version of the updating equation, that is

$$g^s_{t+1}(\boldsymbol{\theta}) = \Psi(\mathbf{y}^s_t, \hat{f}_t(\boldsymbol{\theta}), g^s_t(\boldsymbol{\theta})), \quad s = 1, \ldots, S.$$

We further define the filtered sequence that is the solution to the perturbed equation as $\{\hat{g}^s_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$, to the unperturbed one as $\{\tilde{g}^s_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ and the limit sequence as $\{g^s_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ which, if exists, is as follows,

$$g^s_{t+1}(\boldsymbol{\theta}) = \sum_{i=0}^{\infty}(\gamma_s - \beta_s)^i[\beta_s\psi(\mathbf{y}^s_{t-i}, f_{t-i}(\boldsymbol{\theta}); \boldsymbol{\Lambda}^c_s)]. \tag{D.27}$$

By the triangle inequality we have,

$$\sup_{\boldsymbol{\theta}\in\Theta}|\hat{g}^s_t(\boldsymbol{\theta}) - g^s_t(\boldsymbol{\theta})| \leq \sup_{\boldsymbol{\theta}\in\Theta}|\hat{g}^s_t(\boldsymbol{\theta}) - \tilde{g}^s_t(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}\in\Theta}|\tilde{g}^s_t(\boldsymbol{\theta}) - g^s_t(\boldsymbol{\theta})|. \tag{D.28}$$

For the first term on the right-hand side by the mean value theorem we have,

$$\sup_{\boldsymbol{\theta}\in\Theta}|\hat{g}^s_t(\boldsymbol{\theta}) - \tilde{g}^s_t(\boldsymbol{\theta})| \leq \sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial\Psi(\mathbf{y}^s_t, f^\star_t, g^s_t(\boldsymbol{\theta}))}{\partial f}\right| \sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})\right|,$$

where $f^\star_t$ lies between $\hat{f}_t$ and $f_t$. We can then conclude that $\sup_{\boldsymbol{\theta}\in\Theta}|\hat{g}^s_t(\boldsymbol{\theta}) - \tilde{g}^s_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$ as $t \to \infty$ by Lemma 2.1 in Straumann & Mikosch (2006) since $\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial\Psi(\mathbf{y}^s_t, f^\star_t, g^s_t(\boldsymbol{\theta}))}{\partial f}\right|$ is uniformly bounded and by Proposition 1 $\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})\right| \xrightarrow{e.a.s.} 0$ as $t \to \infty$. The uniform boundedness is established as follows,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial\Psi(\mathbf{y}^s_t, f^\star_t, g^s_t(\boldsymbol{\theta}))}{\partial f}\right| = \sup_{\boldsymbol{\theta}\in\Theta}\left|-\beta_s\frac{1}{N_s}\sum_{i=1}^{N_s}\lambda^c_{i,s} + (\gamma_s - \beta_s)\frac{\partial g^s_{t-1}(\boldsymbol{\theta})}{\partial f}\right| \leq \sup_{\boldsymbol{\theta}\in\Theta}\left|\beta_s\frac{1}{N_s}\sum_{i=1}^{N_s}\lambda^c_{i,s}\right|$$
$$+ \sup_{\boldsymbol{\theta}\in\Theta}|\gamma_s - \beta_s| \sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial g^s_{t-1}(\boldsymbol{\theta})}{\partial f}\right| < \infty,$$

where given the expression for $\frac{\partial g_t^s(\boldsymbol{\theta})}{\partial f}$ in (D.7) and since $\sup_{\boldsymbol{\theta}\in\Theta}|\gamma_s - \beta_s| < 1$, the last inequality follows.

It remains to show that the second term on the right hand side in (D.28) converges e.a.s. to zero and the properties of the limit sequence. These can be shown using the similar arguments as in the proof of Proposition 1. Specifically, using monotone convergence theorem it is easy to verify that the limit sequence is converging almost surely. Then the stationarity and ergodicity of the limit sequence $\{g_t^s(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ follows by Krengel's theorem since it is a measurable function of $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ and $\{f_t\}_{t\in\mathbb{Z}}$ which by Assumption 2 and Proposition 1 are strictly stationary and ergodic. The proof about the moment bounds is similar to the proof in Proposition 1.

Therefore, for the second term on the right hand side in (D.28) we obtain

$$\sup_{\boldsymbol{\theta}\in\Theta}|\tilde{g}_t^s(\boldsymbol{\theta}) - g_t^s(\boldsymbol{\theta})| \leq \sup_{\boldsymbol{\theta}\in\Theta}|\gamma_s - \beta_s|^t \sup_{\boldsymbol{\theta}\in\Theta}|\hat{g}_1^s - g_1^s(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0 \text{ as } \ t \to \infty,$$

where the convergence follows again by Lemma 2.1 in Straumann & Mikosch (2006) given that $\sup_{\boldsymbol{\theta}\in\Theta}|\gamma_s - \beta_s| < 1$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\log^+|g_1^s(\boldsymbol{\theta})| < \infty$. Conversely, if $|\gamma_s - \beta_s| > 1$ the sum in (D.27) would diverge; if $|\gamma_s - \beta_s| = 1$, the sum may diverge.

The uniqueness of the limit sequence $\{g_t^s(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ follows a proof by contradiction as in the proof of Proposition 1. ■

**Corollary TA.2.** *If conditions of Propositions 1 and 2 and Corollary TA.1 are fulfilled, then the limit sequence* $\{\boldsymbol{g}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ *satisfies* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{g}_t(\boldsymbol{\theta})\|^k < \infty$.

**Lemma TA.1.** *Let conditions of Theorem 1 hold. Then*

$$\sup_{\boldsymbol{\theta}\in\Theta^c}\left|Q_T^{(1)}(\boldsymbol{\theta}) - \mathbb{E}[q^{(1)}(\mathbf{y}_t, f_t(\boldsymbol{\theta}), \boldsymbol{\theta})]\right| \xrightarrow{a.s.} 0 \quad \textit{as } T \to \infty.$$

*Proof.* We establish the strong uniform convergence by applying ergodic theorem for separable Banach spaces of Rao (1962) to the sequence $\{q^{(1)}(\mathbf{y}_t, f_t(\cdot), \cdot)\}$. The conditions of the theorem are satisfied, since

1. $\{q^{(1)}(\mathbf{y}_t, f_t(\cdot), \cdot)\}_{t\in\mathbb{Z}}$ is an SE sequence, which follows by application of Krengel's theorem, as $q^{(1)}$ is continuous on the SE sequence $\{(\mathbf{y}_t, f_t)\}_{t\in\mathbb{Z}}$.

2. $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^c}|q^{(1)}(\mathbf{y}_t, f_t(\boldsymbol{\theta}), \boldsymbol{\theta})| < \infty$, since

$$\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^c}|q^{(1)}(\mathbf{y}_t, f_t(\boldsymbol{\theta}), \boldsymbol{\theta})| &= \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^c}\left|\left(\frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{N_s}\sum_{i=1}^{N_s} y_{i,t}^s\right) - f_t(\boldsymbol{\theta})\right)^2\right| \\
&\leq c\mathbb{E}\left|\frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{N_s}\sum_{i=1}^{N_s} y_{i,t}^s\right)\right|^2 + c\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^c}|f_t(\boldsymbol{\theta})|^2 \\
&\leq c\tilde{c}\sum_{s=1}^{S}\tilde{c}_s\sum_{i=1}^{N_s}\mathbb{E}\left|y_{i,t}^s\right|^2 + c\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^c}|f_t(\boldsymbol{\theta})|^2 < \infty,
\end{aligned}$$

where in the last line we used several times Loève's $c_r$ inequality. Therefore, given that $\mathbb{E}\|\mathbf{y}_t\| < \infty$ by Assumption 3 and from Proposition 1 $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^c}|f_t(\boldsymbol{\theta})|^2 < \infty$, the results follows. ■

**Lemma TA.2.** *Let conditions of Theorem 1 hold. Then a plug-in filter $\hat{f}_t(\hat{\boldsymbol{\theta}}_T^{(1)}, \hat{f}_1)$ converges almost surely,*

$$|\hat{f}_t(\hat{\boldsymbol{\theta}}_T^{(1)}, \hat{f}_1) - f_t(\boldsymbol{\theta}_0^{(1)})| \xrightarrow{a.s.} 0 \ \textit{as} \ \ t, T \to \infty.$$

*Proof.* By the triangle inequality, we have

$$|\hat{f}_t(\hat{\boldsymbol{\theta}}_T^{(1)}, \hat{f}_1) - f_t(\boldsymbol{\theta}_0^{(1)})| \leq |\hat{f}_t(\hat{\boldsymbol{\theta}}_T^{(1)}, \hat{f}_1) - \hat{f}_t(\boldsymbol{\theta}_0^{(1)}, \hat{f}_1)| + |\hat{f}_t(\boldsymbol{\theta}_0^{(1)}, \hat{f}_1) - f_t(\boldsymbol{\theta}_0^{(1)})|, \tag{D.29}$$

where the second term in the expression above goes to zero e.a.s. as $t \to \infty$ since the filter is uniformly invertible, see Proposition 1.

For the first term in (D.29), by unfolding it recursively, we notice

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta^c}|\hat{f}_t(\boldsymbol{\theta}, \hat{f}_1) - \hat{f}_t(\boldsymbol{\theta}_0^{(1)}, \hat{f}_1)| \\
&\leq \sup_{\boldsymbol{\theta}\in\Theta^c}|\beta\phi(\mathbf{y}_{t-1}) + (\gamma-\beta)\hat{f}_{t-1}(\boldsymbol{\theta}_0^{(1)}, \hat{f}_1) - \beta_0\phi(\mathbf{y}_{t-1}) - (\gamma_0-\beta_0)\hat{f}_{t-1}(\boldsymbol{\theta}_0^{(1)}, \hat{f}_1)| \\
&+ \sup_{\boldsymbol{\theta}\in\Theta^c}|(\gamma-\beta)(\hat{f}_{t-1}(\boldsymbol{\theta}, \hat{f}_1) - \hat{f}_{t-1}(\boldsymbol{\theta}_0^{(1)}, \hat{f}_1))| \\
&\leq \sum_{i=1}^{t-1}\sup_{\boldsymbol{\theta}\in\Theta^c}|\gamma-\beta|^{i-1}\sup_{\boldsymbol{\theta}\in\Theta^c}|\beta\phi(\mathbf{y}_{t-i}) + (\gamma-\beta)\hat{f}_{t-i}(\boldsymbol{\theta}_0^{(1)}, \hat{f}_1) - \hat{f}_{t-i+1}(\boldsymbol{\theta}_0^{(1)}, \hat{f}_1)|.
\end{aligned}$$

From the strong consistency of $\hat{\boldsymbol{\theta}}_T^{(1)}$, we can conclude that $|\hat{f}_t(\hat{\boldsymbol{\theta}}_T^{(1)}, \hat{f}_1) - \hat{f}_t(\boldsymbol{\theta}_0^{(1)}, \hat{f}_1)| \xrightarrow{a.s.} 0$ as $T \to \infty$ which completes the proof. ■

**Lemma TA.3.** *Let conditions of Theorem 2 hold. Then*

$$\sup_{\boldsymbol{\theta}\in\Theta^{\lambda_c}}\left|Q_T^{(2)}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_T^{(1)}\right) - Q_\infty^{(2)}\left(\boldsymbol{\theta}, \boldsymbol{\theta}_0^{(1)}\right)\right| \quad \textit{as } T \to \infty.$$

*Proof.* The uniform convergence follows by Theorem 3.7 in White (1996) since $\hat{\boldsymbol{\theta}}_T^{(1)} \xrightarrow{a.s.} \boldsymbol{\theta}_0^{(1)}$ as

$T \to \infty$, $\Theta^c$ and $\Theta^{\lambda_c}$ are compact, criterion function is continuous and

$$\sup_{\boldsymbol{\theta}^{(1)} \in \Theta^c} \sup_{\boldsymbol{\theta}^{(2)} \in \Theta^{\lambda_c}} \left| Q_T^{(2)}(\boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(1)}) - Q_\infty^{(2)}(\boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(1)}) \right| \xrightarrow{a.s.} 0 \quad \text{as } T \to \infty.$$

The latter follows again by application of Theorem 6.5 in Rao (1962), as

1. $\{q^{(2)}(\mathbf{y}_t, f_t(\cdot), \cdot\}_{t\in\mathbb{Z}}$ is an SE sequence, which follows by application of Krengel's theorem, as $q^{(2)}$ is continuous on the SE sequence $\{(\mathbf{y}_t, f_t)\}_{t\in\mathbb{Z}}$.

2. trivially $\mathbb{E}\sup_{\boldsymbol{\theta}^{(1)}\in\Theta^c}\sup_{\boldsymbol{\theta}^{(2)}\in\Theta^{\lambda_c}} |q^{(2)}(\mathbf{y}_t, f_t(\boldsymbol{\theta}^{(1)}), \boldsymbol{\theta}^{(2)})| < \infty$.

Therefore, we conclude that

$$\sup_{\boldsymbol{\theta}^{(2)} \in \Theta^{\lambda_c}} \left| \hat{Q}_T^{(2)}(\boldsymbol{\theta}^{(2)}, \hat{\boldsymbol{\theta}}_T^{(1)}) - \mathbb{E}[q^{(2)}(\mathbf{y}_t, f_t(\boldsymbol{\theta}_0^{(1)}), \boldsymbol{\theta}^{(2)})] \right| \xrightarrow{a.s.} 0 \quad \text{as } T \to \infty.$$

■

*Sketch of the proof of Theorem 3:* The proof of this theorem consists of three parts. The first part establishes the consistency of the step III estimator, i.e. $\hat{\boldsymbol{\theta}}_T^g \xrightarrow{a.s.} \boldsymbol{\theta}_0^g$, the second part is about the convergence of the plug-in filter $\|\hat{\boldsymbol{g}}_t(\hat{\boldsymbol{\theta}}_T^g) - \boldsymbol{g}_t(\boldsymbol{\theta}_0^g)\| \xrightarrow{a.s.} 0$ and the last part is about the consistency of the step IV estimator.

The proof of the first and last parts requires the uniform convergence of the corresponding criterion functions to the limit criterion functions as well as identifiable uniqueness of the estimators. The former can be proved using Theorem 3.7 in White (1996) and Theorem 6.5 in Rao (1962). We again need to take into consideration that the function depends on the filtered time-varying parameters rather than on the limit counterparts. Identifiable uniqueness is proved using the same reasoning as in the proofs of Theorems 1 and 2.

Now we turn to the strong consistency of the filter. The reasoning is similar to the proof in Theorem 1, that is by the triangle inequality we have

$$\left\|\hat{\boldsymbol{g}}_t(\hat{\boldsymbol{\theta}}_T^{(3:1)}, \hat{\boldsymbol{g}}_1) - \boldsymbol{g}_t(\boldsymbol{\theta}_0^{(3:1)})\right\| \leq \left\|\hat{\boldsymbol{g}}_t(\hat{\boldsymbol{\theta}}_T^{(3:1)}, \hat{\boldsymbol{g}}_1) - \hat{\boldsymbol{g}}_t(\boldsymbol{\theta}_0^{(3:1)}, \hat{\boldsymbol{g}}_1)\right\| + \left\|\hat{\boldsymbol{g}}_t(\boldsymbol{\theta}_0^{(3:1)}, \hat{\boldsymbol{g}}_1) - \boldsymbol{g}_t(\boldsymbol{\theta}_0^{(3:1)})\right\|,$$

where the second term vanishes to zero e.a.s. as $t \to \infty$ since the filter is uniformly invertible (see Proposition 2). Now we show that the first term also goes to zero a.s.. Repeatedly unfolding the expression, for $s = 1, \dots, S$, we have

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}^{(3:1)}\in\Theta}\left|\hat{g}_t^s\left(\boldsymbol{\theta}^{(3:1)}, \hat{g}_1^s\right) - \hat{g}_t^s\left(\boldsymbol{\theta}_0^{(3:1)}, \hat{g}_1^s\right)\right| \\
&\leq \sup_{\boldsymbol{\theta}^{(3:1)}\in\Theta} |\beta_s\psi(\mathbf{y}_{t-1}^s, \hat{f}_{t-1}(\boldsymbol{\theta}^{(1)}); \boldsymbol{\theta}_s^{(2)}) + (\gamma_s - \beta_s)\,\hat{g}_{t-1}^s(\boldsymbol{\theta}_0^{(3:1)}, \hat{g}_1^s) \\
&- \beta_{s,0}\psi(\mathbf{y}_{t-1}^s, \hat{f}_{t-1}(\boldsymbol{\theta}_0^{(1)}); \boldsymbol{\theta}_{s,0}^{(2)}) - (\gamma_{s,0} - \beta_{s,0})\hat{g}_{t-1}^s(\boldsymbol{\theta}_0^{(3:1)}, \hat{g}_1^s)| \\
&+ \sup_{\boldsymbol{\theta}^{(3:1)}\in\Theta} |(\gamma_s - \beta_s)(\hat{g}_{t-1}^s(\boldsymbol{\theta}^{(3:1)}, \hat{g}_1^s) - \hat{g}_{t-1}^s(\boldsymbol{\theta}_0^{(3:1)}, \hat{g}_1^s)| \\
&\leq \sum_{i=1}^{t-1} \sup_{\boldsymbol{\theta}^{(3:1)}\in\Theta} |\gamma_s - \beta_s|^{i-1} \sup_{\boldsymbol{\theta}^{(3:1)}\in\Theta} |\beta_s\psi(\mathbf{y}_{t-i}^s, \hat{f}_{t-i}(\boldsymbol{\theta}^{(1)}), \boldsymbol{\theta}_s^{(2)}) \\
&+ (\gamma_s - \beta_s)\hat{g}_{t-i}^s(\boldsymbol{\theta}_0^{(3:1)}, \hat{g}_1^s) - \hat{g}_{t-i+1}^s(\boldsymbol{\theta}_0^{(3:1)}, \hat{g}_1^s)|.
\end{aligned}
$$

From the strong consistency of $\hat{\boldsymbol{\theta}}_T^{(1)}$, $\hat{\boldsymbol{\theta}}_T^{(2)}$ and $\hat{\boldsymbol{\theta}}_T^{(3)}$ that was established in Theorems 1–3, we conclude that $|\hat{g}_t^s(\hat{\boldsymbol{\theta}}_T^{(3:1)}) - g_t^s(\boldsymbol{\theta}_0^{(3:1)})| \xrightarrow{a.s.} 0$ as $T \to \infty$. ■

**Lemma TA.4.** *Let conditions of Proposition 1 hold, then*

*a. the limit sequences of the first and second derivatives* $\{f_t'(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ *and* $\{f_t''(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ *exist, are SE and satisfy* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|f_t'(\boldsymbol{\theta})\|^2 < \infty$ *and* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|f_t''(\boldsymbol{\theta})\|^2 < \infty$*. Moreover,*

$$\sup_{\boldsymbol{\theta}\in\Theta}\|\hat{f}_t'(\boldsymbol{\theta}) - f_t'(\boldsymbol{\theta})\| \xrightarrow{e.a.s.} 0, \qquad \sup_{\boldsymbol{\theta}\in\Theta}\|\hat{f}_t''(\boldsymbol{\theta}) - f_t''(\boldsymbol{\theta})\| \xrightarrow{e.a.s.} 0 \quad \text{as } t \to \infty.$$

*If, additionally, conditions of Proposition 2 hold then*

*b. the limit sequences of the first and second derivatives* $\{\boldsymbol{g}_t'(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ *and* $\{\boldsymbol{g}_t''(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ *exist, are SE and satisfy* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{g}_t'(\boldsymbol{\theta})\|^2 < \infty$ *and* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{g}_t''(\boldsymbol{\theta})\|^2 < \infty$*. Moreover,*

$$\sup_{\boldsymbol{\theta}\in\Theta}\|\hat{\boldsymbol{g}}_t'(\boldsymbol{\theta}) - \boldsymbol{g}_t'(\boldsymbol{\theta})\| \xrightarrow{e.a.s.} 0, \qquad \sup_{\boldsymbol{\theta}\in\Theta}\|\hat{\boldsymbol{g}}_t''(\boldsymbol{\theta}) - \boldsymbol{g}_t''(\boldsymbol{\theta})\| \xrightarrow{e.a.s.} 0 \quad \text{as } t \to \infty.$$

*Proof.* The proof follows the same lines as in the proof of Propositions 1 and 2 by backwards unfolding the system of equations (D.3)–(D.6). ■

**Corollary TA.3.** *Let conditions of Corollary TA.1 hold. Then the limit sequences of the first and second derivatives satisfy* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|f_t'(\boldsymbol{\theta})\|^k < \infty$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|f_t''(\boldsymbol{\theta})\|^k < \infty$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{g}_t'(\boldsymbol{\theta})\|^k < \infty$ *and* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{g}_t''(\boldsymbol{\theta})\|^k < \infty$.

**Lemma TA.5.** *Let conditions of Propositions 1 and 2 hold. Then it holds that*

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial \hat{q}_t^{(1)}(\boldsymbol{\theta}^{(1)})}{\partial f}-\frac{\partial q_t^{(1)}(\boldsymbol{\theta}^{(1)})}{\partial f}\right| \xrightarrow{e.a.s.} 0, \tag{D.30}$$

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial \hat{q}_t^{(2)}\left(\boldsymbol{\theta}^{(2:1)}\right)}{\partial \boldsymbol{\theta}^{(2)}}-\frac{\partial q_t^{(2)}\left(\boldsymbol{\theta}^{(2:1)}\right)}{\partial \boldsymbol{\theta}^{(2)}}\right\| \xrightarrow{e.a.s.} 0, \tag{D.31}$$

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial \hat{q}_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial \boldsymbol{g}}-\frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)}))}{\partial \boldsymbol{g}}\right\| \xrightarrow{e.a.s.} 0, \tag{D.32}$$

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial \hat{q}_t^{(4)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(4)}}-\frac{\partial q_t^{(4)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(4)}}\right\| \xrightarrow{e.a.s.} 0, \tag{D.33}$$

*as* $t\to\infty$ *and the limit sequences* $\left\{\frac{\partial q_t^{(1)}(\boldsymbol{\theta}^{(1)})}{\partial f}\right\}_{t\in\mathbb{Z}}$, $\left\{\frac{\partial q_t^{(2)}\left(\boldsymbol{\theta}^{(2:1)}\right)}{\partial \boldsymbol{\theta}^{(2)}}\right\}_{t\in\mathbb{Z}}$, $\left\{\frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)}))}{\partial \boldsymbol{g}}\right\}_{t\in\mathbb{Z}}$, *and* $\left\{\frac{\partial q_t^{(4)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(4)}}\right\}_{t\in\mathbb{Z}}$ *are SE.*

*If, in addition, Assumption 3.a holds then* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial q_t^{(1)}(\boldsymbol{\theta}^{(1)})}{\partial f}\right|^{2r}<\infty$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial q_t^{(2)}(\boldsymbol{\theta}^{(2:1)})}{\partial \boldsymbol{\theta}^{(2)}}\right\|^{r}<\infty$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial q_t^{(3)}(\boldsymbol{\theta}^{(3:1)})}{\partial \boldsymbol{g}}\right\|^{2r}<\infty$, *and* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial q_t^{(4)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(4)}}\right\|^{r}<\infty$.

*Proof.* Given the expressions for the derivatives (D.11) and (D.12), the convergence in (D.30) and (D.32) follows directly from Propositions 1 and 2. Establishing the convergence in (D.31) additionally requires $\sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t^2(\boldsymbol{\theta})-f_t^2(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$ as $t\to\infty$, see (D.9). The latter holds by Corollary TA.15 in Blasques et al. (2022). The conditions of the corollary are satisfied since by Proposition 1 $\sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t(\boldsymbol{\theta})-f_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$ as $t\to\infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\log^+|f_t(\boldsymbol{\theta})|<\infty$.

Finally, (D.33) holds since $\sup_{\boldsymbol{\theta}\in\Theta}\|\hat{f}_t(\boldsymbol{\theta})\hat{\boldsymbol{g}}_t(\boldsymbol{\theta})-f_t(\boldsymbol{\theta})\boldsymbol{g}_t(\boldsymbol{\theta})\| \xrightarrow{e.a.s.} 0$ as well as $\sup_{\boldsymbol{\theta}\in\Theta}\|\hat{\boldsymbol{g}}_t^2(\boldsymbol{\theta})-\boldsymbol{g}_t^2(\boldsymbol{\theta})\| \xrightarrow{e.a.s.} 0$ as $t\to\infty$. The former follows from Propositions 1 and 2 together with Lemma TA.14 in Blasques et al. (2022), while the latter result is a direct implication of Corollary TA.15 in Blasques et al. (2022).

The limit sequences are SE by Krengel's theorem and bounded moments follow by Assumption 3.a, Corollaries TA.1 and TA.2 given the expressions of the first derivatives (D.9)–(D.12). ■

**Lemma TA.6.** *Let conditions of Propositions 1 and 2 hold. Furthermore, let Assumptions 3.a and 4 be satisfied. Then the sequence* $\{\nabla_{\boldsymbol{\theta}}q_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ *is SE and NED of size* $-1$ *on a strongly mixing sequence of size* $-r/(1-r)$ *for some* $r>2$.

*Proof.* By Assumptions 2 and 4, sequence $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ is SE and NED of size $-1$ on a strongly mixing sequence of size $-r/(r-1)$ for some $r>2$. Propositions 1 and 2 together with Corollaries TA.1, TA.2 and TA.3 ensure that the limit sequences $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, $\{f_t'(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, $\{\boldsymbol{g}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and $\{\boldsymbol{g}_t'(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ are SE, which are also jointly SE. Therefore, by Krengel's theorem and continuity of $\nabla_{\boldsymbol{\theta}}q_t(\cdot)$ it follows that $\{\nabla_{\boldsymbol{\theta}}q_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is also SE.

Now we turn to the second part of the lemma about the NED property. We notice that

$$
\begin{aligned}
|f_{t+1}-f^*_{t+1}| &\leq |\beta||\phi(\mathbf{y}_t)-\phi(\mathbf{y}^*_t)| + |\gamma-\beta||f_t-f^*_t| \\
&\leq |\beta|\frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s}\left|y^s_{i,t}-y^{s*}_{i,t}\right| + |\gamma-\beta||f_t-f^*_t|, \\
|g^s_{t+1}-g^{s\,*}_{t+1}| &\leq |\beta_s||\psi(\mathbf{y}^s_t,f_t)-\psi(\mathbf{y}^{s*}_t,f^*_t)| + |\gamma_s-\beta_s||g^s_t-g^{s*}_t| \\
&\leq |\beta_s|\frac{1}{N_s}\sum_{i=1}^{N_s}|y^s_{i,t}-y^{s*}_{i,t}| + |\beta_s|\left|\frac{1}{N_s}\sum_{i=1}^{N_s}\lambda^c_{i,s}\right||f_t-f^*_t| + |\gamma_s-\beta_s||g^s_t-g^{s*}_t|.
\end{aligned}
$$

By assumption $\beta\in\mathbb{R},\beta_s\in\mathbb{R},\mathbf{\Lambda}^c\in\mathbb{R}^N$, and $|\gamma-\beta|<1$ and $|\gamma_s-\beta_s|<1$ for $s=1,\ldots,S$. Moreover, under Assumption 3.a $\mathbf{y}_t$ has $2r$ bounded moments with $r>2$ which by Corollaries TA.1 and 2 implies that $f_t$ and $g^s_t$ also have $2r$ bounded moments. Therefore, by Theorem 6.10 in Pötscher & Prucha (1997) $f_t$ and $g^s_t$ are NED of size $-1$ on a strongly mixing sequence of size $-r/(1-r)$ for some $r>2$. Lemma 6.9 in Pötscher & Prucha (1997) implies that the stacked vector $\boldsymbol{g}_t$ is itself NED.

We proceed similarly with the first order derivatives

$$
\begin{aligned}
\|f'_{t+1}-f'^*_{t+1}\| &\leq \|\boldsymbol{\zeta}_f(\mathbf{y}_t,f_t)-\zeta_f(\mathbf{y}^*_t,f^*_t))\| + |\gamma-\beta|\|f'_t-f'^*_t\| \\
&\leq \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s}|y^s_{i,t}-y^{s*}_{i,t}| + 2|f_t-f^*_t| + |\gamma-\beta|\|f'_t-f'^*_t\|, \\
\|g^{s\prime}_{t+1}-g^{s\,\prime*}_{t+1}\| &\leq \|\boldsymbol{\zeta}_g(\mathbf{y}^s_t,f_t,g^s_t)-\boldsymbol{\zeta}_g(\mathbf{y}^{s*}_t,f^*_t,g^{s*}_t)\| + (\gamma_s-\beta_s)\|g^{s\prime}_t-g^{s\prime*}_t\| \\
&\leq \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s}|y^s_{i,t}-y^{s*}_{i,t}| + \left|\frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s}\lambda^c_{i,s}\right||f_t-f^*_t| \\
&\quad + 2|g^s_t-g^{s*}_t| + |\gamma_s-\beta_s|\|g^{s\prime}_t-g^{s\prime*}_t\|,
\end{aligned}
$$

where in the last step we exploit the norm equivalence.

Using a similar argument as for the time-varying parameters themselves, by Theorem 6.10 and Lemma 6.9 in Pötscher & Prucha (1997) we obtain that $f'_t$ and $\boldsymbol{g}'_t$ are NED of size $-1$ on a strongly mixing sequence of size $-r/(1-r)$ for some $r>2$. This follows since $\mathbf{\Lambda}^c\in\mathbb{R}^N$, $\mathbf{y}_t$, $f_t$ and $\boldsymbol{g}_t$ are NED with $2r$ bounded moments.

By Theorem 17.12 in Davidson (1994), the sequences $\{\nabla_{\boldsymbol{\theta}^{(1)}}q^{(1)}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, $\{\nabla_{\boldsymbol{\theta}^{(2)}}q^{(2)}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, $\{\nabla_{\boldsymbol{\theta}^{(3)}}q^{(3)}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, $\{\nabla_{\boldsymbol{\theta}^{(4)}}q^{(4)}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ are NED. This follows as functions $\nabla_{\boldsymbol{\theta}^{(1)}}q^{(1)}(\mathbf{y}_t,f_t,f'_t,\boldsymbol{\theta})$, $\nabla_{\boldsymbol{\theta}^{(2)}}q^{(2)}(\mathbf{y}_t,f_t,\boldsymbol{\theta})$, $\nabla_{\boldsymbol{\theta}^{(3)}}q^{(3)}(\mathbf{y}_t,f_t,\boldsymbol{g}_t,\boldsymbol{g}'_t,\boldsymbol{\theta})$, $\nabla_{\boldsymbol{\theta}^{(4)}}q^{(4)}(\mathbf{y}_t,f_t,\boldsymbol{g}_t,\boldsymbol{\theta})$ are Lipschitz continuous and they are also functions of NED sequences. By Lemma 6.9 in Pötscher & Prucha (1997) we conclude that the stacked vector $\{\nabla_{\boldsymbol{\theta}}q_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is NED, which finishes the proof. ■

**Lemma TA.7.** *Let conditions of Propositions 1 and 2 hold. Moreover, let Assumptions 3.a and*

*4 be satisfied. Then*

$$\sqrt{T}\nabla_{\boldsymbol{\theta}}\boldsymbol{Q}_T(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{B}(\boldsymbol{\theta}_0)) \quad \text{as } T \to \infty,$$

*with* $\boldsymbol{B}(\boldsymbol{\theta}_0)$ *as defined in Theorem 4.*

*Proof.* By Lemma TA.6 we have $\{\nabla_{\boldsymbol{\theta}} q_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is NED of size $-1$ on a strongly mixing process of size $-r/(r-1)$ for some $r > 2$. Therefore, the proof is based on the central limit theorem for near epoch dependent processes (Pötscher & Prucha, 1997, Theorem 10.2). Below, we verify that all the conditions of the theorem are satisfied.

First, we notice that $\mathbb{E}\left[\nabla_{\boldsymbol{\theta}} q_t(\boldsymbol{\theta}_0)\right] = \nabla_{\boldsymbol{\theta}}\boldsymbol{Q}_\infty(\boldsymbol{\theta}_0) = \mathbf{0}$ where the interchange of the expectation and derivative is permitted since the criterion function is continuously differentiable.

We notice that $\|\nabla_{\boldsymbol{\theta}} q_t(\boldsymbol{\theta})\| = \left(\sum_{i=1}^4 \|\nabla_{\boldsymbol{\theta}} q_t(\boldsymbol{\theta})\|^2\right)^{1/2} \leq \sum_{i=1}^4 \|\nabla_{\boldsymbol{\theta}} q_t(\boldsymbol{\theta})\|$. Hence, by Loève's $c_r$ inequality, there exists a constant $c > 0$ such that,

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} \|\nabla_{\boldsymbol{\theta}} q_t(\boldsymbol{\theta})\|^r \leq c\sum_{i=1}^{4} \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(i)}} q_t^{(i)}(\boldsymbol{\theta})\right\|^r, \tag{D.34}$$

where index $i$ refers to the step of the estimation procedure.

To show that expression (D.34) is finite, we further consider the gradient of each of the steps.

Step I. From (D.8) and Cauchy-Schwartz inequality,

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(1)}} q_t^{(1)}(\boldsymbol{\theta})\right\|^r = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial q_t^{(1)}(\boldsymbol{\theta})}{\partial f} f_t'(\boldsymbol{\theta})\right\|^r \leq \left(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial q_t^{(1)}(\boldsymbol{\theta})}{\partial f}\right|^{2r} \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|f_t'(\boldsymbol{\theta})\right\|^{2r}\right)^{1/2} < \infty,$$

where the bounded moment claim follows by the fact that under Assumption 3.a $\mathbf{y}_t$ has $2r$ bounded moments. Then, Corollary TA.3 implies that $f_t'(\boldsymbol{\theta})$ has $2r$ bounded moments. The derivative $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial q_t^{(1)}(\boldsymbol{\theta})}{\partial f}\right\|^{2r}$ is bounded by Lemma TA.5.

Step III. By Loève's $c_r$ inequality and Cauchy-Schwartz inequality,

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(3)}} q_t^{(3)}(\boldsymbol{\theta})\right\|^r &\leq c_s\sum_{s=1}^{S}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial q_t^{(3)}(\boldsymbol{\theta})}{\partial g^s} g_t^{s\prime}(\boldsymbol{\theta})\right\|^r \\ &\leq c_s\sum_{s=1}^{S}\left(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial q_t^{(3)}(\boldsymbol{\theta})}{\partial g^s}\right|^{2r}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|g_t^{s\prime}(\boldsymbol{\theta})\right\|^{2r}\right)^{1/2} < \infty.\end{aligned}$$

Similarly to step I, the expression is bounded since by Assumption 3.a and Corollary TA.3 $g_t^{s\prime}(\boldsymbol{\theta})$ has $2r$ bounded moments. The derivative $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial q_t^{(3)}(\boldsymbol{\theta})}{\partial g^s}\right|^{2r}$ is also bounded which follows from Lemma TA.5.

Step II. From equation (D.9), Loève's $c_r$ inequality and Cauchy-Schwartz inequality, we have

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(2)}}q_t^{(2)}(\boldsymbol{\theta})\right\|^r = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|2f_t(\boldsymbol{\theta})\left(\mathbf{y}_t - \boldsymbol{\Lambda}^c f_t(\boldsymbol{\theta})\right)\|^r \leq c_r\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|2f_t(\boldsymbol{\theta})\mathbf{y}_t\|^r$$
$$+ c_r\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\boldsymbol{\Lambda}^c 2f_t^2(\boldsymbol{\theta})\right\|^r \leq 2c_r(\mathbb{E}\left\|\mathbf{y}_t\right\|^{2r}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|f_t(\boldsymbol{\theta})|^{2r})^{1/2}$$
$$+ 2c_r\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{\Lambda}^c\|^r\,\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|f_t(\boldsymbol{\theta})|^{2r} < \infty,$$

where the final claim follows from Proposition 1 under Assumption 3.a, $\mathbb{E}\|\mathbf{y}_t\|^{2r} < \infty$, which by Corrolary TA.1 implies that $f_t(\boldsymbol{\theta})$ has $2r$ bounded moments.

Step IV. Considering equation (D.10) and applying Loève's $c_r$ inequality together with the Cauchy-Schwartz inequality, we have

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(4)}}q_t^{(4)}(\boldsymbol{\theta})\right\|^r \leq c_s\sum_{s=1}^{S}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial q_t^{(4)}(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}_s^{(4)}}\right\|^r$$
$$= c_s\sum_{s=1}^{S}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|2(\mathbf{y}_t^s - \boldsymbol{\Lambda}_s^c f_t(\boldsymbol{\theta}) - \boldsymbol{\Lambda}_s^g g_t^s(\boldsymbol{\theta}))g_t^s(\boldsymbol{\theta})\|^r$$
$$\leq c_s(\sum_{s=1}^{S}2c_r(\mathbb{E}\|\mathbf{y}_t^s\|^{2r}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|g_t^s(\boldsymbol{\theta})|^{2r})^{1/2} + c_r\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{\Lambda}_s^c\|^r\,(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|f_t(\boldsymbol{\theta})|^{2r}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|g_t^s(\boldsymbol{\theta})|^{2r})^{1/2}$$
$$+ c_r\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{\Lambda}_s^g\|^r\,\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|g_t^s(\boldsymbol{\theta})|^{2r}) < \infty,$$

where in the last step we used the result of Proposition 2 and Corollary TA.2 under Assumption 3.a.

The gradients of all the steps are uniformly bounded, hence, expression (D.34) is also uniformly bounded, i.e. $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla_{\boldsymbol{\theta}}q_t(\boldsymbol{\theta})\|^r < \infty$. Therefore, all the assumptions of Theorem 10.2 in Pötscher & Prucha (1997) are satisfied which finalizes the proof. ∎

**Lemma TA.8.** *Let conditions of Propositions 1 and 2 hold, then*

$$\sqrt{T}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}}\hat{\boldsymbol{Q}}_T(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}}\boldsymbol{Q}_T(\boldsymbol{\theta})\right\| \xrightarrow{a.s.} \mathbf{0} \quad \textit{as } T\to\infty.$$

*Proof.* We show a.s. convergence by establishing e.a.s. convergence of the time $t$ function contributions, that is

$$\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla_{\boldsymbol{\theta}}\hat{q}_t(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}}q_t(\boldsymbol{\theta})\| \xrightarrow{e.a.s.} 0 \quad \text{as } t\to\infty. \tag{D.35}$$

By Loève's $c_r$ inequality

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla_{\boldsymbol{\theta}}\hat{q}_t(\boldsymbol{\theta})-\nabla_{\boldsymbol{\theta}}q_t(\boldsymbol{\theta})\| \leq c\sum_{i=1}^{4}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(i)}}\hat{q}_t^{(i)}(\boldsymbol{\theta})-\nabla_{\boldsymbol{\theta}^{(i)}}q_t^{(i)}(\boldsymbol{\theta})\right\| \\
&\leq c\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial\hat{q}_t^{(1)}(\boldsymbol{\theta})}{\partial f}\hat{f}_t'(\boldsymbol{\theta})-\frac{\partial q_t^{(1)}(\boldsymbol{\theta})}{\partial f}f_t'(\boldsymbol{\theta})\right\|+c\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial\hat{q}_t^{(2)}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{(2)}}-\frac{\partial q_t^{(2)}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{(2)}}\right\| \\
&+c\tilde{c}\sum_{s=1}^{S}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial\hat{q}_t^{(3)}(\boldsymbol{\theta})}{\partial g^s}\hat{g}_t^{s}{}'(\boldsymbol{\theta})-\frac{\partial q_t^{(3)}(\boldsymbol{\theta})}{\partial g^s}g_t^{s\prime}(\boldsymbol{\theta})\right\| \\
&+c\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial\hat{q}_t^{(4)}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{(4)}}-\frac{\partial q_t^{(4)}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{(4)}}\right\|. \qquad \text{(D.36)}
\end{aligned}
$$

By Lemma TA.5, Propositions 1 and 2 and Lemma TA.14 in Blasques et al. (2022) we obtain that each term in (D.36) converges e.a.s. to 0, hence (D.35) follows. ■

**Lemma TA.9.** *Let Assumptions 1–3 hold. Then*

$$\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{A}_T(\boldsymbol{\theta})-\boldsymbol{A}(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0, \quad as \quad T\to\infty,$$

*where*

$$\boldsymbol{A}(\boldsymbol{\theta})=\begin{bmatrix}\mathbb{E}[\nabla_{\boldsymbol{\theta}^{(1)}\boldsymbol{\theta}^{(1)}}q_t^{(1)}(\boldsymbol{\theta})] & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbb{E}[\nabla_{\boldsymbol{\theta}^{(2)}\boldsymbol{\theta}^{(1)}}q_t^{(2)}(\boldsymbol{\theta})] & \mathbb{E}[\nabla_{\boldsymbol{\theta}^{(2)}\boldsymbol{\theta}^{(2)}}q_t^{(2)}(\boldsymbol{\theta})] & \mathbf{0} & \mathbf{0}\\ \mathbb{E}[\nabla_{\boldsymbol{\theta}^{(3)}\boldsymbol{\theta}^{(1)}}q_t^{(3)}(\boldsymbol{\theta})] & \mathbb{E}[\nabla_{\boldsymbol{\theta}^{(3)}\boldsymbol{\theta}^{(2)}}q_t^{(3)}(\boldsymbol{\theta})] & \mathbb{E}[\nabla_{\boldsymbol{\theta}^{(3)}\boldsymbol{\theta}^{(3)}}q_t^{(3)}(\boldsymbol{\theta})] & \mathbf{0}\\ \mathbb{E}[\nabla_{\boldsymbol{\theta}^{(4)}\boldsymbol{\theta}^{(1)}}q_t^{(4)}(\boldsymbol{\theta})] & \mathbb{E}[\nabla_{\boldsymbol{\theta}^{(4)}\boldsymbol{\theta}^{(2)}}q_t^{(4)}(\boldsymbol{\theta})] & \mathbb{E}[\nabla_{\boldsymbol{\theta}^{(4)}\boldsymbol{\theta}^{(3)}}q_t^{(4)}(\boldsymbol{\theta})] & \mathbb{E}[\nabla_{\boldsymbol{\theta}^{(4)}\boldsymbol{\theta}^{(4)}}q_t^{(4)}(\boldsymbol{\theta})]\end{bmatrix},$$

*and* $\boldsymbol{A}_T(\boldsymbol{\theta}):=\frac{1}{T}\boldsymbol{A}_t(\boldsymbol{\theta})$, *with*

$$\boldsymbol{A}_t(\boldsymbol{\theta}):=\begin{bmatrix}\nabla_{\boldsymbol{\theta}^{(1)}\boldsymbol{\theta}^{(1)}}q_t^{(1)}(\boldsymbol{\theta}^{(1)}) & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \nabla_{\boldsymbol{\theta}^{(2)}\boldsymbol{\theta}^{(1)}}q_t^{(2)}(\boldsymbol{\theta}^{(2:1)}) & \nabla_{\boldsymbol{\theta}^{(2)}\boldsymbol{\theta}^{(2)}}q_t^{(2)}(\boldsymbol{\theta}^{(2:1)}) & \mathbf{0} & \mathbf{0}\\ \nabla_{\boldsymbol{\theta}^{(3)}\boldsymbol{\theta}^{(1)}}q_t^{(3)}(\boldsymbol{\theta}^{(3:1)}) & \nabla_{\boldsymbol{\theta}^{(3)}\boldsymbol{\theta}^{(2)}}q_t^{(3)}(\boldsymbol{\theta}^{(3:1)}) & \nabla_{\boldsymbol{\theta}^{(3)}\boldsymbol{\theta}^{(3)}}q_t^{(3)}(\boldsymbol{\theta}^{(3:1)}) & \mathbf{0}\\ \nabla_{\boldsymbol{\theta}^{(4)}\boldsymbol{\theta}^{(1)}}q_t^{(4)}(\boldsymbol{\theta}) & \nabla_{\boldsymbol{\theta}^{(4)}\boldsymbol{\theta}^{(2)}}q_t^{(4)}(\boldsymbol{\theta}) & \nabla_{\boldsymbol{\theta}^{(4)}\boldsymbol{\theta}^{(3)}}q_t^{(4)}(\boldsymbol{\theta}) & \nabla_{\boldsymbol{\theta}^{(4)}\boldsymbol{\theta}^{(4)}}q_t^{(4)}(\boldsymbol{\theta})\end{bmatrix}. \qquad \text{(D.37)}$$

*with the derivatives expressions presented in equations* (D.13)–(D.22)*.*

*Proof.* The uniform convergence of the Hessian is obtained by the uniform law of large numbers in Rao (1962). Below, we verify the conditions of the theorem.

1. by Krengel's theorem $\{\boldsymbol{A}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is an SE sequence since $\boldsymbol{A}_t(\cdot)$ is continuous on the SE sequences. The latter follows by Assumption 2, Propositions 1 and 2, and Corollary TA.3 which ensure that the sequences $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$, $\{f_t(\cdot)\}_{t\in\mathbb{Z}}$, $\{f_t'(\cdot)\}_{t\in\mathbb{Z}}$, $\{f_t''(\cdot)\}_{t\in\mathbb{Z}}$, $\{\boldsymbol{g}_t(\cdot)\}_{t\in\mathbb{Z}}$, $\{\boldsymbol{g}_t'(\cdot)\}_{t\in\mathbb{Z}}$ and $\{\boldsymbol{g}_t''(\cdot)\}_{t\in\mathbb{Z}}$ are SE, which are also jointly SE.

2. To establish that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{A}_t(\boldsymbol{\theta})\| < \infty$, by Loève's $c_r$ inequality we have

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{A}_t(\boldsymbol{\theta})\| = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left(\sum_{i=1}^{4}\nabla_{\boldsymbol{\theta}^{(i)}\boldsymbol{\theta}^{(i)}}q_t^{(i)}(\boldsymbol{\theta}^{(i:1)})^\top\nabla_{\boldsymbol{\theta}^{(i)}\boldsymbol{\theta}^{(i)}}q_t^{(i)}(\boldsymbol{\theta}^{(i:1)})\right)^{1/2}$$
$$\leq c\sum_{i=1}^{4}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(i)}\boldsymbol{\theta}^{(i)}}q_t^{(i)}(\boldsymbol{\theta}^{(i:1)})\right\|. \quad \text{(D.38)}$$

Let us consider the Hessians of each step.

Step 1. By norm subadditivity and Cauchy-Schwartz inequality

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(1)}\boldsymbol{\theta}^{(1)}}q_t^{(1)}(\boldsymbol{\theta})\right\| = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial^2 q_t^{(1)}(\boldsymbol{\theta})}{\partial f^2}f_t'(\boldsymbol{\theta})f_t'(\boldsymbol{\theta})^\top + \frac{\partial q_t^{(1)}(\boldsymbol{\theta})}{\partial f}f_t''(\boldsymbol{\theta})\right\|$$
$$\leq \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial^2 q_t^{(1)}(\boldsymbol{\theta})}{\partial f^2}f_t'(\boldsymbol{\theta})f_t'(\boldsymbol{\theta})^\top\right\| + \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial q_t^{(1)}(\boldsymbol{\theta})}{\partial f}f_t''(\boldsymbol{\theta})\right\|$$
$$\leq 2\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|f_t'(\boldsymbol{\theta})f_t'(\boldsymbol{\theta})^\top\right\| + \left(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial q_t^{(1)}(\boldsymbol{\theta})}{\partial f}\right|^2\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|f_t''(\boldsymbol{\theta})\right\|^2\right)^{1/2} < \infty.$$

Step 3. By norm subadditivity, Cauchy-Schwartz inequality, and Loève's $c_r$ inequality, we obtain

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(3)}\boldsymbol{\theta}^{(3)}}q_t^{(3)}(\boldsymbol{\theta})\right\| \leq c\sum_{s=1}^{S}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}_s^{(3)}\boldsymbol{\theta}_s^{(3)}}q_t^{(3)}(\boldsymbol{\theta})\right\| =$$
$$c\sum_{s=1}^{S}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial^2 q_t^{(3)}(\boldsymbol{\theta})}{\partial g^{s2}}g_t^{s\prime}(\boldsymbol{\theta})g_t^{s\prime}(\boldsymbol{\theta})^\top + \frac{\partial q_t^{(3)}(\boldsymbol{\theta})}{\partial g^s}g_t^{s\prime\prime}(\boldsymbol{\theta})\right\|$$
$$\leq c\sum_{s=1}^{S}\left(2\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|g_t^{s\prime}(\boldsymbol{\theta})g_t^{s\prime}(\boldsymbol{\theta})^\top\right\| + \left(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial q_t^{(3)}(\boldsymbol{\theta})}{\partial g^s}\right|^2\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|g_t^{s\prime\prime}(\boldsymbol{\theta})\right\|^2\right)^{1/2}\right) < \infty,$$

where in the last lines of Steps 1 and 3 we used (D.23). The last claim for each of the steps is obtained by Lemma TA.5 and Corollary TA.3 since together they ensure that all the terms are uniformly bounded. Particularly, by Lemma TA.5 $\left|\frac{\partial q_t^{(1)}(\boldsymbol{\theta})}{\partial f}\right|^2$ and $\left|\frac{\partial q_t^{(3)}(\boldsymbol{\theta})}{\partial g^s}\right|^2$ are bounded, while by Corollary TA.3 $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|f_t'(\boldsymbol{\theta})\|^2 < \infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|f_t''(\boldsymbol{\theta})\|^2 < \infty$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{g}_t'(\boldsymbol{\theta})\|^2 < \infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{g}_t''(\boldsymbol{\theta})\|^2 < \infty$.

Steps 2 and 4. From (D.15) and (D.22)

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(2)}\boldsymbol{\theta}^{(2)}}q_t^{(2)}(\boldsymbol{\theta})\right\| = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|2(f_t(\boldsymbol{\theta}))^2\boldsymbol{I}_N\right\| < \infty,$$
$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla_{\boldsymbol{\theta}^{(4)}\boldsymbol{\theta}^{(4)}}q_t^{(4)}(\boldsymbol{\theta})\right\| \leq c\sum_{s=1}^{S}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|2(g_t^s(\boldsymbol{\theta}))^2\boldsymbol{I}_{N_s}\right\| < \infty,$$

where in the second line we apply Loève's $c_r$ inequality and exploit the definition of Frobenius norm. The final claim in both lines follows by Proposition 1 and 2 given Assumption 3.a.

Therefore, all the terms in (D.38) are finite. Hence, we conclude $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{A}_t(\boldsymbol{\theta})\| < \infty$. ■

# E Additional Monte Carlo results

In this section we provide additional details on the Monte Carlo simulations.

In Table E.10, we provide the dynamic patters for $f_t$ and $\boldsymbol{g}_t$ used in the simulation design for the analysis of the forecasting group conditional mean.

| Dynamics | Common | Group-specific |
|---|---|---|
| AR(1)+AR(1) | $\kappa\tilde{f}_t + \xi_{t+1}$ | $\boldsymbol{\psi}^\top\tilde{\mathbf{g}}_t + \boldsymbol{\eta}_{t+1}$ |
| AR(1)+Break | $\kappa\tilde{f}_t + \xi_{t+1}$ | $a_s^\star 1(t < T_{break_s}) + b_s^\star 1(t \geq T_{break_s})$ |
| Sine+AR(1) | $1.5\sin(2\pi t/100)$ | $\boldsymbol{\psi}^\top\tilde{\mathbf{g}}_t + \boldsymbol{\eta}_{t+1}$ |
| Sine+Break | $1.5\sin(2\pi t/100)$ | $a_s 1(t < T_{break_s}) + b_s 1(t \geq T_{break_s})$ |
| Fast sine + Steps | $1.5\sin(2\pi t/20)$ | $a_s 1(\sin(2\pi t/T_s) \leq 0) + b_s 1(\sin(2\pi t/T_s) > 0)$ |
| Slow sine + Ramp | $1.5\sin(2\pi t/250)$ | $\mathrm{mod}\ (t/T_{ramp_s})$ |

Table E.10: **Simulation patterns for $f_t$ and $\mathbf{g}_t$.** The moment of break $T_{break_s}$, the period of steps $T_s$, the ramp period $T_{ramp_s}$, the size of the break and steps are different between groups and are randomly chosen: $T_{break_s} \sim U([0,T]), T_s \sim U([100,250]), T_{ramp_s} \sim U([100,200]), a_s \sim U([0,0.2]), b_s \sim U([1.5,2]), a_s^\star = 1.5a_s, b_s^\star = 1.5b_s$, for $s = 1,\dots,S$. The parameters for the AR(1) processes are the same as for the main DGP.

Furthermore, we analyze the filtering of the common and group-specific factors separately. We consider different number of groups $S$ and different values of the parameters for the group specific factors, $\psi_s$. To assess the factors' estimates we regress the simulated factors on the estimated ones and compute the $R^2$ of this regression. For simplicity, in these experiments, we set $\psi_1 = \cdots = \psi_S$ meaning that the group-specific factors have the same persistency. The results for the common factor are provided in Figure E.4. We notice that the $R^2$ for the common factor is high when the number of groups $S$ is large. In this case, the value of the parameter $\psi_s$ does not have any effect on the $R^2$. In contrast, when the number of groups $S$ is small the value of $\psi_s$ plays a role. Particularly, the $R^2$ increases with a decrease in $\psi_s$. This confirms that when the conditional expectation of the group-specific factor is negligible the predicted common factor is closer to the true one which is in line with our discussion in Section 2.5. The results for the group-specific factor reveal that the $R^2$ increases with the increase in $\psi_s$ and number of groups $S$ (Figure E.5). The latter can be explained by the fact that when $S$ is large the prediction for the common factors is more accurate which has a consequent effect on the estimation of the group-specific factors. Overall, if the interest is in predicting accurately common and group-specific factors rather than the conditional group mean, the number of groups $S$ should be large.

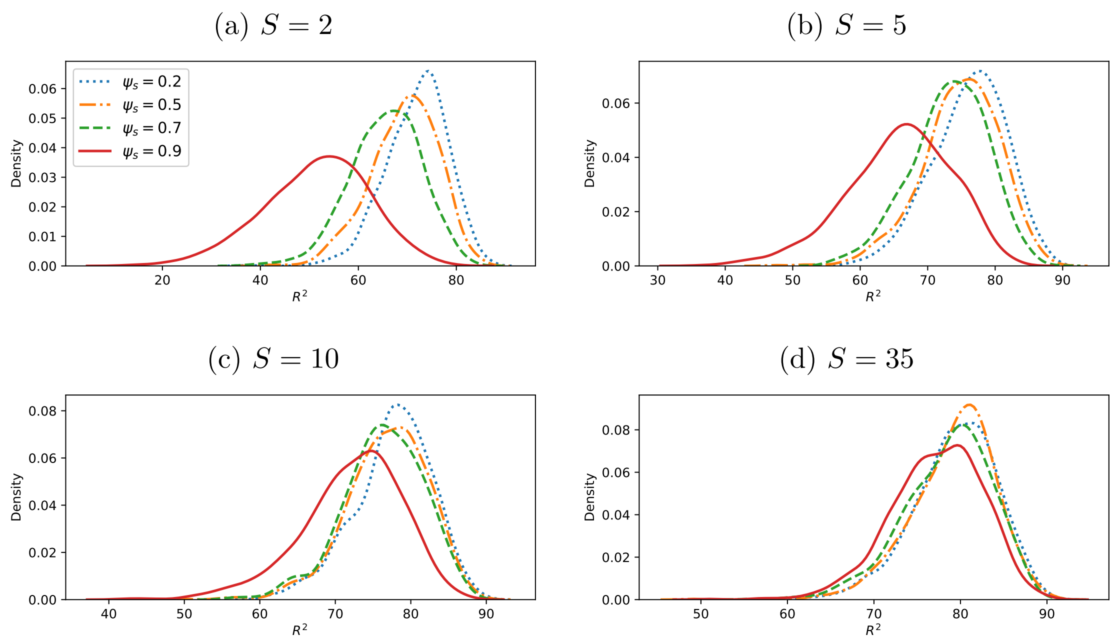



Figure E.4: **Kernel density plot of the $R^2$ of the regression of the simulated common factor $\tilde{f}_t$ on the estimated factor $\hat{f}_t$ for different number of groups $S$ and different values of the parameter $\psi_s$.** The results are based on 1000 Monte Carlo simulations for time series from DGP (4) with $T = 300$ and $N_s = 10$.

Finally, in Figure E.6, we present additional results for the out-of-sample analysis for the setup outlined in Section 3.3. We find that the range of the MSEs is smaller for our model for the one-step-ahead forecasts and the difference becomes smaller once the forecast horizon increases.

(a) $S = 2$ (b) $S = 35$

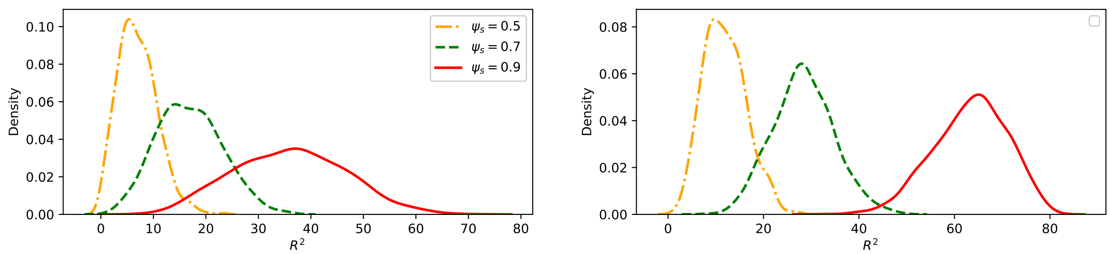



Figure E.5: **Kernel density of the $R^2$ of the regression of $\tilde{g}_t^1$ on $\hat{g}_t^1$.** The results for other group-specific factors are the same and are omitted here. For further details, we refer to Figure E.4.

(a) One-step-ahead (b) Two-step-ahead

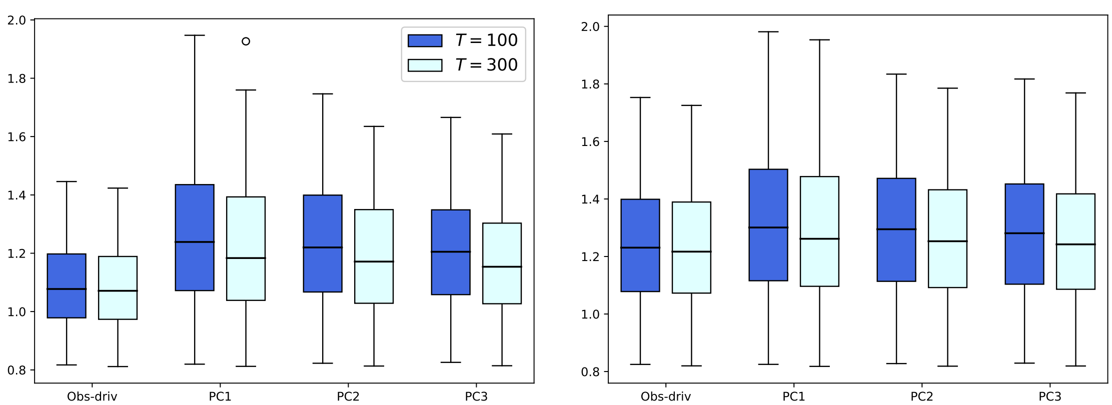



Figure E.6: **Average MSE of the observation driven, PC1, PC2, and PC3 models for the different forecast horizons and different sizes of the rolling windows.** For further explanations, we refer to Table 1.

# References


Al-Osh, M. (1989). A dynamic linear model approach for disaggregating time series data. *Journal of Forecasting*, *8*(2), 85–96.

Blasques, F., van Brummelen, J., Koopman, S. J., & Lucas, A. (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics*, *227*(2), 325–346.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians.* OUP Oxford.

Diebold, F. X., & Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, *182*(1), 119–134.

Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, *74*(1), 119–147.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Springer.

Pötscher, B. M., & Prucha, I. (1997). *Dynamic nonlinear econometric models: Asymptotic theory.* Springer Science & Business Media.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 659–680.

Silva, J. S., & Cardoso, F. (2001). The Chow-Lin method using dynamic models. *Economic Modelling*, *18*(2), 269–280.

Straumann, D., & Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics*, *34*(5), 2449–2495.

White, H. (1996). *Estimation, inference and specification analysis* (No. 22). Cambridge university press.