Tracking Sectoral Economic Conditions*

Daan Opschoor†

Econometric Institute, Erasmus University Rotterdam

November 7, 2023

Job market paper
[Latest version here]

Abstract

We construct a novel set of monthly U.S. sector-level economic conditions indices from a small but diverse set of sectoral economic indicators using mixed-frequency dynamic factor models. The resulting indices are driven by a balanced mix of the underlying indicators and display considerable heterogeneity, particularly in the depths, timing and duration of their downturns. Moreover, the sectoral economic conditions are driven by a common factor that explains most fluctuations in the overall economy and is closely related to aggregate production. Meanwhile, the service-providing sectors are additionally driven by a correction factor that handles the heterogeneous impacts of the financial crisis and covid pandemic. Lastly, sector-level GDP growth nowcasts are constructed, which are found to consistently outperform a simple autoregressive benchmark for almost all sectors, especially during the covid pandemic.

Keywords: Economic conditions, monthly index, sectoral heterogeneity, mixed-frequency dynamic factor model, nowcasting

JEL Classification: C32, C53, E32, E66

*I thank Frank Diebold, Dick van Dijk, Laurent Ferrara, Philip Hans Franses, Daniel Lewis, Robin Lumsdaine, Minchul Shin and Michel van der Wel for their valuable discussions and comments, as well as seminar participants at the Erasmus University Rotterdam (2023). Part of this research was conducted during a visit at the University of Pennsylvania and I thank the Department of Economics for the provided hospitality and stimulating research environment.

†E-mail address: opschoor@ese.eur.nl (Daan Opschoor)

1 Introduction

Monitoring the current state of the economy is a challenging but pivotal task of decision-makers in both the private and public sector, especially during times of economic turmoil. Dating back to Burns and Mitchell (1946), it has long been recognized that examining current economic conditions should revolve around the comovement of different measures of economic activity, rather than just focusing on a single economic indicator only (Lucas, 1977; Diebold and Rudebusch, 1996). Starting with Stock and Watson (1989, 1991), a popular approach to implement this idea is to construct an economic conditions index from multiple indicators using a dynamic factor model (see, among others, Mariano and Murasawa, 2003; Aruoba et al., 2009; Lewis et al., 2022).

These indices generally focus on aggregate economic activity. However, policy makers and investors are not only interested in aggregate activity, but also in the conditions of each underlying sector. For example, they want to know which sectors are hit hardest by economic downturns or to anticipate which sectors are leading or lagging the business cycle. Indeed, there is strong evidence that sectors within an economy exhibit diverse behaviour over the business cycle (see, among others, Fok et al., 2005; Chang and Hwang, 2015; Camacho and Leiva-Leon, 2019). Interestingly, most empirical work on sectoral economic activity focuses on one specific economic indicator only like industrial production or employment, while much less attention has been given to more broadly tracking sectoral economic conditions.

In this paper, we construct a novel set of 20 monthly U.S. sector-level economic conditions indices at the two-digit level of the North American Industry Classification System (NAICS) for the period January 1991 to September 2021. These indices are based on a small but diverse set of sectoral economic indicators observed at monthly and quarterly frequencies. Specifically, we include output and labor-market variables as well as sales, revenue and personal income. We obtain the indices by estimating sector-specific mixed frequency dynamic factor models that summarize the (co)variation in the underlying sectoral indicators into a single index.

The resulting sector-level indices are generally driven by a balanced mix of the underlying variables, highlighting the importance of combining different measures of sectoral economic activity. Moreover, the indices show that sectors display considerable heterogeneity over the business cycle, emphasizing the relevance of sectoral disaggregation.

Specifically, we find substantial differences across and within recession periods in the presence and depth of sector-level downturns, as well as in their timing and duration. Most sectors experience deeper drops in economic activity during the covid pandemic than during the financial crisis and especially the dot-com bubble. In particular, service-providing sectors like health care, arts and accommodation display the deepest drops during the pandemic, while they are only mildly affected by the other recessions. Meanwhile, sectors that are closely related to production like construction and manufacturing experience drops in economic activity that are roughly similar in magnitude during the pandemic as during the financial crisis. Following the cycle phase methodology of Harding and Pagan (2006), it becomes clear that the timing of the peaks before recessions are quite dispersed, while the troughs at the end of recessions are highly concentrated, especially after the financial crisis and covid pandemic. Consequently, the sectoral growth cycles exhibit considerable differences in the lengths of their phases.

Next, we examine the comovement of sectoral economic conditions with the overall state of the economy. Most sectors are highly cyclical and closely follow aggregate economic conditions, particularly manufacturing, wholesale trade and retail trade. This comovement between sectoral and aggregate economic activity is generally stronger during recessions than during expansions. Based on the synchronization measure of Harding and Pagan (2006), the sector most in sync with the overall economy is manufacturing, highlighting its dominant role in the aggregate growth cycle. On the other hand, several other sectors like agriculture, utilities and healthcare move more independently with almost no evidence of synchronization with the aggregate economy.

Monitoring the economic state of each sector is the most obvious application of the estimated indices. However, several other purposes can be imagined. In this paper, we consider two of these further uses in detail. First, we demonstrate the possibility to revisit existing analyses that have been based on one type of sectoral economic activity only, but which can now be generalized to broader economic conditions. Specifically, we revisit the analyses in Foerster et al. (2011) and Andreou et al. (2019). Both studies aim to investigate whether common shocks to sectoral economic conditions are able to explain economy-wide fluctuations. The former study focuses on industrial production data, ignoring the service-providing sectors, whereas the latter extends this by additionally including annual GDP data for non-industrial production sectors. Instead, our analysis uses a broader measure of economic activity rather than only output-related variables and is based on monthly indices for all sectors, including the non-industrial production ones. Corroborating Foerster et al. (2011) and Andreou et al. (2019), we find that the first common factor of the sector-level indices explains most of the variation in aggregate economic conditions and is closely related to the production-related sectors. In addition, the second common factor serves as a correction factor for the service-providing sectors that are more severely hit by the covid pandemic and less severely by the financial crisis than accounted for by the first common factor only.

Second, the mixed-frequency dynamic factor models that generate the sectoral indices simultaneously produce estimates of latest sector-level GDP growth, which are used as inputs for the indices. To examine the accuracy of these estimates, we conduct a nowcasting exercise of sector-level GDP growth. These GDP figures are typically published with a three-month delay, while other sectoral information is released much earlier, indicating that there is much to gain in terms of improving nowcasts. Indeed, the nowcasts from the mixed-frequency dynamic factor model are generally more accurate than the ones from a simple autoregressive benchmark, with an average improvement of 22% in terms of root mean squared forecast errors. This relative outperformance is consistent throughout the out-of-sample period from 2010Q1 to 2021Q2, but particularly pronounced during the covid pandemic. In fact, the nowcasts made during the covid pandemic are often close to the realized GDP growth values. This emphasizes the accuracy and potential of these nowcasts and justifies their use as inputs for the sector-level indices.

This paper is closely related to and builds on two strands of literature. First, it adds to the vast literature on constructing economic conditions indices using dynamic factor models, initiated by the seminal work of Stock and Watson (1989, 1991). Prominent advances in this line of research are incorporating mixed-frequency data (Mariano and Murasawa, 2003; Nunes, 2005; Aruoba et al., 2009), moving to higher frequencies such as weekly (Lewis et al., 2022; Wegmüller et al., 2023) or constructing indices at the U.S. state-level (Crone and Clayton-Matthews, 2005; Baumeister et al., 2022). We contribute to this literature by constructing a set of monthly U.S. sector-level economic conditions indices, which has to the best of our knowledge not been pursued yet. Somewhat related work by Carriero and Marcellino (2011) constructs five sector-level confidence indices on the current status of the economy based on survey-data for European countries, but these indices are not composed of any real economic activity data. In addition, some studies produce an index for just one sector only like the weekly retail trade index of Brave et al. (2021) or the transportation services index from the U.S. Department of Transportation (Young et al., 2014). Yet, these indices are often based on sector-specific sources of economic activity, making them not directly comparable to the economic activity of other sectors. The benefit of the estimated sectoral indices in this paper is that they are based on a similar set of economic indicators for all sectors, making them directly comparable.

Second, this paper contributes to the literature on the heterogeneous dynamics of sectoral economic activity. So far, most of this work focuses on specific economic variables like industry-level production (Fok et al., 2005; Chang and Hwang, 2015; Foerster et al., 2011; Guisinger et al., 2021; Brunner and Hipp, 2023; Graeve and Schneider, 2023), sector-level employment (Camacho and Leiva-Leon, 2019; Anderson et al., 2020) or sector-level GDP (Karadimitropoulou and León-Ledesma, 2013; Li and Martin, 2019; Böhm et al., 2022). However, sectoral economic conditions are not just described by one type of activity only, but by the comovement of several activities, just as for total economic conditions. Moreover, focusing on one source of sectoral economic activity often comes with its limitations. For example, using solely industry-level production data ignores a large part of the economy related to the service sectors. In addition, sector-level GDP is only available on an annual basis or, from 2005 onwards, on a quarterly basis (at least in the U.S.), restricting its usage to assess sectoral activity on a higher frequency or longer time span. We contribute to this literature by combining various sectoral economic indicators into a single sector-level index, providing an economic activity measure that is available at a monthly frequency with a long time span for all sectors. This makes it possible to more easily conduct sectoral analyses, as demonstrated in this paper by revisiting the results in Foerster et al. (2011) and Andreou et al. (2019) using a more complete measure of economic activity.

The rest of the paper is organized as follows. Section 2 describes the mixed-frequency dynamic factor model and its estimation framework to construct the sectoral indices. Section 3 discusses the sector-level economic data. Section 4 presents the results of tracking sector-level economic conditions, while Section 5 illustrates alternative applications of the indices. Section 6 concludes.

2 Constructing sectoral economic conditions indices

2.1 Mixed-frequency dynamic factor model

To estimate the sector-level economic condition indices, we construct for each sector $i = 1, \dots, S$ a mixed-frequency dynamic factor model. For notational simplicity, we suppress the dependence on i, but keep in mind that everything is sector-specific. Let $\boldsymbol{Y}_t = (\boldsymbol{Y}_t^{M\prime}, \boldsymbol{Y}_t^{Q\prime})'$ denote a vector with N observed time series that track the economic conditions for a specific sector, including K monthly observed series in $\boldsymbol{Y}_t^M = (Y_{1,t}^M, \dots, Y_{K,t}^M)'$ (for example, employment or industrial production) and L quarterly observed series in $\boldsymbol{Y}_t^Q = (Y_{1,t}^Q, \dots, Y_{L,t}^Q)'$ (for example, GDP). The number of monthly and quarterly series are allowed to differ across sectors. The quarterly variables are only observed every third month in the quarter (that is, on $t = 3, 6, 9, \dots$), while they are considered missing during the other months. Following Lewis et al. (2022) and Baumeister et al. (2022), we focus on the year-on-year percentage changes in sector-specific economic activity.[1] In particular, we compute annual growth rates of $\boldsymbol{Y}_t$, resulting in the vector $\boldsymbol{y}_t = (\boldsymbol{y}_t^{M\prime}, \boldsymbol{y}_t^{Q\prime})' = (y_{1,t}^M, \dots, y_{K,t}^M, y_{1,t}^Q, \dots, y_{L,t}^Q)'$ with $y_{k,t}^M = \Delta_{12} \log Y_{k,t}^M$ for $k = 1, \dots, K$ and $y_{l,t}^Q = \Delta_{12} \log Y_{l,t}^Q$ for $l = 1, \dots, L$.

Starting with the monthly observed series, we assume that $\boldsymbol{y}_t^M$ has a factor model representation given by

$$\boldsymbol{y}_t^M = \boldsymbol{\lambda}^M f_t + \boldsymbol{\varepsilon}_t^M, \tag{1}$$

for $t = 1, \dots, T$, where f_t denotes the sector-specific latent factor summarizing its economic conditions, $\boldsymbol{\lambda}^M = (\lambda_1^M, \dots, \lambda_K^M)'$ the corresponding factor loadings and $\boldsymbol{\varepsilon}_t^M = (\varepsilon_{1,t}^M, \dots, \varepsilon_{K,t}^M)'$ the idiosyncratic components that are uncorrelated with f_t at all leads and lags. In fact, we assume that the whole observation vector $\boldsymbol{y}_t$ can be summarized by a single factor only, but given the moderately small dimension of $\boldsymbol{y}_t$ for each sector (that is, four to six series) this is a plausible assumption. Moreover, following Bańbura et al. (2011) and Cascaldi-Garcia et al. (2023), we assume that both f_t and the elements in $\boldsymbol{\varepsilon}_t^M$

[1] We opt for annual growth rates to get indices that are less sensitive to volatile monthly changes (Sarantis, 2001; Nunes, 2005) and therefore smoother and more persistent. Moreover, taking annual growth rates eliminates the seasonal component in the seasonally unadjusted series (Lewis et al., 2022), which in our case are the sector-level continued unemployment insurance claims and fuel sales.

follow a stationary first-order autoregression, that is,

$$\begin{pmatrix} f_t \\ \boldsymbol{\varepsilon}_t^M \end{pmatrix} = \begin{pmatrix} \phi & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^M \end{pmatrix} \begin{pmatrix} f_{t-1} \\ \boldsymbol{\varepsilon}_{t-1}^M \end{pmatrix} + \begin{pmatrix} \eta_t \\ \boldsymbol{\nu}_t^M \end{pmatrix}, \tag{2}$$

where $\boldsymbol{\Psi}^M$ is a $K \times K$ diagonal matrix and the error terms $\eta_t \sim$ i.i.d. $\mathcal{N}(0, \omega^2)$ and $\boldsymbol{\nu}_t^M \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^M)$ are uncorrelated, with $\boldsymbol{\Sigma}^M$ also a $K \times K$ diagonal matrix. In other words, we assume an exact factor model structure for $\boldsymbol{y}_t^M$ in which all cross-sectional dependence is captured by f_t, which is a common assumption in this literature (see, among others, Stock and Watson, 1989, 1991; Mariano and Murasawa, 2003; Aruoba et al., 2009). By explicitly modelling the dynamics of the idiosyncratic components, we allow for possible variable-specific shocks such as strikes in the transportation sector or sudden drops in oil production due to a hurricane (see, for example, Rothstein, 1997; Cruz and Krausmann, 2008). Similarly as Stock and Watson (1991), we set $\omega^2 = 1$ to identify the scale of f_t. In addition, we normalize the series in $\boldsymbol{y}_t$ to have mean zero and variance one such that the parameters are conveniently scaled for interpretation and equations (1) and (2) do not require constants.

Moving to the quarterly observed series, we follow Mariano and Murasawa (2003) and Nunes (2005), and assume that $Y_{l,t}^Q$ is the geometric mean of a latent monthly variable $Y_{l,t}^*$ and its two lags, that is,

$$\log Y_{l,t}^Q = \frac{1}{3}\big(\log Y_{l,t}^* + \log Y_{l,t-1}^* + \log Y_{l,t-2}^*\big),$$

for $l = 1, \ldots, L$.[2] Consequently, taking year-on-year differences results in

$$y_{l,t}^Q = \frac{1}{3}(y_{l,t}^* + y_{l,t-1}^* + y_{l,t-2}^*), \tag{3}$$

where $y_{l,t}^* = \Delta_{12} \log Y_{l,t}^*$. We only observe $y_{l,t}^Q$ for $t = 3, 6, 9, \ldots$, while $y_{l,t}^*$ is never observed.

Next, we follow Nunes (2005) and introduce the cumulator variable c_t of Harvey (1989, Section 6.3.3) that facilitates the temporal aggregation of the monthly latent factor f_t

[2] The standard accounting identity for quarterly variables such as GDP prescribes to take the arithmetic mean instead of the geometric mean. However, taking the arithmetic mean results in a non-linear temporal aggregation constraint (Proietti and Moauro, 2006), while the geometric mean keeps this constraint linear. Moreover, Mitchell et al. (2005) and Camacho and Perez-Quiros (2010) argue that the latter is a good first-order approximation of the former for quarterly GDP.

towards the quarterly observed variables in $\boldsymbol{y}_t^Q$. In particular, the quarterly series of each sector can be expressed as

$$\boldsymbol{y}_t^Q = \boldsymbol{\lambda}^Q c_t + \boldsymbol{\varepsilon}_t^Q, \tag{4}$$

for $t = 1, \ldots, T$, with $\boldsymbol{\lambda}^Q = (\lambda_1^Q, \ldots, \lambda_L^Q)'$ and $\boldsymbol{\varepsilon}_t^Q = (\varepsilon_{1,t}^Q, \ldots, \varepsilon_{L,t}^Q)'$ being the corresponding factor loadings and idiosyncratic shocks, respectively. Similarly as for the monthly series, we assume that the elements in $\boldsymbol{\varepsilon}_t^Q$ follow a first-order autoregressive process. The corresponding error terms are again independent and normally distributed, that is, $\boldsymbol{\nu}_t^Q \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^Q)$ with $L \times L$ diagonal matrix $\boldsymbol{\Sigma}^Q$, implying an exact factor model structure for $\boldsymbol{y}_t^Q$ as well. Furthermore, the cumulator variable is given by

$$c_t = \xi_t c_{t-1} + \frac{1}{3} f_t, \tag{5}$$

where ξ_t is equal to zero for the first month of the quarter (that is, for $t = 1, 4, 7 \ldots$) and one otherwise. Hence, every third month in the quarter, when $\boldsymbol{y}_t^Q$ is observed, the cumulator variable becomes equal to

$$c_t = \frac{1}{3}(f_t + f_{t-1} + f_{t-2}),$$

which resembles the expression under the geometric mean assumption in equation (3). In other words, the year-on-year growth rate of the unobserved monthly counterpart of the quarterly variable ($y_{l,t}^*$) is embodied by the monthly latent factor f_t.

We opt for this cumulator-based temporal aggregation scheme of Harvey (1989, Section 6.3.3) as it remains tractable in the case of multiple quarterly series. By contrast, the monthly-to-yearly growth rate aggregation scheme of Mariano and Murasawa (2003) requires the contemporaneous and first 12 lags of the latent factor, as well as the contemporaneous and first 12 lags of the idiosyncratic component for each quarterly series.[3] Clearly, this would lead to a huge dimensional state vector for multiple quarterly variables, slowing down the filtering/smoothing recursions and increasing the computation time. Meanwhile, the cumulator-based approach only requires to increase the state-vector dimension by one, independent of the number of quarterly variables, keeping the dimen-

[3]Specifically, Mariano and Murasawa (2003) define $y_{l,t}^* = \Delta \ln Y_{l,t}^*$ in equation (3), resulting in $y_{l,t}^Q = \frac{1}{3}(y_{l,t}^* + 2y_{l,t-1}^* + 3(y_{l,t-2}^* + \ldots + y_{l,t-11}^*) + 2y_{l,t-12}^* + y_{l,t-13}^*)$ for yearly growth rates.

sion sufficiently low and facilitating tractable estimation.

After establishing the temporal aggregation scheme, we put the mixed-frequency dynamic factor model in state-space form. To model the dynamics of the idiosyncratic components, we follow the approach of Bańbura and Modugno (2014) and include them in the state vector. Consequently, the observation equation is given as

$$\begin{pmatrix} \boldsymbol{y}_t^M \\ \boldsymbol{y}_t^Q \end{pmatrix} = \begin{pmatrix} \boldsymbol{\lambda}^M & \mathbf{0} & \boldsymbol{I}_K & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\lambda}^Q & \mathbf{0} & \boldsymbol{I}_L \end{pmatrix} \begin{pmatrix} f_t \\ c_t \\ \boldsymbol{\varepsilon}_t^M \\ \boldsymbol{\varepsilon}_t^Q \end{pmatrix} + \begin{pmatrix} \boldsymbol{e}_t^M \\ \boldsymbol{e}_t^Q \end{pmatrix}, \tag{6}$$

where $\boldsymbol{I}_n$ denotes an n-dimensional identity matrix. Following Bańbura and Modugno (2014), an artificial error term $\boldsymbol{e}_t = (\boldsymbol{e}_t^{M\prime}, \boldsymbol{e}_t^{Q\prime})' \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \tau \boldsymbol{I}_N)$ is introduced with τ a small pre-fixed number (that is, 10^{-4}) such that the complete data log-likelihood can be written in its exact form (see Appendix A for further details). Finally, by plugging the dynamics of f_t into the cumulator variable expression in (5), the state equation can be written as

$$\begin{pmatrix} f_t \\ c_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix} = \begin{pmatrix} \phi & 0 & \mathbf{0} \\ \frac{1}{3}\phi & \xi_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix} \begin{pmatrix} f_{t-1} \\ c_{t-1} \\ \boldsymbol{\varepsilon}_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & \mathbf{0} \\ \frac{1}{3} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_N \end{pmatrix} \begin{pmatrix} \eta_t \\ \boldsymbol{\nu}_t \end{pmatrix}, \tag{7}$$

where $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_t^{M\prime}, \boldsymbol{\varepsilon}_t^{Q\prime})'$, $\boldsymbol{\nu}_t = (\boldsymbol{\nu}_t^{M\prime}, \boldsymbol{\nu}_t^{Q\prime})'$, $\boldsymbol{\Psi}$ is an $N \times N$ diagonal matrix and $\boldsymbol{\upsilon}_t = (\eta_t, \boldsymbol{\nu}_t')' \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ with

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{pmatrix},$$

where $\boldsymbol{\Sigma}$ is also an $N \times N$ diagonal matrix. Together, observation equation (6) and state equation (7) constitute the sector-specific mixed-frequency dynamic factor model.

2.2 Estimation

Given the model in equations (6)-(7), we need to estimate the unknown parameters in $\boldsymbol{\Theta} = \{\boldsymbol{\lambda}^M, \boldsymbol{\lambda}^Q, \phi, \boldsymbol{\Psi}, \boldsymbol{\Sigma}\}$ and latent factors in $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T)'$ with $\boldsymbol{z}_t = (f_t, c_t, \boldsymbol{\varepsilon}_t')'$

for $t = 1, \ldots, T$. Since $\boldsymbol{Z}$ is unobserved, it is generally not possible to derive a closed-form estimator of $\boldsymbol{\Theta}$. Instead, one could directly maximize the likelihood function with respect to $\boldsymbol{\Theta}$ in combination with the Kalman filter to estimate $\boldsymbol{Z}$ (see, for example, Engle and Watson, 1981; Stock and Watson, 1989, 1991). However, a moderately large dimension of the observation vector $\boldsymbol{y}_t$ already results in a large number of parameters to estimate, making direct optimization of the likelihood computationally cumbersome. Therefore, we opt for the expectation-maximization (EM) algorithm of Dempster et al. (1977), which has been adapted by Shumway and Stoffer (1982) and Watson and Engle (1983) for dynamic factor models in state-space form.

The idea of the EM algorithm is to focus on the complete data log-likelihood of $\boldsymbol{Z}$ and $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)'$ and to iterate between estimating the latent factors $\boldsymbol{Z}$ by means of the Kalman smoother (E-step) and estimating the parameters in $\boldsymbol{\Theta}$ by maximizing the expected complete data log-likelihood (M-step). Under some regularity conditions, Dempster et al. (1977) show that by iterating between these two steps the EM algorithm converges to a local maximum of the likelihood. Moreover, the M-steps have analytic expressions for the unknown parameters in $\boldsymbol{\Theta}$, making the algorithm computationally fast and stable, even in high dimensions. To handle possible missing observations in $\boldsymbol{Y}$, we follow Bańbura and Modugno (2014) and integrate out the missing data from the likelihood function. For further details on the EM algorithm, see Appendix A.1.

The expressions of the M-steps look generally similar to the ones in, for example, Shumway and Stoffer (1982) and Bańbura and Modugno (2014). However, there are two notable differences from the standard M-steps used to estimate state-space models. First, Opschoor and van Dijk (2023) show that the low-noise observation equation in the serially-correlated idiosyncratic-component framework of Bańbura and Modugno (2014) leads to extremely slow convergence in estimating the factor loadings $\boldsymbol{\lambda}^M$ and $\boldsymbol{\lambda}^Q$, resulting in poor estimates of parameters and latent factors. To tackle this issue, we augment the M-steps of the factor loadings with the ones of the overrelaxed adaptive EM (AEM) algorithm of Salakhutdinov and Roweis (2003) as recommended by Petersen et al. (2005). Indeed, Opschoor and van Dijk (2023) show that using the adaptive M-steps for the loadings speeds up convergence and leads to more accurate parameter and latent factor estimates. Second, the factor persistence parameter ϕ corresponds to both f_t and c_t, implying that it is bounded to certain restrictions. Moreover, the state equation of the cumulator variable contains a time-varying indicator ξ_t, which also needs to be handled. To do this, we follow the derivation in Holmes (2013) for time-varying system matrices with linear constraints to obtain the constrained M-step of ϕ. For further details on the derivations of the M-steps and the AEM algorithm, see Appendix A.2, and for a convergence comparison between the EM and AEM algorithm, see Appendix A.3.

After obtaining estimates of the parameters in $\boldsymbol{\Theta}$ and latent factors in $\boldsymbol{Z}$, we follow the approach of Baumeister et al. (2022) and construct for each sector $i = 1, \ldots, S$ the monthly economic conditions index (ECI) as

$$ECI_t = (\hat{\boldsymbol{\lambda}}'\hat{\boldsymbol{\lambda}})^{-1}\hat{\boldsymbol{\lambda}}'(\boldsymbol{y}_t^S - \bar{\boldsymbol{y}}^S), \tag{8}$$

for $t = 1, \ldots, T$, where $\hat{\boldsymbol{\lambda}} = (\hat{\boldsymbol{\lambda}}^{M'}, \hat{\boldsymbol{\lambda}}^{Q'})'$ contains the estimated factor loadings, $\boldsymbol{y}_t^S$ contains the N observed variables with the missing values replaced by the smoothed values from the Kalman smoother and $\bar{\boldsymbol{y}}^S = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}_t^S$.[4] The signs of the estimated loadings (and factors) are chosen such that the majority of the loadings are consistent with the cyclicality of the corresponding economic indicator. Compared to using f_t, the approach in (8) leads to a more robust measure of economic conditions that minimizes the effect of data revisions (Baumeister et al., 2022), while it also facilitates an easy way to decompose the ECI_t into its underlying series. For the sake of comparison, we standardize the indices to have variance one such that they are presented into standard deviation units from their historical averages.

3 Data

We distinguish 20 sectors in the U.S. economy based on the two-digit level of the North American Industry Classification System (NAICS). To monitor the economic activity of each of these sectors, we collect a consistent set of variables that are available for most sectors and that are widely used to construct economic conditions indices of the overall economy (see, for example, Stock and Watson, 1989, 1991; Mariano and Murasawa, 2003; Aruoba et al., 2009). Specifically, we use output-related variables such as gross domestic product (GDP) and industrial production (IP), labor-market variables such

[4]Note that for the missing values of the quarterly variables we do not use the cumulator variable c_t for $t = 1, 2, 4, 5, \ldots$, but f_t for $t = 1, 4, \ldots$ and $\frac{1}{2}(f_t + f_{t-1})$ for $t = 2, 5, \ldots$. This makes the transition between observed quarterly variables smoother.

as employment and unemployment insurance (UI) claims, as well as sales, revenue and personal income.

Table 1 shows the selection of sector-level economic indicators and their availability. For a complete overview of the sectors and which series are available for which sector, see Appendix B. Regarding the coverage, 12 sectors consist of five underlying variables, while four sectors consist of four variables and the other four sectors of six. Specifically, real GDP, employment, continued UI claims and personal income are available for all sectors and can be obtained from the U.S. Bureau of Economic Analysis, U.S. Bureau of Labor Statistics and U.S. Department of Labor. Meanwhile, industrial production is only available for a few goods-producing and service-providing sectors from the Federal Reserve Board (that is, mining, manufacturing, information and utilities).[5] The production of utilities includes electric power generation, transmission and distribution (NAICS 2211) and natural gas distribution (NAICS 2212), whereas the IP index assigned to the information sector is only related to newspaper and book publishers (NAICS 5111). Still, the publishers industry accounts for roughly 26% of information sector GDP as well as 32% of the total number of employees in the information sector. Furthermore, sales data is available for five sectors, including manufacturing, retail trade, wholesale trade and accommodation and food services from the U.S. Census Bureau, and mining from the U.S. Energy Information Administration. Lastly, revenue data is available for 11

Table 1: Overview of sector-level economic variables

Variables	**Availability**	**Frequency**	**Starting dates**		**Publ. delay (in months)**	**Seasonal adjustment**	**Source**
			Earliest	**Latest**			
Real GDP	20	Q	2005Q1	2005Q1	3	SA	BEA
Industrial production	4	M	1919M1	1972M1	1	SA	FRB
Employment level	20	M	1939M1	1990M1	0	SA	BLS
Continued UI claims	20	M	2003M12	2005M2	0	NSA	DOL
Sales	5	M	1981M1	1992M1	1/2	SA/NSA	CB/EIA
Revenue	11	Q	2003Q4	2012Q3	2	SA	CB
Personal income	20	Q	1998Q1	1998Q1	3	SA	BEA

Notes: This table shows the availability (in number of sectors) and details of sector-level economic variables at the two-digit NAICS level, including their frequencies (quarterly (Q) and monthly (M)), earliest and latest starting dates across the sectors, publication delays (in months) and seasonal adjustments indicating whether the raw series are available in seasonally-adjusted (SA) form or only in non-seasonally adjusted (NSA) form. The data sources are the U.S. Bureau of Economic Analaysis (BEA), the Federal Reserve Board (FRB), the U.S. Bureau of Labor Statistics (BLS), the U.S. Department of Labor (DOL), the U.S. Energy Information Administration (EIA) and the U.S. Census Bureau (CB).

[5] There also exists an IP index for logging (NAICS 1133), which belongs to the agriculture, forestry, fishing and hunting sector. However, this industry is relatively small with only 2.5% of the total number of employees in the agriculture sector. Hence, we do not include this IP index as economic indicator in the agriculture sector.

service-providing sectors from the Quarterly Service Survey of the U.S. Census Bureau.

The various sector-level economic indicators are observed at a monthly or quarterly frequency. In particular, real GDP, revenue and personal income are observed at a quarterly frequency, while industrial production, employment, continued UI claims and sales are observed at a monthly frequency. The raw series are generally available in seasonally adjusted form, except for continued UI claims and fuel sales of crude oil and petroleum products corresponding to the mining sector. In fact, the continued UI claims are only available at the disaggregate U.S. state level, so they should be summed up again to constitute the overall sector in the U.S. economy.

The starting dates of the series are quite heterogeneous across indicators and sectors, ranging from January 1919 for IP in the mining sector to the third quarter of 2012 for revenue in the real estate, rental and leasing sector (see Appendix B).[6] Consequently, for all sectors, we start the estimation sample from January 1991 onwards, which is the first date that all sectors have at least one available variable. The remaining missing data at the beginning of the sample (or anywhere else) is easily dealt with in the estimation framework using the EM algorithm and the Kalman filter/smoother. The end of the sample is also prone to missing data due to publication delays (also known as the ragged-edge), ranging from only a few days for sectoral employment or UI claims to three months for sectoral real GDP or personal income data. These publication delays should be taken into account in nowcasting or forecasting exercises (like the one conduced in Section 5.2) to mimic the real-time data availability.

All series are transformed into yearly growth rates by means of taking the year-on-year differences of the logarithms of the series as already described in Section 2. This transformation eliminates possible seasonal components in the raw series such as increases (decreases) in UI claims for the construction sector during fall and winter (spring and summer). Moreover, the labor-market related variables are prone to outliers due to sudden incidental drops or hikes in their level (for example, due to strikes in the transportation sector or hiring peaks at the U.S. government during the decennial census periods). As a result, we remove the observations in employment and UI claims growth that are more

[6] The continued UI claims have different starting dates across states, hence we take the earliest date that all states have available data for that specific sector. This is done to circumvent a sudden increase in the number of total claims when a new state would become available. Note that we do not include the territories (Puerto Rico and the Virgin Islands) and the Federal District (District of Columbia). Also, we do not include the state Washington as these UI claims are only available from January 2017 onwards.

than three local standard deviations away from their local median, computed using the five preceding, current and five succeeding observations (and this procedure is not applied to the first five and last five observations of the series).

Due to the extreme observations during the covid pandemic, several studies like Maroz et al. (2021) and Schorfheide and Song (2022) have argued to ignore the observations from March 2020 to June 2020 in the estimation of the model parameters. Consequently, we also construct sector-level indices in which these observations are considered missing during estimation. However, the resulting estimates are often indistinguishable from the full-sample based ones with correlations ranging from 0.93 to 1.00 (see Appendix C). The only exception is for the agriculture, forestry, fishing and hunting sector with a correlation of -0.24, which is due to weak comovement in its underlying series (see Appendix D). Hence, for the agriculture sector, we ignore the covid observations in the estimation of the parameters, while for the other sectors we use all data.

4 Tracking sectoral economic conditions

4.1 Heterogeneous sectoral economic conditions

Applying the mixed-frequency dynamic factor in (6)-(7) to the sectoral economic data described in the previous section, we obtain the sector-level economic conditions indices (ECIs) from equation (8). Figure 1 presents the monthly ECIs of the four largest non-governmental U.S. sectors, as measured by their average GDP level over the full sample, namely (i) manufacturing, (ii) real estate, rental and leasing, (iii) professional, scientific and technical services and (iv) health care and social assistance. For comparison, we also include an aggregate ECI that is constructed in a similar fashion as the sector-level ones.[7] Since the ECIs are standardized, they are presented into standard deviation units from their historical average, meaning that positive (negative) values are associated with higher-than-average (lower-than-average) growth. The gray shaded areas indicate NBER

[7]The aggregate ECI is based on the same key aggregate economic variables as used by Aruoba et al. (2009), namely initial UI claims, employment, industrial production, personal income, manufacturing and trade sales and GDP (see Appendix B for further details). To be consistent with the sectoral data, we take the year-on-year difference of the logarithms of the series (except for initial UI claims which are used in levels), after which they are normalized to have mean zero and variance one for estimation.

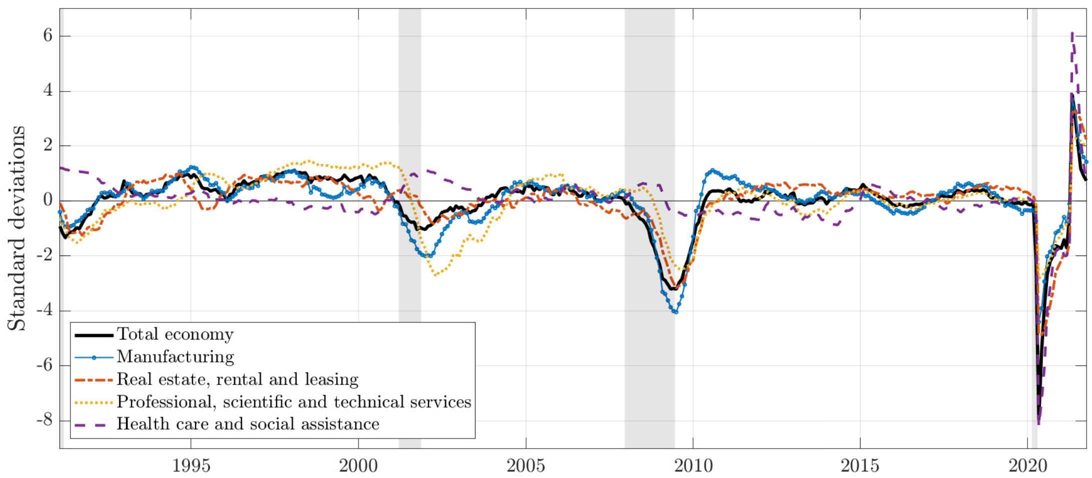

Figure 1: Aggregate and sector-level economic conditions indices of the four largest non-governmental U.S. sectors with gray shaded NBER recession periods.

recession periods.[8]

Figure 1 highlights that there is quite some heterogeneity in economic activity across the four sectors. Manufacturing and professional services are highly cyclical and closely follow aggregate economic conditions with substantial drops in economic activity during the NBER recession periods. Meanwhile, health care seems to move more independently from the overall economy. Moreover, there are differences in the presence and depth of the downturns. For example, the professional services sector is strongly impacted by both the dot-com bubble in 2000-2002 and financial crisis in 2007-2008, while the real estate sector is only severely impacted by the financial crisis and much less so by the dot-com bubble. All four sectors are strongly and negatively affected by the pandemic, albeit to different degrees. In particular, the professional services and manufacturing sectors have drops in economic activity of a roughly similar magnitude as during the financial crisis, while the health care sector experiences a much more extreme drop in activity. Furthermore, the manufacturing sector seems to be most in sync with the overall economy, whereas the real estate and professional services sectors are somewhat more sluggish, especially around the early 1990s recession and dot-com bubble. At the same time, the real estate sector seems to be leading the overall economy at the onset of the financial crisis.

[8] Note that these NBER recession periods are related to the classical business cycle (that is, economic activity in (log-)levels), while we study cycles in growth rates (see Harding and Pagan, 2005, for further details). Hence, the corresponding peaks are not directly comparable, whereas the troughs of the aggregate ECI should agree with the troughs of the NBER.

To fully examine the downturns of sectoral economic activity across recessions, Figure 2 shows the troughs (expressed by the minimum ECIs) across sectors over three NBER recession periods (± one year to account for possible leads/lags). These three periods correspond to the dot-com bubble, financial crisis and covid pandemic. Figure 2 shows that there is considerable heterogeneity in the troughs, both within and across sectors. First, comparing across sectors shows that some sectors are only marginally hit by specific recession periods, while the same episodes have a much more stringent effect on other sectors. For instance, the construction, arts and accommodation sectors are barely hit by the dot-com bubble with drops in economic activity around -0.5 standard deviations from the long-run mean. Meanwhile, sectors like transportation and management experience more severe drops of around -2.5 standard deviations. Similarly for the financial crisis, several sectors such as manufacturing, wholesale trade and retail trade exhibit downturns going as low as -3 to -4 standard deviations, whereas others such as educational services and health care are generally unaffected. Still, during the covid pandemic, almost all sectors display deep drops in economic activity, especially the service-providing sectors like health care, arts, accommodation and other services sectors, as is also highlighted by Maroz et al. (2021).

Second, within sectors, there is clear heterogeneity across the different recession periods. In particular, most sectors as well as the total economy experience the deepest

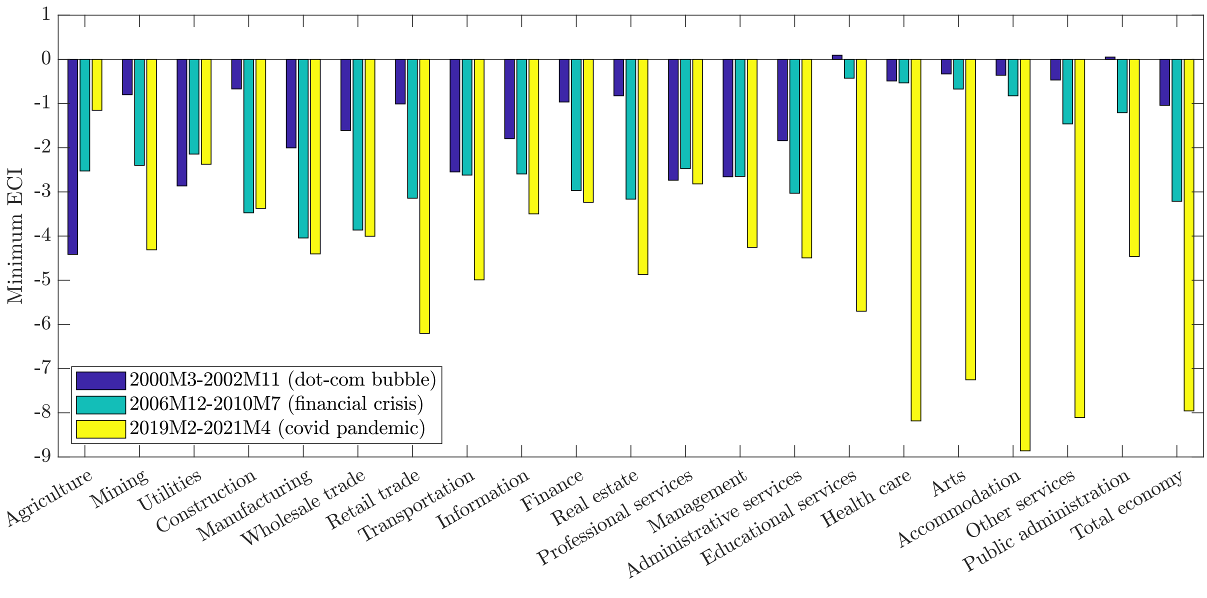

Figure 2: Troughs in sectoral and aggregate economic conditions indices over the NBER recession periods (± one year) corresponding to the dot-com bubble, financial crisis and covid pandemic.

drops during the covid pandemic, followed by the financial crisis and then the dot-com bubble. The most striking difference is observed within the service-providing sectors like health care, arts, accommodation and other services. These sectors experience drops in economic activity around -7 to -8 standard deviations during the covid pandemic, whereas they are hardly affected by the other recessions. For most other sectors, the differences across recession periods are often closer to each other, with the professional services sector even displaying roughly similar declines in economic activity. Overall, the sectoral ECIs facilitate an easy way to assess the impact of various recession periods on the economic activity of different sectors in the economy.

An alternative approach to examining the dynamics of the sectoral economic conditions is by looking at their cycles (see, among others, Harding and Pagan, 2002, 2006; Stock and Watson, 2014; Chang and Hwang, 2015) To do this, we follow Harding and Pagan (2006) and construct, for each sector (and the total economy), a binary variable S_t for $t = 1, \ldots, T$. These binary variables indicate the cycle phases, that is, whether the sector (or total economy) is in an expansion phase ($S_t = 1$) or contraction phase ($S_t = 0$), which are separated by turning points (that is, peaks and troughs). To identify these turning points (and consequently S_t), we use the non-parametric dating algorithm of Bry and Boschan (1971) for monthly observations and apply it on the sectoral and aggregate ECIs.

The resulting sectoral and total economy growth cycles are presented in Figure 3, where we additionally distinguish between lower- or higher-than-average growth (that is, a negative or positive ECI). Notably, the cycles display clear differences across sectors. Some sectoral cycles like the ones of manufacturing and wholesale trade are generally in sync with the overall economy, while others like agriculture and utilities show more distinctive cyclical behaviour. Unsurprisingly, most sectors display contractions with lower-than-average growth during the NBER recession periods, although some sectors like professional services and management are somewhat lagging (as already noted in Figure 1). On the other hand, the retail trade, construction, finance and real estate sectors are leading the financial crisis with below-average-growth as early as June 2006. The timing of the troughs at the end of the contractions during the NBER recessions are highly concentrated among the sectors, especially after the financial crisis and covid pandemic, whereas the foregoing peaks are generally more dispersed. This concurs with

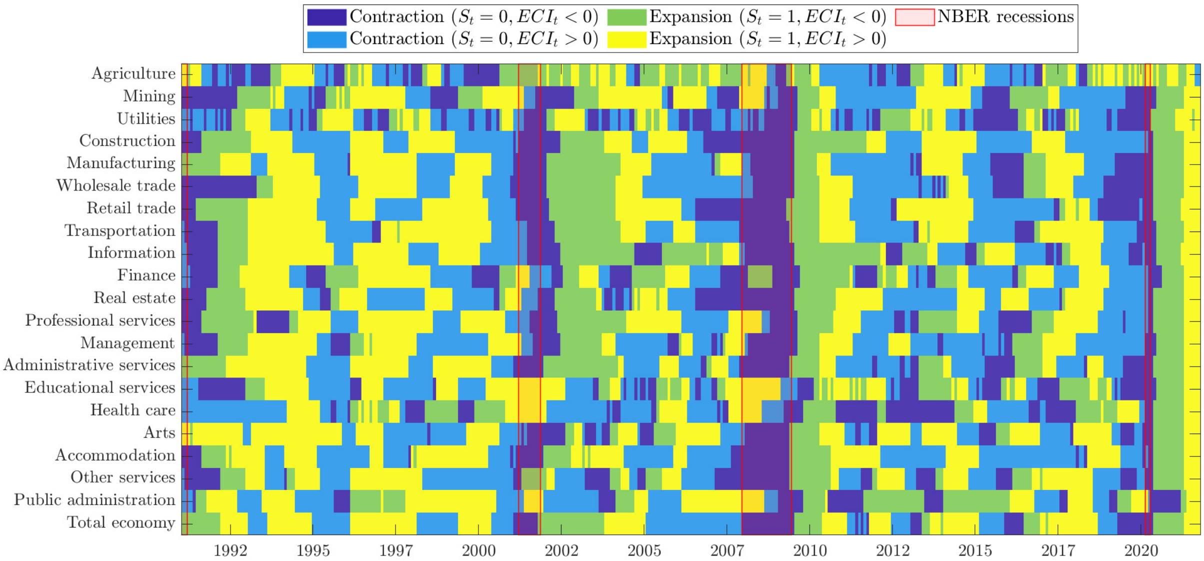

Figure 3: Sectoral and total economy growth cycles based on the Bry and Boschan (1971) algorithm also differentiating between positive and negative economic condition indices with red shaded NBER recession periods.

Chang and Hwang (2015), who also find asymmetric concentration of peak and trough clusters among industries, but then based on industry-level production levels. In sum, we find clear differences in the timing, duration and synchronization of the sectoral growth cycles.

4.2 Comovement of sectoral and aggregate economic conditions

As already observed in Figure 1, some sector-level economic conditions are likely to be closely related to the overall state of the economy. To examine the strength of this comovement, we compute the correlations between the sectoral and aggregate ECIs, which are shown in Figure 4 over the full-sample (January 1991 to September 2021) and pre-covid sample (January 1991 to December 2019). Two observations stand out. First, it becomes clear that most sectors are highly correlated with the overall state of the economy, resulting in an average full-sample correlation of 0.69. The administrative services sector has the highest full-sample correlation of 0.93, closely followed by the manufacturing and retail trade sectors with correlations of 0.91. Still, some sectors move more independently from aggregate conditions. The agriculture sector has a full-sample correlation close to zero, concurring with Da-Rocha and Restuccia (2006) that the agriculture

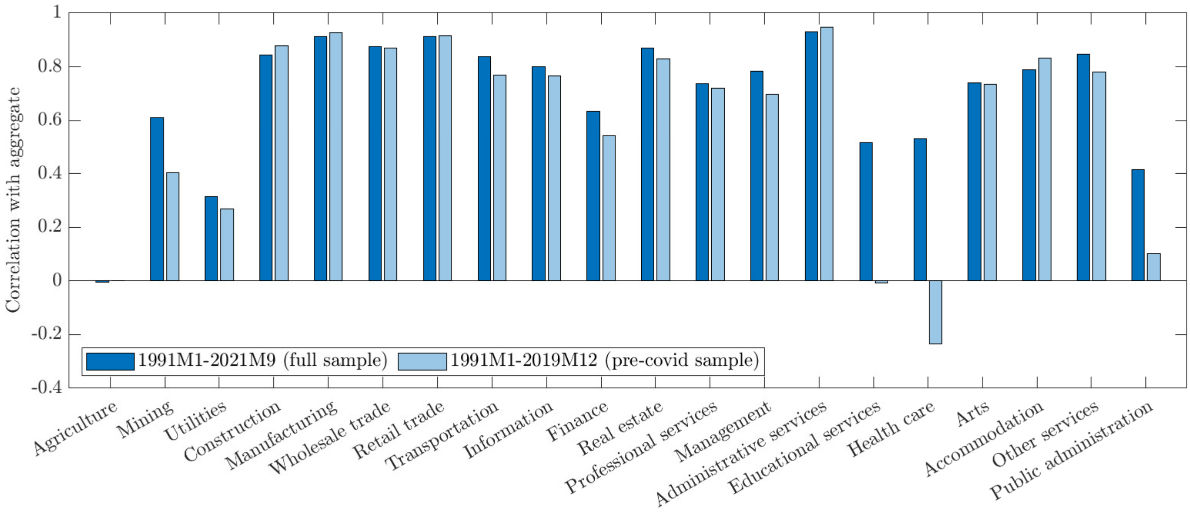

Figure 4: Correlation between sector-level and aggregate economic condition indices over the full-sample (January 1991 - September 2021) and pre-covid-sample (January 1991 - December 2019).

sector is only weakly correlated with the rest of the economy, while the utilities and public administration sectors have correlations of 0.32 and 0.41, respectively.

Second, by comparing the correlations across the two samples, we find that for several sectors the correlation coefficient is largely influenced by the pandemic period, resulting in a somewhat lower average correlation of 0.59 during the pre-covid sample. In particular, the mining, educational services, health care and public administration sectors have substantially lower correlations when the covid pandemic is excluded. In fact, leaving out the pandemic period for the educational services and health care sectors results in correlations of -0.01 and -0.23, respectively. This implies that, during the pre-covid sample, the educational service sector is acyclical and the health care sector even (moderately) counter-cyclical. These differences in correlations between the two samples can be explained by the fact that the Pearson correlation coefficient is sensitive to extreme observations like the ones observed in the beginning of the pandemic. To account for this, we also compute Spearmans's rank correlations, which are more robust to outliers. These results are given in Appendix E and are qualitatively similar, but with slightly weaker correlations. The average rank correlations for the full and pre-covid sample are 0.55 and 0.49, respectively.

To assess whether the comovement is (a)symmetric over the business cycle, Figure 5 distinguishes between correlations computed over recession periods (as indicated by the

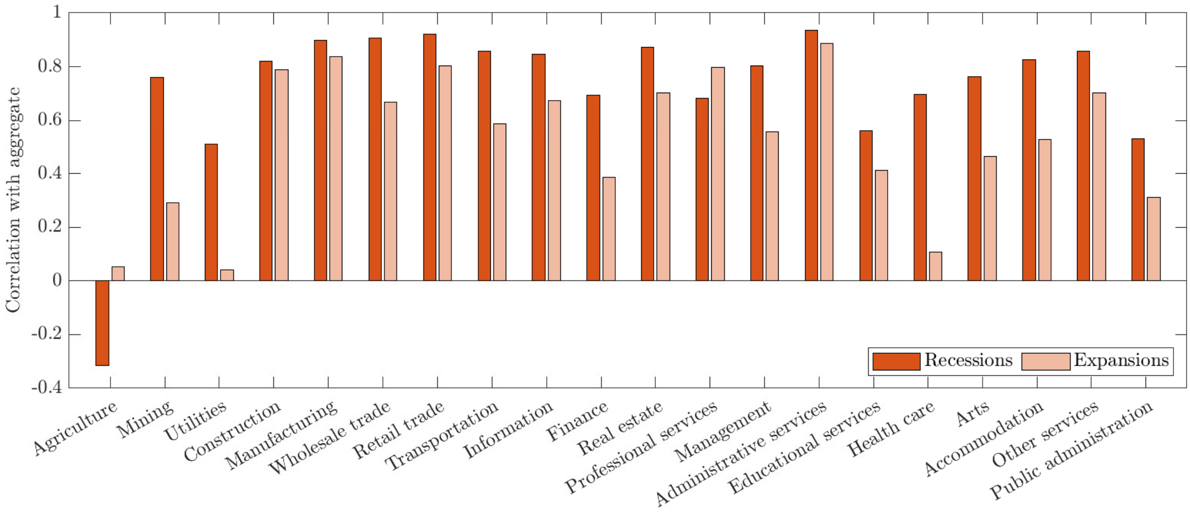

Figure 5: Correlation between sector-level and aggregate economic condition indices over recession periods (as indicated by the NBER recession dates (± one year)) and expansion periods.

NBER recession dates (± one year)) and correlations computed over expansion periods. Clearly, the correlations during expansions are often lower than the ones during recessions, with an average of 0.53 relative to 0.72. This implies that the comovement between sectoral and aggregate economic activity is generally stronger during recessions than during expansions. These asymmetries are most pronounced in sectors that are weakly related to the overall economy like agriculture, utilities and health care. On the other hand, sectors that are closely related to aggregate economic conditions like administrative services, manufacturing and retail trade are more symmetric, displaying the highest correlations during both business cycle phases. Again, we also compute Spearman's rank correlations to account for extreme observations. These show roughly similar degrees of asymmetry, but with somewhat weaker average correlations of 0.44 for expansions and 0.65 for recessions (see Appendix E).

The comovement between sectoral and aggregate economic conditions can also be examined by the synchronization of their cycles (see, among others, Harding and Pagan, 2006; Chang and Hwang, 2015). Given the specific cycles S_t constructed in the previous section, the synchronization measure of Harding and Pagan (2006) can be computed as the correlation between the sectoral cycle and total economy cycle.[9] Figure 6 presents

[9] An alternative way of measuring the degree of synchronization is by means of the concordance index of Harding and Pagan (2002), which measures the fraction of time that the cycles are in the same phase. Yet, the concordance index is just a monotonic transformation of the correlation-based synchronization measure (Harding and Pagan, 2006), so we only focus on the latter.

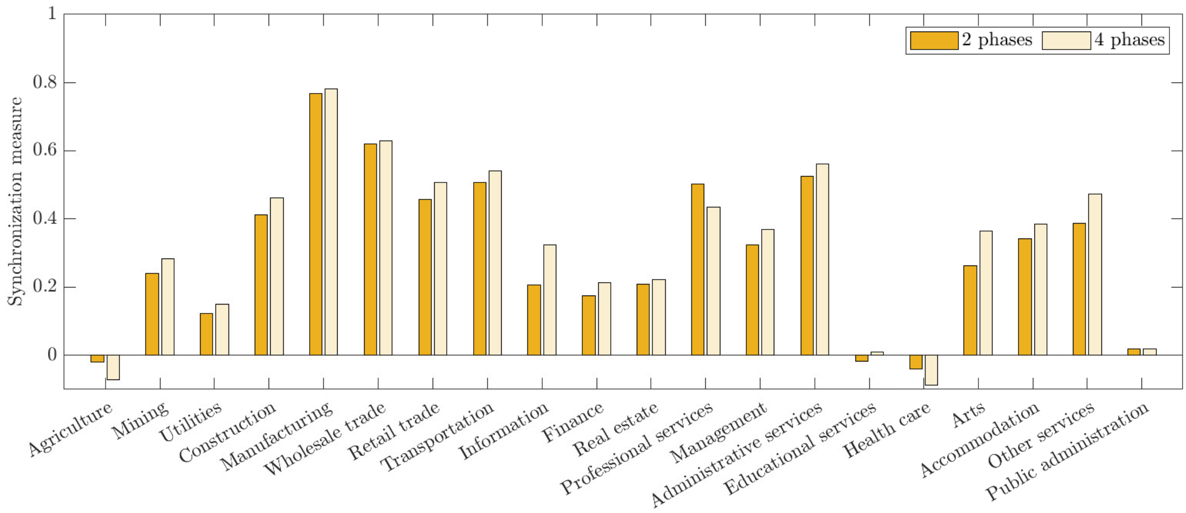

Figure 6: Correlation-based synchronization measures between sectoral and total economy growth cycles based on the two phases determined by the Bry and Boschan (1971) algorithm and the four phases also differentiating between positive and negative economic condition indices.

these correlation-based synchronization measures for the two phases (contraction and expansion) and the four phases that additionally distinguish between positive and negative ECIs (just as in Figure 3). There is a large dispersion in the synchronization measures, with an average two-phase and four-phase correlation of 0.30 and 0.33, respectively. The sector that is most in sync with the total economy is manufacturing with a two-phase correlation of 0.76, which is consistent with Figures 1 and 3. This provides evidence that manufacturing is a dominant sector in the aggregate growth cycle. The manufacturing sector is followed by wholesale trade and administrative services with correlations of 0.62 and 0.52, respectively. By contrast, the measures of agriculture, educational services, health care and public administration are close to zero, indicating almost no evidence of synchronization with the overall economy. The four-phase comovement is often close to the two-phase one, with only slightly higher correlations. Excluding the covid pandemic or using rank correlations generally returns similar results (see Appendix E), highlighting the robustness of the synchronization measures.

4.3 Drivers of sectoral economic conditions

To examine which economic indicators drive the sectoral ECIs, Figure 7 presents the monthly ECIs (black lines) and their decomposition of underlying series (stacked colored bars) for a selection of four sectors. Specifically, we show the sectoral indices of two goods-producing sectors (that is, mining and construction) and two service-providing sectors (that is, retail trade and information). The estimates of the other 16 sectoral ECIs are shown in Appendix F and generally exhibit the same features discussed below.

Figure 7 indicates that the sectoral ECIs are composed of a balanced mix of different underlying sector-specific economic series. There are clear contributions by both labor-market related variables (employment, UI claims) and output-related variables (GDP, industrial production). Unsurprisingly, the goods-producing sectors like mining and construction (and information to some extent) are largely driven by output-related variables, while a service-providing sector like retail trade is more driven by labor-related variables. The other economic series (sales, revenue and personal income) play substantial roles as well for various sectors. For instance, personal income is an important driver of economic activity in the construction sector, while sales and revenue play prominent roles in the retail trade and information sectors, respectively.

When comparing the underlying drivers of each sector, three observations stand out. First, the drivers show heterogeneous behaviour across recession periods, just as in the total sector-level indices. For example, the information sector exhibits a strong drop in labor-market related variables during the dot-com bubble, but much less so in output-related variables, while it is the other way around during the financial crisis. During the covid pandemic, however, both output-related and labor-market related variables contribute to poor economic conditions in the information sector. In a similar fashion, losses in sales play only a minor role in the retail trade sectors during the dot-com bubble and covid pandemic, but a much larger role during the financial crisis.

Second, the underlying drivers unveil sector-specific periods of heightened and reduced economic activity. In particular, the mining sector exhibits a two standard deviations decline in economic activity in 2014-2015 largely due to reductions in production, which could be attributed to the oil price slump during this period (Baffes et al., 2015; Baumeister and Kilian, 2016). Indeed, the U.S. state-level economic indices of Baumeister et al. (2022) show a similar kind of drop in economic activity for oil-producing states like North Dakota and New Mexico. Meanwhile, the information sector exhibited a prolonged period of high economic activity in the late 1990s across all underlying economic series, which is largely fueled by the dot-com bubble. After the burst of the dot-come bubble, however,

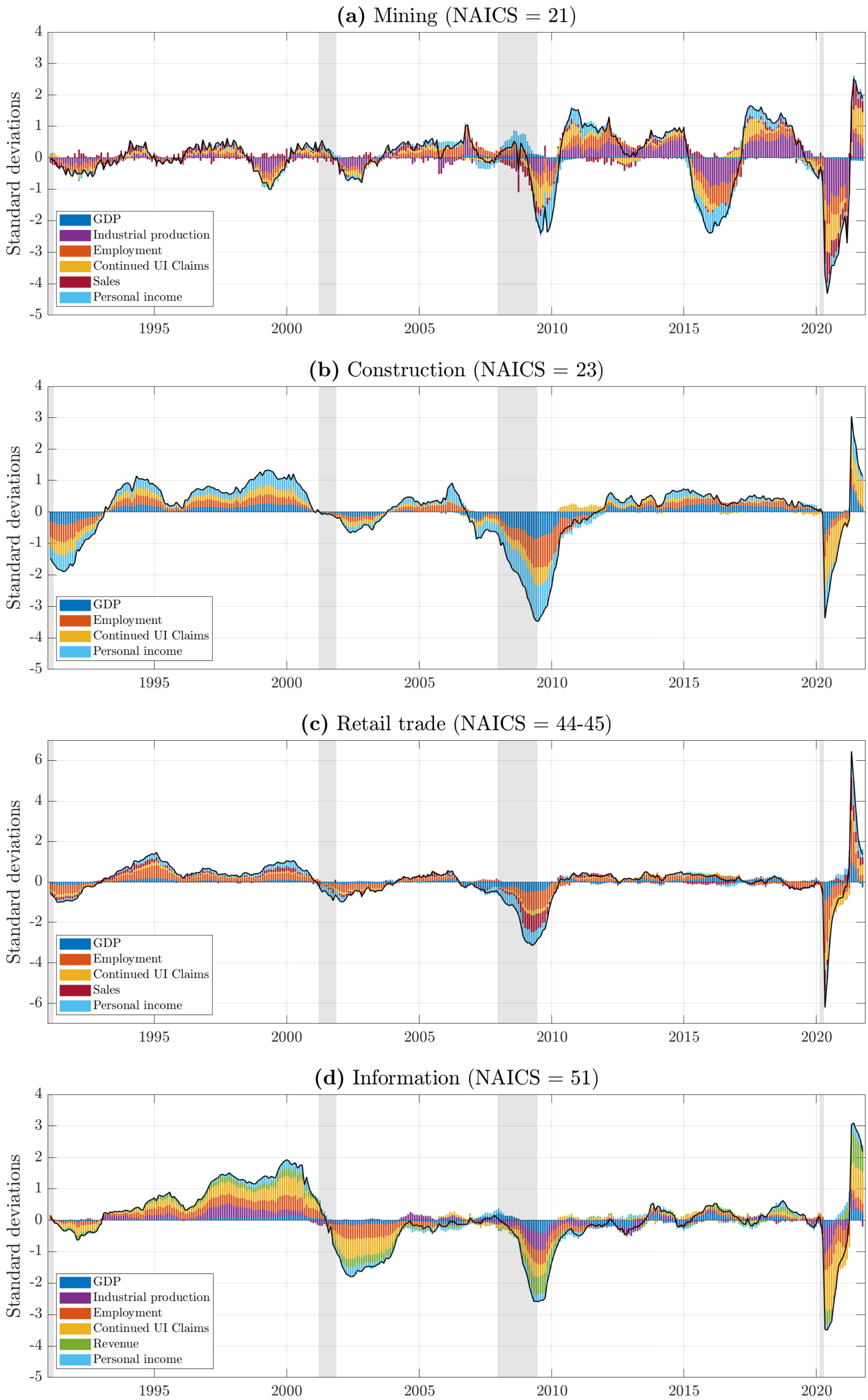

Figure 7: Sector-level economic conditions indices (black lines) and their drivers for a selection of four sectors with the contribution of each data category in the colored bars and with gray shaded NBER recession periods.

the information sector dropped drastically in economic activity and, in fact, has never reached the pre-burst level again (except for the rebound after the initial covid shock).

Third, some drivers are leading or lagging the corresponding sectoral economic conditions during or after a recession. Specifically, during the financial crisis, GDP growth seems to lag the economic conditions in the information sector, while sales appear to lead in the mining sector. Moreover, continued UI claims seem to lead the recovery in the construction and information sector, making the ECI (black line) lie between the stacked contributions (colored bars) instead of lying on the border.

The well-balanced composition of the ECIs noted before can also be seen in Figure 8, which shows the factor loading estimates ($\hat{\boldsymbol{\lambda}}$) across sectors and variables. In particular, the loadings have roughly similar absolute magnitudes across the different series, implying that the various sources of economic activity all seem to contribute to the sectoral ECIs. In line with the sign restrictions, the loadings are generally consistent with the cyclicality of the corresponding economic variables. Specifically, GDP, IP, employment, sales, revenue and personal income are all pro-cyclical variables and, indeed, have positive loadings, while continued UI claims are counter-cyclical and have negative loadings. There are a few exceptions of opposite loadings, but these are generally small in magnitude. The only exception is the larger negative loading of GDP growth for the finance sector, which could be explained by its weaker comovement of the underlying economic variables and especially the negative comovement between its employment and GDP growth (see Appendix D).

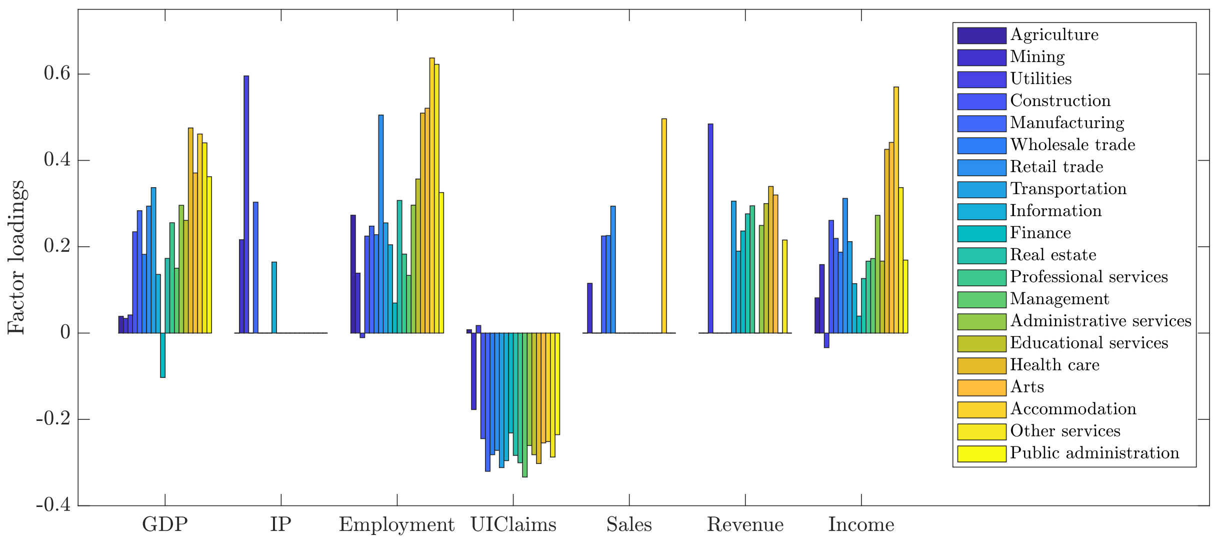

Figure 8: Factor loading estimates ($\hat{\boldsymbol{\lambda}}$) across sectors and variables

5 Alternative applications of sectoral ECIs

5.1 Common factors to sectoral economic conditions

Monitoring economic activity at the sector level is the most obvious use of the constructed indices. However, they can be used for several other purposes as well. A prominent example is to examine whether the comovement of various economic indicators over the business cycle is driven by common aggregate shocks or sector-specific shocks (see, among others, Long and Plosser, 1987; Foerster et al., 2011; Andreou et al., 2019; Graeve and Schneider, 2023). To answer this question, most existing studies focus solely on output-related variables like industrial production (IP) data (Long and Plosser, 1987; Foerster et al., 2011; Graeve and Schneider, 2023), ignoring the service-providing sectors, while Andreou et al. (2019) additionally include annual GDP growth data of non-IP sectors (that is, agriculture, construction, service-providing sectors and public administration). Yet, none of the existing studies, to the best of our knowledge, focus on the comovement of different types of economic activity at the aggregate and sectoral level in answering this question.

Given the estimated sectoral ECIs, we therefore revisit the analysis in Foerster et al. (2011) and Andreou et al. (2019), but then based on a broader measure of economic activity rather than only output-related variables (that is, IP and GDP growth) and based on monthly indices for all sectors in the economy.[10] Moreover, by having a more extensive sample, we can examine the effect of the covid pandemic on the explanatory power of the common factors on aggregate and sectoral economic activity. Following Foerster et al. (2011), we decompose the 20 monthly sectoral ECIs, collected in $\boldsymbol{x}_t = (ECI_{1,t}, \ldots, ECI_{20,t})'$, into a factor model structure with two common factors, that is,

$$\boldsymbol{x}_t = \boldsymbol{\Gamma} \boldsymbol{g}_t + \boldsymbol{u}_t, \tag{9}$$

where $\boldsymbol{\Gamma}$ denotes the 20×2 factor loading matrix and $\boldsymbol{g}_t$ the 2×1 vector with common factors, each following a stationary first-order autoregression with the corresponding error

[10] Both Foerster et al. (2011) and Andreou et al. (2019) focus on quarterly IP data as monthly IP data is considered too noisy for their analysis. However, our sectoral ECIs are estimated based on yearly growth rates of the underlying variables, which already smooths out the volatile month-to-month fluctuations.

term being i.i.d. standard normally distributed.[11] Furthermore, we assume that the observation error vector $\boldsymbol{u}_t \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{\Pi})$ with $\boldsymbol{\Pi}$ being a 20×20 diagonal covariance matrix. Similarly as for the mixed-frequency dynamic factor model, this dynamic factor model is estimated by means of the EM algorithm with the key differences of having no mixed-frequencies and no autoregressive dynamics for $\boldsymbol{u}_t$.

Figure 9 shows the smoothed estimates of the common factors in $\boldsymbol{g}_t$, while Figure 10 shows the corresponding factor loading estimates in $\boldsymbol{\Gamma}$. The first common factor closely resembles the aggregate ECI in Figure 1 with a correlation of 0.87, albeit with a less steep drop during the pandemic. In fact, the first common factor has a correlation of 0.91 with year-on-year aggregate IP growth, suggesting that the driving component in the economy is closely tied to total production. Indeed, leaving out the service-providing sectors and only considering a single common factor in the estimation generates a factor estimate that is similar to the one obtained from the complete set of sectors, with a correlation of 0.93 (see Appendix H). Furthermore, Figure 10 shows that the loadings related to the first common factor are well-balanced with almost all values being between 0.23 and 0.38. The only exceptions are the agriculture and utilities sector with loadings of 0.00 and 0.10, respectively, which is consistent with their low correlation between sectoral and aggregate economic activity in Figure 4.

Moving to the second common factor, Figure 9 indicates close to perfect negative comovement with the first common factor before the pandemic period, with a correlation of -1.00 during the pre-covid sample (January 1991 - December 2019). During the pandemic period, however, the comovement becomes positive with a correlation of 0.78 over the covid sample (January 2020 - September 2021). In other words, for sectors with a large loading, the second common factor implies a less severe downturn during the recession periods of 2001 and 2007-2009 and a more severe downturn during the pandemic in 2020. Indeed, Figure 10 shows that the service-providing sectors that are considerably hit by the covid recession period like health care, educational services, arts and accommodation have the largest loadings. Meanwhile, the sectors that have drops in economic activity of a roughly similar magnitude during the financial crisis as during the pandemic like the construction, manufacturing, wholesale trade and professional services

[11] We also experiment with three and four common factors, but these generally lead to qualitatively similar results in terms of how much of the total variation can be explained by the common factors. These results are given in Appendix G.

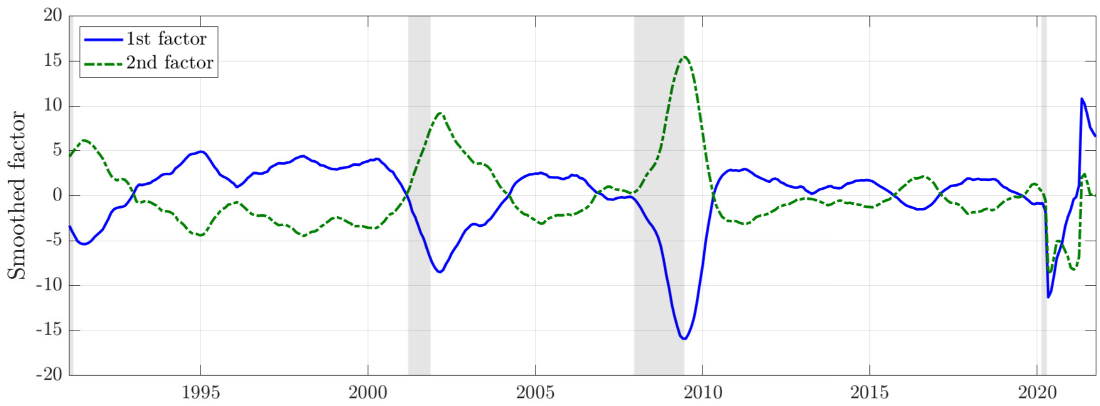

Figure 9: Smoothed estimates of common factors ($\boldsymbol{g}_t$) with gray shaded NBER recession periods.

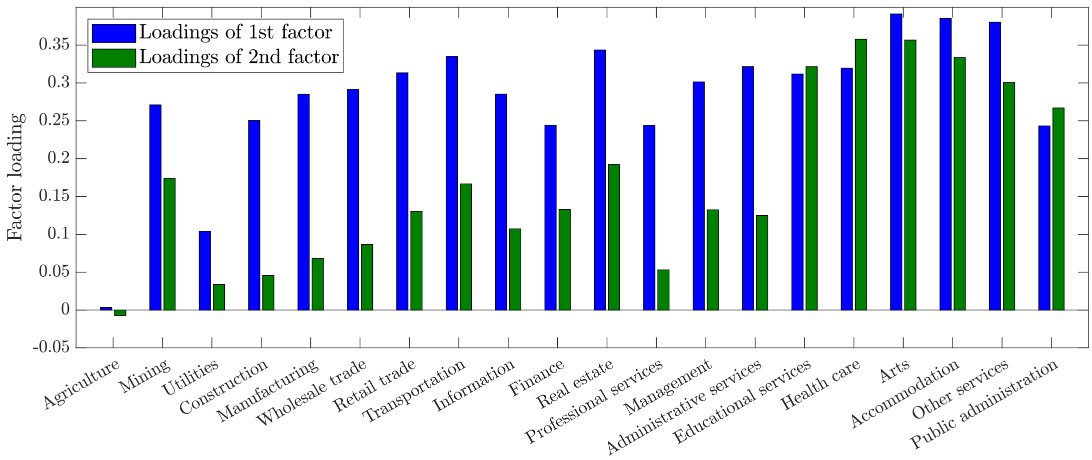

Figure 10: Factor loading estimates across sectors of common factors ($\boldsymbol{\Gamma}$).

sectors have much smaller loadings. This implies that the second common factor serves as a *correction* factor for the service-providing sectors that are more severely hit by the covid pandemic and less severely by the financial crisis than accounted for by the first common factor only.

As a robustness check, we also estimate the common factors and loadings using data up to December 2019, excluding the covid pandemic. These results are given in Appendix H and show that the first common factor remains similar to its full-sample counterpart with a correlation of 0.98. However, the structure and dynamics of the second common factor completely changes, shifting away from the hardest hit sectors by the pandemic towards the hardest hit sectors by the financial crisis like construction, manufacturing, finance and real estate. Therefore, we consider the full sample and pre-covid sample separately in the subsequent analysis.

To investigate the explanatory power of the estimated factors on aggregate and sectoral economic activity, we follow Andreou et al. (2019) and regress the aggregate and sectoral ECIs on (i) the first common factor, (ii) the second common factor and (iii) both common factors. Table 2 displays the adjusted R^2 of these regressions based on factor estimates over the full sample and pre-covid sample. We first focus on the full-sample results. Panel A presents the results for the aggregate ECI and aggregate IP growth, where the former represents a broad measure of economic activity and the latter only production-related activity. Despite that neither are used in the estimation of the factors, the first common factor explains about 76% of the full-sample variation in the aggregate ECI and 77% of the full-sample variation in aggregate IP. Adding the second common factor to the first one leads to increments of the adjusted R^2 of 14% for aggregate ECI and 3% for aggregate IP. In other words, the second common factor only adds explanatory power for the broad economic measure that also includes the non-IP sectors, while it only marginally adds explanatory power for total IP. This also aligns with the loadings in Figure 10. Overall, this implies that the variation in aggregate economic activity is largely driven by a small number of common factors among sectoral economic activity. Moreover, Panel B shows that the first common factor explains 91% of the variability in the manufacturing sector as well as most of the fluctuations in the service-providing sectors that are closely related to IP like wholesale trade (85%), retail trade (76%), transportation and warehousing (72%) and administrative and waste management services (86%). This suggests that the first common factor among all sectors can be interpreted as an IP-related factor, which corroborates with the findings in Andreou et al. (2019).

The second common factor only seems to explain part of the full-sample variation in the sectors that are part of or have a close link to IP like manufacturing (42%), construction (40%), wholesale trade (36%) and professional services (33%). However, for these sectors, regressing the sectoral ECIs on both common factors shows that the second common factor does not add much on top of the first factor in terms of explanatory power, indicating that these sectors are solely driven by the first common factor. On the other hand, for sectors with variation that cannot be explained by the second common factor alone like educational services (7%) and accommodation and food services (1%), it becomes clear that jointly considering both common factors has a huge impact on the

Table 2: Adjusted R^2's of regressing sector-level and aggregate economic conditions indices on smoothed common factor estimates.

	Full sample (1991M1 - 2021M9)			**Pre-covid sample (1991M1 - 2019M12)**		
	$\bar{R}^2(1)$	$\bar{R}^2(2)$	$\bar{R}^2(1+2)$	$\bar{R}^2(1)$	$\bar{R}^2(2)$	$\bar{R}^2(1+2)$
Panel A: Total economy						
Aggregate ECI	0.76	0.21	0.90	0.95	0.03	0.95
Aggregate IP	0.77	0.33	0.80	0.78	0.01	0.86
Panel B: Sectors						
Agriculture, forestry, fishing, and hunting	0.00	0.00	0.00	0.00	0.02	0.02
Mining	0.32	0.03	0.50	0.22	0.03	0.28
Utilities	0.10	0.04	0.11	0.06	0.03	0.10
Construction	0.78	0.40	0.79	0.82	0.34	1.00
Manufacturing	0.91	0.42	0.93	0.90	0.02	1.00
Wholesale trade	0.85	0.36	0.89	0.81	0.00	0.86
Retail trade	0.76	0.24	0.86	0.88	0.08	0.89
Transportation and warehousing	0.72	0.18	0.88	0.73	0.00	0.74
Information	0.69	0.25	0.76	0.60	0.04	0.60
Finance and insurance	0.34	0.07	0.44	0.30	0.21	0.43
Real estate, rental and leasing	0.65	0.12	0.86	0.76	0.17	0.83
Professional, scientific, and technical services	0.69	0.33	0.71	0.54	0.01	0.55
Management of companies and enterprises	0.67	0.20	0.77	0.59	0.01	0.59
Administrative and waste management services	0.86	0.30	0.95	0.89	0.01	0.90
Educational services	0.07	0.07	0.69	0.00	0.07	0.07
Health care and social assistance	0.04	0.15	0.80	0.08	0.00	0.08
Arts, entertainment, and recreation	0.24	0.02	0.99	0.54	0.00	0.54
Accommodation and food services	0.28	0.01	0.95	0.77	0.10	0.79
Other services (except public administration)	0.38	0.00	0.91	0.59	0.15	0.65
Public administration	0.03	0.08	0.45	0.00	0.02	0.02

Notes: This table shows the adjusted R^2's ($\bar{R}^2$'s) of regressing the aggregate ECI and aggregate IP (Panel A) and sectoral ECIs (Panel B) on the estimates of the first common factor ($\bar{R}^2(1)$), second common factor ($\bar{R}^2(2)$) or both ($\bar{R}^2(1+2)$). The results are shown for full-sample (January 1991 - September 2021) based estimated factors and regressions and pre-covid sample (January 1991 - December 2019) ones. The factors are estimated from the 20 sectoral ECIs using the dynamic factor model given in equation (9).

explanatory power. For instance, for the arts, entertainment and recreation sector, the first and second common factor, on their own, explain about 24% and 2%, respectively, while jointly they explain about 99% of its variation in economic activity. This confirms the correction effect that the second common factor has on top of the first one for sectors that are hit hardest during the pandemic, but much less so during the financial crisis.

Lastly, to assess the effect of the pandemic on the explanatory power, we also focus on the pre-covid sample (that is, the results where the factors and regressions are estimated with pre-covid sample data only). Compared to the full-sample, the first common factor explains considerably more of the variation in total economic activity (+19%), while it explains roughly the same for aggregate IP (−1%) and manufacturing (−1%). Furthermore, for most sectors (except construction and manufacturing), the two common factors together explain less variation during the pre-covid sample than during the full sample, especially for the service-providing sectors that are hit hardest during the pandemic. On its own, the second common factor only seems to explain variation of the construction (34%), finance and insurance (21%) and real estate, rental and leasing

(17%) sectors, which are sectors that are closely tied to the origins of the financial crisis of 2007-2008. In other words, the second common factor estimated up till December 2019 can be characterized as a factor capturing the real estate and financial cycle (see also Appendix H). Overall, the structure and explanatory power of the second common factor has changed the most by the covid pandemic, particularly for the hardest hit sectors, while the explanatory power of the first common factor remained roughly similar.

5.2 Nowcasting sectoral GDP growth

As a by-product of the construction of the sectoral ECIs from the mixed frequency dynamic factor models, we obtain estimates of the latest underling sector-level economic indicators from the Kalman smoother. These estimates provide an alternative approach to assessing the current economic state of each sector (Nunes, 2005) and, consequently, to validate the corresponding sectoral ECIs. To examine the accuracy of these estimates, we conduct an expanding-window nowcasting exercise for sector-level year-on-year GDP growth, where at each point in time we take into account the publication delays of the series (see Table 1) and impose this ragged edge structure onto the data.[12] In particular, sector-level GDP is released with a considerable delay of three months, while most other sectoral economic series are released much earlier. This indicates that there is much to gain in terms of using earlier released sector-level information for nowcasting.

Following Bańbura and Modugno (2014), we construct a sequence of nowcasts for each target quarter, ranging from the first month of the previous quarter (that is, Q(-1)M1) to the second month of the subsequent quarter (that is, Q(+1)M2), just before the first official figures are released. Since the sector-level GDP series are only available since the first quarter of 2005, the evaluation period runs from the first quarter of 2010 to the second quarter of 2021. This means that the first estimation window is from January 1991 to October 2009 to produce Q(-1)M1 for 2010Q1 with at least close to four years of GDP growth series in the sample. After the initial estimation window, the sample is expanded by one month and the sector-specific mixed-frequency dynamic factor models are re-estimated again. This process is repeated to produce all the nowcasts. As a benchmark, we compute nowcasts from a univariate first-order autoregressive (AR) model for sector-

[12]Note that this exercise is pseudo real-time since there is generally no vintage data for the sector-level economic series.

level GDP growth, also taking into account the publication delays.

Table 3 shows the relative root mean squared forecast errors (RMSFE) of the sequence of sector-level GDP nowcasts for the mixed-frequency dynamic factor model compared to the AR(1) benchmark. Panel A shows the nowcasting results for the complete out-of-sample period that includes the covid pandemic, while Panel B excludes the covid pandemic and displays results up to 2019Q4. Three observations stand out from Table 3. First, the nowcasts from the mixed-frequency dynamic factor model are generally (significantly) more accurate than the ones of the AR(1) benchmark, as indicated by the large number of relative RMSFEs that are below one. In fact, this goes as low as 0.20 for Q(+1)M2 in the arts, entertainment, and recreation sector (Panel A). Moreover, this observation holds for both including (Panel A) and excluding (Panel B) the covid pandemic. For the complete out-of-sample period (Panel A), the relative average performance across sequence of nowcasts and sectors is 0.78, meaning that on average the model-based nowcasts are 22% more accurate than the benchmark. Excluding the covid pandemic (Panel B) gives a roughly similar average of 0.76. For completeness, we also show the relative accuracy of the nowcasts made for the total economy, which are also well below one in both Panel A and B with a relative average RMSFE of 0.58 and 0.57, respectively.[13]

Second, the relative RMFSEs within each construction quarter Q(-1), Q(0) and Q(+1) are generally close to each other. This implies that the economic data releases within a quarter do not considerably change the relative performance of the models compared to the benchmark, with some exceptions like Q(0) for the health care sector in Panel A. Yet, across blocks, there can be substantial differences. For instance, the manufacturing sector in Panel A has a relative RMSFE around 0.70 in Q(-1), while it is around 0.45 in Q(0) and around 0.35 in Q(+1). This can be explained by the fact that the GDP figures of the previous quarter are released every third month in the quarter, which updates both the benchmark and model-based nowcasts. The largest relative improvements are often observed from Q(0) to Q(+1), meaning that whenever the target quarter has past, enough information has become available for the mixed-frequency dynamic factor model to provide relatively better nowcasts than before.

Third, when comparing Panel A with Panel B, it become clear that for some sectors

[13] Note that the nowcasts for Q(+1) are not shown for the total economy as the publication delay is only one month for the total GDP figures.

Table 3: Relative nowcasting performance of sector-level GDP growth based on a mixed-frequency dynamic factor model compared to an AR(1) benchmark.

	Q(-1)			Q(0)			Q(+1)		Average
	M1	**M2**	**M3**	**M1**	**M2**	**M3**	**M1**	**M2**	
Panel A: Evaluation period including covid pandemic (2010Q1 - 2021Q2)									
Agriculture, forestry, fishing, and hunting	0.89**	0.89**	0.96	0.95	0.96	1.04**	0.83**	0.83**	0.91
Mining	0.80***	0.80***	0.81***	0.97	0.98	1.04	0.65***	0.65***	0.83
Utilities	0.84*	0.84*	0.84*	0.96	0.98	1.01	0.79	0.86	0.88
Construction	0.80	0.77*	0.81*	0.75	0.74	0.71*	0.54**	0.55**	0.72
Manufacturing	0.73**	0.68**	0.73**	0.48**	0.43**	0.47**	0.35**	0.35**	0.55
Wholesale trade	0.83	0.83	0.84	1.07	1.10	1.10	0.77	0.79	0.90
Retail trade	0.94	0.94	0.98	0.83	0.84	0.73*	0.60*	0.62*	0.82
Transportation and warehousing	0.84*	0.79**	0.87***	0.44**	0.33**	0.38*	0.33**	0.30**	0.55
Information	0.91	0.98	0.96	1.13	1.22	1.20	0.92	0.94	1.02
Finance and insurance	0.73**	0.73*	0.73*	0.72	0.74	0.75	0.59**	0.62*	0.70
Real estate, rental and leasing	0.90	0.89	0.88	1.07	1.12	1.17	0.82	0.86	0.95
Professional, scientific, and technical services	0.80	0.75*	0.78*	0.77	0.72	0.73	0.53**	0.51**	0.71
Management of companies and enterprises	1.18	1.16	1.16	1.50*	1.47	1.54	1.21	1.18	1.28
Administrative and waste management services	0.83	0.79	0.78*	0.63*	0.57**	0.54**	0.41**	0.36**	0.63
Educational services	0.74***	0.74**	0.73***	0.49*	0.49*	0.43*	0.32**	0.33**	0.56
Health care and social assistance	1.06	1.19	1.19	1.01	0.85	0.59*	0.62*	0.61*	0.90
Arts, entertainment, and recreation	0.92	0.84*	0.92*	0.40**	0.35**	0.23*	0.27*	0.20*	0.52
Accommodation and food services	0.89**	1.32	0.94	0.46*	0.28**	0.27**	0.31*	0.32*	0.60
Other services (except public administration)	1.11	1.08	1.02	0.89	0.77	0.60*	0.50**	0.57*	0.83
Public administration	0.83	0.83	1.01	0.78*	0.69**	0.59*	0.53***	0.50***	0.73
Total economy	0.80**	0.76***	0.73***	0.29*	0.38*	0.34*	-	-	0.58
Panel B: Evaluation period excluding covid pandemic (2010Q1 - 2019Q4)									
Agriculture, forestry, fishing, and hunting	0.90**	0.90*	0.94	0.96	0.97	1.07**	0.76**	0.76**	0.90
Mining	0.80***	0.80***	0.80***	0.94	0.95	1.01	0.63**	0.63**	0.81
Utilities	0.84*	0.84*	0.84*	0.96	0.98	1.00	0.78	0.86	0.88
Construction	0.70	0.67*	0.75	0.92	0.93	0.95	0.69	0.70	0.77
Manufacturing	0.58**	0.53**	0.54*	0.62	0.60	0.86	0.48*	0.48*	0.57
Wholesale trade	0.49**	0.44**	0.48**	0.60*	0.60*	0.87	0.54**	0.54**	0.54
Retail trade	1.09	1.07	1.07	1.26***	1.27***	1.18***	0.96	0.96	1.10
Transportation and warehousing	0.77**	0.70***	0.74**	0.81**	0.77**	0.88*	0.63**	0.63***	0.74
Information	0.73**	0.72**	0.74**	0.86*	0.86*	0.98	0.71**	0.77*	0.78
Finance and insurance	0.73*	0.73*	0.73*	0.70	0.71	0.73	0.57**	0.59*	0.69
Real estate, rental and leasing	0.64**	0.64**	0.60**	0.75	0.77	0.84	0.57*	0.57*	0.66
Professional, scientific, and technical services	0.55**	0.47***	0.54**	0.63**	0.54**	0.67**	0.42**	0.39***	0.52
Management of companies and enterprises	0.93*	0.91**	0.98	1.08	0.99	1.02	0.84*	0.85*	0.95
Administrative and waste management services	0.49**	0.45**	0.45**	0.51*	0.49**	0.63*	0.38**	0.36**	0.46
Educational services	0.71**	0.76**	0.78***	0.98	1.01	1.00	0.65***	0.64***	0.80
Health care and social assistance	0.72***	0.72***	0.73**	0.90	0.87	0.88*	0.67***	0.62***	0.75
Arts, entertainment, and recreation	0.74***	0.74***	0.84***	0.87**	0.87**	0.93	0.80***	0.80***	0.82
Accommodation and food services	0.60**	0.59***	0.55**	0.71**	0.70**	0.79**	0.47**	0.46**	0.60
Other services (except public administration)	0.74***	0.72***	0.69***	0.86**	0.83**	0.87**	0.51***	0.51***	0.71
Public administration	1.15*	1.17*	0.94	1.05	1.01	1.11	0.76**	0.78**	1.00
Total economy	0.53**	0.47**	0.44**	0.83*	0.76**	0.80*	-	-	0.57

Notes: This table shows the relative root mean squared forecast errors (RMFSE) of nowcasting sector-level year-to-year GDP growth based on a mixed-frequency dynamic factor model (DFM) from 2010Q1 to 2021Q2 compared to an AR(1) benchmark. Panel A shows the results including the covid pandemic, while Panel B excludes it. For each target quarter, the nowcasts construction dates range from the first month of the previous quarter (that is, Q(-1)M1) to the second month of the subsequent quarter (that is, Q(+1)M2). The final column shows the relative average RMSFEs across projections based on the mixed-frequency DFM compared to the ones of the AR benchmark. A value smaller (larger) than one means that the mixed-frequency DFM produces more (less) accurate nowcasts than the AR benchmark. The asterisks $*$, $**$ and $***$ indicate significance at the 10%, 5% and 1% level, respectively, based on the modified Diebold-Mariano test of Harvey et al. (1997).

the relative performance is largely influenced by the covid pandemic. Specifically, for some sectors like retail trade and public administration, the relative performance deteriorates when excluding the covid pandemic, meaning that the model-based nowcasts are relatively more accurate compared to the benchmark during the covid pandemic than during the pre-covid period. For example, based on the full out-of-sample period (Panel A), the relative RMSFEs for the retail trade sector range from 0.60 to 0.94 with an average of 0.82, while in the pre-covid period (Panel B) they are between 0.96 and 1.27 with an average of 1.10. On the other hand, for some sectors like wholesale trade and management, it is the other way around with a relative performance that is better in case the covid pandemic is excluded. There are also sectors like mining and finance that have relative performance that is robust to including or excluding the covid pandemic, indicating that their nowcasts are not largely influenced by the pandemic. For a more extensive treatment and examination of macroeconomic modelling and forecasting during the covid pandemic, see, among others, Ng (2021), Schorfheide and Song (2022) and Carriero et al. (2022).

Next, we examine the evolution of the relative nowcasting accuracy. To this end, Figure 11 shows the cumulative sum of squared forecast error difference plots (Welch and Goyal, 2008; Pettenuzzo and Timmermann, 2017) between the several projected nowcasts from the mixed-frequency dynamic factor model and the ones of the benchmark for the four largest non-governmental U.S. sectors. Negative (positive) values imply that the mixed-frequency dynamic factor model generates more (less) accurate nowcasts than the benchmark. Figure 11 indicates that the differences are generally below zero and decreasing, especially for the manufacturing and professional services sectors. This means that over time the model-based nowcasts become more accurate relative to the benchmark. For the forecasts and nowcasts made in Q(-1)M2 and Q(0)M1, the model-based performance deteriorates compared to the benchmark during the covid pandemic, particularly for the health care sector. Meanwhile, for the nowcasts and backcasts made in Q(0)M3 and Q(+1)M2, the relative model-based performance improves during the pandemic, except for the real estate sector. Overall, the model-based nowcast gains are consistent throughout the out-of-sample period and especially pronounced for the nowcasts and backcasts during the covid pandemic.

Lastly, we demonstrate the potential of the sector-level GDP nowcasts (and consequently of the sectoral ECIs) during a period of economic turmoil like the onset of the covid pandemic. Specifically, Figure 12 shows the GDP nowcasts made for the second quarter of 2020 (that is, the first complete quarter in the pandemic) across the different construction dates and sectors, as well as the realized GDP growth rates. Clearly, the nowcasts made in the first quarter of 2020 (that is, in Q(-1)) do not convey any information yet on the pandemic, resulting in 2020Q2 GDP nowcasts values that are often positive and far off from the realized values. However, when entering the pandemic in

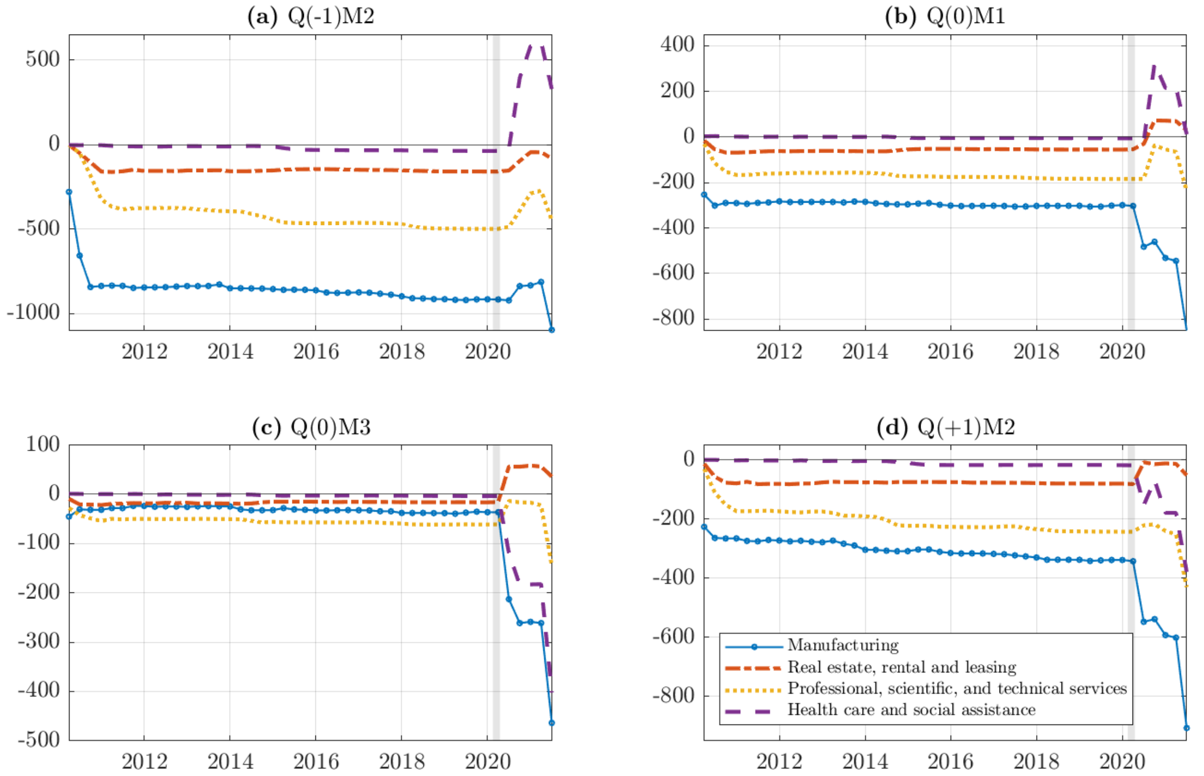

Figure 11: Cumulative sum of squared forecast error difference plots between the nowcasts from the mixed-frequency dynamic factor model and the ones of the AR(1) benchmark for the four largest non-governmental U.S. sectors with gray shaded NBER recession periods.

the second quarter of 2020, more information becomes available like drops in employment and production levels and increases in the number of continued UI claims, which pushes the GDP nowcasts downward. The sectors that are severely hit by the pandemic like arts and accommodation (as indicated by Figure 2) have the deepest drops in both nowcasted and observed GDP growth rates. In fact, the nowcasts constructed from Q(0)M1 onwards are generally close to the corresponding realized values. This indicates that, once enough information has become available, the mixed-frequency dynamic factor model provides an accurate signal of the severity of the drop in output. Such early signals can be used by decision makers in both the private and public sector to anticipate current and near-future economic conditions. Overall, this illustrates the accuracy and usefulness of the sector-level GDP nowcasts that are in turn used as inputs for the sectoral ECIs to more broadly track the economic state of each sector.

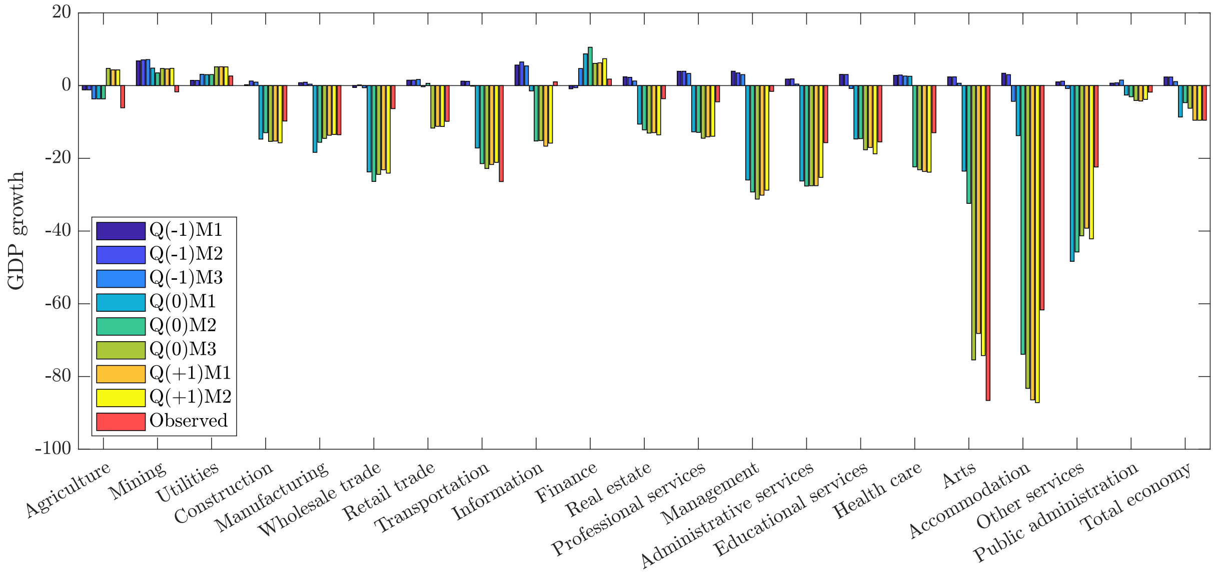

Figure 12: Nowcasts and realized values of 2020Q2 sector-level year-on-year GDP growth rates across sequence of construction dates and sectors.

6 Conclusion

This paper constructs a novel set of 20 monthly U.S. state-level economic conditions indices at the two-digit North American Industry Classification System (NAICS) level from January 1991 to September 2021. In particular, these indices are composed of a small but diverse set of sectoral economic indicators and are estimated using mixed-frequency dynamic factor models. The resulting indices show substantial heterogeneity in dynamics across sectors, particularly during recessions periods, emphasizing the relevance of sectoral disaggregation. Moreover, the indices are generally driven by balanced mix of the underlying indicators, meaning that the various sources of economic activity all seem to contribute to describing sectoral economic conditions. This highlights that sectoral conditions are not just described by one type of economic activity only, but by the comovement of different indicators of sectoral activity.

Given the estimated indices, it becomes straightforward to (re)examine questions concerning the dynamics of sectoral economic activity. Specifically, we revisit the analyses in Foerster et al. (2011) and Andreou et al. (2019) by examining whether the common factors that drive sectoral economic conditions are able to explain the variation in aggregate conditions. Confirming their results, we find that the first common factor drives aggregate economic activity and is closely related to production-related sectors. At the same time, we complement their results by showing that the second common factor serves as a correction factor that handles the differing impacts of the financial crisis and especially covid pandemic on service-providing sectors. Alternatively, the indices could be used to examine spillovers between sectors (Li and Martin, 2019; Guisinger et al., 2021; Brunner and Hipp, 2023) or regime switches within and across sectors (Cooper, 1998; Bidarkota, 1999; Fok et al., 2005; Korenok et al., 2009), but then based on a more complete measure of sectoral activity available for all sectors in the economy (including the service-providing ones). These applications are left for further research.

Lastly, the estimated indices are based on smoothed estimates of the underlying economic indicators when no observed values are available yet due to publication delays. To examine the accuracy of these estimates and consequently of the indices, we conduct a nowcasting exercise of sector-level GDP growth, which are typically published with a three-month delay. For most sectors, the mixed-frequency dynamic factor model outperforms the benchmark, with an average improvement of 22% in terms of root mean squared error. These improvements are most pronounced during the covid pandemic, highlighting the usefulness of these nowcasts as inputs for the sector-level economic conditions indices during times of economic turmoil.

References

Anderson, H., G. Caggiano, F. Vahid, and B. Wong (2020): "Sectoral Employment Dynamics in Australia and the COVID-19 Pandemic," *Australian Economic Review*, 53, 402–414.

Andreou, E., P. Gagliardini, E. Ghysels, and M. Rubin (2019): "Inference in Group Factor Models With an Application to Mixed-Frequency Data," *Econometrica*, 87, 1267–1305.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009): "Real-Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 27, 417–427.

Baffes, J., M. A. Kose, F. Ohnsorge, and M. Stocker (2015): "The Great Plunge in Oil Prices: Causes, Consequences, and Policy Responses," Policy Research Note. PRN/15/01, World Bank, Washington, DC.

Baumeister, C. and L. Kilian (2016): "Understanding the Decline in the Price of Oil since June 2014," *Journal of the Association of Environmental and Resource Economists*, 3, 131–158.

Baumeister, C., D. Leiva-León, and E. Sims (2022): "Tracking Weekly State-Level Economic Conditions," *The Review of Economics and Statistics*, Forthcoming.

Bańbura, M., D. Giannone, and L. Reichlin (2011): "Nowcasting," in *The Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements and D. F. Hendry, Oxford: Oxford University Press, 63–90.

Bańbura, M. and M. Modugno (2014): "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data," *Journal of Applied Econometrics*, 29, 133–160.

Bidarkota, P. V. (1999): "Sectoral Investigation of Asymmetries in the Conditional Mean Dynamics of the Real U.S. GDP," *Studies in Nonlinear Dynamics & Econometrics*, 3, publisher: De Gruyter.

Brave, S. A., M. Fogarty, D. Aaronson, E. Karger, and S. Krane (2021):

"Tracking U.S. Consumers in Real Time with a New Weekly Index of Retail Trade," Tech. rep., Federal Reserve Bank of Chicago.

Brunner, F. and R. Hipp (2023): "Estimating Large-Dimensional Connectedness Tables: The Great Moderation through the Lens of Sectoral Spillovers," *Quantitative Economics*, 14, 1021–1058.

Bry, G. and C. Boschan (1971): *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, National Bureau of Economic Research.

Burns, A. F. and W. C. Mitchell (1946): *Measuring Business Cycles*, National Bureau of Economic Research.

Böhm, H., J. Schaumburg, and L. Tonzer (2022): "Financial Linkages and Sectoral Business Cycle Synchronization: Evidence from Europe," *IMF Economic Review*, 70, 698–734.

Camacho, M. and D. Leiva-Leon (2019): "The Propagation of Industrial Business Cycles," *Macroeconomic Dynamics*, 23, 144–177.

Camacho, M. and G. Perez-Quiros (2010): "Introducing the Euro-STING: Short-Term Indicator of Euro Area Growth," *Journal of Applied Econometrics*, 25, 663–694.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2022): "Addressing COVID-19 Outliers in BVARs with Stochastic Volatility," *The Review of Economics and Statistics*, Forthcoming.

Carriero, A. and M. Marcellino (2011): "Sectoral Survey-based Confidence Indicators for Europe," *Oxford Bulletin of Economics and Statistics*, 73, 175–206.

Cascaldi-Garcia, D., T. R. Ferreira, D. Giannone, and M. Modugno (2023): "Back to the Present: Learning about the Euro Area through a Now-casting Model," *International Journal of Forecasting*, Forthcoming.

Chang, Y. and S. Hwang (2015): "Asymmetric Phase Shifts in U.S. Industrial Production Cycles," *Review of Economics and Statistics*, 97, 116–133.

Cooper, S. J. (1998): "Multiple Regimes in U.S. Output Fluctuations," *Journal of Business & Economic Statistics*, 16, 92–100, publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Crone, T. M. and A. Clayton-Matthews (2005): "Consistent Economic Indexes for the 50 States," *Review of Economics and Statistics*, 87, 593–603.

Cruz, A. M. and E. Krausmann (2008): "Damage to Offshore Oil and Gas Facilities Following Hurricanes Katrina and Rita: An Overview," *Journal of Loss Prevention in the Process Industries*, 21, 620–626.

Da-Rocha, J. M. and D. Restuccia (2006): "The Role of Agriculture in Aggregate Business Cycles," *Review of Economic Dynamics*, 9, 455–482.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1–38.

Diebold, F. X. and G. D. Rudebusch (1996): "Measuring Business Cycles: A Modern Perspective," *The Review of Economics and Statistics*, 78, 67–77.

Engle, R. and M. Watson (1981): "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates," *Journal of the American Statistical Association*, 76, 774–781.

Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011): "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 119, 1–38.

Fok, D., D. v. Dijk, and P. H. Franses (2005): "A Multi-Level Panel STAR Model for US Manufacturing Sectors," *Journal of Applied Econometrics*, 20, 811–827.

Graeve, F. D. and J. D. Schneider (2023): "Identifying Sectoral Shocks and Their Role in Business Cycles," *Journal of Monetary Economics*, Forthcoming.

Guisinger, A. Y., M. T. Owyang, and D. Soques (2021): "Industrial Connectedness and Business Cycle Comovements," *Econometrics and Statistics*, Forthcoming.

Harding, D. and A. Pagan (2002): "Dissecting the Cycle: A Methodological Investigation," *Journal of Monetary Economics*, 49, 365–381.

——— (2005): "A Suggested Framework for Classifying the Modes of Cycle Research," *Journal of Applied Econometrics*, 20, 151–159.

——— (2006): "Synchronization of Cycles," *Journal of Econometrics*, 132, 59–79.

Harvey, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Harvey, D., S. Leybourne, and P. Newbold (1997): "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, 13, 281–291.

Holmes, E. E. (2013): "Derivation of an EM Algorithm for Constrained and Unconstrained Multivariate Autoregressive State-Space (MARSS) Models," Tech. Rep. arXiv:1302.3919, arXiv.

Karadimitropoulou, A. and M. León-Ledesma (2013): "World, Country, and Sector Factors in International Business Cycles," *Journal of Economic Dynamics and Control*, 37, 2913–2927.

Korenok, O., B. Mizrach, and S. Radchenko (2009): "A Note on Demand and Supply Factors in Manufacturing Output Asymmetries," *Macroeconomic Dynamics*, 13, 263–277.

Lewis, D. J., K. Mertens, J. H. Stock, and M. Trivedi (2022): "Measuring Real Activity Using a Weekly Economic Index," *Journal of Applied Econometrics*, 37, 667–687.

Li, N. and V. L. Martin (2019): "Real Sectoral Spillovers: A Dynamic Factor Analysis of the Great Recession," *Journal of Monetary Economics*, 107, 77–95.

Long, J. B. and C. I. Plosser (1987): "Sectoral vs. Aggregate Shocks in the Business Cycle," *American Economic Review*, 77, 333–336.

Lucas, R. E. (1977): "Understanding Business Cycles," *Carnegie-Rochester Conference Series on Public Policy*, 5, 7–29.

Mariano, R. S. and Y. Murasawa (2003): "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," *Journal of Applied Econometrics*, 18, 427–443.

Maroz, D., J. H. Stock, and M. W. Watson (2021): "Comovement of Economic Activity During the Covid Recession," Tech. rep.

Mitchell, J., R. J. Smith, M. R. Weale, S. Wright, and E. L. Salazar (2005): "An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth," *The Economic Journal*, 115, F108–F129.

Ng, S. (2021): "Modeling Macroeconomic Variations After COVID-19," Tech. Rep. arXiv:2103.02732, arXiv.

Nunes, L. C. (2005): "Nowcasting Quarterly GDP Growth in a Monthly Coincident Indicator Model," *Journal of Forecasting*, 24, 575–592.

Opschoor, D. and D. J. van Dijk (2023): "Slow Expectation-Maximization Convergence in Low-Noise Dynamic Factor Models," Tinbergen Institute Discussion Paper 2023-018/III, Tinbergen Institute.

Petersen, K. B., O. Winther, and L. K. Hansen (2005): "On the Slow Convergence of EM and VBEM in Low-Noise Linear Models," *Neural Computation*, 17, 1921–1926.

Pettenuzzo, D. and A. Timmermann (2017): "Forecasting Macroeconomic Variables Under Model Instability," *Journal of Business & Economic Statistics*, 35, 183–201.

Proietti, T. and F. Moauro (2006): "Dynamic Factor Analysis with Non-Linear Temporal Aggregation Constraints," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 55, 281–300.

Rothstein, R. (1997): "Union Strength in the United States: Lessons from the UPS Strike," *International Labour Review*, 136, 469–492.

Salakhutdinov, R. and S. Roweis (2003): "Adaptive Overrelaxed Bound Optimization Methods," in *Proceedings of the Twentieth International Conference on Interna-*

tional Conference on Machine Learning, Washington, DC, USA: AAAI Press, ICML'03, 664–671.

Sarantis, N. (2001): "Nonlinearities, Cyclical Behaviour and Predictability in Stock Markets: International Evidence," *International Journal of Forecasting*, 17, 459–482.

Schorfheide, F. and D. Song (2022): "Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic," *International Journal of Central Banking*, Forthcoming.

Shumway, R. H. and D. S. Stoffer (1982): "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis*, 3, 253–264.

Stock, J. H. and M. W. Watson (1989): "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*, 4, 351–394.

——— (1991): "A Probability Model of the Coincident Economic Indicators," in *Leading Economic Indicators: New Approaches and Forecasting Records*, ed. by G. H. Moore and K. Lahiri, Cambridge: Cambridge University Press, 63–90.

——— (2014): "Estimating Turning Points Using Large Data Sets," *Journal of Econometrics*, 178, 368–381.

Watson, M. W. and R. F. Engle (1983): "Alternative Algorithms for the Estimation of Dynamic Factor, Mimic and Varying Coefficient Regression Models," *Journal of Econometrics*, 23, 385–400.

Wegmüller, P., C. Glocker, and V. Guggia (2023): "Weekly Economic Activity: Measurement and Informational Content," *International Journal of Forecasting*, 39, 228–243.

Welch, I. and A. Goyal (2008): "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508.

Young, P., K. Notis, and T. Firestine (2014): "Transportation Services Index and the Economy—Revisited," Tech. rep., U.S. Department of Transportation, Bureau of Transportation Statistics.

Tracking Sectoral Economic Conditions

Online Appendix

Daan Opschoor

November 7, 2023

Contents

A Estimation of mixed-frequency dynamic factor model

This section discusses the details of the expectation-maximization (EM) algorithm in the estimation of the mixed-frequency dynamic factor model (DFM). For each sector i, the measurement equation of the mixed-frequency DFM in (6) is given by

$$\underbrace{\begin{pmatrix} \boldsymbol{y}_t^M \\ \boldsymbol{y}_t^Q \end{pmatrix}}_{\boldsymbol{y}_t} = \underbrace{\begin{pmatrix} \boldsymbol{\lambda}^M & \mathbf{0} & \boldsymbol{I}_K & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\lambda}^Q & \mathbf{0} & \boldsymbol{I}_L \end{pmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{pmatrix} f_t \\ c_t \\ \boldsymbol{\varepsilon}_t^M \\ \boldsymbol{\varepsilon}_t^Q \end{pmatrix}}_{\boldsymbol{z}_t} + \underbrace{\begin{pmatrix} \boldsymbol{e}_t^M \\ \boldsymbol{e}_t^Q \end{pmatrix}}_{\boldsymbol{e}_t},$$

while the state equation in (7) is given by

$$\underbrace{\begin{pmatrix} \boldsymbol{x}_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix}}_{\boldsymbol{z}_t} = \underbrace{\begin{pmatrix} \boldsymbol{B}_t & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}}_{\boldsymbol{\Phi}_t} \underbrace{\begin{pmatrix} \boldsymbol{x}_{t-1} \\ \boldsymbol{\varepsilon}_{t-1} \end{pmatrix}}_{\boldsymbol{z}_{t-1}} + \underbrace{\begin{pmatrix} \boldsymbol{g} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_N \end{pmatrix}}_{\boldsymbol{G}} \underbrace{\begin{pmatrix} \eta_t \\ \boldsymbol{\nu}_{i,t} \end{pmatrix}}_{\boldsymbol{v}_t},$$

with $\boldsymbol{x}_t = (f_t, c_t)'$, $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_t^{M\prime}, \boldsymbol{\varepsilon}_t^{Q\prime})'$, $\boldsymbol{g} = (1, \frac{1}{3})'$ and

$$\boldsymbol{B}_t = \begin{pmatrix} \phi & 0 \\ \frac{1}{3}\phi & \xi_t \end{pmatrix}. \tag{A.1}$$

Hence, the complete model in state-space form, including the distributional assumptions on the error terms, can be compactly written as

$$\boldsymbol{y}_t = \boldsymbol{\Lambda} \boldsymbol{z}_t + \boldsymbol{e}_t, \qquad \boldsymbol{e}_t \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I}_N), \tag{A.2}$$

$$\boldsymbol{z}_t = \boldsymbol{\Phi}_t \boldsymbol{z}_{t-1} + \boldsymbol{G} \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \tag{A.3}$$

for $t = 1, \ldots, T$, where, following Bańbura and Modugno (2014), τ is pre-fixed at a very small value (that is, 10^{-4}).

A.1 Expectation-maximization algorithm

Given the model in (A.2)-(A.3), we need to estimate the unknown parameters in $\boldsymbol{\Theta} = \{\boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}, \boldsymbol{\Omega}\}$ and the latent states in $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T)'$ with the expectation-maxization

(EM) algorithm of Dempster et al. (1977). The idea of the EM algorithm is to focus on the complete data log-likelihood of $\boldsymbol{Z}$ and $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)'$, denoted as $\ell(\boldsymbol{Z}, \boldsymbol{Y}; \boldsymbol{\Theta})$. As not all elements in $\boldsymbol{Y}$ are observed, we follow Bańbura and Modugno (2014) and Spånberg (2022) and integrate out the missing data from the likelihood function. we define $\boldsymbol{W}_t$ as an N-dimensional diagonal matrix with the n-th diagonal element $w_{n,t}$ being an indicator function that equals one if $y_{n,t}$ is observed and zero otherwise for all $n = 1, \ldots, N$ and $t = 1, \ldots, T$. Moreover, let $P = \sum_{t=1}^{T} \sum_{n=1}^{N} w_{n,t}$ denote the total number of available observations. Then, the complete data log-likelihood of the model in (A.2)-(A.3) can be written as

$$\begin{aligned}\ell(\boldsymbol{Z}, \boldsymbol{Y}; \boldsymbol{\Theta}) = &-\frac{1}{2}\log|\boldsymbol{V}_0| - \frac{1}{2}(\boldsymbol{z}_0 - \boldsymbol{\mu}_0)'\boldsymbol{V}_0^{-1}(\boldsymbol{z}_0 - \boldsymbol{\mu}_0) \\ &- \frac{P}{2}\log\tau - \frac{1}{2\tau}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{\Lambda}\boldsymbol{z}_t)'\boldsymbol{W}_t(\boldsymbol{y}_t - \boldsymbol{\Lambda}\boldsymbol{z}_t) \\ &- \frac{T}{2}\log|\boldsymbol{G}\boldsymbol{\Omega}\boldsymbol{G}'| - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{z}_t - \boldsymbol{\Phi}\boldsymbol{z}_{t-1})'(\boldsymbol{G}\boldsymbol{\Omega}\boldsymbol{G}')^{-1}(\boldsymbol{z}_t - \boldsymbol{\Phi}\boldsymbol{z}_{t-1}),\end{aligned}$$

where $\boldsymbol{\mu}_0$ and $\boldsymbol{V}_0$ follow from the initial conditions that are specified as $\boldsymbol{z}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{V}_0)$ with $\boldsymbol{V}_0$ being a diagonal matrix. The expectation step (E-step) is conducted by taking the expectation of the complete data log-likelihood conditional on the observed data $\boldsymbol{Y}$ based on the j-th iteration of the parameter estimates, denoted as $\boldsymbol{\Theta}^{(j)}$, that is,

$$\mathcal{L}(\boldsymbol{\Theta}; \boldsymbol{\Theta}^{(j)}) = \mathbb{E}_{(j)}\Big(\ell(\boldsymbol{Z}, \boldsymbol{Y}; \boldsymbol{\Theta})\Big|\boldsymbol{Y}\Big), \tag{A.4}$$

which can be computed by a pass of the Kalman filter and smoother (see, for example, Durbin and Koopman, 2012, for a textbook treatment), where $\mathbb{E}_{(j)}(\cdot)$ indicates that the expectation is taken conditional on $\boldsymbol{\Theta}^{(j)}$. Next, the maximization step (M-step) is conducted by maximizing the expected complete data log-likelihood with respect to $\boldsymbol{\Theta}$, that is,

$$\boldsymbol{\Theta}^{(j+1)} = \arg\max_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\Theta}; \boldsymbol{\Theta}^{(j)}). \tag{A.5}$$

Fortunately, there exist analytic solutions to the maximization problem in (A.5) for the unknown parameters in $\boldsymbol{\Theta}$, making the algorithm computationally fast, even in high dimensions. These are derived in Section A.2. Finally, the E-step in (A.4) and M-step in

(A.5) are iterated until convergence, where Dempster et al. (1977) show that under some regularity conditions the algorithm converges to a local optimum.

To determine the convergence, we follow the stopping rule of Doz et al. (2012), which states that the algorithm is stopped at the first iteration $j < J$ that satisfies

$$\frac{\left|\ell\big(\boldsymbol{Y};\boldsymbol{\Theta}^{(j)}\big) - \ell\big(\boldsymbol{Y};\boldsymbol{\Theta}^{(j-1)}\big)\right|}{\frac{1}{2}\left|\ell\big(\boldsymbol{Y};\boldsymbol{\Theta}^{(j)}\big) + \ell\big(\boldsymbol{Y};\boldsymbol{\Theta}^{(j-1)}\big)\right|} < \epsilon,$$

where J is the maximum number of iterations set equal to 5,000, ϵ is the pre-specified tolerance level set equal to 10^{-8} and $\ell(\boldsymbol{Y};\boldsymbol{\Theta}^{(j)})$ is the prediction error log-likelihood of $\boldsymbol{Y}$ evaluated at the j-th parameter iteration. Moreover, the EM algorithm is initialized with the estimates from the two-step approach of Doz et al. (2011) based on principal component analysis and ordinary least squares.

A.2 M-step derivations

In order to derive the M-step for the parameter of interest in $\boldsymbol{\Theta}$, we need to take the partial derivative of $\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{\Theta}^{(j)})$ with respect to that parameter and set it equal to zero.

The M-step of the diagonal elements in $\boldsymbol{\Psi}$ that are part of $\boldsymbol{\Phi}$ can be derived by taking the partial derivative with respect to ψ_n for each $n = 1,\dots,N$, that is,

$$\begin{aligned}\frac{\partial}{\partial \psi_n}\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{\Theta}^{(j)}) &= \frac{\partial}{\partial \psi_n}\left(-\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[(\boldsymbol{z}_t - \boldsymbol{\Phi}\boldsymbol{z}_{t-1})'\boldsymbol{\Omega}^{-1}(\boldsymbol{z}_t - \boldsymbol{\Phi}\boldsymbol{z}_{t-1})\middle|\boldsymbol{Y}\right]\right)\\ &= \frac{\partial}{\partial \psi_n}\left(-\frac{1}{2\sigma_n^2}\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[(\varepsilon_{n,t} - \psi_n\varepsilon_{n,t-1})^2\middle|\boldsymbol{Y}\right]\right)\\ &= \frac{1}{\sigma_n^2}\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[\varepsilon_{n,t-1}(\varepsilon_{n,t} - \psi_n\varepsilon_{n,t-1})\middle|\boldsymbol{Y}\right],\end{aligned}$$

where $\varepsilon_{n,t}$ denotes the n-th element of $\boldsymbol{\varepsilon}_t$ and σ_n^2 the n-th diagonal element of $\boldsymbol{\Sigma}$. Note that we discard the superscript M or Q as the expressions are similar across frequencies. Setting the partial derivative equal to zero and rearranging gives

$$\psi_n^{(j+1)} = \left(\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[\varepsilon_{n,t-1}\varepsilon_{n,t}\middle|\boldsymbol{Y}\right]\right)\left(\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[\varepsilon_{n,t-1}^2\middle|\boldsymbol{Y}\right]\right)^{-1}, \tag{A.6}$$

for $n = 1,\dots,N$, where $\mathbb{E}_{(j)}\left[\varepsilon_{n,t-1}\varepsilon_{n,t}\middle|\boldsymbol{Y}\right]$ and $\mathbb{E}_{(j)}\left[\varepsilon_{n,t-1}^2\middle|\boldsymbol{Y}\right]$ can be obtained by the

Kalman smoother.

Moving to the diagonal elements in $\boldsymbol{\Sigma}$, denoted as σ_n^2, that are part of $\boldsymbol{\Omega}$, we get

$$\begin{aligned}
\frac{\partial}{\partial \sigma_n^2}\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{\Theta}^{(j)}) &= \frac{\partial}{\partial \sigma_n^2}\Bigg(-\frac{T}{2}\log|\boldsymbol{G\Omega G}'| \\
&\qquad -\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_j\left[(\boldsymbol{z}_t-\boldsymbol{\Phi z}_{t-1})'(\boldsymbol{G\Omega G}')^{-1}(\boldsymbol{z}_t-\boldsymbol{\Phi z}_{t-1})\Big|\boldsymbol{Y}\right]\Bigg) \\
&= \frac{\partial}{\partial \sigma_n^2}\Bigg(-\frac{T}{2}\log\sigma_n^2 - \frac{1}{2\sigma_n^2}\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[(\varepsilon_{n,t}-\psi_{n,i}\varepsilon_{n,t-1})^2\Big|\boldsymbol{Y}_i\right]\Bigg) \\
&= -\frac{T}{2\sigma_n^2} + \frac{1}{2\sigma_n^4}\left(\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[\varepsilon_{n,t}^2 - 2\psi_n\varepsilon_{n,t}\varepsilon_{n,t-1} + \psi_n^2\varepsilon_{n,t-1}^2\big|\boldsymbol{Y}\right]\right).
\end{aligned}$$

for each $n=1,\ldots,N$. Setting equal to zero and rearranging gives

$$\begin{aligned}
\sigma_n^{2\,(j+1)} &= \frac{1}{T}\left(\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[\varepsilon_{n,t}^2\big|\boldsymbol{Y}\right] - 2\psi_n\mathbb{E}_{(j)}\left[\varepsilon_{n,t}\varepsilon_{n,t-1}\big|\boldsymbol{Y}\right] + \psi_n^2\mathbb{E}_{(j)}\left[\varepsilon_{n,t-1}^2\big|\boldsymbol{Y}\right]\right) \\
&= \frac{1}{T}\left(\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[\varepsilon_{n,t}^2\big|\boldsymbol{Y}\right] - \psi_n^{(j+1)}\mathbb{E}_{(j)}\left[\varepsilon_{n,t}\varepsilon_{n,t-1}\big|\boldsymbol{Y}\right]\right),
\end{aligned}$$

for each $n=1,\ldots,N$, where the second equality follows from setting ψ_n equal to the new iteration value $\psi_n^{(j+1)}$ derived in equation (A.6) and the conditional expectations are again obtained by the Kalman smoother.

Turning to the factor loadings corresponding to the monthly series λ_k^M in $\boldsymbol{\Lambda}$ for $k = 1,\ldots,K$, we get

$$\begin{aligned}
\frac{\partial}{\partial \lambda_k^M}\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{\Theta}^{(j)}) &= \frac{\partial}{\partial \lambda_k^M}\left(-\frac{1}{2\tau}\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[(\boldsymbol{y}_t-\boldsymbol{\Lambda z}_t)'\boldsymbol{W}_t(\boldsymbol{y}_t-\boldsymbol{\Lambda z}_t)\Big|\boldsymbol{Y}\right]\right) \\
&= \frac{\partial}{\partial \lambda_k^M}\left(-\frac{1}{2\tau}\sum_{t=1}^{T}w_{k,t}^M\mathbb{E}_{(j)}\left[(y_{k,t}^M-\lambda_k^M f_t-\varepsilon_{k,t}^M)^2\Big|\boldsymbol{Y}\right]\right) \\
&= \frac{1}{\tau}\left(\sum_{t=1}^{T}w_{k,t}^M\mathbb{E}_{(j)}\left[f_t(y_{k,t}^M-\lambda_k^M f_t-\varepsilon_{k,t}^M)\big|\boldsymbol{Y}\right]\right),
\end{aligned}$$

where $w_{k,t}^M$ denotes the availability indicator of the k-th monthly series. Setting equal to zero and rearranging gives

$$\lambda_k^{M(j+1)} = \frac{\sum_{t=1}^{T} w_{k,t}^{M}\Big(\mathbb{E}_{(j)}\left[f_t\middle|\boldsymbol{Y}\right]y_{k,t}^{M} - \mathbb{E}_{(j)}\left[f_t\varepsilon_{k,t}^{M}\middle|\boldsymbol{Y}\right]\Big)}{\sum_{t=1}^{T} w_{k,t}^{M}\mathbb{E}_{(j)}\left[f_t^2\middle|\boldsymbol{Y}\right]}, \tag{A.7}$$

for $k = 1, \dots, K$, where the conditional expectations can be obtained by the Kalman smoother. Similarly for the factor loadings corresponding to the quarterly series λ_l^Q, we obtain

$$\lambda_l^{Q(j+1)} = \frac{\sum_{t=1}^{T} w_{l,t}^{Q}\Big(\mathbb{E}_{(j)}\left[c_t\middle|\boldsymbol{Y}\right]y_{l,t}^{Q} - \mathbb{E}_{(j)}\left[c_t\varepsilon_{l,t}^{Q}\middle|\boldsymbol{Y}\right]\Big)}{\sum_{t=1}^{T} w_{l,t}^{Q}\mathbb{E}_{(j)}\left[c_t^2\middle|\boldsymbol{Y}\right]}, \tag{A.8}$$

for $l = 1, \dots, L$, where $w_{l,t}^Q$ denotes the availability indicators of the l-th quarterly series and the conditional expectations are obtained by the Kalman smoother.

However, Opschoor and van Dijk (2023) show that the framework of Bańbura and Modugno (2014), in which the idiosyncratic components are included in the state vector, leads to extremely slow EM convergence in estimating $\boldsymbol{\lambda}^M$ and $\boldsymbol{\lambda}^Q$ with the conventional M-steps in (A.7) and (A.8). To overcome this issue, Opschoor and van Dijk (2023) advocate to use the overrelaxed adaptive EM (AEM) algorithm of Salakhutdinov and Roweis (2003) for estimating the factor loadings, which they show speeds up convergence and leads to more accurate estimates of factor loadings and latent factors. The new M-step of λ_k^M and λ_l^Q under this adaptive scheme is

$$\tilde{\lambda}_k^{M(j+1)} = \tilde{\lambda}_k^{M(j)} + \rho_j\Big(\lambda_k^{M(j+1)} - \tilde{\lambda}_k^{M(j)}\Big), \tag{A.9}$$

for $k = 1, \dots, K_i$, and

$$\tilde{\lambda}_l^{Q(j+1)} = \tilde{\lambda}_l^{Q(j)} + \rho_j\Big(\lambda_l^{Q(j+1)} - \tilde{\lambda}_l^{Q(j)}\Big), \tag{A.10}$$

for $l = 1, \dots, L$, respectively, where $\tilde{\lambda}_k^{M(j)}$ and $\tilde{\lambda}_l^{Q(j)}$ denote the j-th factor loading iterations of the AEM algorithm and $\lambda_k^{M(j+1)}$ and $\lambda_l^{Q(j+1)}$ are obtained from (A.7) and (A.8), respectively. Moreover, ρ_j is the adaptive factor that is able to grow over the iterations, increasing the learning rate relative to the conventional M-step. We follow Salakhutdinov and Roweis (2003) and use the update $\rho_{j+1} = \alpha\rho_j$ with $\rho_1 = 1$ and $\alpha = 1.1$. To ensure monotonic increments, we follow Salakhutdinov and Roweis (2003) and reset $\rho_{j+1} = 1$,

$\tilde{\lambda}_k^{M(j+1)} = \lambda_k^{M(j+1)}$ for all $k = 1, \ldots, K$, and $\tilde{\lambda}_l^{Q(j+1)} = \lambda_l^{Q(j+1)}$ for all $l = 1, \ldots, L$ in case the likelihood does not improve in iteration $j+1$, after which the algorithm continues. A comparison of the convergence in estimating the sector-specific mixed-frequency DFMs using the standard M-step and adaptive M-step for the loadings is given in Appendix A.3.

Moving to the factor persistence ϕ in $\boldsymbol{\Phi}_t$, recall that ϕ appears twice in $\boldsymbol{\Phi}_t$ and is thus bounded by certain restrictions. To incorporate this restriction, we follow Holmes (2013) and write the matrix $\boldsymbol{B}_t$ in equation (A.1) in its vectorized form, that is,

$$\operatorname{vec}(\boldsymbol{B}_t) = \boldsymbol{a}_t + \boldsymbol{d}\phi,$$

where $\operatorname{vec}(\boldsymbol{B}_t) = (\phi, \frac{1}{3}\phi, 0, \xi_t)'$, $\boldsymbol{a}_t = (0, 0, 0, \xi_t)'$ and $\boldsymbol{d} = (1, \frac{1}{3}, 0, 0)'$. Furthermore, the covariance matrix of the error term $\boldsymbol{g}\eta_t$ is equal to $\boldsymbol{g}\boldsymbol{g}'$, which is not invertable. Instead, we follow Holmes (2013) and pre-multiply the left- and right-hand side of the state equation corresponding to $\boldsymbol{x}_t$ with $\boldsymbol{\Xi} = (\boldsymbol{g}'\boldsymbol{g})^{-1}\boldsymbol{g}'$, which only requires the invertability of $\boldsymbol{g}'\boldsymbol{g}$ and does not affect the argument of ϕ that maximizes the expected complete data log-likelihood. Consequently, we can write the partial derivative with respect to ϕ as

$$\begin{aligned}
\frac{\partial}{\partial\phi}\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{\Theta}^{(j)}) &= \frac{\partial}{\partial\phi}\left(-\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[(\boldsymbol{x}_t - \boldsymbol{B}_t\boldsymbol{x}_{t-1})'\boldsymbol{\Xi}'\boldsymbol{\Xi}(\boldsymbol{x}_t - \boldsymbol{B}_t\boldsymbol{x}_{t-1})\middle|\boldsymbol{Y}\right]\right)\\
&= \frac{\partial}{\partial\phi}\left(-\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[\big(\boldsymbol{x}_t - \boldsymbol{\Upsilon}_{t-1}\operatorname{vec}(\boldsymbol{B}_t)\big)'\boldsymbol{Q}\big(\boldsymbol{x}_t - \boldsymbol{\Upsilon}_{t-1}\operatorname{vec}(\boldsymbol{B}_t)\big)\middle|\boldsymbol{Y}\right]\right)\\
&= \frac{\partial}{\partial\phi}\left(-\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_{(j)}\left[\big(\boldsymbol{x}_t - \boldsymbol{\Upsilon}_{t-1}(\boldsymbol{a}_t + \boldsymbol{d}\phi)\big)'\boldsymbol{Q}\big(\boldsymbol{x}_t - \boldsymbol{\Upsilon}_{t-1}(\boldsymbol{a}_t + \boldsymbol{d}\phi)\middle|\boldsymbol{Y}\right]\right)\\
&= \frac{\partial}{\partial\phi}\left(-\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_{(j)}\Big[-\boldsymbol{x}_t'\boldsymbol{Q}\boldsymbol{\Upsilon}_{t-1}\boldsymbol{d}\phi + \boldsymbol{a}_t'\boldsymbol{\Upsilon}_{t-1}'\boldsymbol{Q}\boldsymbol{\Upsilon}_{t-1}\boldsymbol{d}\phi\right.\\
&\qquad\left. - \boldsymbol{d}'\boldsymbol{\Upsilon}_{t-1}'\boldsymbol{Q}\boldsymbol{x}_t\phi + \boldsymbol{d}'\boldsymbol{\Upsilon}_{t-1}'\boldsymbol{Q}\boldsymbol{\Upsilon}_{t-1}\boldsymbol{a}_t\phi + \boldsymbol{d}'\boldsymbol{\Upsilon}_{t-1}'\boldsymbol{Q}\boldsymbol{\Upsilon}_{t-1}\boldsymbol{d}\phi^2\middle|\boldsymbol{Y}\Big]\right)\\
&= \sum_{t=1}^{T}\mathbb{E}_{(j)}\left[\boldsymbol{d}'\boldsymbol{\Upsilon}_{t-1}'\boldsymbol{Q}\boldsymbol{x}_t - \boldsymbol{d}'\boldsymbol{\Upsilon}_{t-1}'\boldsymbol{Q}\boldsymbol{\Upsilon}_{t-1}\boldsymbol{a}_t - \boldsymbol{d}'\boldsymbol{\Upsilon}_{t-1}'\boldsymbol{Q}\boldsymbol{\Upsilon}_{t-1}\boldsymbol{d}\phi\middle|\boldsymbol{Y}\right]\Big)
\end{aligned}$$

where the second equality follows from defining $\boldsymbol{\Upsilon}_{t-1} = (\boldsymbol{x}_{t-1}' \otimes \boldsymbol{I}_2)$ and $\boldsymbol{Q} = \boldsymbol{\Xi}'\boldsymbol{\Xi}$. Setting equal to zero and rearranging gives

$$\phi^{(j+1)} = \frac{\sum_{t=1}^{T} \boldsymbol{d}' \Big(\operatorname{vec}\Big(\boldsymbol{Q}\mathbb{E}_{(j)}\big[\boldsymbol{x}_t\boldsymbol{x}_{t-1}'\big|\boldsymbol{Y}\big]\Big) - \Big(\mathbb{E}_{(j)}\big[\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\big|\boldsymbol{Y}\big] \otimes \boldsymbol{Q}\Big)\boldsymbol{a}_t\Big)}{\sum_{t=1}^{T} \boldsymbol{d}'\Big(\mathbb{E}_{(j)}\big[\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\big|\boldsymbol{Y}\big] \otimes \boldsymbol{Q}\Big)\boldsymbol{d}},$$

where we use that

$$\begin{aligned}
\mathbb{E}_{(j)}\Big[\boldsymbol{\Upsilon}_{t-1}'\boldsymbol{Q}\boldsymbol{\Upsilon}_{t-1}\Big|\boldsymbol{Y}\Big] &= \mathbb{E}_{(j)}\Big[(\boldsymbol{x}_{t-1}' \otimes \boldsymbol{I}_2)'\boldsymbol{Q}(\boldsymbol{x}_{t-1}' \otimes \boldsymbol{I}_2)\Big|\boldsymbol{Y}\Big] \\
&= \mathbb{E}_{(j)}\Big[(\boldsymbol{x}_{t-1} \otimes \boldsymbol{I}_2)(\boldsymbol{x}_{t-1}' \otimes \boldsymbol{Q})\Big|\boldsymbol{Y}\Big] \\
&= \mathbb{E}_{(j)}\Big[\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}' \otimes \boldsymbol{Q}\Big|\boldsymbol{Y}\Big] \\
&= \mathbb{E}_{(j)}\Big[\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\Big|\boldsymbol{Y}\Big] \otimes \boldsymbol{Q},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}_{(j)}\Big[\boldsymbol{\Upsilon}_{t-1}'\boldsymbol{Q}\boldsymbol{x}_t\Big|\boldsymbol{Y}\Big] &= \mathbb{E}_{(j)}\Big[(\boldsymbol{x}_{t-1}' \otimes \boldsymbol{I}_2)'\boldsymbol{Q}\boldsymbol{x}_t\Big|\boldsymbol{Y}\Big] \\
&= \mathbb{E}_{(j)}\Big[(\boldsymbol{x}_{t-1} \otimes \boldsymbol{Q})\boldsymbol{x}_t\Big|\boldsymbol{Y}\Big] \\
&= \mathbb{E}_{(j)}\Big[\operatorname{vec}\big(\boldsymbol{Q}\boldsymbol{x}_t\boldsymbol{x}_{t-1}'\big)\Big|\boldsymbol{Y}\Big] \\
&= \operatorname{vec}\Big(\boldsymbol{Q}\mathbb{E}_{(j)}\big[\boldsymbol{x}_t\boldsymbol{x}_{t-1}'\big|\boldsymbol{Y}\big]\Big).
\end{aligned}$$

Lastly, turning to the parameters corresponding to the initial conditions $\boldsymbol{\mu}_0$ and $\boldsymbol{V}_0$, we get

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\mu}_0}\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{\Theta}^{(j)}) &= \frac{\partial}{\partial \boldsymbol{\mu}_0}\Bigg(-\frac{1}{2}\mathbb{E}_{(j)}\Big[(\boldsymbol{z}_0 - \boldsymbol{\mu}_0)'\boldsymbol{V}_0^{-1}(\boldsymbol{z}_0 - \boldsymbol{\mu}_0)\Big|\boldsymbol{Y}\Big]\Bigg) \\
&= \boldsymbol{V}_0^{-1}\Big(\mathbb{E}_{(j)}\big[\boldsymbol{z}_0\big|\boldsymbol{Y}\big] - \boldsymbol{\mu}_0\Big),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{V}_0}\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{\Theta}^{(j)}) &= \frac{\partial}{\partial \boldsymbol{V}_0}\Bigg(-\frac{1}{2}\log|\boldsymbol{V}_0| - \frac{1}{2}\mathbb{E}_{(j)}\Big[(\boldsymbol{z}_0 - \boldsymbol{\mu}_0)'\boldsymbol{V}_0^{-1}(\boldsymbol{z}_0 - \boldsymbol{\mu}_0)\Big|\boldsymbol{Y}\Big]\Bigg) \\
&= -\frac{1}{2}\boldsymbol{V}_0^{-1} + \frac{1}{2}\boldsymbol{V}_0^{-1}\mathbb{E}_{(j)}\Big[(\boldsymbol{z}_0 - \boldsymbol{\mu}_0)(\boldsymbol{z}_0 - \boldsymbol{\mu}_0)'\Big|\boldsymbol{Y}\Big]\boldsymbol{V}_0^{-1}.
\end{aligned}$$

Setting equal to zero and rearranging gives

$$\boldsymbol{\mu}_0^{(j+1)} = \mathbb{E}_{(j)}\big[\boldsymbol{z}_0 \big| \boldsymbol{Y}\big],$$

and, also taking into account that $\boldsymbol{V}_0$ should be diagonal,

$$\boldsymbol{V}_0^{(j+1)} = \text{diag}\left(\mathbb{E}_{(j)}\Big[(\boldsymbol{z}_0 - \boldsymbol{\mu}_0)(\boldsymbol{z}_0 - \boldsymbol{\mu}_0)' \Big| \boldsymbol{Y}\Big]\right),$$

which can both be obtained by the Kalman smoother.

A.3 Convergence comparison of EM and AEM algorithms

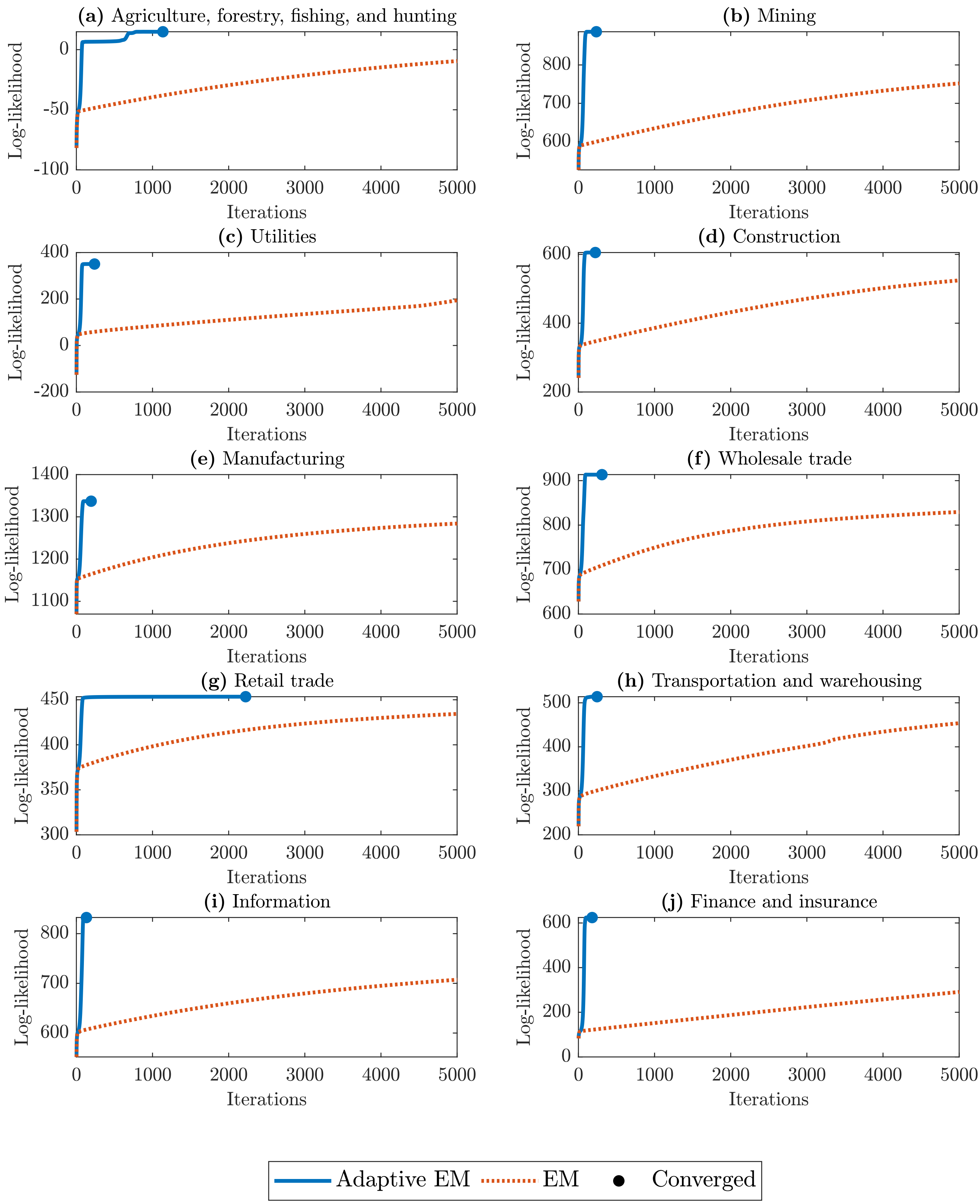

Figure A.1: Log-likelihood values over iterations until convergence (based on tolerance of $\epsilon = 10^{-8}$) or reaching maximum number of iterations ($J = 5,000$) of the EM and AEM algorithm in estimating sector-specific mixed-frequency dynamic factor models.

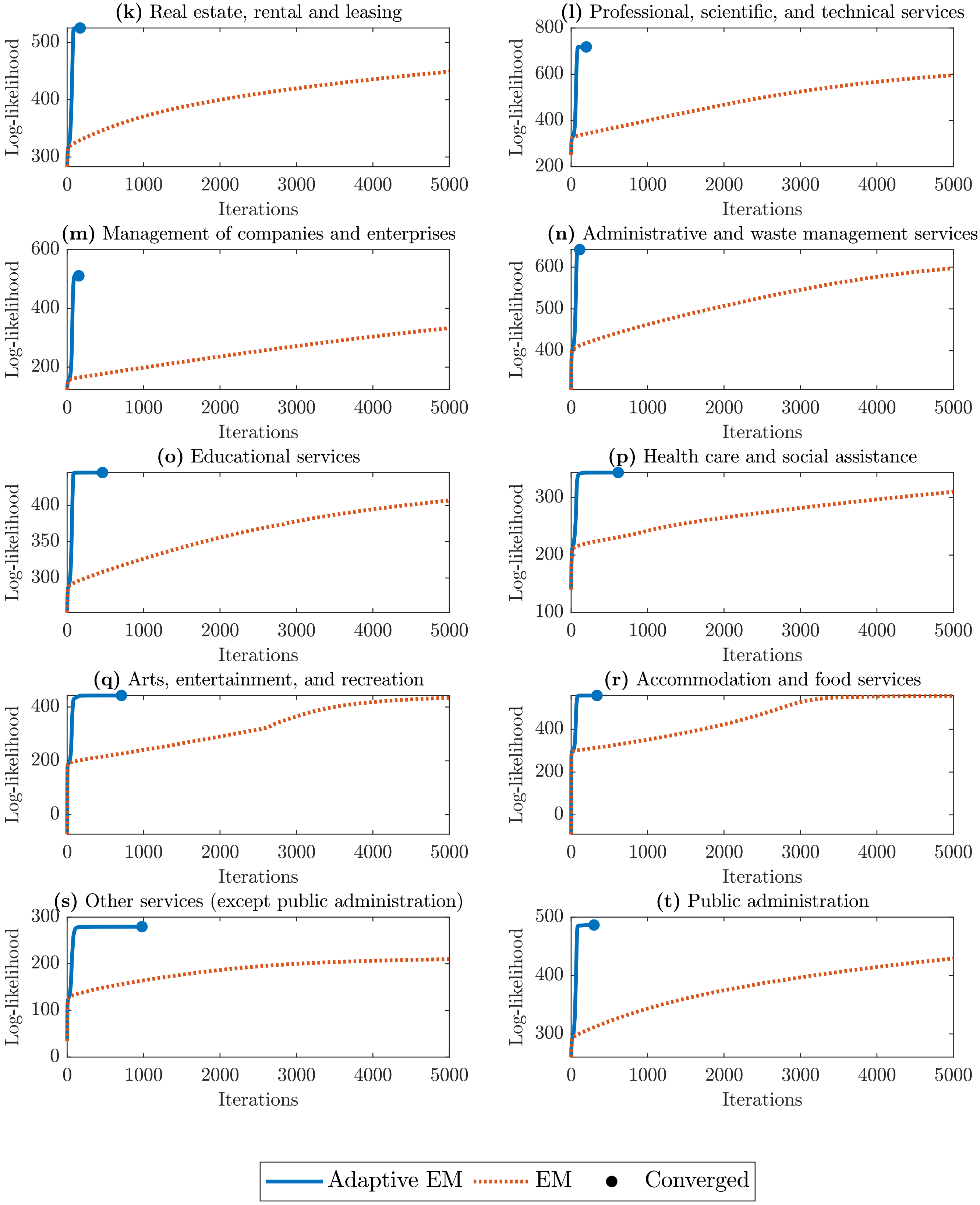

Figure A.1: Log-likelihood values over iterations until convergence (based on tolerance of $\epsilon = 10^{-8}$) or reaching maximum number of iterations ($J = 5,000$) of the EM and AEM algorithm in estimating sector-specific mixed-frequency dynamic factor models (continued).

B Description of data

Table B.1: Overview of sector-level and aggregate economic series

NAICS Code	Sector	Variables	Frequency	Starting date	Source
11	Agriculture, forestry, fishing, and hunting	Real GDP	Quarterly	2005Q1	BEA
11	Agriculture, forestry, fishing, and hunting	Employment level	Monthly	1948M1	FRED
11	Agriculture, forestry, fishing, and hunting	Continued UI claims	Monthly	2005M2	DOL
11	Agriculture, forestry, fishing, and hunting	Personal income	Quarterly	1998Q1	BEA
21	Mining	Real GDP	Quarterly	2005Q1	BEA
21	Mining	Industrial production	Monthly	1919M1	FRB
21	Mining	Employment level	Monthly	1939M1	BLS
21	Mining	Continued UI claims	Monthly	2003M12	DOL
21	Mining	Fuel sales of total crude oil and petroleum products	Monthly	1981M1	EIA
21	Mining	Personal income	Quarterly	1988Q1	BEA
22	Utilities	Real GDP	Quarterly	2005Q1	BEA
22	Utilities	Industrial production	Monthly	1939M1	FRB
22	Utilities	Employment level	Monthly	1964M1	BLS
22	Utilities	Continued UI claims	Monthly	2003M12	DOL
22	Utilities	Revenue	Quarterly	2010Q1	CB
22	Utilities	Personal income	Quarterly	1998Q1	BEA
23	Construction	Real GDP	Quarterly	2005Q1	BEA
23	Construction	Employment level	Monthly	1939M1	BLS
23	Construction	Continued UI claims	Monthly	2003M12	DOL
23	Construction	Personal income	Quarterly	1998Q1	BEA
31-33	Manufacturing	Real GDP	Quarterly	2005Q1	BEA
31-33	Manufacturing	Industrial production	Monthly	1972M1	FRB
31-33	Manufacturing	Employment level	Monthly	1939M1	BLS
31-33	Manufacturing	Continued UI claims	Monthly	2003M12	DOL
31-33	Manufacturing	Manufacturers sales	Monthly	1992M1	CB
31-33	Manufacturing	Personal income	Quarterly	1998Q1	BEA

Table B.1: Continued

NAICS Code	Sector	Variables	Frequency	Starting date	Source
42	Wholesale trade	Real GDP	Quarterly	2005Q1	BEA
42	Wholesale trade	Employment level	Monthly	1939M1	BLS
42	Wholesale trade	Continued UI claims	Monthly	2004M10	DOL
42	Wholesale trade	Wholesalers sales	Monthly	1992M1	CB
42	Wholesale trade	Personal income	Quarterly	1998Q1	BEA
44-45	Retail trade	Real GDP	Quarterly	2005Q1	BEA
44-45	Retail trade	Employment level	Monthly	1939M1	BLS
44-45	Retail trade	Continued UI claims	Monthly	2005M2	DOL
44-45	Retail trade	Retailers sales	Monthly	1992M1	CB
44-45	Retail trade	Personal income	Quarterly	1998Q1	BEA
48-49	Transportation and warehousing	Real GDP	Quarterly	2005Q1	BEA
48-49	Transportation and warehousing	Employment level	Monthly	1972M1	BLS
48-49	Transportation and warehousing	Continued UI claims	Monthly	2005M2	DOL
48-49	Transportation and warehousing	Revenue	Quarterly	2010Q1	CB
48-49	Transportation and warehousing	Personal income	Quarterly	1998Q1	BEA
51	Information	Real GDP	Quarterly	2005Q1	BEA
51	Information	Industrial production	Monthly	1972M1	FRB
51	Information	Employment level	Monthly	1939M1	BLS
51	Information	Continued UI claims	Monthly	2005M2	DOL
51	Information	Revenue	Quarterly	2003Q4	CB
51	Information	Personal income	Quarterly	1998Q1	BEA
52	Finance and insurance	Real GDP	Quarterly	2005Q1	BEA
52	Finance and insurance	Employment level	Monthly	1990M1	BLS
52	Finance and insurance	Continued UI claims	Monthly	2003M12	DOL
52	Finance and insurance	Revenue	Quarterly	2009Q3	CB
52	Finance and insurance	Personal income	Quarterly	1998Q1	BEA
53	Real estate and rental and leasing	Real GDP	Quarterly	2005Q1	BEA
53	Real estate and rental and leasing	Employment level	Monthly	1990M1	BLS
53	Real estate and rental and leasing	Continued UI claims	Monthly	2005M2	DOL
53	Real estate and rental and leasing	Revenue	Quarterly	2012Q3	CB
53	Real estate and rental and leasing	Personal income	Quarterly	1998Q1	BEA

Table B.1: Continued

NAICS Code	Sector	Variables	Frequency	Starting date	Source
54	Professional, scientific, and technical services	Real GDP	Quarterly	2005Q1	BEA
54	Professional, scientific, and technical services	Employment level	Monthly	1939M1	BLS
54	Professional, scientific, and technical services	Continued UI claims	Monthly	2003M12	DOL
54	Professional, scientific, and technical services	Revenue	Quarterly	2006Q3	CB
54	Professional, scientific, and technical services	Personal income	Quarterly	1998Q1	BEA
55	Management of companies and enterprises	Real GDP	Quarterly	2005Q1	BEA
55	Management of companies and enterprises	Employment level	Monthly	1990M1	BLS
55	Management of companies and enterprises	Continued UI claims	Monthly	2005M2	DOL
55	Management of companies and enterprises	Personal income	Quarterly	1998Q1	BEA
56	Administrative and waste management services	Real GDP	Quarterly	2005Q1	BEA
56	Administrative and waste management services	Employment level	Monthly	1990M1	BLS
56	Administrative and waste management services	Continued UI claims	Monthly	2005M2	DOL
56	Administrative and waste management services	Revenue	Quarterly	2006Q3	CB
56	Administrative and waste management services	Personal income	Quarterly	1998Q1	BEA
61	Educational services	Real GDP	Quarterly	2005Q1	BEA
61	Educational services	Employment level	Monthly	1990M1	BLS
61	Educational services	Continued UI claims	Monthly	2005M2	DOL
61	Educational services	Revenue	Quarterly	2010Q1	CB
61	Educational services	Personal income	Quarterly	1998Q1	BEA
62	Health care and social assistance	Real GDP	Quarterly	2005Q1	BEA
62	Health care and social assistance	Employment level	Monthly	1990M1	BLS
62	Health care and social assistance	Continued UI claims	Monthly	2005M2	DOL
62	Health care and social assistance	Revenue	Quarterly	2009Q1	CB
62	Health care and social assistance	Personal income	Quarterly	1998Q1	BEA
71	Arts, entertainment, and recreation	Real GDP	Quarterly	2005Q1	BEA
71	Arts, entertainment, and recreation	Employment level	Monthly	1990M1	BLS
71	Arts, entertainment, and recreation	Continued UI claims	Monthly	2005M2	DOL
71	Arts, entertainment, and recreation	Revenue	Quarterly	2009Q1	CB
71	Arts, entertainment, and recreation	Personal income	Quarterly	1998Q1	BEA

Table B.1: Continued

NAICS Code	Sector	Variables	Frequency	Starting date	Source
72	Accommodation and food services	Real GDP	Quarterly	2005Q1	BEA
72	Accommodation and food services	Employment level	Monthly	1990M1	BLS
72	Accommodation and food services	Continued UI claims	Monthly	2005M2	DOL
72	Accommodation and food services	Revenue	Quarterly	1992M1	CB
72	Accommodation and food services	Personal income	Quarterly	1998Q1	BEA
81	Other services (except public administration)	Real GDP	Quarterly	2005Q1	BEA
81	Other services (except public administration)	Employment level	Monthly	1939M1	BLS
81	Other services (except public administration)	Continued UI claims	Monthly	2003M12	DOL
81	Other services (except public administration)	Revenue	Quarterly	2009Q1	CB
81	Other services (except public administration)	Personal income	Quarterly	1998Q1	BEA
92	Public administration	Real GDP	Quarterly	2005Q1	BEA
92	Public administration	Employment level	Monthly	1939M1	BLS
92	Public administration	Continued UI claims	Monthly	2003M12	DOL
92	Public administration	Personal income	Quarterly	1998Q1	BEA
Total		Real GDP	Quarterly	1947Q1	BEA
Total		Industrial production	Monthly	1919M1	FRB
Total		Employment level	Monthly	1939M1	BLS
Total		Initial UI claims	Monthly	1967M1	DOL
Total		Manufacturing and trade sales	Monthly	1967M1	CB
Total		Personal income (less transfer payments)	Monthly	1959M1	BEA

Notes: This table describes the details of the available sector-level economic series at the two-digit NAICS level and the aggregate economic series. The data sources are the U.S. Bureau of Economic Analaysis (BEA), the Federal Reserve Board (FRB), the U.S. Bureau of Labor Statistics (BLS), the U.S. Department of Labor (DOL), the U.S. Energy Information Administration (EIA) and the U.S. Census Bureau (CB).

C Robustness to covid observations

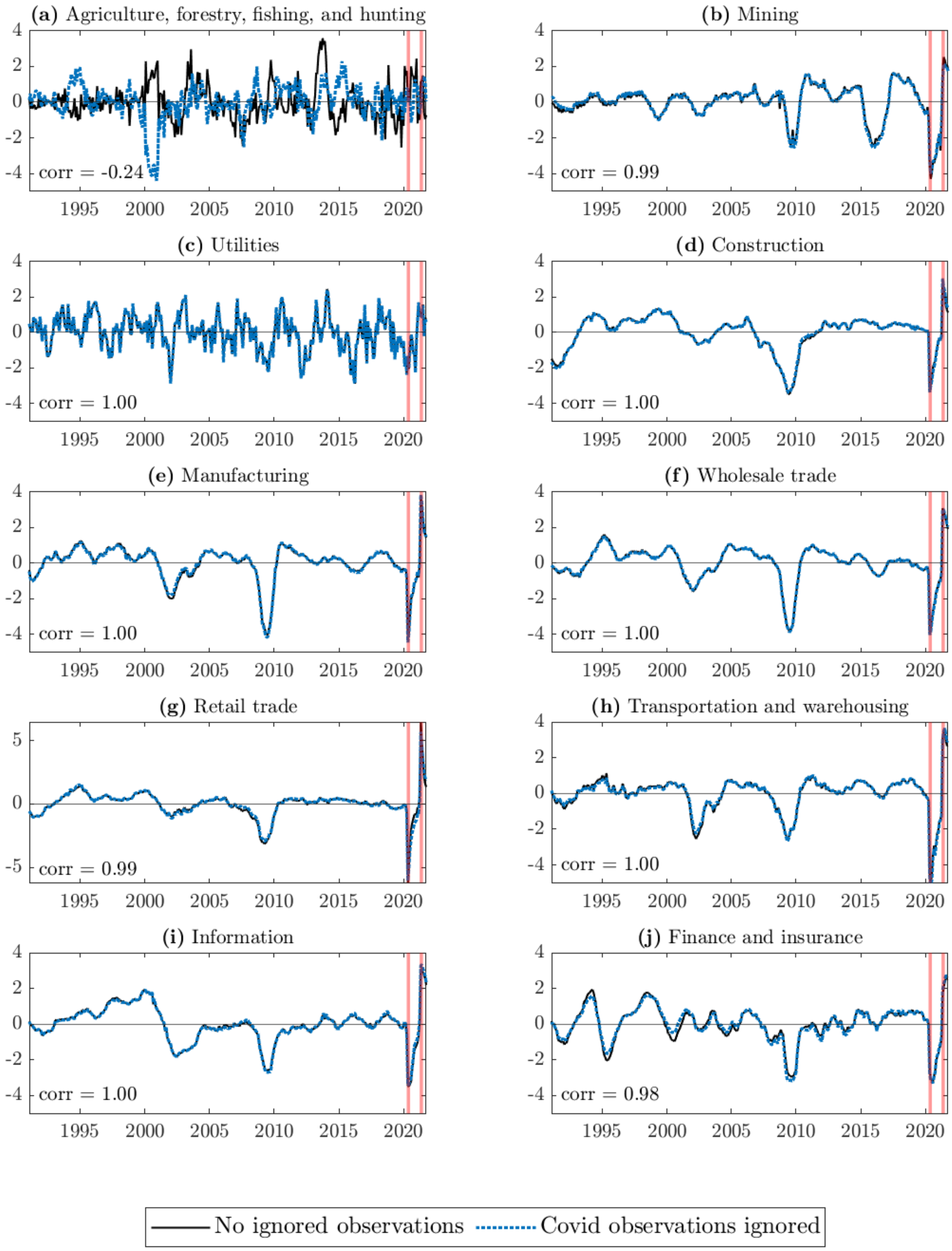

Figure C.1: Sector-level economic condition indices estimated with and without ignoring the covid observations (March 2020 to June 2020) as determined by Maroz et al. (2021) and Schorfheide and Song (2022) and indicated with red shaded areas, including the correlations between the two series.

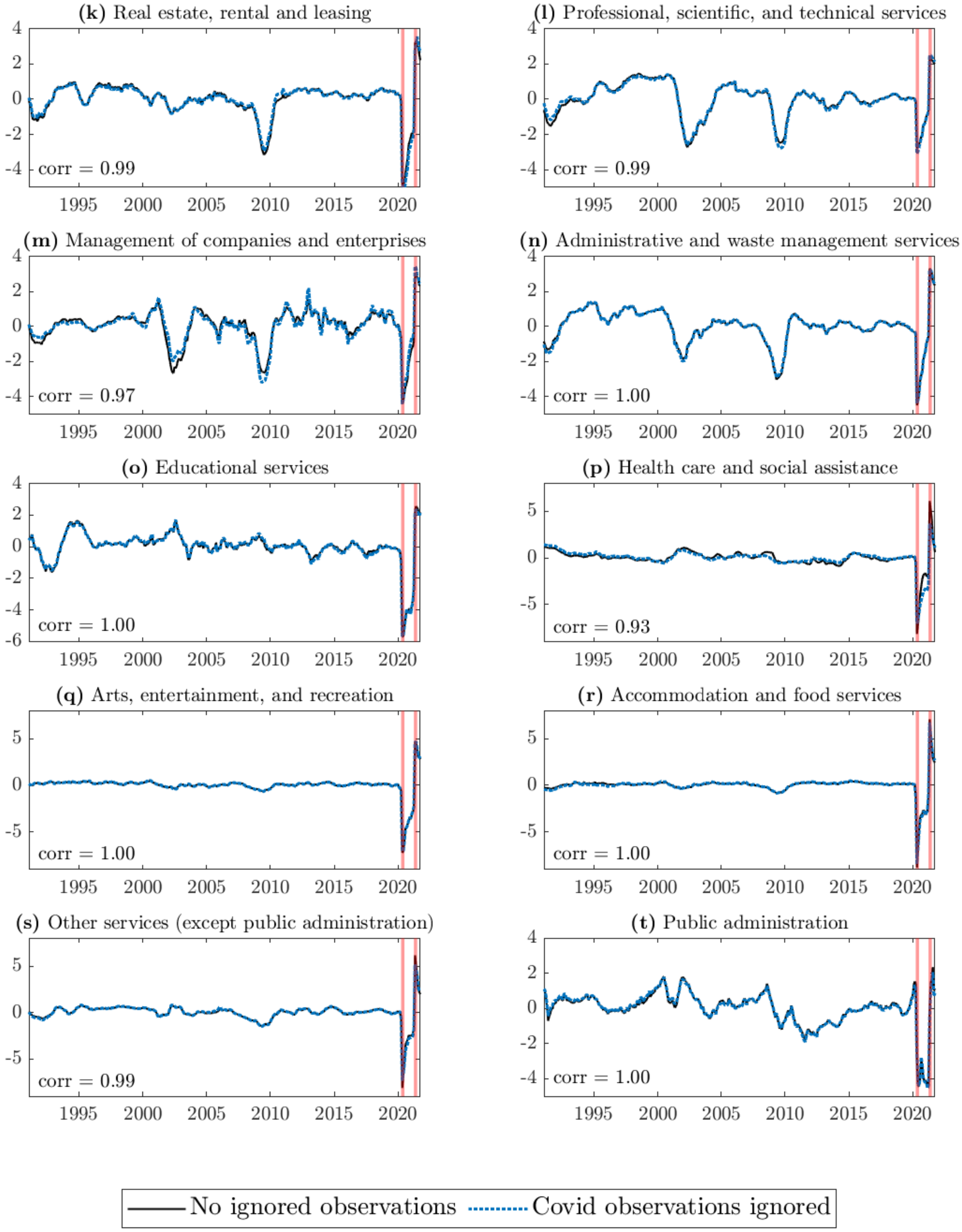

Figure C.1: Sector-level economic condition indices estimated with and without ignoring the covid observations (March 2020 to June 2020) as determined by Maroz et al. (2021) and Schorfheide and Song (2022) and indicated with red shaded areas, including the correlations between the two series (continued).

D Cross-correlations of sectoral economic series

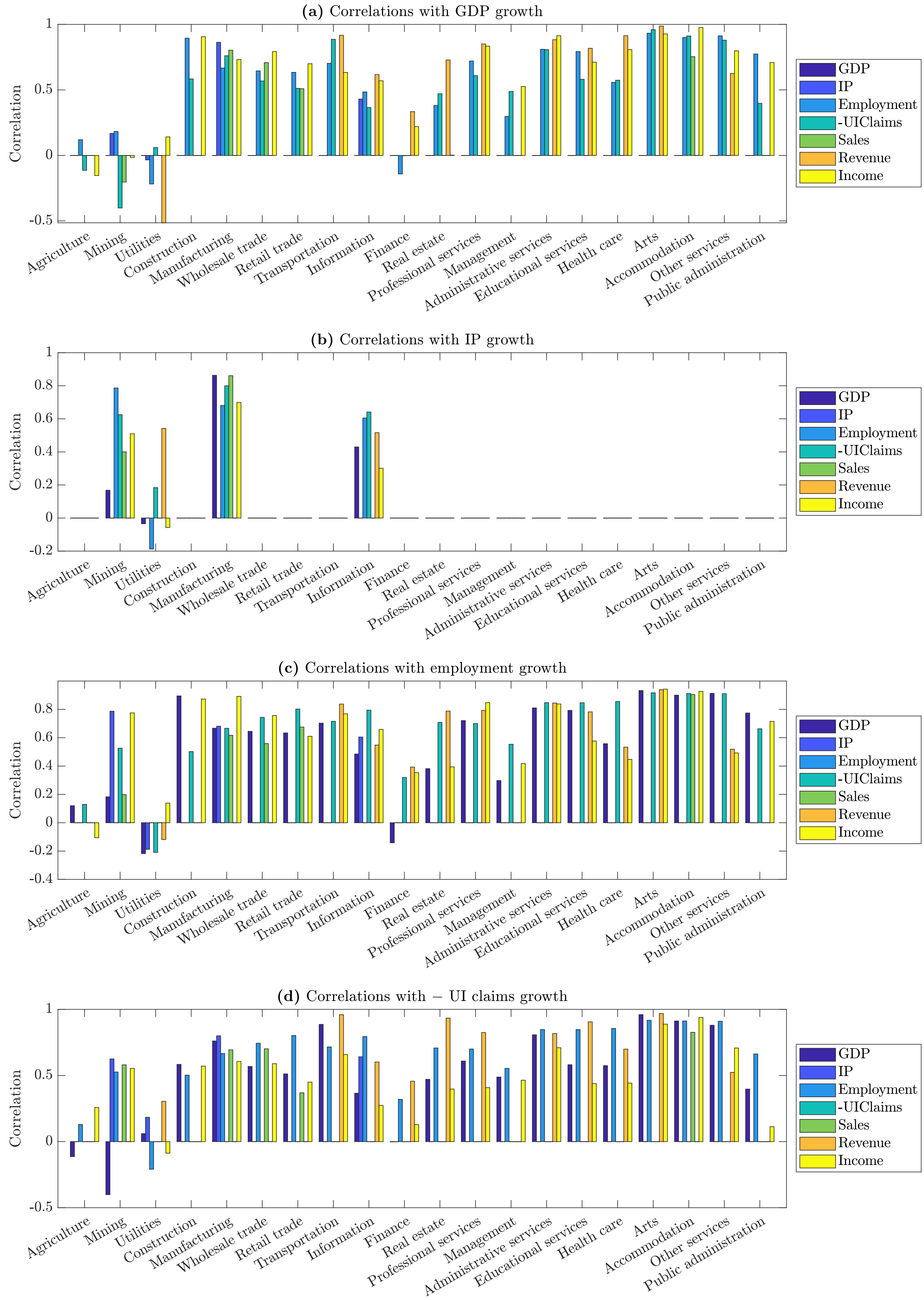

Figure D.1: Cross-correlations of sector-level economic growth series based on the intersection of available observations.

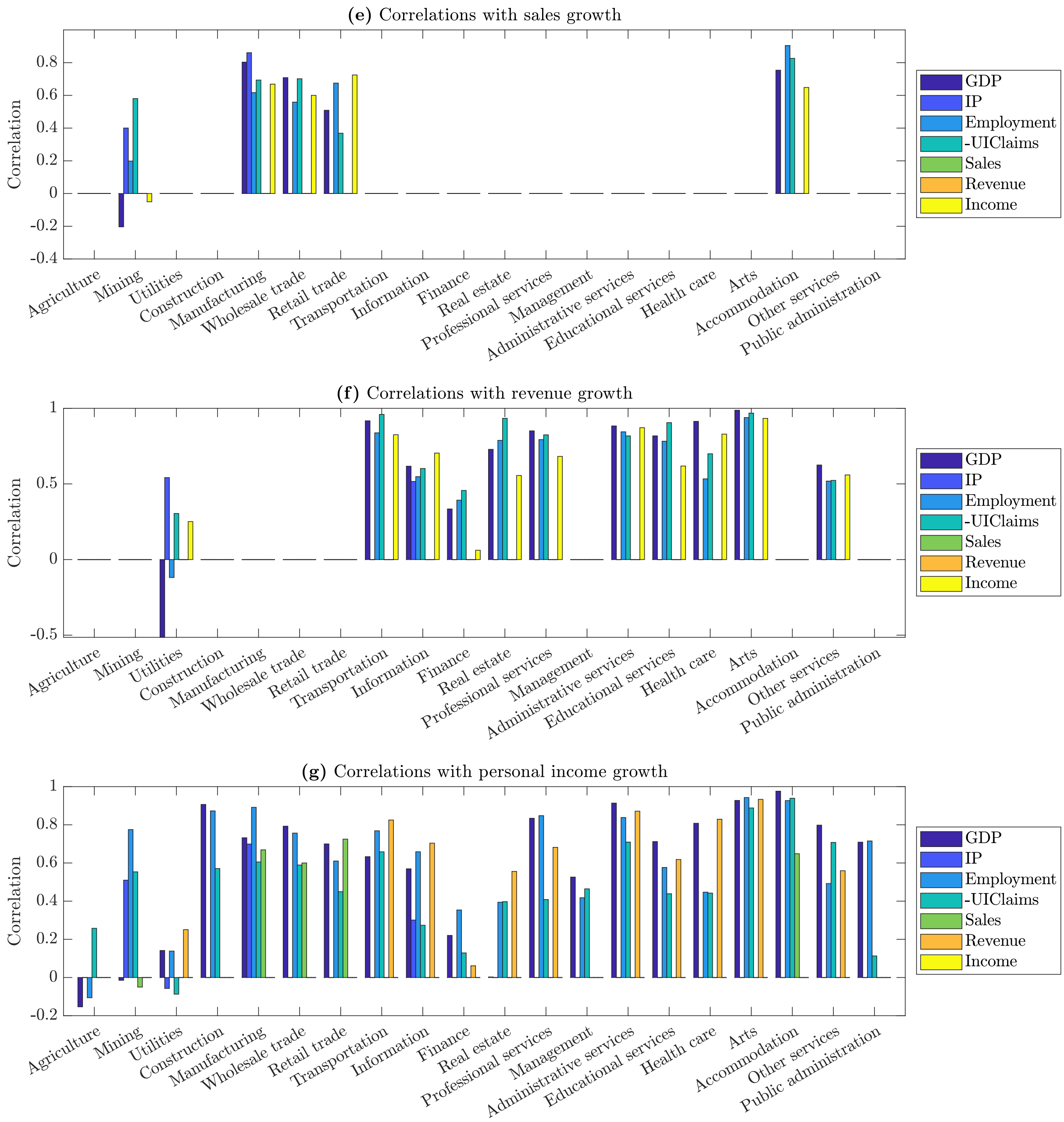

Figure D.1: Cross-correlations of sector-level economic growth series based on the intersection of available observations (continued).

E Robustness analysis of comovement measures

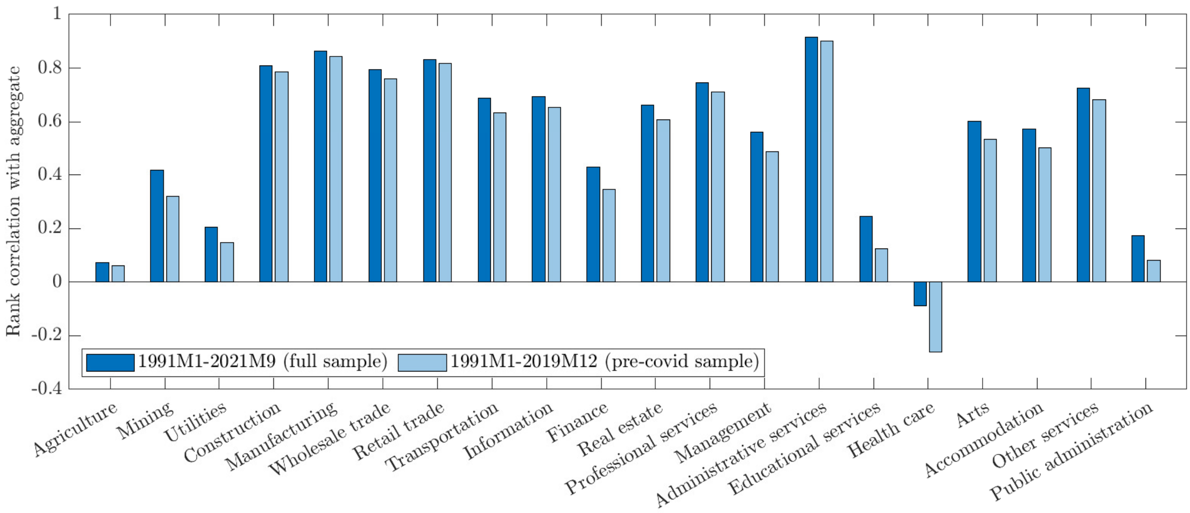

Figure E.1: Spearman's rank correlation between sector-level and aggregate economic condition indices over full-sample (January 1991 - September 2021) and pre-covid-sample (January 1991 - December 2019).

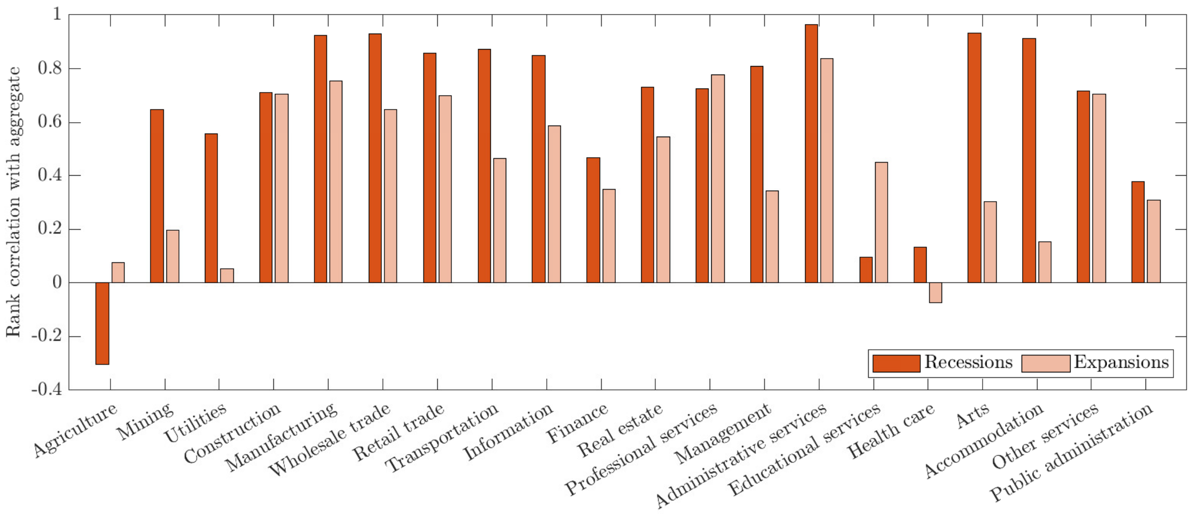

Figure E.2: Spearman's rank correlation between sector-level and aggregate economic condition indices over recession periods (as indicated by the NBER recession dates (± one year)) and expansion periods.

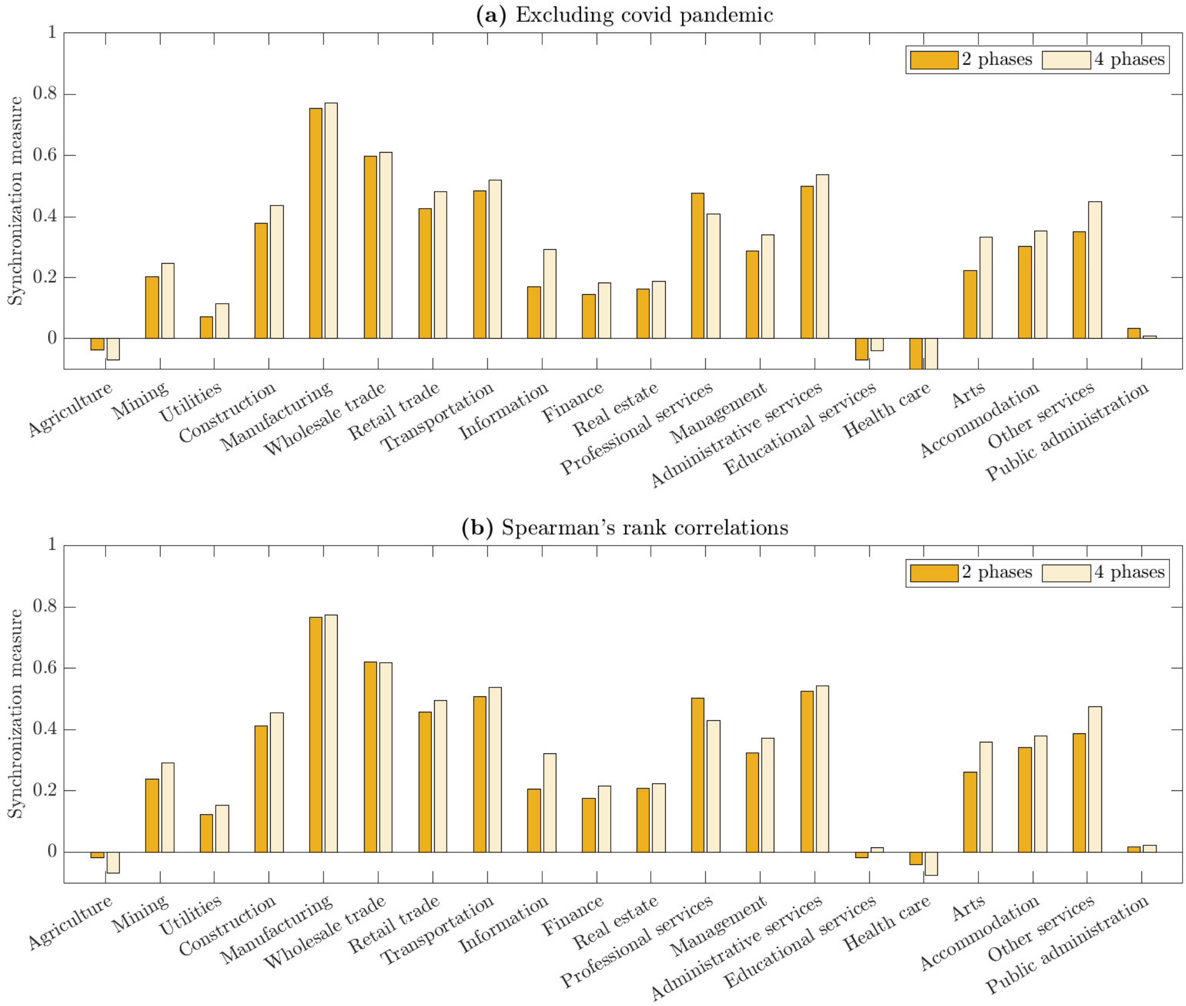

Figure E.3: Correlation-based synchronization measures (pre-covid sample) and rank-correlation-based synchronization measures between sectoral and total economy growth cycles based on the two phases determined by the Bry and Boschan (1971) algorithm and the four phases with additional distinction between positive and negative economic condition indices.

F Drivers of sectoral economic conditions

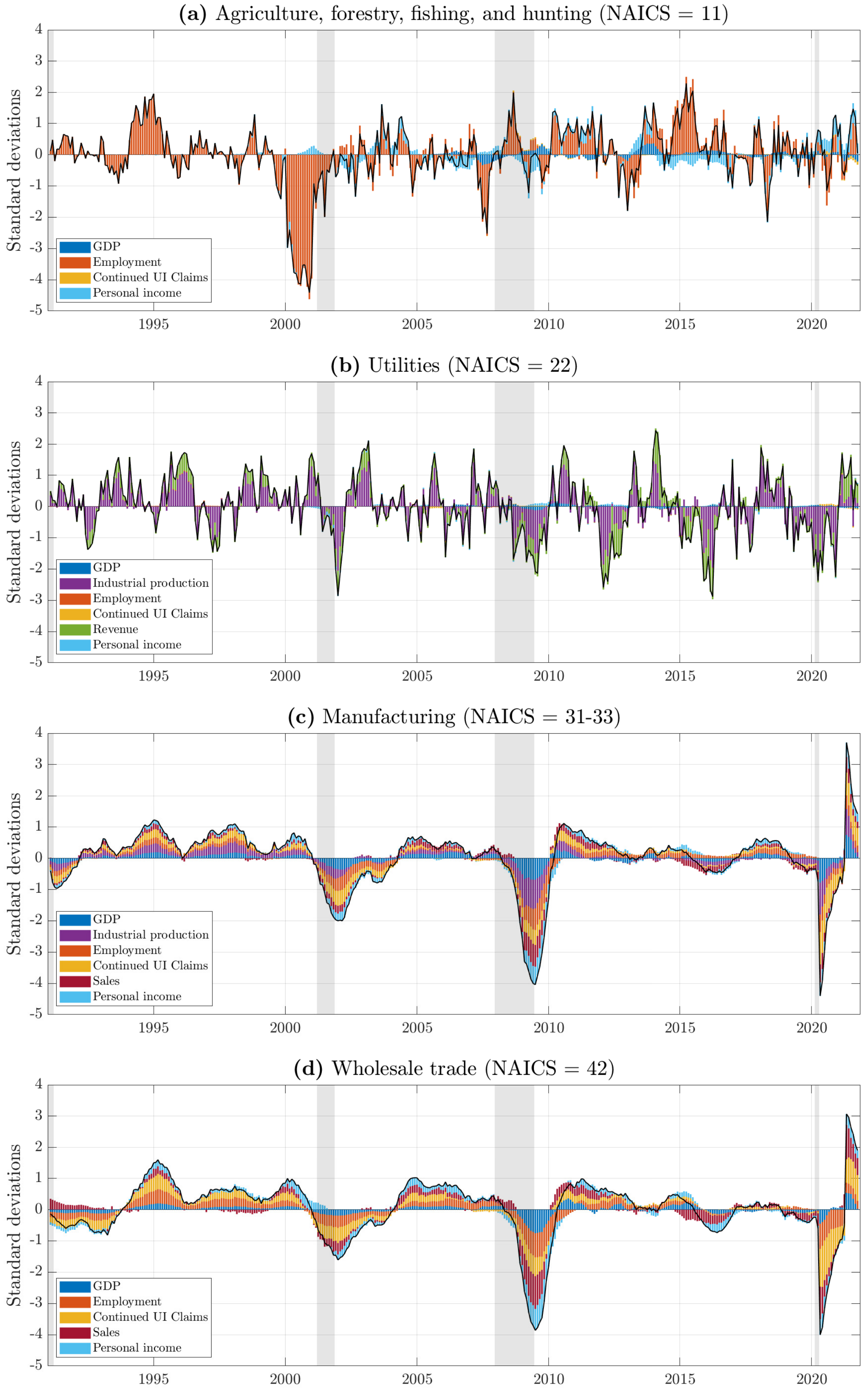

Figure F.1: Sector-level economic conditions indices and their drivers for a selection of sectors with gray shaded NBER recession periods.

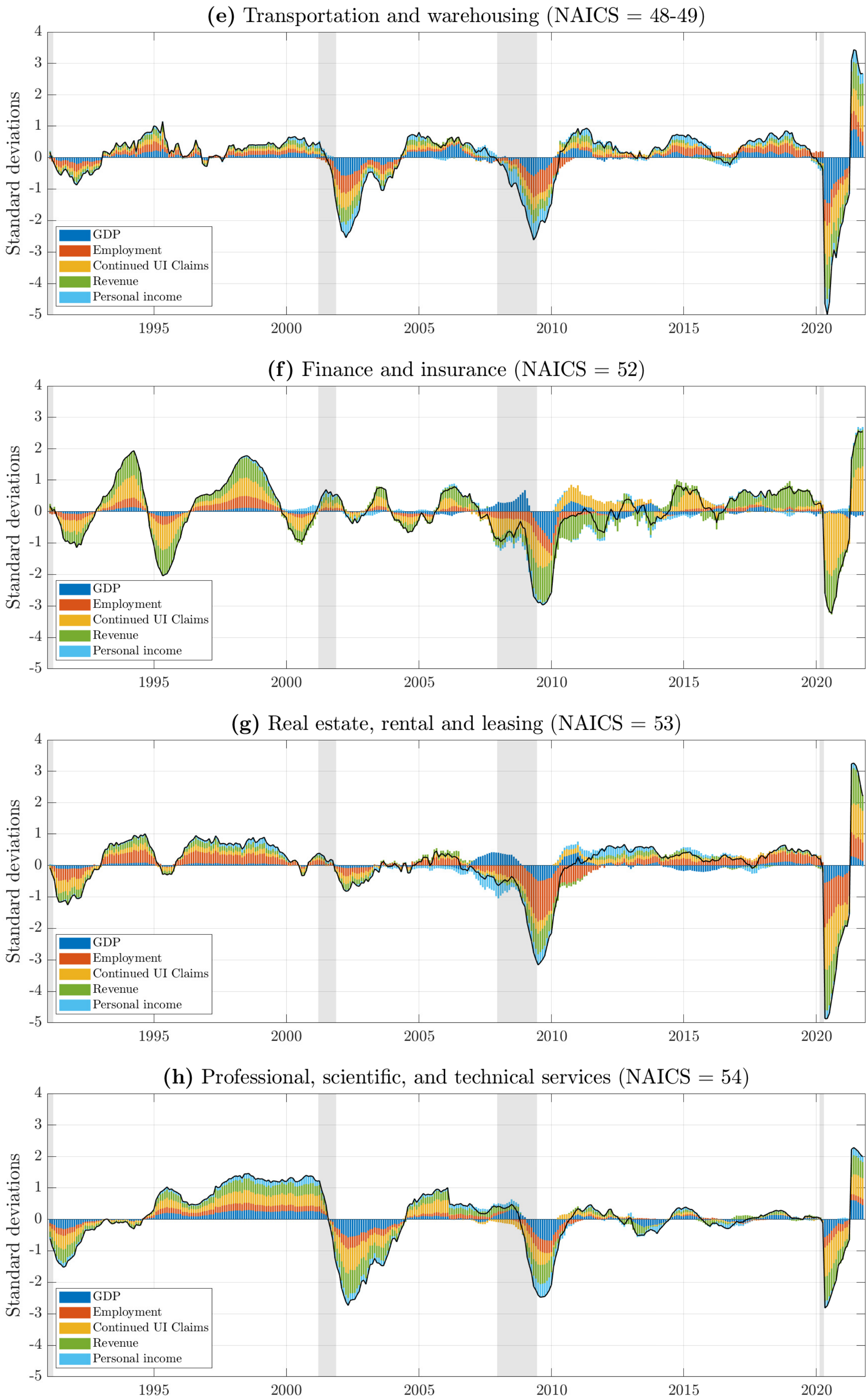

Figure F.1: Sector-level economic conditions indices and their drivers for a selection of sectors with gray shaded NBER recession periods (continued).

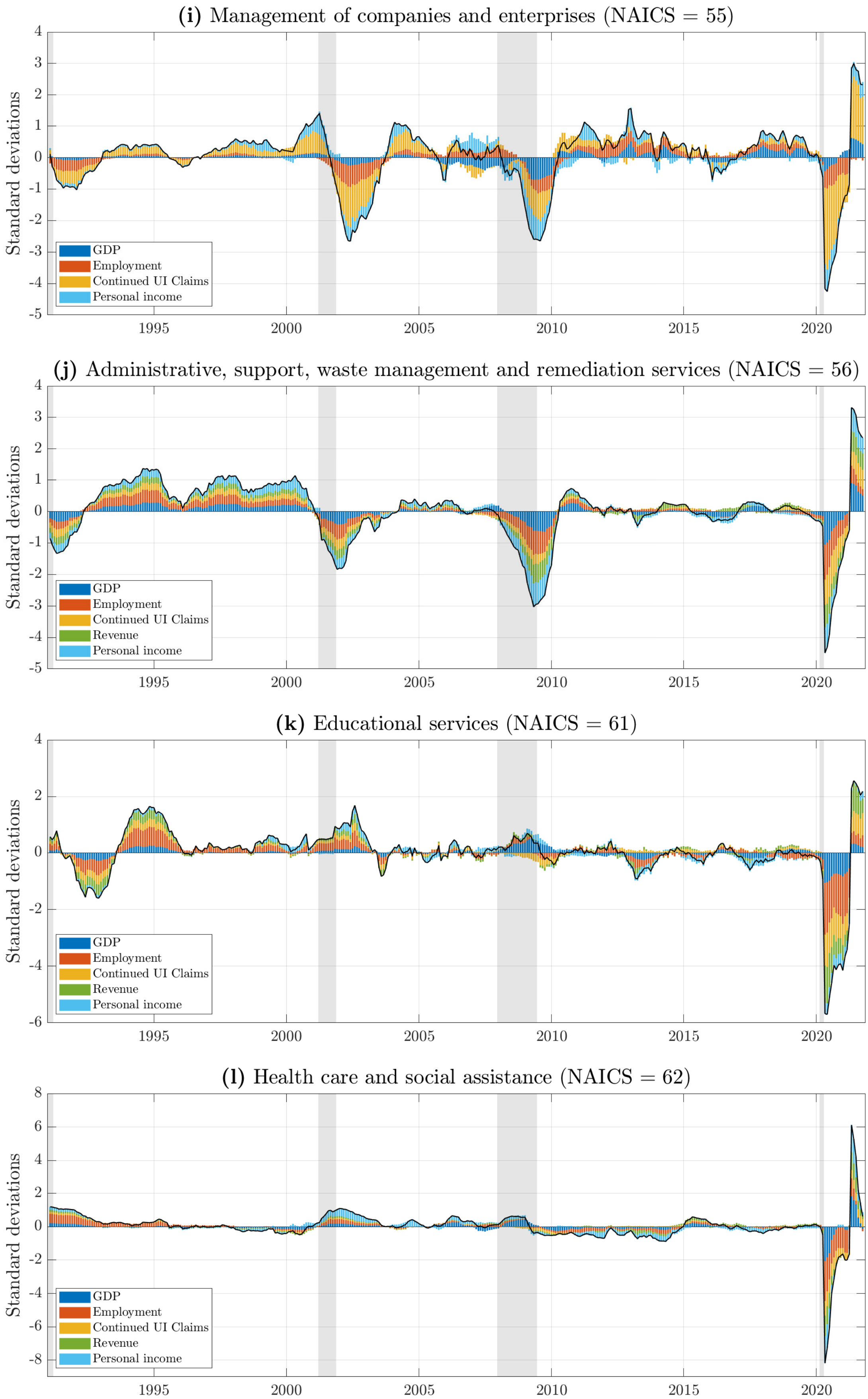

Figure F.1: Sector-level economic conditions indices and their drivers for a selection of sectors with gray shaded NBER recession periods (continued).

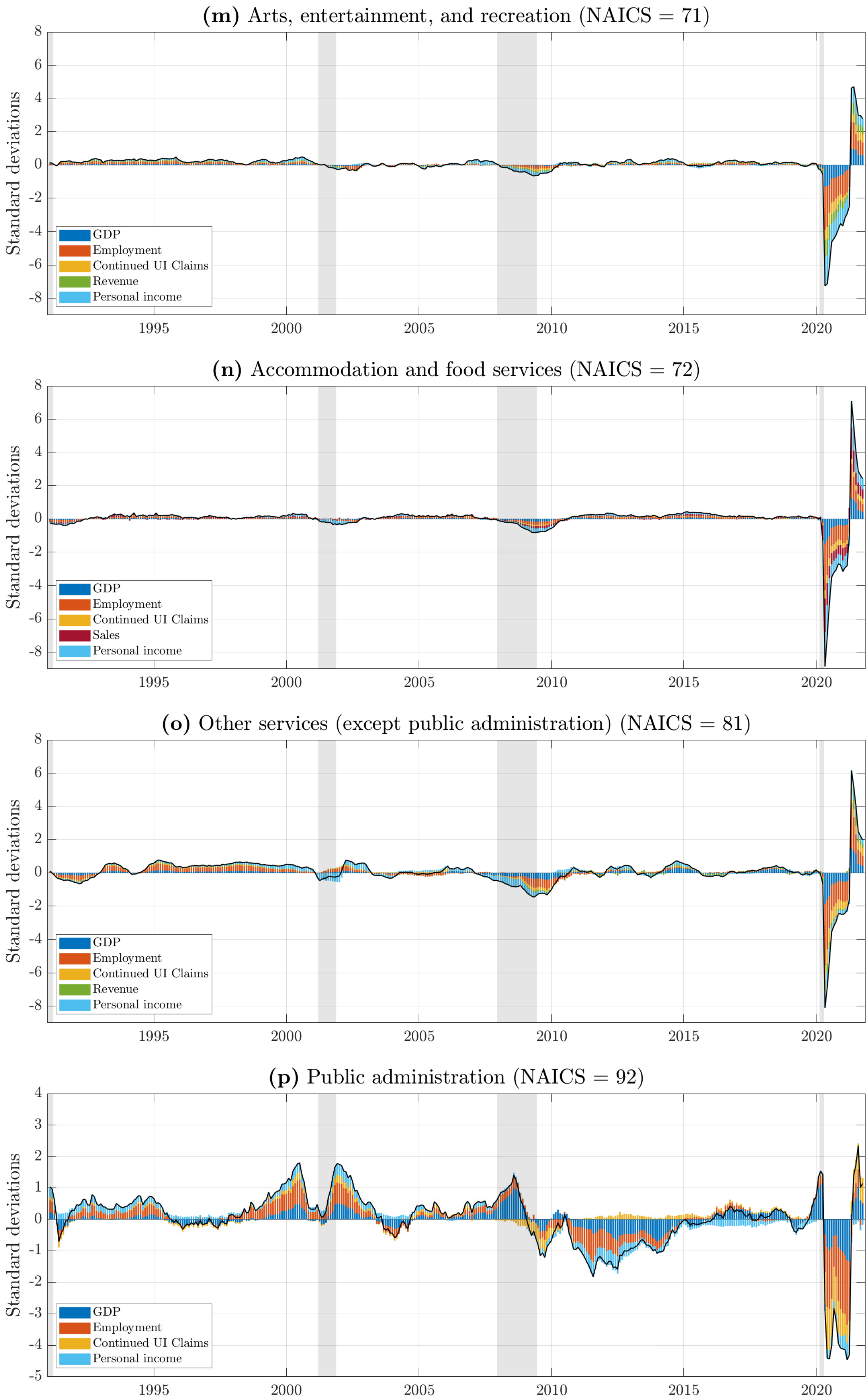

Figure F.1: Sector-level economic conditions indices and their drivers for a selection of sectors with gray shaded NBER recession periods (continued).

G Robustness analysis to number of common factors

Table G.1: Adjusted R^2's of regressing sector-level and aggregate economic conditions indices on <u>three</u> smoothed common factor estimates.

	Full sample (1991M1 - 2021M9)				**Pre-covid sample (1991M1 - 2019M12)**			
	$\bar{R}^2(1)$	$\bar{R}^2(2)$	$\bar{R}^2(3)$	$\bar{R}^2(1{+}2{+}3)$	$\bar{R}^2(1)$	$\bar{R}^2(2)$	$\bar{R}^2(3)$	$\bar{R}^2(1{+}2{+}3)$
Panel A: Total economy								
Aggregate ECI	0.93	0.29	0.08	0.93	0.94	0.00	0.18	0.95
Aggregate IP	0.78	0.44	0.01	0.88	0.83	0.05	0.15	0.86
Panel B: Sectors								
Agriculture, forestry, fishing, and hunting	0.00	0.01	0.01	0.02	0.00	0.03	0.05	0.08
Mining	0.46	0.06	0.00	0.52	0.24	0.07	0.12	0.30
Utilities	0.09	0.05	0.01	0.13	0.08	0.03	0.00	0.12
Construction	0.75	0.40	0.50	1.00	0.73	0.20	0.24	1.00
Manufacturing	0.86	0.53	0.01	1.00	0.96	0.07	0.23	1.00
Wholesale trade	0.83	0.42	0.03	0.89	0.85	0.03	0.20	0.86
Retail trade	0.87	0.31	0.13	0.88	0.85	0.02	0.24	0.89
Transportation and warehousing	0.85	0.19	0.03	0.85	0.73	0.01	0.31	0.76
Information	0.71	0.22	0.10	0.72	0.59	0.01	0.07	0.63
Finance and insurance	0.45	0.07	0.21	0.55	0.26	0.15	0.05	0.43
Real estate, rental and leasing	0.86	0.14	0.16	0.90	0.70	0.08	0.27	0.83
Professional, scientific, and technical services	0.59	0.29	0.06	0.62	0.55	0.00	0.10	0.55
Management of companies and enterprises	0.73	0.20	0.04	0.73	0.58	0.00	0.34	0.64
Administrative and waste management services	0.92	0.33	0.07	0.93	0.90	0.00	0.18	0.90
Educational services	0.32	0.06	0.00	0.68	0.00	0.09	0.04	0.13
Health care and social assistance	0.30	0.08	0.01	0.73	0.07	0.01	0.39	0.39
Arts, entertainment, and recreation	0.64	0.01	0.00	1.00	0.55	0.00	0.11	0.55
Accommodation and food services	0.68	0.00	0.00	0.92	0.71	0.03	0.35	0.81
Other services (except public administration)	0.76	0.01	0.03	0.91	0.54	0.09	0.11	0.66
Public administration	0.18	0.05	0.00	0.45	0.00	0.10	0.69	1.00

Notes: This table shows the adjusted R^2's ($\bar{R}^2$'s) of regressing the aggregate ECI and aggregate IP (Panel A) and sectoral ECIs (Panel B) on the estimates of the first common factor ($\bar{R}^2(1)$), second common factor ($\bar{R}^2(2)$), third common factor ($\bar{R}^2(3)$), or all of them ($\bar{R}^2(1+2+3)$). The results are shown for full-sample (January 1991 - September 2021) based estimated factors and regressions and pre-covid sample (January 1991 - December 2019) ones. The factors are estimated from the 20 sectoral ECIs using the dynamic factor model given in equation (9).

Table G.2: Adjusted R^2's of regressing sector-level and aggregate economic conditions indices on four smoothed common factor estimates.

	Full sample (1991M1 - 2021M9)					Pre-covid sample (1991M1 - 2019M12)				
	$\bar{R}^2(1)$	$\bar{R}^2(2)$	$\bar{R}^2(3)$	$\bar{R}^2(4)$	$\bar{R}^2(1{+}2{+}3{+}4)$	$\bar{R}^2(1)$	$\bar{R}^2(2)$	$\bar{R}^2(3)$	$\bar{R}^2(4)$	$\bar{R}^2(1{+}2{+}3{+}4)$
Panel A: Total economy										
Aggregate ECI	0.84	0.37	0.20	0.00	0.94	0.93	0.48	0.00	0.05	0.95
Aggregate IP	0.60	0.56	0.12	0.02	0.90	0.82	0.25	0.03	0.01	0.86
Panel B: Sectors										
Agriculture, forestry, fishing, and hunting	0.00	0.01	0.00	0.00	0.02	0.00	0.01	0.00	0.08	0.10
Mining	0.45	0.16	0.00	0.12	0.58	0.22	0.06	0.07	0.03	0.35
Utilities	0.06	0.08	0.00	0.00	0.14	0.08	0.01	0.03	0.00	0.10
Construction	0.74	0.33	0.69	0.01	1.00	0.70	0.80	0.06	0.05	1.00
Manufacturing	0.67	0.69	0.14	0.00	1.00	0.96	0.29	0.04	0.03	1.00
Wholesale trade	0.69	0.55	0.15	0.00	0.89	0.85	0.28	0.01	0.03	0.86
Retail trade	0.80	0.35	0.27	0.01	0.90	0.83	0.57	0.00	0.04	0.89
Transportation and warehousing	0.82	0.33	0.11	0.08	0.90	0.74	0.39	0.03	0.06	0.76
Information	0.67	0.29	0.21	0.01	0.72	0.64	0.45	0.00	0.60	1.00
Finance and insurance	0.52	0.07	0.25	0.01	0.56	0.25	0.26	0.13	0.01	0.46
Real estate, rental and leasing	0.91	0.19	0.25	0.02	0.91	0.69	0.59	0.03	0.05	0.83
Professional, scientific, and technical services	0.54	0.39	0.18	0.05	0.68	0.58	0.35	0.01	0.32	0.75
Management of companies and enterprises	0.74	0.36	0.13	0.27	1.00	0.58	0.39	0.04	0.05	0.64
Administrative and waste management services	0.81	0.43	0.20	0.00	0.93	0.90	0.46	0.01	0.07	0.90
Educational services	0.44	0.02	0.00	0.02	0.71	0.00	0.00	0.11	0.02	0.12
Health care and social assistance	0.39	0.02	0.02	0.01	0.80	0.06	0.36	0.58	0.00	1.00
Arts, entertainment, and recreation	0.75	0.01	0.00	0.06	0.99	0.57	0.26	0.00	0.14	0.60
Accommodation and food services	0.78	0.02	0.01	0.04	0.94	0.69	0.52	0.01	0.00	0.81
Other services (except public administration)	0.84	0.04	0.05	0.00	0.94	0.53	0.42	0.06	0.03	0.67
Public administration	0.25	0.03	0.00	0.00	0.48	0.00	0.01	0.28	0.12	0.42

Notes: This table shows the adjusted R^2's ($\bar{R}^2$'s) of regressing the aggregate ECI and aggregate IP (Panel A) and sectoral ECIs (Panel B) on the estimates of the first common factor ($\bar{R}^2(1)$), second common factor ($\bar{R}^2(2)$), third common factor ($\bar{R}^2(3)$), fourth common factor ($\bar{R}^2(4)$), or all of them ($\bar{R}^2(1+2+3+4)$). The results are shown for full-sample (January 1991 - September 2021) based estimated factors and regressions and pre-covid sample (January 1991 - December 2019) ones. The factors are estimated from the 20 sectoral ECIs using the dynamic factor model given in equation (9).

H Robustness analysis of estimated common factors

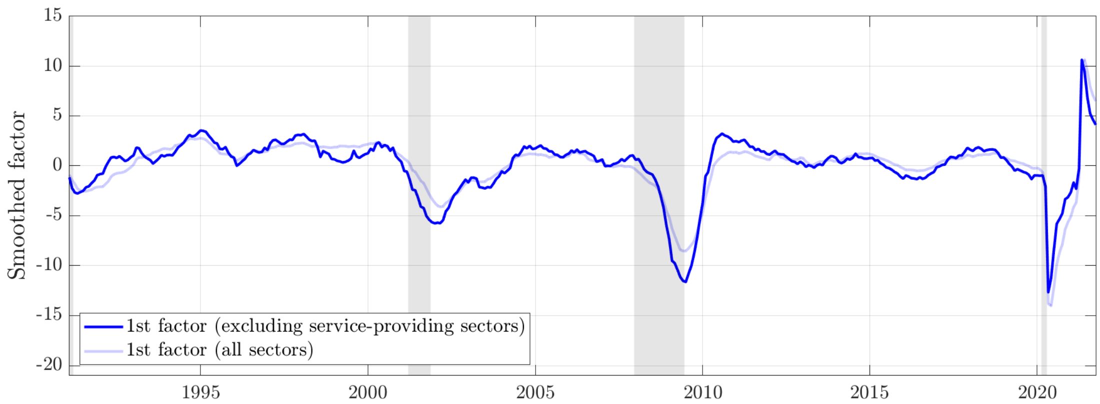

Figure H.1: Smoothed estimate of a single common factor (g_t) estimated by excluding the service-providing sectors with gray shaded NBER recession periods.

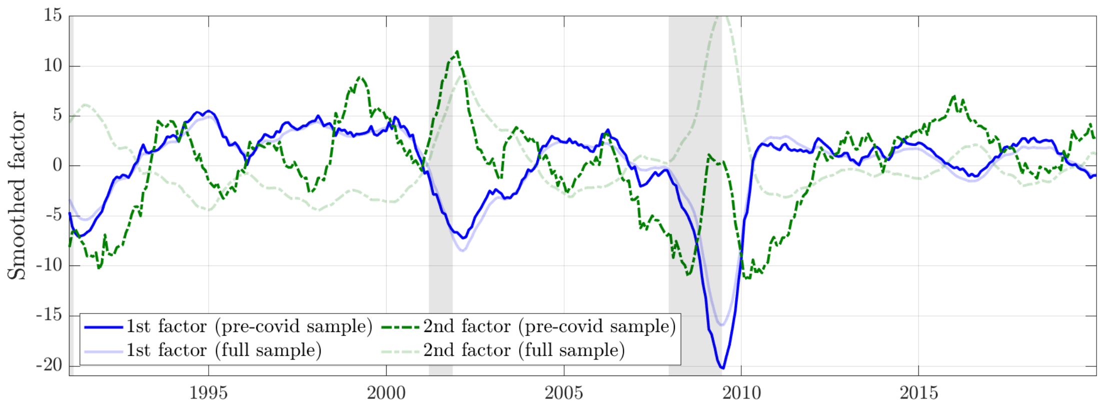

Figure H.2: Smoothed estimates of common factors ($\boldsymbol{g}_t$) estimated up to December 2019 (excluding the covid pandemic) with gray shaded NBER recession periods.

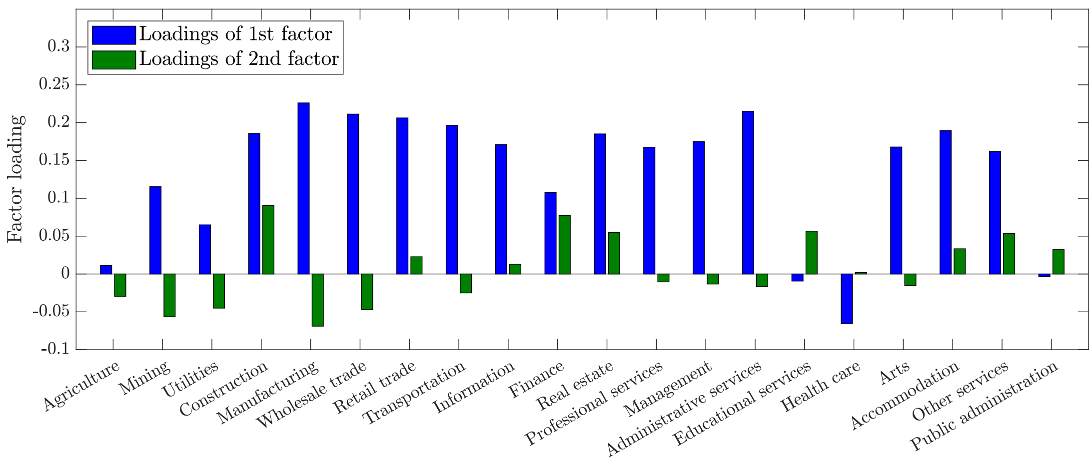

Figure H.3: Factor loading estimates across sectors of common factors ($\boldsymbol{\Gamma}$) estimated up to December 2019 (excluding the covid pandemic).

References

Bańbura, M. and M. Modugno (2014): "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data," *Journal of Applied Econometrics*, 29, 133–160.

Bry, G. and C. Boschan (1971): *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, National Bureau of Economic Research.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1–38.

Doz, C., D. Giannone, and L. Reichlin (2011): "A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering," *Journal of Econometrics*, 164, 188–205.

——— (2012): "A Quasi-Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models," *The Review of Economics and Statistics*, 94, 1014–1024.

Durbin, J. and S. J. Koopman (2012): *Time Series Analysis by State Space Methods*, Oxford: Oxford University Press, 2 ed.

Holmes, E. E. (2013): "Derivation of an EM Algorithm for Constrained and Unconstrained Multivariate Autoregressive State-Space (MARSS) Models," Tech. Rep. arXiv:1302.3919, arXiv.

Maroz, D., J. H. Stock, and M. W. Watson (2021): "Comovement of Economic Activity During the Covid Recession," Tech. rep.

Opschoor, D. and D. J. van Dijk (2023): "Slow Expectation-Maximization Convergence in Low-Noise Dynamic Factor Models," Tinbergen Institute Discussion Paper 2023-018/III, Tinbergen Institute.

Salakhutdinov, R. and S. Roweis (2003): "Adaptive Overrelaxed Bound Optimization Methods," in *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, Washington, DC, USA: AAAI Press, ICML'03, 664–671.

Schorfheide, F. and D. Song (2022): "Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic," *International Journal of Central Banking*, Forthcoming.

Spånberg, E. (2022): "Maximum a Posteriori Estimation of Dynamic Factor Models with Incomplete Data," Tech. Rep. arXiv:2207.01922, arXiv.