Unfolding Regional Business Cycles: Factor Models for Three-Way State-Level Tensors

Sebastian Fossati*

This version: March 6, 2026

PRELIMINARY AND INCOMPLETE
DO NOT DISTRIBUTE

Abstract

This paper develops a new approach to characterize regional business cycle dynamics using high-dimensional state-level data. Regional cycles are estimated using a three-way tensor decomposition subject to orthogonal time components and non-negative factor loadings. These constraints act as inductive biases that resolve scale, sign, and rotation indeterminacy, and produce what is known in the machine learning literature as a "parts-based" representation of the data. A model with four factors captures over 80% of the variation in the data while reducing dimensionality by more than 90%. The estimated regional cycles correspond to distinct economic forces and align closely with well documented regional specializations. These patterns emerge directly from the data without the use of covariates or geographic clusters.

Keywords: regional cycles; tensor decomposition; non-negative factorization; inductive bias
JEL Codes: C38; E32; R12

*Email: sfossati@ualberta.ca. Department of Economics, University of Alberta, Edmonton (AB), Canada. ORCID: https://orcid.org/0009-0007-8081-2197

1 Introduction

Regional economic activity is inherently three-dimensional, with variation across time, geographic units, and many different indicators. For example, state-level economic activity is generally measured using indicators such as employment, income, or the unemployment rate, observed across multiple states, and using time series data. As a result, the unit of observation is the value of indicator k in state s at time t, $X_{t,s,k}$, and the underlying dataset is a three-way state-level tensor. Yet, since traditional factor models can only be estimated in two-dimensional datasets, the existing literature on regional business cycles has so far worked with two of the three dimensions of the tensor, fixing the third dimension at a single geographic unit (e.g., Baumeister et al., 2024) or a specific indicator such as employment growth rates (e.g., Hamilton and Owyang, 2012; Owyang et al., 2008). Alternatively, other studies work with the state-level coincident indexes of economic activity (Federal Reserve Bank of Philadelphia, 2025; Crone and Clayton-Matthews, 2005), effectively compressing the three-way data into a two-way matrix of state-by-time indexes (e.g., Aguiar-Conraria et al., 2017; Leiva-Leon, 2017; Crone, 2005; Owyang et al., 2005).

This paper brings two distinct innovations to the literature on regional business cycles. First, instead of estimating factors from a two-dimensional panel, state-level economic activity is represented as a three-way tensor with modes corresponding to time, states, and economic indicators. To estimate a three-way factor model, I adopt the PARAFAC decomposition ("parallel factors", Harshman, 1970), also known as the CP decomposition ("canonical polyadic", Carroll and Chang, 1970), which has a long tradition in psychometrics and chemometrics (Kolda and Bader, 2009) and has recently gained attention in economics and finance for analyzing multi-dimensional

datasets (e.g., Lettau, 2022). The CP/PARAFAC decomposition extends principal component analysis (PCA) to higher-order tensors by decomposing a three-way tensor $\mathcal{X} \in \mathbb{R}^{T \times S \times K}$ (in our case, time $\times$ states $\times$ indicators) into a sum of R rank-one components:

$$\mathcal{X} \approx \sum_{r=1}^{R} a_r \circ b_r \circ c_r \tag{1}$$

where $\circ$ denotes the outer product, and $a_r \in \mathbb{R}^T$, $b_r \in \mathbb{R}^S$, and $c_r \in \mathbb{R}^K$ represent the r-th time profile, state loading vector, and indicator loading vector, respectively.

Next, I propose a new approach to identify the latent factors as distinct regional cycles that capture comovement across states and indicators. More specifically, I apply the PARAFAC decomposition to the three-way state-level tensor subject to two sets of restrictions that jointly resolve rotation indeterminacy, identify the latent factors, and enhance interpretability. First, I require the latent factors to be orthogonal. This is not new in the factor model literature where orthogonality is often imposed to achieve identification (see e.g., Bai and Ng, 2008; Despois and Doz, 2023). In addition to helping with identification, orthogonality implies that the factors span mutually uncorrelated directions in the time domain, simplifying interpretation of the factors as distinct temporal patterns. Second, I constrain the state and indicator loadings to be non-negative. This restriction makes the decomposition analogous to (semi) non-negative matrix factorization, NMF in the machine learning (ML) literature, where constraining factor loadings to be non-negative forces the model to represent the data as additive combinations of interpretable building blocks, rather than as differences of positive and negative components as it is generally the case in PCA and other unconstrained factor models (Lee and Seung, 1999; Despois and Doz, 2023).

Taken together, these constraints force each factor to capture a distinct temporal

pattern of regional business cycle dynamics (e.g., a manufacturing cycle, a commodity-driven cycle, etc.). State loadings indicate how strongly each state is exposed to each regional cycle, while indicator loadings show how each indicator reflects the underlying latent cycle. The restrictions I impose act as inductive biases, i.e., structured assumptions that restrict the set of admissible representations and thereby select a particular decomposition (a set of factors and loadings) among the many that fit the data similarly well (Mitchell, 1997; Kahou et al., 2024). This type of decomposition yields a "parts-based" representation of the data, as it is known in the ML literature (Lee and Seung, 1999). Furthermore, non-negative decompositions combined with orthogonality constraints have been shown to induce sparsity in the loadings matrix, which generally yields factors that are easier to interpret (Ding et al., 2006).

Applying this approach to three quarterly state-level indicators for the 48 contiguous U.S. states from 1988Q1 to 2019Q4, I find that a parsimonious four-component model captures 81% of the variation in the data while reducing dimensionality by over 90%. This substantial dimensionality reduction suggests that U.S. regional fluctuations are highly structured and strongly comoving. The majority of state-level variation is explained by exposure to a small number of common latent factors rather than by distinct state-specific dynamics. The four regional cycles I identify correspond to distinct economic forces and align closely with well-documented regional specializations and historical episodes. The first factor captures a *credit/housing cycle* concentrated in financial centers in the northeast (New York, Massachusetts) and housing-boom states (Arizona, Florida, Nevada), with pronounced downturns in the early 1990s and 2008-09. The second factor reflects a *western growth cycle* driven by rapid population growth and construction in the 1990s, loading strongly on Mountain and Pacific states (Arizona, Utah, Oregon, Nevada). The third factor represents a *manufacturing cy-*

cle concentrated in the Midwest Rust Belt (Michigan, Ohio, Tennessee), exhibiting a structural decline after 2001 consistent with trade shocks and deindustrialization. The fourth factor identifies an *energy cycle* tied to commodity price fluctuations, loading on energy-producing states (North Dakota, Wyoming) and featuring a sharp downturn during the 2015 oil price collapse. The estimated state loadings exhibit sharp selectivity and some degree of sparsity across these cycles. For example, manufacturing states load heavily on the manufacturing cycle but minimally on the energy cycle, while the opposite holds for energy states. Northeast states load strongly on the credit/housing cycle but very weakly on the western growth cycle. And so on. These regional patterns are *learned* directly from the data without the use of covariates, or imposing *ex ante* geographic clusters or labels.

The existing literature on regional business cycles has adopted two main approaches to measuring and characterizing regional cycles. One approach is to take the unit of analysis to be individual states (or countries, counties, metropolitan areas, etc.). After modeling each unit separately, for example using a factor model, regional business cycles are characterized by noting similarities in the estimated factors (e.g., Carlino and DeFina, 2004; Crone and Clayton-Matthews, 2005; Crone, 2005; Owyang et al., 2005, 2008; Arias et al., 2016; Aguiar-Conraria et al., 2017; Baumeister et al., 2024). This approach, however, does not model regional cycles directly; rather, it infers them indirectly by comparing state-specific cycles estimated separately. In contrast, in this paper I treat regional cycles as primitive latent factors that are common across states and indicators, and propose a factor model to estimate them directly from a three-way state-level tensor of economic activity data.

An alternative approach is to group states into exogenously defined regions and consider each region as a unit of analysis. In this case, regional cycles are characterized by

the common dynamics within each region, often using factor models. Del Negro (2002) cluster states into regions based on those adopted by the BEA (Bureau of Economic Analysis). Kose et al. (2003) group countries into regions based on geographic proximity and other factors. Alternatively, some studies define regions endogenously. In Hamilton and Owyang (2012) states are grouped into a small number of clusters that share common business cycle characteristics. Similarly, Francis et al. (2017), Berger and Wortmann (2022), and Francis et al. (2022) endogenously cluster countries into groups based on similarities in their business cycle dynamics, while Prüser and Schmidt (2021) cluster house price cycles across U.S. states. Under this approach, cluster membership is discrete, with each state assigned to one or more clusters either exogenously or endogenously. In this paper, I replace discrete cluster assignments with graded membership through the state and indicator loadings. A state can load strongly on multiple factors simultaneously (or not at all), with loadings representing intensities of association with each regional cycle.

The rest of the paper proceeds as follows. Section 2 describes the data, estimation methodology, and discusses the constrained three-way decompositions that identify the latent regional cycles. Section 3 presents the empirical results, characterizing the spatial patterns in state loadings and temporal patterns in the estimated regional cycles. Section 4 concludes.

2 Data and Methods

2.1 High-dimensional regional data

Regional economic activity is inherently multi-dimensional, with variation across time, geographic units, and many different economic indicators. For example, our empirical

analysis is based on quarterly time series data, measured at the state-level, using three indicators. The indicators are: (1) total nonfarm payroll employment (employment); (2) real personal income (income); and (3) the unemployment rate (unemployment). These indicators are standard inputs in state coincident indexes, which use DFMs to summarize real economic activity at the state level (e.g., Crone and Clayton-Matthews, 2005; Baumeister et al., 2024). At the state level, we exclude Alaska and Hawaii, in line with much of the regional business cycle literature, and focus on the 48 contiguous states. The sample period is 1988Q1-2019Q4, ending before the COVID-19 pandemic in order to avoid structural breaks in the early 2020s. As a result, we have a balanced dataset of $T = 128$ quarters, $S = 48$ states, and $K = 3$ indicators. The resulting dataset contains $T \times S \times K = 18,432$ quarter-state-indicator observations. The raw quarterly series are obtained from standard U.S. statistical agencies: employment and unemployment from the Bureau of Labor Statistics, and personal income is from the Bureau of Economic Analysis.

The three indicators are transformed into stationary time series by taking quarterly percentage changes for employment and income, and first differences for unemployment. In addition, unemployment is sign-reversed so that positive values for all three indicators correspond to local expansions. Figure 1 displays the (transformed) quarterly time series of employment, income, and unemployment for each state. We observe substantial comovement across states and indicators, reflecting the influence of common business cycle fluctuations on local economic activity. However, we also see significant heterogeneity in the amplitude and timing of the cycles across states, as well as differences in the dynamics of the three indicators within each state.

[Figure 1 about here]

The transformed series are then standardized to have mean zero within each state and unit variance across all states and quarters for each indicator. We discuss preprocessing in more detail in Section 2.3.3. The standardized data can be represented as a three-dimensional $T \times S \times K$ tensor $\mathcal{X} \in \mathbb{R}^{T \times S \times K}$, with elements $X_{t,s,k}$ and modes corresponding to time t, states s, and indicators k.

The three-way tensor $\mathcal{X}$ can be viewed in several ways. Fibers are one-dimensional vectors and correspond to rows and columns of a matrix. Fibers are obtained by fixing all but one index of the tensor. For example, fixing state s and indicator k yields a time fiber $X_{\cdot,s,k}$, which is the time series of indicator k in state s (each time series plotted in Figure 1 corresponds to a fiber). Slices are two-dimensional matrices and are obtained by fixing all but two indices. For example, $X_{\cdot,\cdot,k}$ denotes the slice corresponding to indicator k, which is a $T \times S$ matrix of time series for indicator k across all states (i.e., a state-by-time panel). Similarly, for each state s we have an indicator-by-time slice $X_{\cdot,s,\cdot}$, which is a $T \times K$ matrix of time series for all indicators in state s. Finally, for each quarter t we have a state-by-indicator slice $X_{t,\cdot,\cdot}$, which is a $S \times K$ matrix of cross-sectional data for all states and indicators at quarter t. As a result, the three-way tensor can be viewed as a time series of state-by-indicator matrices, i.e., a matrix-variate time series (Tsay, 2024). Three-way tensors can be written as matrices by unfolding (or stacking) two-dimensional slices along one mode. For example, unfolding along the time mode (mode-1 unfolding) yields a $T \times (S \cdot K)$ matrix where each row corresponds to a quarter and each column corresponds to a state-indicator pair.[1]

Previous research using factor models for regional business cycle analysis mostly works with two-dimensional matrices rather than three-way tensors, fixing the third

[1]See, Kolda and Bader (2009) and Lettau (2022) for a comprehensive review of tensor notation, operations, and graphical representations.

dimension at a selected indicator or geographic unit. For example, Crone and Clayton-Matthews (2005) and Baumeister et al. (2024) construct rich three-way datasets, but then proceed to estimate state-level coincident indexes of economic activity using DFMs applied separately to each state panel (i.e., each factor model is estimated on the slice $X_{\cdot,s,\cdot}$). Other studies on regional business cycles use the outcomes of these DFMs (i.e., the coincident indexes) as measures of state-level economic activity, effectively compressing the three-way data into a two-way matrix of state-by-time indexes (e.g., Crone, 2005; Owyang et al., 2005; Aguiar-Conraria et al., 2017; Leiva-Leon, 2017). Owyang et al. (2008) and Hamilton and Owyang (2012) work with employment growth rates for each state (i.e., the slice $X_{\cdot,\cdot,1}$). Similarly, Owyang et al. (2009) estimate a DFM for regional business cycles using a large panel formed by stacking the employment and income slices ($X_{\cdot,\cdot,1}$ and $X_{\cdot,\cdot,2}$, respectively) into a matrix. As a result, this literature does not exploit the inherent three-way structure of the data: factor models are estimated separately for each state, or using one indicator and discarding the others, or with the indicator dimension flattened into the cross-section.

2.2 High-dimensional decompositions

2.2.1 CP/PARAFAC decomposition

To exploit the three-way structure of the data, in this paper I work directly with the standardized tensor $\mathcal{X}$, and seek a low-rank decomposition that summarizes the main patterns of variation across time, states, and indicators. To this end, I adopt the PARAFAC decomposition ("parallel factors", Harshman, 1970). PARAFAC is a multi-way decomposition for high-dimensional data and is equivalent to the CP decomposition ("canonical polyadic", Carroll and Chang, 1970). See Kolda and Bader

(2009) for a comprehensive review of tensor decompositions and their applications (see also Harshman and Lundy, 1994; Bro, 1997; Lettau, 2022).

The PARAFAC model approximates the standardized tensor $\mathcal{X}$ with R components such that

$$\mathcal{X} = \sum_{r=1}^{R} a_r \circ b_r \circ c_r + \mathcal{E}, \tag{2}$$

where R is an integer, $a_r \in \mathbb{R}^T$, $b_r \in \mathbb{R}^S$, and $c_r \in \mathbb{R}^K$ are the rank-one vectors for component r, the error $\mathcal{E}$ has the same dimensions as the data matrix, and $\circ$ denotes the outer product. Elementwise, equation (2) is written as

$$X_{t,s,k} = \sum_{r=1}^{R} a_r(t)\, b_r(s)\, c_r(k) + \varepsilon_{t,s,k}, \tag{3}$$

for $t = 1, \ldots, T$, $s = 1, \ldots, S$, and $k = 1, \ldots, K$, where $a_r(t)$ is the time profile of component r, $b_r(s)$ is the state loading, $c_r(k)$ is the indicator loading, and $\varepsilon_{t,s,k}$ is the error term. I interpret each component r as a latent regional cycle that drives comovement across states and indicators (more on identification in Section 2.2.4). The time profile $a_r(t)$ traces the evolution of that regional cycle over time, while $b_r(s)$ measures how strongly each state comoves with that cycle, and $c_r(k)$ captures how each indicator reflects the underlying latent cycle.

Working with the three-way tensor has at least two advantages worth emphasizing. First, it allows us to distinguish between heterogeneity across states and heterogeneity across indicators. For example, a regional cycle can load differently on each indicator, rather than forcing all indicators to be combined ex-ante into a single state index or simply being dropped from the analysis. Second, PARAFAC delivers decompositions that are generally unique (up to trivial scaling and permutation of components) under mild conditions (Harshman, 1970; Harshman and Lundy, 1994; Kolda and Bader, 2009), which mitigates the rotation indeterminacy that affects matrix factor models

(such as PCA) and facilitates structural interpretation of the factors as distinct regional cycles. However, the PARAFAC solution may not exist for some datasets, and the algorithm may fail to converge to a stable solution.

2.2.2 PARAFAC2 decomposition

A limitation of the PARAFAC decomposition is that it imposes a common time profile $a_r(t)$ for each component r across all indicators. In our context, indicators likely respond differently in terms of timing and amplitude to common shocks. The indicator dimension may therefore exhibit phase differences relative to a single common time pattern. To accommodate such differences, we also consider a PARAFAC2 decomposition. PARAFAC2 generalizes the CP/PARAFAC decomposition by allowing the factor matrix in one mode to vary across slices, subject to a common cross-product constraint (Kiers et al., 1999; Kolda and Bader, 2009). In our application, the PARAFAC2 model approximates each entry of the tensor as

$$X_{t,s,k} = \sum_{r=1}^{R} a_{r,k}(t)\, b_r(s)\, c_r(k) + \varepsilon_{t,s,k}, \tag{4}$$

so that, relative to PARAFAC, we now allow the time profile $a_{r,k}(t)$ to depend on the indicator k, while the state loadings $b_r(s)$ and indicator loadings $c_r(k)$ remain common across indicators. The sequences $\{a_{r,k}(t)\}_t$ are constrained by the PARAFAC2 coupling condition

$$A_k^\top A_k = \Phi \quad \text{for all } k, \tag{5}$$

where $A_k = \{a_{r,k}(t)\}_{t,r}$ is the $T \times R$ matrix of time profiles for indicator k, and Φ is a positive-definite $R \times R$ matrix that does not depend on k. This condition ensures that the latent components remain comparable across indicators and can also produce unique decomposition under mild conditions (Kolda and Bader, 2009). In this context,

the standard PARAFAC model arises as a special case of PARAFAC2 when $A_k \equiv A$ for all k, i.e. when all indicators share the same time profile for each component.

A notable advantage of PARAFAC2 is that can be applied to a collection of matrices with varying sizes in one mode, e.g., when different indicators have different sample lengths due to data availability (Kolda and Bader, 2009). In our application, however, we have a balanced tensor with the same number of time periods for all indicators, so PARAFAC2 mainly serves to relax the common time profile restriction of PARAFAC.

2.2.3 Preprocessing

Standardization in three-way tensors is not as straightforward as in two-way matrices, where each series is typically demeaned and standardized to have unit variance (see Bro and Smilde, 2003, for a thorough discussion of two- and three-way centering and scaling). In our context, we want to remove indicator-specific means within each state but scale the data across all states *and* quarters at the indicator level. Let $x_{t,s,k}$ denote indicator k in state s at quarter t. First, I remove indicator-specific means for each indicator, so that the analysis focuses on cyclical deviations around each indicator's average growth. Specifically, for each state-indicator pair I compute the within-state mean

$$\bar{x}_{s,k} = \frac{1}{T} \sum_{t=1}^{T} x_{t,s,k} \tag{6}$$

and construct demeaned series $\tilde{x}_{t,s,k} = x_{t,s,k} - \bar{x}_{s,k}$. Next, I scale across time and states at the indicator level by dividing by the cross-section-and-time standard deviation of the demeaned series for each indicator k:

$$X_{t,s,k} \equiv \frac{\tilde{x}_{t,s,k}}{\sigma_k}, \tag{7}$$

with

$$\sigma_k = \left(\frac{1}{TS} \sum_{t=1}^{T} \sum_{s=1}^{S} \tilde{x}_{t,s,k}^2 \right)^{1/2}. \tag{8}$$

The resulting standardized tensor $X_{t,s,k}$ thus has mean zero within each state for each indicator and unit variance across all states and quarters for each indicator. Note that this pre-processing preserves within indicator variation across states (important for capturing regional heterogeneity), while it ensures that no single indicator dominates purely because of scale differences, making the decomposition comparable to standard PCA and factor-model analyses which typically operate on demeaned and standardized data if the indicators have different units or volatilities (see, e.g., Stock and Watson, 1989, 2002; Greenacre et al., 2022).

2.2.4 Identification

Factor models are subject to various forms of indeterminacy, namely scale, column sign changes, column permutations, and rotation. For example, any orthogonal rotation of the factor space yields an observationally equivalent representation of the data. Although PARAFAC/PARAFAC2 decompositions are unique under mild conditions (Kolda and Bader, 2009), some degree of rotational indeterminacy may still be present in practice. Other forms of indeterminacy, such as scale and sign changes, require additional constraints to be resolved. Finally, permutation indeterminacy is trivial to resolve by ordering the factors according to some criterion, e.g., the amount of variance explained by each factor. To address these issues, I impose additional constraints on the factor matrices in order to resolve scale, sign, and rotation indeterminacies, and to enhance interpretability of the latent regional cycles.

First, I use the normalization $A^\top A/T = I_R$, where I_R is the identity matrix. This normalization, very common in two-dimensional factor models, sets the variance of

the latent factors to one and, as a result, resolves *scale* indeterminacy (e.g., Stock and Watson, 2002; Owyang et al., 2009; Despois and Doz, 2023). The normalization also imposes orthogonality in the latent factors, which implies that the time profiles $a_r(t)$ span mutually uncorrelated directions in the time domain, sharply reducing the rotation indeterminacy, and simplifying the interpretation of the latent cycles due to their distinct temporal patterns.

A novel feature of our approach is the imposition of non-negativity constraints on the state and indicator loadings. Specifically, I constrain the elements of B and C to be non-negative, such that $b_r(s) \geq 0$ and $c_r(k) \geq 0$ for all s, k, r. Given that all indicators have been transformed so that higher values correspond to better local economic conditions, non-negative loadings imply that an increase in the time profile $a_r(t)$ unambiguously reflects an improvement in the latent regional cycle for every state and indicator that loads on it. These constraints fix the sign of the state and indicator loadings and, as a result, resolve *sign* indeterminacy. In addition, this representation makes the decomposition analogous to (semi) NMF in the ML literature, where constraining factor loadings to be non-negative forces the model to represent the data as additive combinations of interpretable building blocks rather than as differences of positive and negative components as it is generally the case in PCA, yielding a "parts-based" representation of the data (Lee and Seung, 1999; Ding et al., 2008). Furthermore, NMF combined with orthogonality constraints has been shown to induce sparsity in the loadings matrix, which can yield factors that are easier to interpret (Ding et al., 2006).

[FORMALIZE RESULTS]

2.2.5 Estimation

The PARAFAC model is estimated by minimizing a least-squares objective subject to the constraints described above. Specifically, we solve the optimization problem

$$\min_{A,B,C} \sum_{t,s,k} \left(X_{t,s,k} - \sum_{r=1}^{R} a_r(t)\, b_r(s)\, c_r(k) \right)^2, \tag{9}$$
$$s.t. \quad A^\top A/T = I_R, \quad B \geq 0, \quad C \geq 0,$$

where $B \geq 0$ and $C \geq 0$ denote elementwise non-negativity constraints on the state and indicator loadings, respectively. The PARAFAC2 model is estimated by minimizing a similar least-squares objective, but allowing for indicator-specific time profiles $a_{r,k}(t)$ and incorporating the coupling constraint $A_k^\top A_k = \Phi$ for all k. We implement these minimization problems using an alternating least squares (ALS) algorithm adapted to constrained tensor factorizations (Kiers et al., 1999; Helwig, 2017). ALS cycles through updating A, B, and C in turn, solving at each step a constrained least-squares problem conditional on the other factors. Because the objective is non-convex, we run the algorithm from multiple random starting values and retain the solution with the lowest objective function value.

2.2.6 Normalization and interpretation

Because the PARAFAC/PARAFAC2 models are only defined up to scaling of each triplet (a_r, b_r, c_r) by constants whose product is one, we adopt the following normalization for figures and tables. State and indicator loadings are rescaled so that $\max_s b_r(s) = 1$ and $\max_k c_r(k) = 1$ for each component r. This normalization fixes the scale of the components and allows $b_r(s)$ and $c_r(k)$ to be interpreted directly as relative

amplitudes (e.g., a state with $b_r(s) = 0.5$ reacts with half the amplitude of the most exposed state to changes in component r). Each time component $a_r(t)$ or $a_{r,k}(t)$ is rescaled appropriately so that the reconstructed tensor $\hat{X}_{t,s,k}$ remains the same. With this normalization, the time components $a_r(t)$ can be interpreted as latent regional business cycle factors whose fluctuations are modulated across states and indicators by the non-negative loadings $b_r(s)$ and $c_r(k)$.

3 Empirical Results

3.1 Three-way decompositions

Both the PARAFAC and PARAFAC2 models are estimated on the standardized tensor $\mathcal{X}$. For each candidate rank $R = 1, \ldots, 8$ and for each set of constraints (unconstrained, orthogonal time profiles, orthogonal time profiles plus non-negativite loadings), we run the constrained ALS algorithm from 100 random starting values, for both PARAFAC and PARAFAC2. For the preferred specification selected on the basis of these diagnostic runs, we re-estimate the model with 500 random starts and up to 500 iterations to reduce the risk of convergence to a poor local minimum. All analyses were performed in `R` (R Core Team, 2025, version 4.5.2) using the package `multiway` (Helwig, 2025, version 1.0-7).[2]

3.1.1 Model and rank selection

Model comparison is based on the explained variation (R^2). For each candidate model we compute the fraction of variation in the standardized tensor explained by the rank-R

[2]Other `R` packages used include: `tidyverse`, `cowplot`, `usmap`, `RColorBrewer`, `scales`, `tinytable`, `reshape2`, `fixest`, and `pacman`.

approximation,

$$R^2 = 1 - \frac{\sum_{t,s,k} \hat{\varepsilon}^2_{t,s,k}}{\sum_{t,s,k} X^2_{t,s,k}}, \tag{10}$$

where $\hat{\varepsilon}_{t,s,k} = X_{t,s,k} - \hat{X}_{t,s,k}$, $\hat{X}_{t,s,k}$ denotes the fitted value from the tensor decomposition. This measure is analogous to the fraction of variance explained by the first R components in PCA and is standard in the tensor-decomposition literature. We summarize the evolution of R^2 as functions of R separately for each model (PARAFAC vs PARAFAC2) and constraint set, producing a tensor scree plot analogous to the eigenvalue scree plots used in PCA literature.

Figure 2 displays the tensor scree plot for the PARAFAC and PARAFAC2 models under the three sets of constraints. For both models, we observe that imposing orthogonality on the time profiles or orthogonality plus non-negativity on the state and indicator loadings has negligible effects on the fit of the models. This suggests that the constraints do not force the model to deviate from the data in a material way, and that the underlying structure of the data is well captured by orthogonal time profiles and non-negative loadings. Comparing the two models, we see that PARAFAC2 consistently achieves a higher R^2 than PARAFAC for each rank and constraint set, indicating that allowing for indicator-specific time profiles yields notizable improvements in fit relative to the more restrictive PARAFAC model with common time profiles.

[Figure 2 about here]

These models achieve substantial dimensionality reduction. The original dataset consists of $T \times S \times K = 18,432$ observations, whereas the four-component PARAFAC decomposition (with common time profiles across indicators) summarizes the data in terms of 1 time profile, 48 non-negative state loadings, and 3 indicator weights for each component. Up to normalizations, this is of order $R(T + S + K)$, which in our case is

716 free parameters. As a result, this model captures 75% of the variation in the data and reduces the dimensionality of the data by *more* than 96%. Similarly, the four-component PARAFAC2 decomposition summarizes the data in terms of 3 indicator-specific time profiles, 48 non-negative state loadings, and 3 indicator weights for each component. Up to normalizations and ignoring the common cross-product constraint, this is of order $R(T \times K + S + K)$, which in our case is $1,740$ free parameters. As a result, this model captures 81% of the variation in the data and reduces the dimensionality of the data by *more* than 90%.

I choose the four-component PARAFAC2 model as the preferred specification for the rest of the paper as it offers both a good fit (R^2 over 80%) and a high degree of parsimony (over 90% dimension reduction). Regardless of the specific model chosen, these results suggest that U.S. regional fluctuations are highly structured and strongly co-moving. A substantial share of the variation in the state-level cycles can be explained by just four common latent factors, and differences across states are largely explained by how strongly each state loads on those forces, not by distinct state-specific dynamics. State level cyclical fluctuations are, to a large extent, driven by the same underlying shocks and propagation channels.

3.2 Regional business cycle factors

We begin by describing the spatial patterns captured by the state loadings ($b_r(s)$) for each of the four factors of the preferred specification, followed by a discussion of the temporal patterns captured by the time profiles ($a_{r,k}(t)$ for PARAFAC2), and the relative exposure of each indicator to the underlying regional cycles as captured by the indicator loadings ($c_r(k)$).

3.2.1 Spatial patterns of state loadings

Figure 3 shows the estimated state loadings $b_r(s)$ for each of the four factors of the PARAFAC2 specification. State loadings capture how strongly each state comoves with each latent regional cycle. Panel A (left) displays the state loadings for factors 1 and 2, while Panel B (right) displays the state loadings for factors 3 and 4. To facilitate interpretation of the spatial patterns captured by the state loadings, we group states into the four Census regions: Northeast, Midwest, South, and West (U.S. Census Bureau, 2025). The geographic distribution of states within each region is shown in Figure A.1 of the appendix.

[Figure 3 about here]

The first factor loads more strongly on finance-heavy states in the Northeast, such as New York, Connecticut, Massachusetts, and New Hampshire, as well as the "sand states" in the South and West regions, including Arizona, California, Florida, and Nevada (Abel and Deitz, 2011). However, we observe substantial heterogeneity within regions with many states exhibiting weak exposure to this factor. In particular, resource oriented states, including North Dakota, South Dakota, and Louisiana, exhibit no exposure to this factor. This cross-sectional pattern is consistent with a regional cycle tied to financial conditions and real-estate/housing-driven demand, which historically has had outsized effects in finance-intensive coastal metro economies and in states that experienced particularly strong housing cycles (Lamm and O'Keefe, 1997; Saulsbury and Curry, 1997).

The second factor loads more strongly on states in the South and West regions, particularly in the Mountain and Pacific divisions, including Arizona, Idaho, Colorado, Oregon, Utah, Washington, and Nevada. In contrast to the first factor, the second

factor exhibits strong clustering along Census regions, with the Northeast exhibiting very weak exposure, the Midwest and South exhibiting moderate exposure, and the West exhibiting the strongest exposure. Western states experienced rapid population growth and inmigration in the 1990s, which suggests a regional cycle tied to local demand growth, construction, and small-business/service expansion (Perry and Mackun, 2001; Perry, 2002; Abel and Deitz, 2011; Haughwout et al., 2013).

The third factor loads more strongly on states in the Midwest and South regions, such as Michigan, Ohio, Indiana, and Tennessee. These states have a high concentration of manufacturing and automotive industries and, as a result, this factor likely captures the influence of the manufacturing sector on regional economic activity (Lavelle and Lombardi, 2010; Abel and Deitz, 2011). Finally, the fourth factor loads more strongly on energy-producing states, mainly Wyoming, North Dakota, Oklahoma, Louisiana, Texas, and West Virginia. As a result, this factor likely captures the influence of the energy sector on regional economic activity (Brown, 2017).

Several states stand out for their extreme loadings on the factors, illustrating the economic forces captured by each regional cycle. For example, Nevada is maximally exposed to factors 1 and 2, reflecting its heavy reliance on the tourism and hospitality sectors, but exhibits no exposure to factor 4. In contrast, North Dakota exhibits notably low loadings on factors 1 and 2 (essentially zero) but is maximally exposed to factor 4. Wyoming, another energy-oriented economy, is also maximally exposed to factor 4, exhibits moderate exposure to factor 2, but has very low exposure to factors 1 and 3. Michigan is maximally exposed to factor 3, reflecting its heavy reliance on the automotive and manufacturing sectors, has moderate exposure to factors 1 and 2, and no exposure to factor 4. Based on these spatial patterns, we proceed to name the four factors as: (1) a *credit/housing cycle*; (2) a *western growth cycle*; (3) a *manufacturing*

cycle; and (4) an *energy cycle*. Figure 4 shows the geographic distribution of state loadings for each of the four factors.

[Figure 4 about here]

3.2.2 Temporal patterns of time profiles

Next, we proceed to describe the temporal patterns captured by the time profiles of each of the four factors. Figure 5 shows the estimated time profiles $a_{r,k}(t)$ for each indicator and each of the four factors. Panel A (top-left) displays the time profile for factor 1 (the credit/housing cycle) for each of the three indicators: employment, income, and unemployment. Factor 1 exhibits a clear business cycle pattern, with pronounced downturns around 1990-91 and 2008-09, aligning closely with major episodes in which credit tightening and real-estate balance-sheet stress were central propagation mechanisms. For example, the early-1990s contraction is commonly linked to regional banking and commercial real estate stress, especially in the Northeast and parts of California (Lamm and O'Keefe, 1997; Saulsbury and Curry, 1997). Similarly, the 2008-09 collapse is usually interpreted through the lens of housing prices, leverage, and credit constraints. However, we note that the 2008-09 downturn was not as pronounced in this factor, consistent with evidence that the Great Recession did not impact states in the Northeast as severely as other regions due to the region's relatively low incidence of sub-prime mortgages (Abel and Deitz, 2010). In addition, the strong exposure of Arizona, Nevada, and Florida is also consistent with evidence that "boom-bust" housing markets were concentrated in a subset of states and metros that experienced large swings in construction and employment during the Great Recession (Abel and Deitz, 2011).

[Figure 5 about here]

Panel B (top-right) displays the time profile for factor 2 (the western growth cycle) for each of the three indicators. Factor 2 exhibits a pronounced peak around 1990, followed by a gradual decline through the early 2000s, and then a sharp downturn during the Great Recession of 2008-09. A "boom" around 1990 can be understood as a period when these states were less exposed to the early-1990s credit-real-estate stress, while the large decline in 2008 is consistent with the housing/construction collapse hitting supply-elastic, fast-growing markets particularly hard (Abel and Deitz, 2011; Haughwout et al., 2013).

Panel C (bottom-left) displays the time profile for factor 3 (the manufacturing cycle) for each of the three indicators. This factor exhibits pronounced downturns in 1990-91, 2001, and 2008-09, and a notably weak trajectory after 2001. Manufacturing and durables are highly recession-sensitive, and the early-2000s period is associated with a large and persistent decline in manufacturing employment (Congressional Budget Office, 2008; Harris, 2020; Mattoon, 2016). The "jobless recovery" after 2001 is also consistent with sectoral and structural headwinds facing industrial states (Lavelle and Lombardi, 2010) and in longer-run mechanisms such as import competition that generated persistent local labor-market effects in manufacturing-intensive regions (Autor et al., 2013).

Finally, Panel D (bottom-right) displays the time profile for factor 4 (the energy cycle) for each of the three indicators. This factor exhibits a clear energy "boom-bust" cycle, including a clear downturn around 2015. This timing aligns with the mid-2014 oil price collapse and the subsequent retrenchment in drilling activity and energy-sector employment and investment, which disproportionately affected energy-producing states and regions.

Comparing the time profiles across indicators, we see that the PARAFAC2 model allows for some heterogeneity in the timing and amplitude of the cycles across indicators, while the PARAFAC model imposes a common time profile across indicators. Overall, all three components exhibit similar temporal patterns across indicators. However, there are some differences in the amplitude and timing of the cycles across indicators, specially for the unemployment indicator, which justifies the use of PARAFAC2 to allow for indicator-specific time components. In contrast, PARAFAC imposes a common time profile across indicators, which may be too restrictive in this context.

3.2.3 Indicator loadings

Coming soon.

[Figure 6 about here]

3.3 Model assessment

Coming soon.

[Figure 7 about here]

3.4 Forecast error variance decomposition

Coming soon.

4 Conclusion

Coming soon.

References

Abel, J. R. and Deitz, R. (2010). Bypassing the Bust: The Stability of Upstate New York's Housing Markets during the Recession. *Current Issues in Economics and Finance*, 16(3):1–9.

Abel, J. R. and Deitz, R. (2011). Local Hangovers: How the Housing Boom and Bust Affected Jobs in Metro Areas. Liberty Street Economics, Federal Reserve Bank of New York. Accessed 2026-01-26.

Aguiar-Conraria, L., Brinca, P., Gudjonsson, H. V., and Soares, M. J. (2017). Business cycle synchronization across US states. *The BE Journal of Macroeconomics*, 17(1):1–15.

Arias, M. A., Gascon, C. S., and Rapach, D. E. (2016). Metro business cycles. *Journal of Urban Economics*, 94:90–108.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, 103(6):2121–2168.

Bai, J. and Ng, S. (2008). Large Dimensional Factor Analysis. *Foundations and Trends in Econometrics*, 3(2):89–163.

Baumeister, C., Leiva-León, D., and Sims, E. (2024). Tracking weekly state-level economic conditions. *Review of Economics and Statistics*, 106(2):483–504.

Berger, T. and Wortmann, M. (2022). Global vs. group-specific business cycles: the importance of defining the groups. *Macroeconomic Dynamics*, 26(1):49–71.

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2):149–171.

Bro, R. and Smilde, A. K. (2003). Centering and scaling in component analysis. *Journal of Chemometrics*, 17(1):16–33.

Brown, J. P. (2017). Identifying State-Level Recessions. *Economic Review-Federal Reserve Bank of Kansas City*, 102(1):85–108.

Carlino, G. A. and DeFina, R. H. (2004). How Strong is Co-Movement in Employment Over the Business Cycle? Evidence from State/Sector Data. *Journal of Urban Economics*, 55(2):298–315.

Carroll, J. D. and Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n–way generalization of the eckart–young decomposition. *Psychometrika*, 35(3):283–319.

Congressional Budget Office (2008). Factors Underlying the Decline in Manufacturing Employment Since 2000. Economic and Budget Issue Brief, Congressional Budget Office. Accessed 2026-01-26.

Crone, T. M. (2005). An Alternative Definition of Economic Regions in the United States Based on Similarities in State Business Cycles. *The Review of Economics and Statistics*, 87(4):617–626.

Crone, T. M. and Clayton-Matthews, A. (2005). Consistent economic indexes for the 50 states. *Review of Economics and Statistics*, 87(4):593–603.

Del Negro, M. (2002). Asymmetric shocks among us states. *Journal of International Economics*, 56(2):273–297.

Despois, A. and Doz, C. (2023). Identifying and Interpreting the Factors in Factor Models via Sparsity: Different Approaches. *Journal of Applied Econometrics*, 38(3):394–416.

Ding, C., Li, T., Peng, W., and Park, H. (2006). Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 126–135.

Ding, C. H., Li, T., and Jordan, M. I. (2008). Convex and semi-nonnegative matrix factorizations. *IEEE transactions on pattern analysis and machine intelligence*, 32(1):45–55.

Federal Reserve Bank of Philadelphia (2025). State Coincident Indexes. https://www.philadelphiafed.org/surveys-and-data/regional-economic-analysis/state-coincident-indexes. Accessed 2026-01-26.

Francis, N., Owyang, M. T., and Savascin, O. (2017). An endogenously clustered factor approach to international business cycles. *Journal of Applied Econometrics*, 32(7):1261–1276.

Francis, N., Owyang, M. T., and Soques, D. (2022). Business Cycles Across Space and Time. *Journal of Money, Credit and Banking*, 54(6):1529–1568.

Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., and Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1):100.

Hamilton, J. D. and Owyang, M. T. (2012). The Propagation of Regional Recessions. *The Review of Economics and Statistics*, 94(4):935–947.

Harris, K. (2020). Forty Years of Falling Manufacturing Employment. Beyond the Numbers: Employment & Unemployment 16, U.S. Bureau of Labor Statistics. Accessed 2026-01-26.

Harshman, R. A. (1970). *Foundations of the PARAFAC Procedure: Models and Conditions for an Explanatory Multimodal Factor Analysis*. PhD thesis, University of California, Los Angeles.

Harshman, R. A. and Lundy, M. E. (1994). Parafac: Parallel factor analysis. *Computational Statistics & Data Analysis*, 18(1):39–72.

Haughwout, A., Peach, R. W., Sporn, J., and Tracy, J. (2013). The Supply Side of the Housing Boom and Bust of the 2000s. In *Housing and the Financial Crisis*, pages 69–104. University of Chicago Press.

Helwig, N. E. (2017). Estimating latent trends in multivariate longitudinal data via Parafac2 with functional and structural constraints. *Biometrical Journal*, 59(4):783–803.

Helwig, N. E. (2025). *multiway: Component Models for Multi-Way Data*. R package version 1.0-7.

Kahou, M. E., Fernández-Villaverde, J., Gómez-Cardona, S., Perla, J., and Rosa, J. (2024). Spooky Boundaries at a Distance: Inductive Bias, Dynamic Models, and Behavioral Macro. Technical report, National Bureau of Economic Research.

Kiers, H. A., Ten Berge, J. M., and Bro, R. (1999). Parafac2—part i. a direct fitting algorithm for the parafac2 model. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 13(3-4):275–294.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, 51(3):455–500.

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

Lamm, B. and O'Keefe, J. (1997). Banking Problems in the Northeast. In *History of the Eighties: Lessons for the Future (Vol. 1). An Examination of the Banking Crises of the 1980s and Early 1990s*, chapter 10. Federal Deposit Insurance Corporation.

Lavelle, M. and Lombardi, B. (2010). Automotive Industry Outlook: Understanding the Impact on Workers and Communities. Chicago Fed Letter 270a, Federal Reserve Bank of Chicago. Accessed 2026-01-26.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non–negative matrix factorization. *Nature*, 401(6755):788–791.

Leiva-Leon, D. (2017). Measuring business cycles intra-synchronization in us: A regime-switching interdependence framework. *Oxford Bulletin of Economics and Statistics*, 79(4):513–545.

Lettau, M. (2022). High dimensional factor models with an application to mutual fund characteristics. Technical report, National Bureau of Economic Research.

Mattoon, R. (2016). Where are the Jobs? Midwest Jobs Recovery by State and Sector. Midwest Economy Blog, Federal Reserve Bank of Chicago. Accessed 2026-01-26.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill, New York.

Owyang, M. T., Piger, J. M., and Wall, H. J. (2005). Business Cycle Phases in U.S. States. *The Review of Economics and Statistics*, 87(4):604–616.

Owyang, M. T., Piger, J. M., and Wall, H. J. (2008). A State-Level Analysis of the Great Moderation. *Regional Science and Urban Economics*, 38:578–589.

Owyang, M. T., Rapach, D. E., and Wall, H. J. (2009). States and the Business Cycle. *Journal of Urban Economics*, 65(2):181–194.

Perry, M. (2002). Population growth in the 1990s: Patterns within the United States. *Population Research and Policy Review*, 21(1):55–71.

Perry, M. and Mackun, P. (2001). Census 2000 Brief. Technical Report TR-1234, U.S. Census Bureau.

Prüser, J. and Schmidt, T. (2021). Regional composition of national house price cycles in the us. *Regional Science and Urban Economics*, 87:103645.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. R version 4.5.2.

Saulsbury, V. and Curry, T. (1997). Banking Problems in the California. In *History of the Eighties: Lessons for the Future (Vol. 1). An Examination of the Banking Crises of the 1980s and Early 1990s*, chapter 11. Federal Deposit Insurance Corporation.

Stock, J. H. and Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, 4:351–394.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Tsay, R. S. (2024). Matrix-Variate Time Series Analysis: A Brief Review and Some New Developments. *International Statistical Review*, 92(2):246–262.

U.S. Census Bureau (2025). Census Regions and Divisions of the United States. https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf. Accessed 2026-01-26.

Tables and Figures

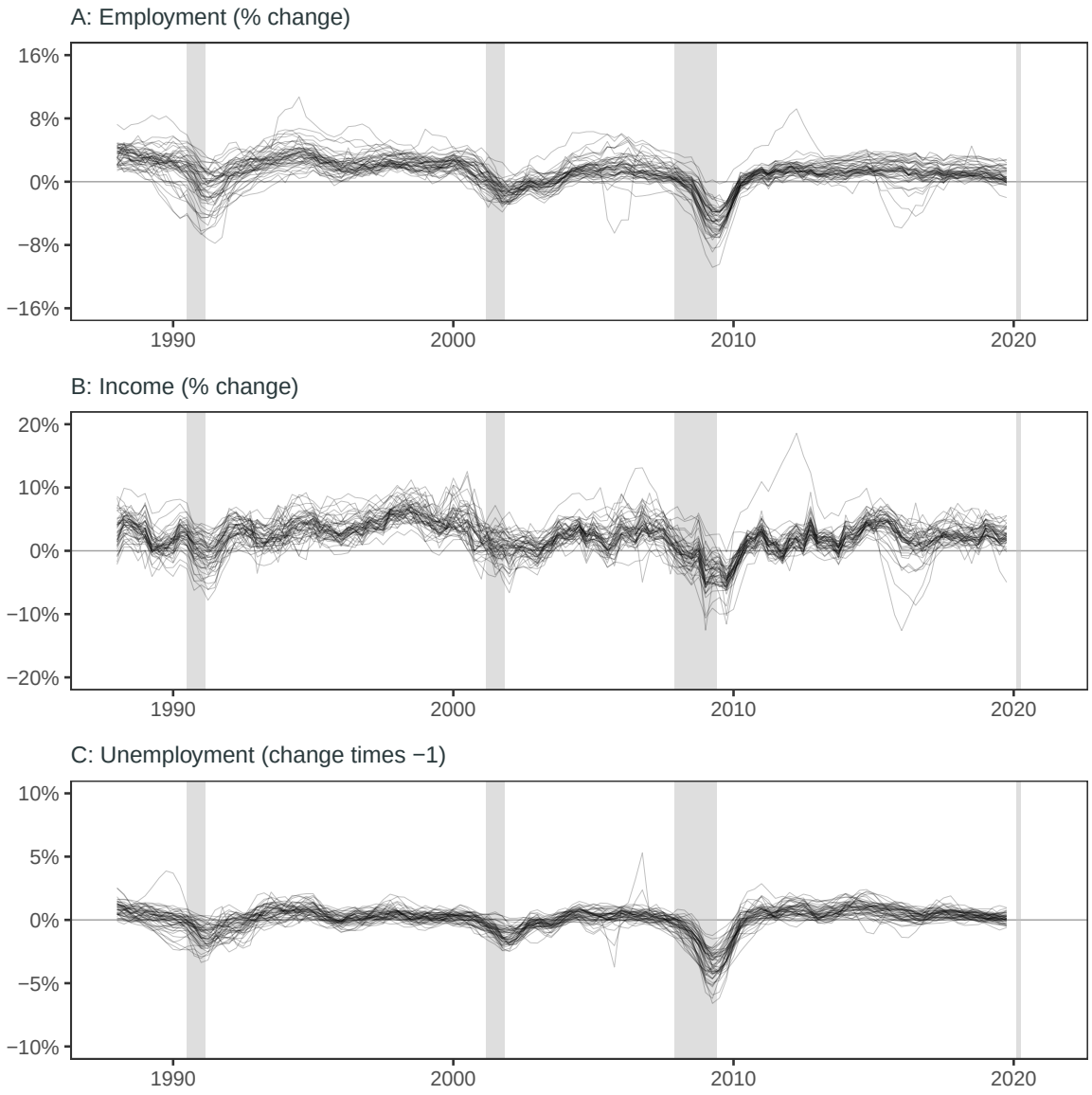


Figure 1: This figure plots the three state-level indicators used in the analysis. The indicators are transformed into stationary time series by taking quarterly percentage changes for employment and income, and first differences for unemployment. Unemployment is sign-reversed so that positive values for all three indicators correspond to local expansions. Shaded areas denote NBER recessions.

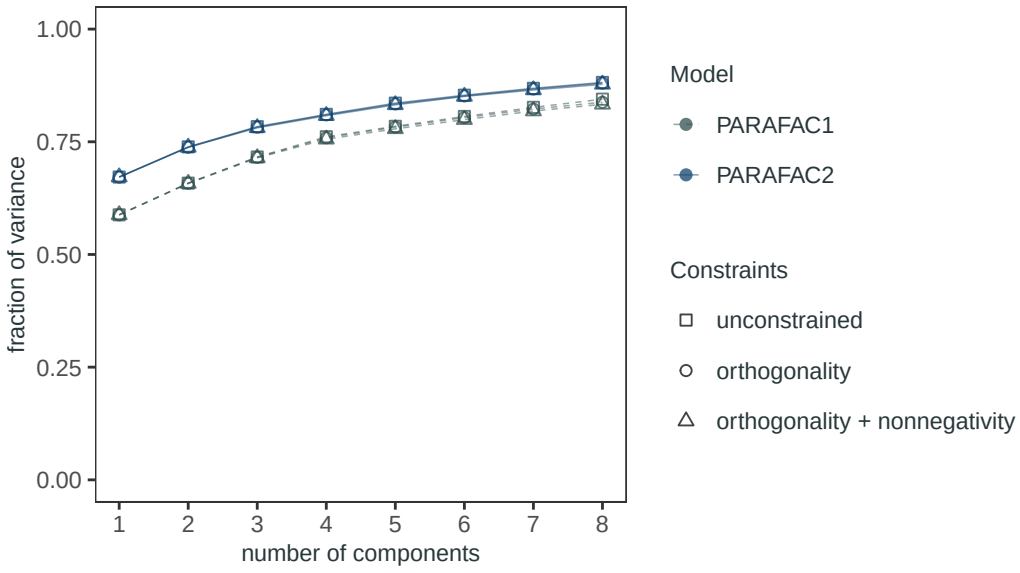


Figure 2: Tensor scree plots for PARAFAC and PARAFAC2 models under the three sets of constraints. The gray lines corresponds to the PARAFAC model. The blue lines corresponds to the PARAFAC2 model. The y-axis shows the proportion of variance explained (R^2) by each model as a function of the rank R (shown on the x-axis).

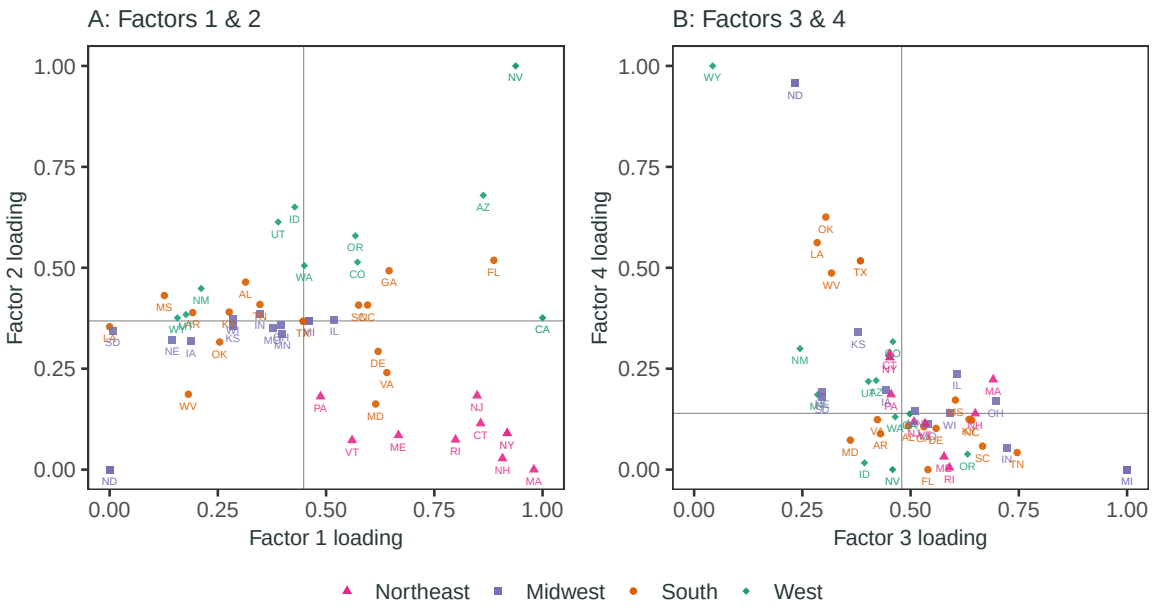


Figure 3: This figure shows the estimated state loadings (normalized) for each of the four regional cycles of the preferred PARAFAC2 specification. Panel A (left) displays the state loadings for factors 1 and 2. Panel B (right) displays the state loadings for factors 3 and 4. States are grouped into the four Census regions: Northeast, Midwest, South, and West.

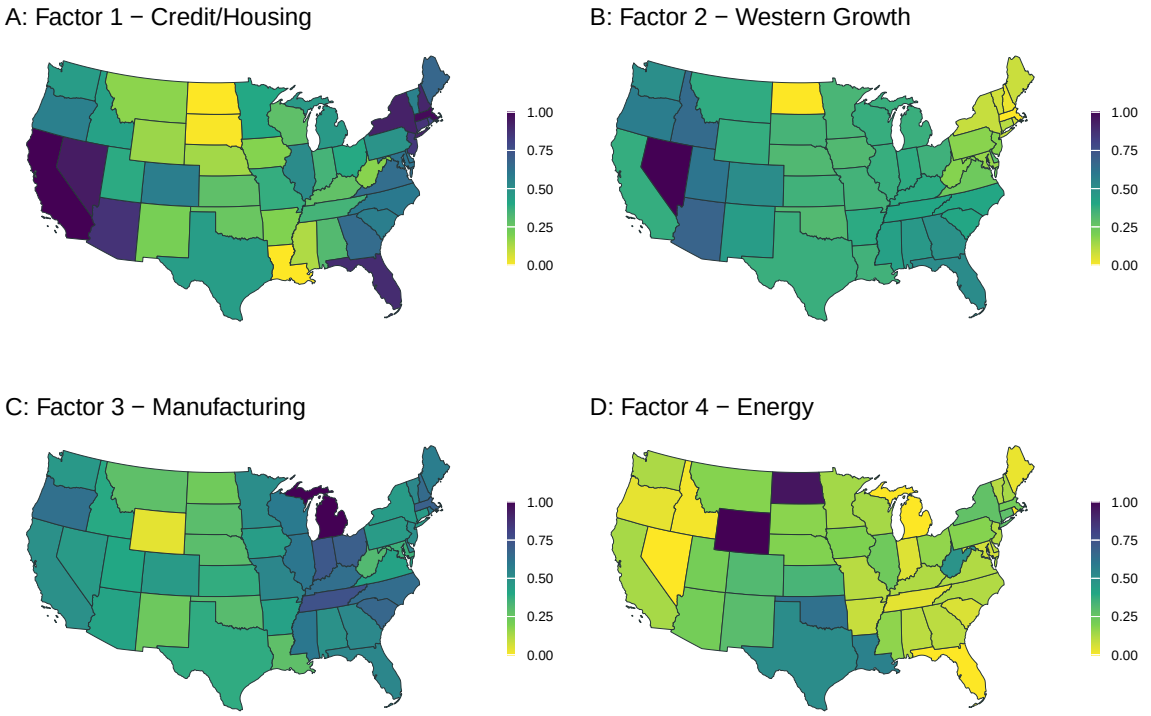


Figure 4: This figure shows the geographic distribution of the normalized state loadings for each of the four regional cycles of the preferred PARAFAC2 specification.

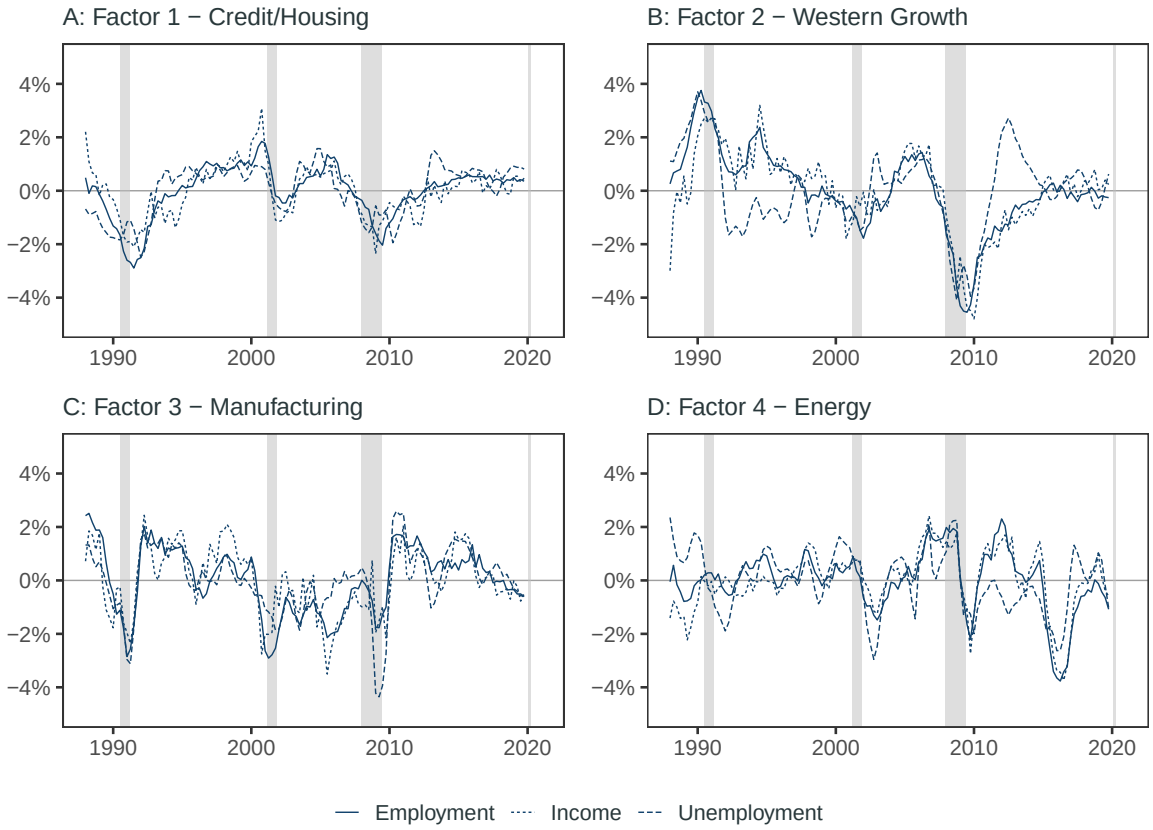


Figure 5: This figure shows the estimated time profiles for each of the four regional cycles of the preferred PARAFAC2 specification. Panel A (top-left) displays factor 1 (the credit/housing cycle) for each of the three indicators: employment, income, and unemployment. Panel B (top-right) displays factor 2 (the western growth cycle) for each of the three indicators. Panel C (bottom-left) displays factor 3 (the manufacturing cycle) for each of the three indicators. Panel D (bottom-right) displays factor 4 (the energy cycle) for each of the three indicators. Shaded areas denote NBER recessions.

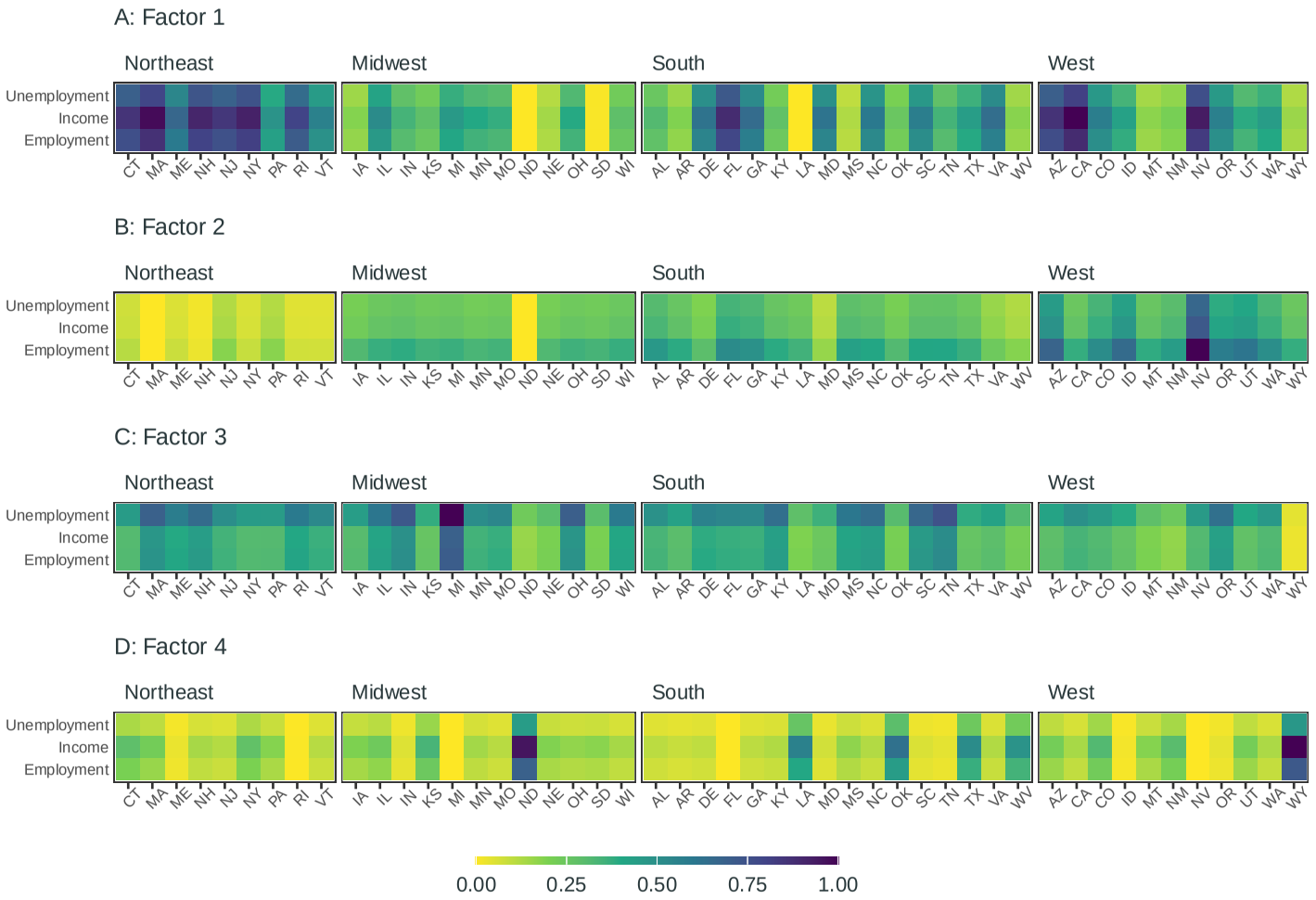


Figure 6: This figure shows the predicted contributions $b_r(s) \times c_r(k)$ for each state s, indicator k, and each component r of the preferred PARAFAC2 specification.



Figure 7: This figure shows the R^2 between the actual and fitted values for each state and each indicator of the preferred PARAFAC2 specification.

Online Appendix

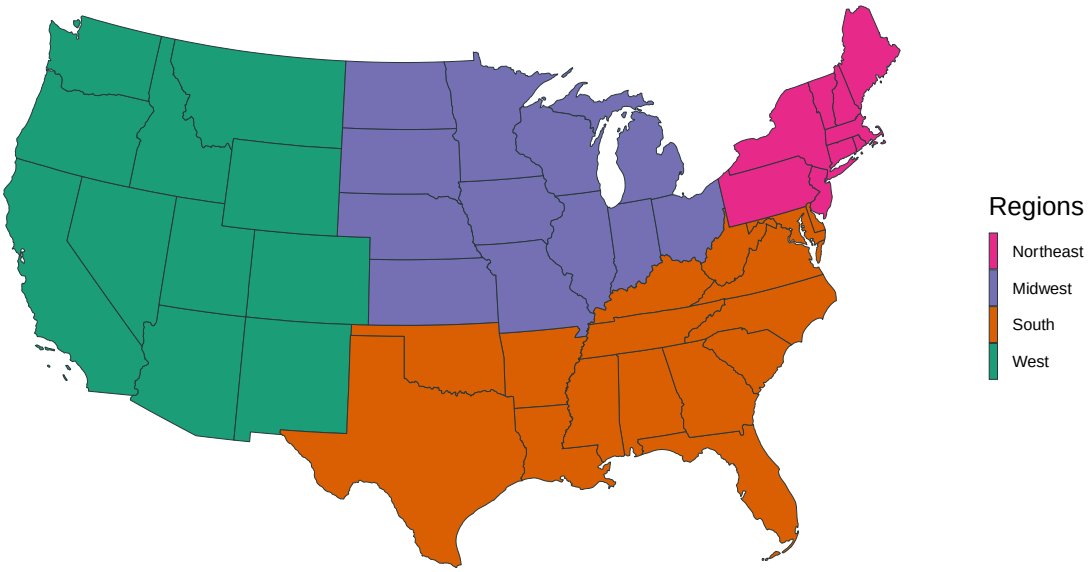


Figure A.1: Census regions of the United States. The Northeast region includes the following states: CT, ME, MA, NH, NJ, NY, PA, RI, and VT. The Midwest region includes: IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, and WI. The South region includes: AL, AR, DE, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, and WV. The West region includes: AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA, and WY. We exclude Alaska (AK) and Hawaii (HI) from the analysis.
Source: U.S. Census Bureau.

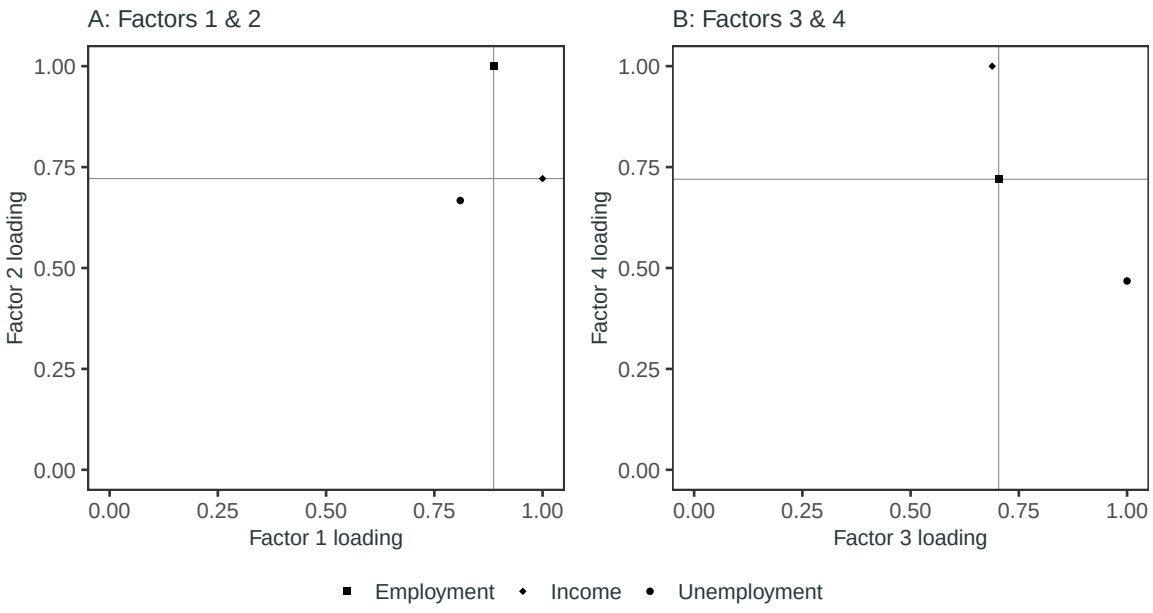


Figure A.2: This figure shows the estimated indicator loadings (normalized) for each of the four regional cycles of the preferred PARAFAC2 specification. Panel A (left) displays the indicator loadings for components 1 and 2. Panel B (right) displays the indicator loadings for components 3 and 4.

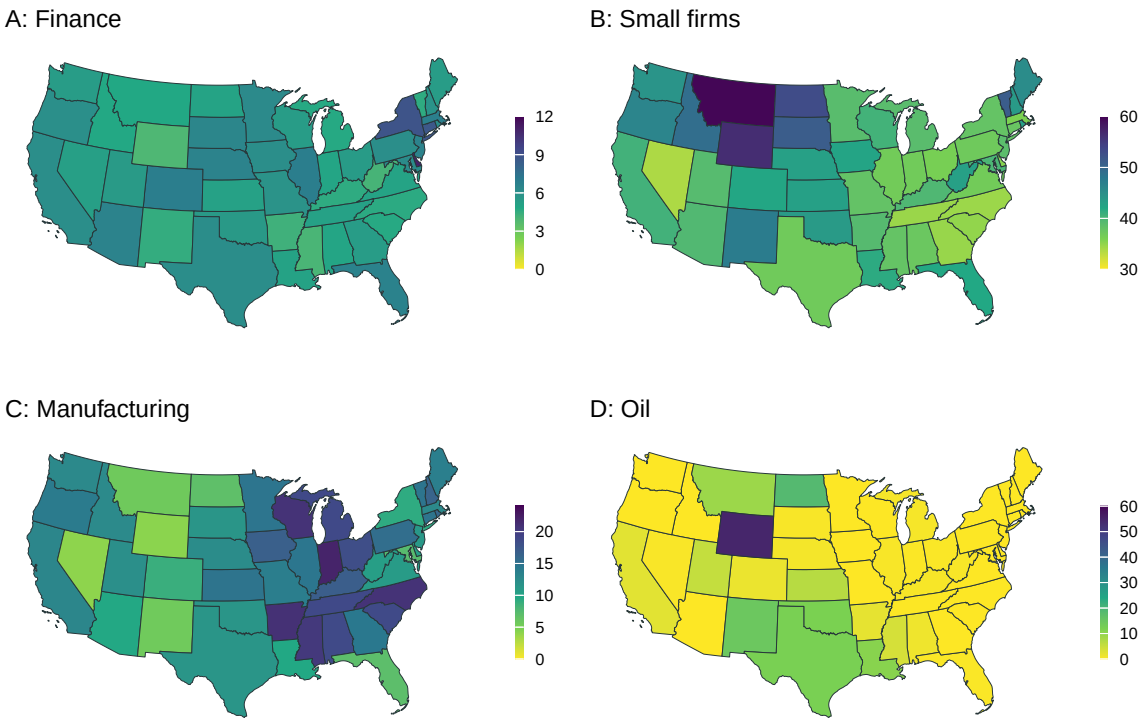


Figure A.3: Covariates plot.

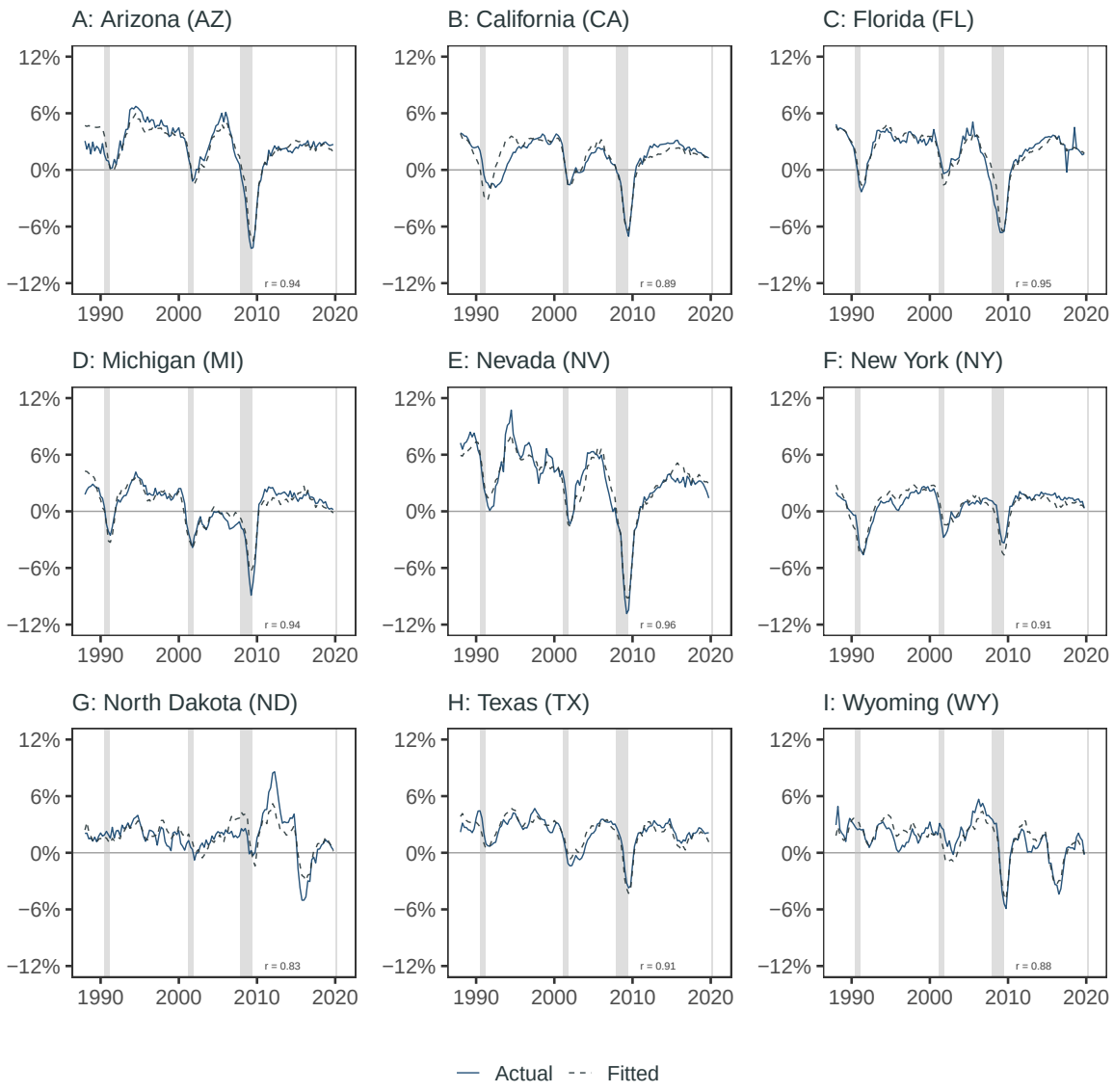


Figure A.4: This figure shows the fitted values for the employment indicator of the preferred PARAFAC2 specification for selected states. Shaded areas denote NBER recessions.

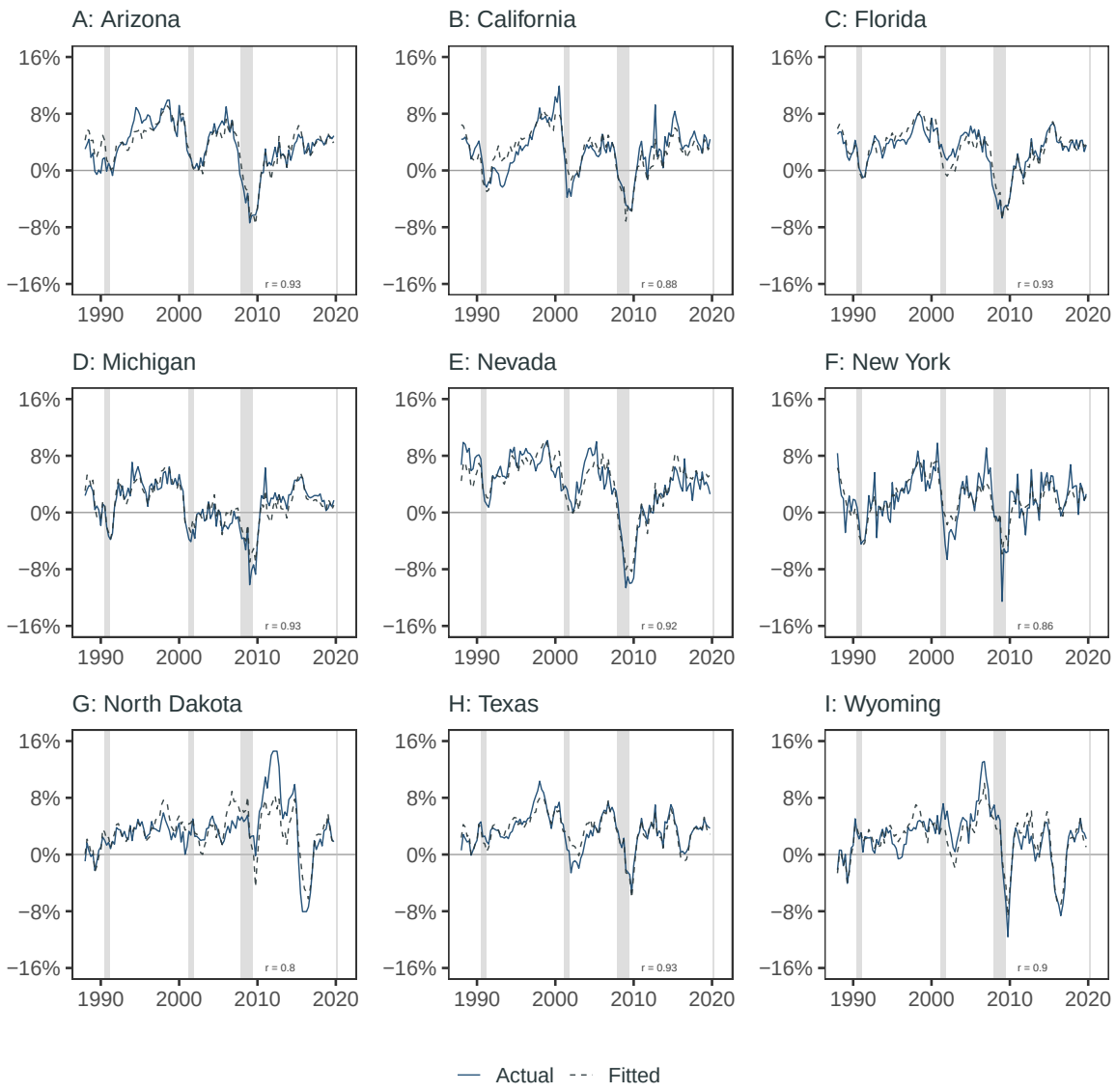


Figure A.5: This figure shows the fitted values for the income indicator of the preferred PARAFAC2 specification for selected states. Shaded areas denote NBER recessions.

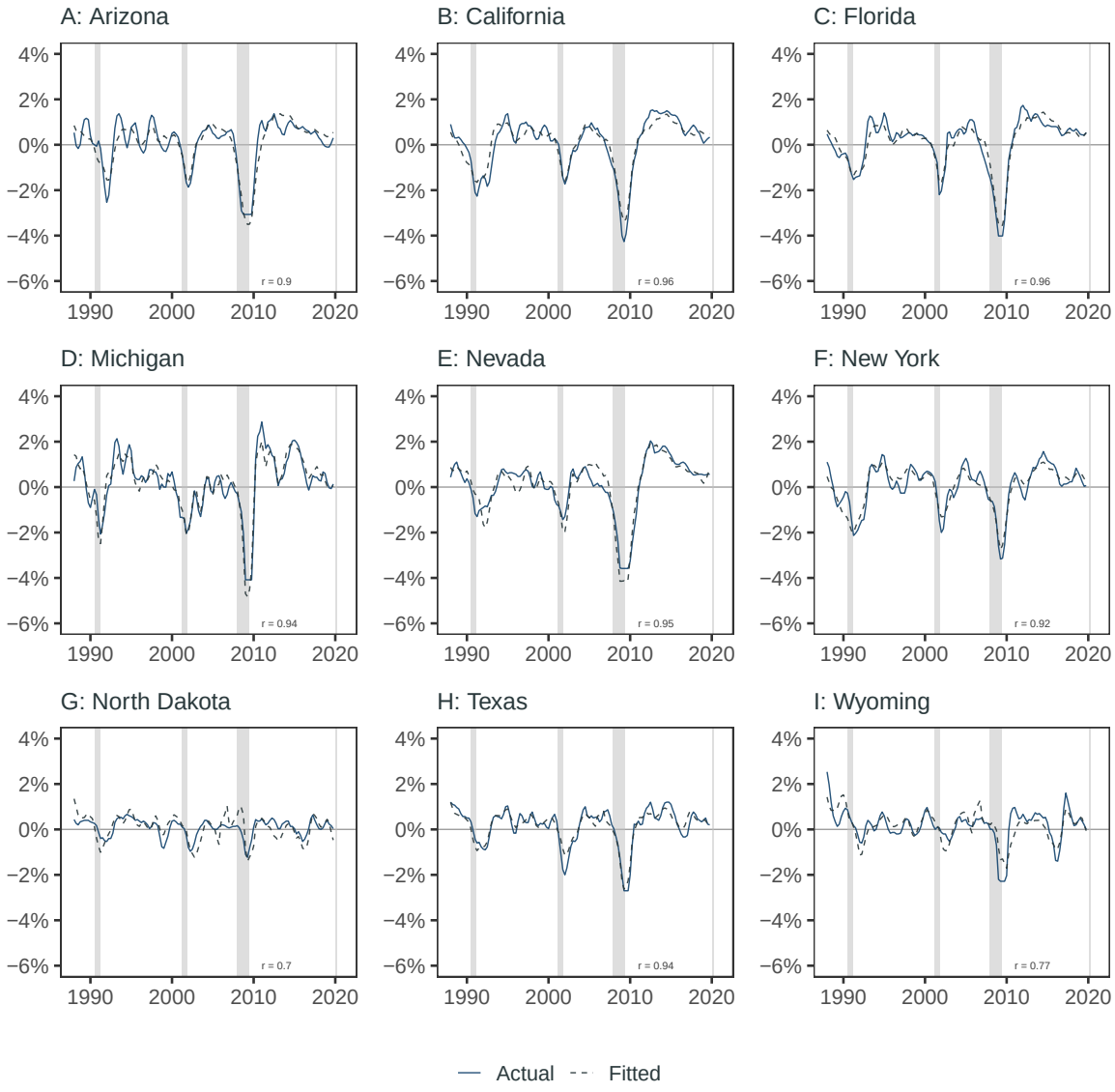


Figure A.6: This figure shows the fitted values for the unemployment indicator of the preferred PARAFAC2 specification for selected states. Shaded areas denote NBER recessions.

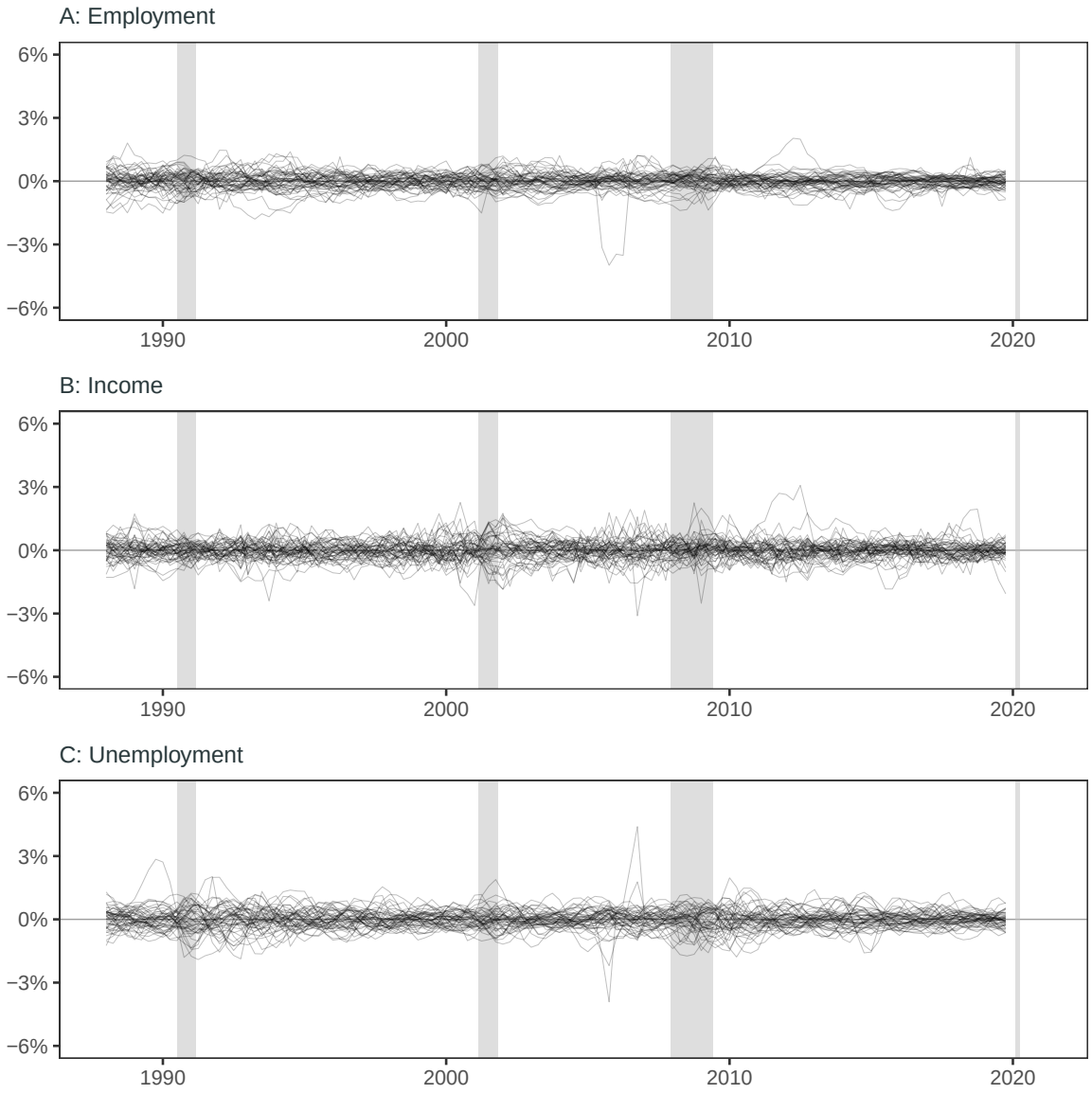


Figure A.7: This figure shows the residuals for each of the three indicators of the preferred PARAFAC2 specification. Shaded areas denote NBER recessions.

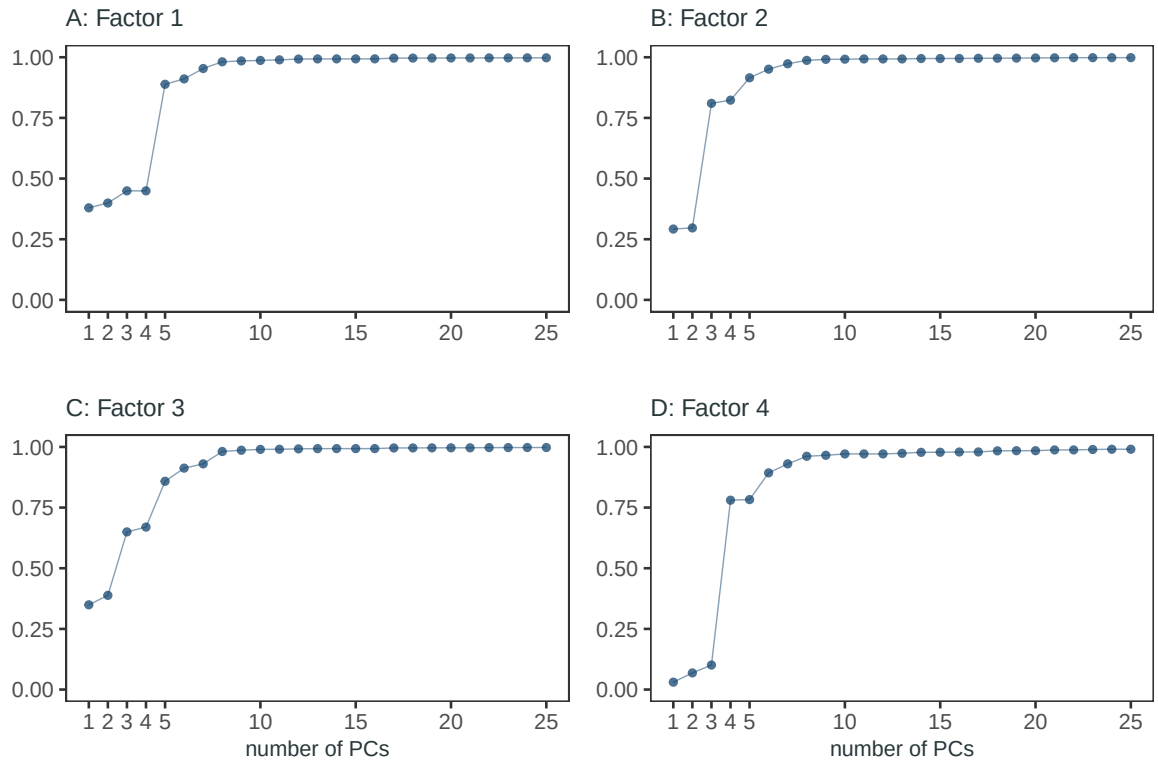


Figure A.8: This figure shows the correlation betweem the estimated regional cycles from the preferred PARAFAC2 specification and the first 25 components from PCA.