

DSSR

Discussion Paper No. 150

Density-valued VAR Models with Latent Factors

Yasumasa Matsuda and Michel F. C. Haddad

April, 2026

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Density-valued VAR Models with Latent Factors

Yasumasa Matsuda*

Graduate School of Economics and Management
Tohoku University, Japan

Michel F. C. Haddad†

Department of Business Analytics and Applied Economics
Queen Mary University of London, UK

Abstract

We propose a density-valued vector autoregressive model with latent factors for multivariate time series of density functions. Motivated by weekly regional distributions of SARS-CoV-2 cycle threshold (Ct) values in Brazil, we study their distributional dynamics across regions. The Ct value is the number of amplification cycles required for the viral signal to cross a detection threshold (lower Ct values correspond to higher viral load). We estimate each regional density by a B-spline mixture, mapping the mixture weights to a Euclidean space by a generalized logit transform equipped with an isometric inner product, and model the transformed series by a cross-regional VAR with latent factors. This decomposition allows for the separation between strong common movements and directed idiosyncratic dynamics. Directed edges are identified from the idiosyncratic VAR component using one-sided tests with Benjamini–Yekutieli false discovery rate control. Simulations show that increasing the number of estimated factors does not mechanically eliminate genuine idiosyncratic dependence; rather, it mainly removes spuriously detected edges driven by common factor movements. In the real-world data application, the full sample yields only a weak directed network, whereas a substantial network emerges once the first six months are excluded and the density prior is kept weak. The estimated links suggest directed predictive relations from the northern region toward southeastern metropolitan areas.

Keywords: B-spline mixtures; Density-valued time series; Generalized logit transform; Granger causality; Viral load surveillance.

1 Introduction

The cycle threshold (Ct) value from a polymerase chain reaction (PCR) test is the number of amplification cycles required for the viral signal to cross a detection threshold. It is widely used as a proxy for viral load, with lower Ct values corresponding to higher viral load (Jones et al., 2021; Lin et al., 2022). In epidemic surveillance, however, the object of interest is often not a single summary such as the weekly mean Ct value, but its entire distribution within a region over time. Changes in this distribution may reflect shifts in the composition of tested individuals, the timing of testing, and the intensity of local transmission. When such distributional information is observed for many regions over time, a crucial empirical question is whether one can detect directional lead–lag dependence in the evolution of regional distributions.

*yasumasa.matsuda.a4@tohoku.ac.jp

†m.haddad@qmul.ac.uk

Ct values have increasingly been used in epidemic surveillance as population-level indicators that may contain information beyond the diagnosis of individual cases. In particular, aggregate Ct measures have been studied as early signals of variant emergence and changing transmission intensity. For example, [Harrison et al. \(2023\)](#) reported that shifts in population-level Ct values preceded subsequent increases in case counts during the emergence of major SARS-CoV-2 variants, suggesting that Ct-based summaries may complement conventional surveillance indicators. The epidemiological relevance of Ct values is also supported by their association with transmission dynamics and disease severity. [Baggio et al. \(2021\)](#) documented comparable SARS-CoV-2 viral loads in children and adults during early infection, while [Liu et al. \(2020\)](#) found that severe COVID-19 cases tended to exhibit lower Ct values and longer viral shedding; see also [Rao et al. \(2020\)](#) for a broader review of the relation between Ct values and clinical outcomes. At the same time, most existing work relies on scalar summaries of Ct values, such as means or other aggregate indicators. When the aim is to study how viral load evolves and propagates across regions, such descriptive statistics may be too restrictive. This is because they discard potentially important information about dispersion, skewness, and tail behavior in the regional Ct-value distributions.

A further relevant challenge arises in the statistical analysis of multivariate time series of density functions. A density function is not an ordinary vector-valued observation. Instead, it must satisfy nonnegativity and unit-integral constraints, and therefore does not belong to a linear space under the usual operations. In regional epidemic surveillance one often expects at least two distinct sources of dependence. One is pervasive nationwide comovement driven by common shocks, such as aggregate epidemic waves or nationwide policy changes. The other is the directed lagged dependence across regions, reflecting heterogeneous local propagation. A useful statistical framework must therefore handle both the nonlinear structure of density-valued observations and the separation of common and idiosyncratic dynamic dependence.

Methods for density-valued data ([Petersen and Müller, 2016](#)), factor models ([Bai, 2003, 2009](#)), and Granger-causality networks ([Granger, 1969](#)) have largely developed in parallel. Consequently, there is still no unified methodology for multivariate time series of density functions that separates strong common movements from directed idiosyncratic dependence across regions. This separation is essential in practice, because without controlling for pervasive common factors, apparent cross-regional predictability may simply reflect nationwide shocks rather than genuine directional dependence. Related work is reviewed in [Section 2](#).

In this paper, we propose a density-valued vector autoregressive (VAR) model with latent factors for multivariate time series of density functions observed across regions, motivated by regional Ct-value distributions. For each week and region, we first estimate the Ct-value density by a B-spline mixture and represent it by its mixture-weight vector. We then map these weights from the simplex to a Euclidean coordinate space using a generalized logit transform, equipping that space with an inner product that is isometric to the corresponding spline-mixture function space under the L^2 geometry. This representation enables us to model the transformed series by a cross-regional VAR with latent factors.

In this specification, the factor part captures common nationwide movements, while the VAR part describes directed lagged dependence in the idiosyncratic distributional dynamics. Based on the estimated VAR coefficients, we construct a directed network using one-sided t -tests with Benjamini–Yekutieli false discovery rate control ([Benjamini and Yekutieli, 2001](#)), so that selected edges represent Granger-predictive relations after partialling out common factors. An additional advantage of the proposed decomposition is that stability is required only for the idiosyncratic VAR component, not for the observed series as a whole. The common factor component may exhibit trend-like nonstationary movements. Thus, the framework can still be applied when the observed multivariate density-valued series exhibits pervasive nonstationary comovement.

The present paper also provides a formal basis for testing directed lagged dependence after removing latent common factors. We develop the asymptotic theory for the factor-adjusted VAR estimator in the transformed Euclidean representation. Under a joint asymptotic regime in which both the basis dimension and the time dimension diverge while the number of regions is fixed, we establish its rate of convergence and, under stronger conditions, derive the asymptotic normality for coefficient-wise inference.

Our contribution is threefold. First, we develop a tractable statistical framework for multivariate time series of density functions that combines B-spline density estimation, a generalized logit transformation of simplex-valued weights, an isometric inner-product structure, latent common factors capturing nationwide movements, and cross-regional VAR dependence. Second, we provide an inferential procedure for identifying directed edges under multiplicity control, thereby separating common-factor comovement from idiosyncratic predictive links. Third, our simulations and real-world data analysis clarify how detected edges should be interpreted when the factor dimension, the sample period, and the density prior are varied. In particular, the simulation study shows that genuine idiosyncratic dependence remains detectable after factor adjustment, whereas the empirical application shows that a substantial directed network emerges only after the early pandemic period is excluded and the nationwide prior is retained weak.

2 Related Work

2.1 Density-valued time series

A basic difficulty in modeling density-valued data is that the space of density functions is not a linear space in the usual sense. Although densities can be equipped with various metrics, ordinary addition and scalar multiplication do not generally preserve nonnegativity and unit-integral constraints. For this reason, statistical methodology for density-valued data has developed along two main lines: object-oriented approaches and transformation-based approaches.

Object-oriented approaches treat densities directly as structured random objects endowed with an intrinsic geometry or metric, rather than embedding them first into a finite-dimensional Euclidean space. Typical examples are frameworks based on the Wasserstein metric or the Fisher–Rao metric, which view the space of densities as a nonlinear geometric object and formulate statistical analysis through the corresponding tangent spaces or transport maps (Panaretos and Zemel, 2020; Srivastava et al., 2011; Petersen et al., 2022). In the time-series setting, Zhang et al. (2022) proposed Wasserstein autoregressive models, providing a rigorous way to describe temporal dependence of density-valued observations directly in Wasserstein space. These approaches are attractive when the density itself, together with its intrinsic geometry, is the primary object of inference.

Transformation-based approaches instead map densities into a linear space, and then apply conventional statistical tools in that representation space. Representative examples include the transformation-based functional data analysis of densities in Petersen and Müller (2016), Bayes-space and simplicial representations for density functions (van den Boogaart et al., 2010; Hron et al., 2016), the panel-of-densities analysis in Kutta et al. (2025), and the generalized-logit spline-mixture approach in Matsuda and Iwafuchi (2025). A common advantage of these methods is that they provide explicit coordinates in which linear operations, regression structures, and standard inferential tools can be developed while retaining essential distributional information.

The present paper belongs to this transformation-based line of research. This is not solely driven by computational convenience. Our main objective is to estimate cross-regional VAR dependence, remove pervasive common factor movements, and conduct coefficient-wise inference for directed edges in multivariate time series of density functions. While one might in principle formulate

dynamic dependence directly in an object-oriented framework as well, it is considerably harder to derive the regression-type asymptotic theory needed for VAR coefficient estimation and edge selection. Conversely, an explicit Euclidean coordinate representation makes the model amenable to both factor adjustment and asymptotic analysis.

At the same time, existing work in either stream of research has paid limited attention to settings where density-valued observations are observed across multiple units and where pervasive common movements must be separated from directed lagged dependence in the idiosyncratic component. This is the setting considered in the present paper.

2.2 Latent factor components

Factor models provide a standard framework for separating common movements from unit-specific variation in high-dimensional data. In macroeconomics and finance, dynamic factor models have been widely used to summarize comovements across many time series by a small number of latent factors (Stock and Watson, 2002; Forni et al., 2000, 2005). The statistical foundations of factor estimation and factor-number selection are developed in, among others, Bai and Ng (2002) and Bai (2003). In panel settings, Bai (2009) extended this logic through interactive fixed-effects models with heterogeneous unit-level loadings.

Several types of factor models could in principle be considered for the present problem, including spectral factor formulations as in Forni et al. (2000, 2005) and dynamic factor formulations as in Stock and Watson (2002). In the present paper, however, our main inferential target is not the common component itself, but the coefficient matrix of the factor-adjusted idiosyncratic VAR component. We adopt a Bai-type formulation for two reasons. First, our empirical setting may exhibit pervasive common movements that need not be stationary in the usual time-series sense, and the Bai-type framework allows such common variation to be treated under relatively weak second-moment conditions. Second, our main inferential target is the coefficient matrix of the factor-adjusted idiosyncratic VAR component, and the Bai-type framework leads naturally to regression-style asymptotic arguments for that object.

At the same time, our setting differs fundamentally from the standard interactive fixed-effects regression framework. In Bai (2009), the primary object is a regression model with regressors that are treated as exogenous or conditionally mean independent, and the latent factor component enters as an unobserved common effect. In our framework, by contrast, the post-adjustment component of interest is itself dynamic and is modeled through a cross-regional VAR. Thus, the factor structure is not introduced merely to absorb nuisance heterogeneity. Rather, it is introduced in order to separate pervasive common movements from the idiosyncratic dynamic component whose lagged dependence is the main object of inference.

The present paper builds on this literature by applying a Bai-type factor framework to transformed density-valued time series. Once density-valued observations are mapped into an isometric Euclidean coordinate system, the resulting model can be written in a factor-augmented representation in which the common component captures pervasive nationwide movements, while the remaining component is modeled by a cross-regional VAR. This makes the Bai-type algebraic framework useful, but the interpretation is importantly different from that in conventional scalar panel applications.

The difference is not only in the type of observed data but also in the role played by dynamics. First, the observed objects here are not scalar outcomes but transformed representations of densities. Second, the regressors in the idiosyncratic equation are generated by lagged dependent variables through a VAR structure. This means that, unlike in standard interactive fixed-effects regressions, the regressor structure is intrinsically dynamic and does not satisfy the same exogene-

ity interpretation. For this reason, the present model is not a direct application of Bai (2009). Rather, it is a factor-augmented dynamic regression setting in which the Bai-type decomposition remains useful, but the subsequent asymptotic arguments must be adapted to account for the VAR dependence.

This distinction is important for our substantive objective. In much of the factor-model literature, the common component is itself the main object of interpretation, and the idiosyncratic term is treated largely as residual variation. In our framework, the logic is reversed. The common component is removed because it reflects pervasive nationwide movements that may obscure regional transmission patterns, while the idiosyncratic component is retained and modeled explicitly. The main object of interest is therefore the directed lagged dependence embedded in the idiosyncratic VAR coefficient matrix.

A further advantage of this formulation is that it aligns with our asymptotic regime. In our problem, the number of regions is fixed, while the spline basis dimension and the time length diverge. After the isometric transformation, this produces a high-dimensional representation that differs from the conventional large-panel setting of the factor literature. The resulting asymptotic theory is therefore tailored to transformed density-valued time series with fixed regional dimension and growing basis dimension, rather than to standard scalar panels with many cross-sectional units. In this sense, our approach is Bai-type in algebraic structure, but it is designed for a substantively and theoretically different dynamic environment.

2.3 Network dependence and Granger-causality analysis

A large literature studies directed dependence in multivariate time series through vector autoregressive models and Granger-causality analysis. In this framework, directed edges are typically defined by whether past values of one unit may predict current values of another unit, conditional on the remaining variables in the system (Granger, 1969; Lütkepohl, 2005). Related developments include graphical approaches to multivariate time-series dependence (Eichler, 2007) and network measures of connectedness and spillovers in large systems (Billio et al., 2012; Diebold and Yilmaz, 2014).

Our paper is related to this literature, but differs from standard applications in two essential respects. First, the objects entering the VAR are not scalar observations such as levels, growth rates, or means, but transformed representations of density functions. The resulting directed edges therefore describe predictive relations in the evolution of regional density functions, rather than in a conventional scalar panel. Second, the network is constructed only after removing latent common factors. This distinction is crucial in our setting, because strong nationwide epidemic waves may induce pervasive comovement across regions, and such comovement should not be confused with genuine directed lagged dependence. These two features jointly determine the interpretation of the network in our setting.

Accordingly, our procedure measures Granger-causality through the coefficient matrix of the idiosyncratic VAR component in a factor-adjusted density-valued VAR model. Existing Granger-causality analyses based on ordinary VAR models are not designed to assess lead-lag relations between density functions themselves, nor do they directly separate such relations from common factor-driven distributional comovement. In this sense, the proposed network is not simply a distributional analogue of a standard scalar VAR network, but a device for identifying predictive dependence in density dynamics after common movements have been partialled out.

From an inferential perspective, our procedure is based on coefficient-wise t -tests for VAR edges under multiple testing. This requires asymptotic justification for the VAR coefficient estimators after factor adjustment. Building on the Bai-type framework discussed above, we develop such

inference in the transformed density-valued setting, where the generalized logit transformation, together with the induced inner-product structure, provides an isometric Euclidean representation of density-valued observations. To control multiplicity in edge selection, we use the false discovery rate procedure of [Benjamini and Yekutieli \(2001\)](#).

Taken together, these three strands of literature motivate the present framework. Transformation-based methods provide a workable Euclidean representation for density-valued observations, factor models provide a way to remove pervasive common movements, and Granger-causality analysis provides a language for directed lagged dependence. The contribution of this paper is to combine them into a single framework suited to multivariate time series of density functions.

3 Method

The methodology introduced in this section applies more generally to multivariate time series of density functions observed across units, despite the fact that the empirical application in this paper focuses on regional Ct-value distributions. Let C denote the number of units, and suppose that at each time t and unit $c = 1, \dots, C$, we observe a density supported on a compact interval $[a, b]$. This section first explains how such density-valued time series are constructed in practice, and then proposes a new density-valued VAR model with latent factors by extending the interactive fixed-effects framework of [Bai \(2009\)](#) to density-valued time series. Moreover, we develop estimation and asymptotic theory for the resulting estimators.

3.1 Density Estimation by B-spline Mixtures

In practice, density-valued time series are observed through samples drawn from an underlying density. We begin by describing a general procedure for estimating a density at each time point and for each unit. Suppose that we observe independent samples x_1, \dots, x_n from an unknown density f supported on a compact interval $[a, b]$. We approximate f by a finite mixture of normalized B-spline basis functions. Let $\{B_j(\cdot)\}_{j=1}^{J+1}$ be B-splines of degree k defined on a knot sequence covering $[a, b]$. Normalizing each basis yields the density components

$$\phi_j(x) = \frac{B_j(x)}{\int_a^b B_j(u) du}, \quad j = 1, \dots, J + 1.$$

The unknown density is approximated by

$$f(x; w) = \sum_{j=1}^{J+1} w_j \phi_j(x), \quad w \in \mathcal{S}_{J+1} := \left\{ w_j > 0, \sum_{j=1}^{J+1} w_j = 1 \right\}, \quad (1)$$

where $w = (w_1, \dots, w_{J+1})'$ is the mixture-weight vector. The log-likelihood is

$$\ell(w) = \sum_{i=1}^n \log \left(\sum_{j=1}^{J+1} w_j \phi_j(x_i) \right),$$

which we maximize using the EM algorithm ([Dempster et al., 1977](#); [McLachlan and Peel, 2000](#)). In the E-step, we compute the posterior responsibilities

$$\tau_{i,j} = \frac{w_j \phi_j(x_i)}{\sum_{\ell=1}^{J+1} w_\ell \phi_\ell(x_i)}, \quad i = 1, \dots, n, \quad j = 1, \dots, J + 1,$$

which represent the conditional probabilities that observation x_i is associated with the j th spline component. In the M-step, we update the mixture weights by

$$w_j = \frac{1}{n} \sum_{i=1}^n \tau_{i,j}, \quad j = 1, \dots, J+1.$$

Starting from an initial value for w , we iterate the E-step and M-step until convergence.

To stabilize estimation in small samples, we also consider a Dirichlet-regularized version of the procedure. Specifically, suppose that

$$w \sim \text{Dir}(\alpha_1^{(0)}, \dots, \alpha_{J+1}^{(0)}).$$

Then the posterior parameters are

$$\alpha_j^{\text{post}} = \gamma \alpha_j^{(0)} + \sum_{i=1}^n \tau_{i,j}, \quad j = 1, \dots, J+1, \quad (2)$$

where γ controls the prior strength. The corresponding posterior mean is

$$w_j = \frac{\alpha_j^{\text{post}}}{\sum_{\ell=1}^{J+1} \alpha_\ell^{\text{post}}}, \quad j = 1, \dots, J+1. \quad (3)$$

The E-step remains unchanged, while the M-step is replaced by the posterior-mean update in (2) and (3). Starting from an initial value for w , we iterate the modified E-step and M-step until convergence.

Applying this procedure separately to each time t and unit c , we obtain a sequence of weight vectors $w_{t,c} \in \mathcal{S}_{J+1}$. These vectors serve as the observed density-valued time series used in the subsequent analysis. In the Ct-value application, the units correspond to regions.

3.2 Generalized logit and softmax transforms

Let $w_{t,c} \in \mathcal{S}_{J+1}$ denote the estimated B-spline mixture weights for unit $c = 1, \dots, C$ at time t . To analyze density-valued time series using linear time-series methods, we map these simplex-valued weights into a finite-dimensional linear space by the generalized logit transform of [Matsuda and Iwafuchi \(2025\)](#). This transformation is central to our approach, because it provides an explicit coordinate representation of density-valued observations while retaining a one-to-one correspondence with the underlying spline-mixture representation.

For $p = (p_1, \dots, p_{J+1})' \in \mathcal{S}_{J+1}$ with base component p_{J+1} , define, for $\delta > 0$,

$$\text{logit}_\delta(p) = \left(\log \frac{\delta + p_1}{\delta + p_{J+1}}, \dots, \log \frac{\delta + p_J}{\delta + p_{J+1}} \right)' \in \mathbb{R}^J.$$

When $\delta = 0$, logit_δ reduces to the usual multinomial logit transform from \mathcal{S}_{J+1} to \mathbb{R}^J . For $\delta > 0$, the shift by δ improves numerical stability when some components of p are close to zero. It is convenient to extend the domain to

$$\mathcal{S}_{J+1}^\delta = \left\{ p \in \mathbb{R}^{J+1} : p_j > -\delta, \sum_{j=1}^{J+1} p_j = 1 \right\},$$

which strictly contains the probability simplex \mathcal{S}_{J+1} .

The inverse map of logit_δ is given by, for $b \in \mathbb{R}^J$,

$$\text{softmax}_\delta(b) = \left(\frac{((J+1)\delta+1)e^{b_1}}{1+\sum_{j=1}^J e^{b_j}} - \delta, \dots, \frac{((J+1)\delta+1)e^{b_J}}{1+\sum_{j=1}^J e^{b_j}} - \delta, \frac{(J+1)\delta+1}{1+\sum_{j=1}^J e^{b_j}} - \delta \right)' \in \mathcal{S}_{J+1}^\delta.$$

When $\delta = 0$, softmax_δ reduces to the standard softmax map to \mathcal{S}_{J+1} . It is straightforward to verify that $\text{softmax}_\delta(b) \in \mathcal{S}_{J+1}^\delta$, since each component is strictly larger than $-\delta$ and

$$\sum_{m=1}^{J+1} (\text{softmax}_\delta(b))_m = \frac{((J+1)\delta+1) \left(1 + \sum_{j=1}^J e^{b_j}\right)}{1 + \sum_{j=1}^J e^{b_j}} - (J+1)\delta = 1.$$

Moreover,

$$\text{softmax}_\delta(\text{logit}_\delta(p)) = p \quad \text{for } p \in \mathcal{S}_{J+1}^\delta, \quad \text{logit}_\delta(\text{softmax}_\delta(b)) = b \quad \text{for } b \in \mathbb{R}^J.$$

Hence, logit_δ and softmax_δ define a bijection between \mathcal{S}_{J+1}^δ and \mathbb{R}^J .

Using this bijection, we transport the usual addition and scalar multiplication in \mathbb{R}^J to \mathcal{S}_{J+1}^δ . Specifically, for $w_1, w_2 \in \mathcal{S}_{J+1}^\delta$ and $\alpha \in \mathbb{R}$, define

$$\begin{aligned} w_1 \oplus w_2 &:= \text{softmax}_\delta(\text{logit}_\delta(w_1) + \text{logit}_\delta(w_2)), \\ \alpha \otimes w_1 &:= \text{softmax}_\delta(\alpha \text{logit}_\delta(w_1)). \end{aligned}$$

Thus, via the generalized logit transform, we may identify \mathcal{S}_{J+1}^δ with \mathbb{R}^J and equip it with a linear structure.

Next we introduce an inner product on the transformed space so that the coordinate representation is isometric to its image in the associated spline-mixture function space under the $L^2[a, b]$ inner product. Let $e_i \in \mathbb{R}^J$, $i = 1, \dots, J$, be the canonical unit vector, and define

$$\hat{e}_i = \text{softmax}_\delta(e_i) \in \mathcal{S}_{J+1}^\delta.$$

We identify any weight vector $p = (p_1, \dots, p_{J+1})' \in \mathcal{S}_{J+1}^\delta$ with the associated spline-mixture function

$$\Psi(p)(s) := \sum_{m=1}^{J+1} p_m \phi_m(s), \quad s \in [a, b].$$

For brevity, write $\hat{e}_i(s) := \Psi(\hat{e}_i)(s)$. Define the Gram matrix

$$(H_J)_{ij} = \int_a^b \hat{e}_i(s) \hat{e}_j(s) ds, \quad i, j = 1, \dots, J.$$

This induces the inner product

$$\langle u, v \rangle_H = u' H_J v, \quad u, v \in \mathbb{R}^J.$$

By construction, this is the pullback of the $L^2[a, b]$ inner product of the associated spline-mixture functions through the map $b \mapsto \Psi(\text{softmax}_\delta(b))$. Hence the transformed coordinate space is isometric to its image in the spline-mixture function space.

The introduction of $\delta > 0$ improves numerical stability when some mixture weights are close to, or equal to, zero. At the same time, because \mathcal{S}_{J+1}^δ strictly contains the probability simplex, the associated spline-mixture functions may take negative values for some choices of transformed

coordinates. This is a potentially controversial feature. In our framework, however, \mathcal{S}_{J+1} remains the empirical starting point, and the enlargement to \mathcal{S}_{J+1}^δ is used as a local device to obtain a stable bijection and linear structure for statistical modeling. The gain is computational and inferential tractability, while interpretation remains anchored in the original density representation through the inverse transformation.

We represent the density-valued time series for unit c by the transformed series

$$Y_{t,c} = \text{logit}_\delta(w_{t,c}) \in \mathbb{R}^J.$$

Stacking across units, we define

$$Y_t = (Y'_{t,1}, \dots, Y'_{t,C})' \in \mathbb{R}^{CJ}. \quad (4)$$

For the stacked vectors $Y_t \in \mathbb{R}^{CJ}$, we introduce the block-diagonal metric

$$\mathcal{H} = I_C \otimes H_J. \quad (5)$$

In the Ct-value application, the units correspond to regions.

3.3 Density-valued VAR models with latent factors

3.3.1 Model description

Suppose we observe the transformed density-valued time series Y_t defined in (4). Recall that each $Y_{t,c} \in \mathbb{R}^J$ is obtained by applying the generalized logit map to the spline-mixture weights of the density at time t and unit c , and that this coordinate space is equipped with an inner product making it isometric to the corresponding spline-mixture image under the L^2 geometry. A natural starting point is the density-valued VAR model without latent factors,

$$Y_{t,c} = \sum_{k=1}^p \sum_{d=1}^C V_{k,cd} Y_{t-k,d} + u_{t,c}, \quad c = 1, \dots, C, \quad (6)$$

where $V_k \in \mathbb{R}^{C \times C}$ captures cross-regional autoregressive dependence and $u_{t,c} \in \mathbb{R}^J$ is a disturbance term. This specification directly represents lagged dependence across regions in the transformed density-valued series.

However, in the present setting one expects strong contemporaneous comovement across regions, driven for example by nationwide epidemic waves or other aggregate shocks. If such common movements remain in the disturbance term $u_{t,c}$, then a direct VAR fit may attribute part of this pervasive comovement to cross-regional lagged dependence, making the resulting network difficult to interpret. One possible response is to impose a structural VAR decomposition on the contemporaneous dependence (Lütkepohl, 2005; Kilian and Lütkepohl, 2017). In practice, however, such an approach requires additional identifying restrictions, and the resulting analysis may depend on assumptions about contemporaneous ordering or other structural constraints.

To separate common movements from directed idiosyncratic dynamics without imposing such structural restrictions, we augment (6) with latent factors and consider the following density-valued VAR model with latent factors:

$$Y_{t,c} = \sum_{k=1}^p \sum_{d=1}^C V_{k,cd} Y_{t-k,d} + \Lambda_c f_t + \varepsilon_{t,c}, \quad c = 1, \dots, C. \quad (7)$$

Here, $V_k \in \mathbb{R}^{C \times C}$ captures cross-regional autoregressive dependence, $\Lambda_c \in \mathbb{R}^{J \times r}$ is the loading matrix for unit c , and $f_t \in \mathbb{R}^r$ denotes the vector of latent common factors. Thus, the model decomposes the transformed density-valued series into a common factor component and an idiosyncratic component with cross-regional VAR dependence.

Defining

$$\Lambda = (\Lambda'_1, \dots, \Lambda'_C)' \in \mathbb{R}^{CJ \times r},$$

the model can be written in matrix form as

$$Y_t = \sum_{k=1}^p X_{t-k} \text{Vec}(V'_k) + \Lambda f_t + \varepsilon_t, \quad (8)$$

where

$$X_t = I_C \otimes (Y_{t,1}, \dots, Y_{t,C}).$$

From this matrix representation, (7) can be viewed as a density-valued dynamic extension of the interactive fixed-effects framework of Bai (2009). Relative to standard factor models, it allows density-valued observations together with cross-regional VAR dependence in the idiosyncratic component.

3.3.2 Estimation

We estimate V_k , Λ , and f_t in (7) under the normalization

$$\begin{aligned} (CJ)^{-1} \Lambda' \mathcal{H} \Lambda &= I_r, \\ \frac{1}{T-p} \sum_{t=p+1}^T f_t f_t' &= \text{diag}(v_1, \dots, v_r). \end{aligned}$$

For notational simplicity, we assume that Y_t has zero mean. For nonzero-mean processes, we replace Y_t with

$$Y_t - \frac{1}{T} \sum_{s=1}^T Y_s.$$

Our estimation procedure alternates between PCA estimation of the latent factor component and projected least-squares estimation of the VAR component. Given current VAR coefficients, we extract the factor structure from the residual covariance matrix; given the resulting loading space, we partial out the factor component and update the VAR coefficients by least squares. This iterative scheme parallels Bai (2009), adapted here to the whitened transformed representation of density-valued observations.

Let $\mathcal{K} = I_C \otimes K_J$, where $K_J K_J' = H_J$. Define

$$\tilde{Y}_t = \mathcal{K}' Y_t, \quad \tilde{\Lambda} = \mathcal{K}' \Lambda, \quad \tilde{X}_t = \mathcal{K}' X_t, \quad \tilde{\varepsilon}_t = \mathcal{K}' \varepsilon_t.$$

Then the model becomes

$$\tilde{Y}_t = \sum_{k=1}^p \tilde{X}_{t-k} \text{Vec}(V'_k) + \tilde{\Lambda} f_t + \tilde{\varepsilon}_t,$$

which rewrites the model in a coordinate space equipped with the standard inner product.

Step 1: Residualizing the VAR component. Given current values of V_k , compute

$$\tilde{R}_t = \tilde{Y}_t - \sum_{k=1}^p \tilde{X}_{t-k} \text{Vec}(V_k').$$

Step 2: PCA for the factor component. Using these residuals, update the loading space by PCA. Form the covariance matrix

$$\tilde{S} = \frac{1}{T-p} \sum_{t=p+1}^T \tilde{R}_t \tilde{R}_t'.$$

Let $\hat{\Lambda} \in \mathbb{R}^{CJ \times r}$ collect the top r eigenvectors of \tilde{S} , normalized such that

$$(CJ)^{-1} \hat{\Lambda}' \hat{\Lambda} = I_r.$$

Step 3: FWL estimation of V_k . Define the projection matrix

$$M_{\hat{\Lambda}} = I_{CJ} - (CJ)^{-1} \hat{\Lambda}' \hat{\Lambda}.$$

The projected model is

$$M_{\hat{\Lambda}} \tilde{Y}_t = \sum_{k=1}^p M_{\hat{\Lambda}} \tilde{X}_{t-k} \text{Vec}(V_k') + M_{\hat{\Lambda}} \tilde{\varepsilon}_t, \quad t = p+1, \dots, T.$$

Let $\tilde{W}_t = (\tilde{X}_{t-1}, \dots, \tilde{X}_{t-p})$. Then

$$\begin{pmatrix} \text{vec}(\hat{V}_1') \\ \vdots \\ \text{vec}(\hat{V}_p') \end{pmatrix} = \left(\sum_{t=p+1}^T \tilde{W}_t' M_{\hat{\Lambda}} \tilde{W}_t \right)^{-1} \left(\sum_{t=p+1}^T \tilde{W}_t' M_{\hat{\Lambda}} \tilde{Y}_t \right).$$

Step 4: Iteration and factor recovery. Repeat Steps 1–3 until convergence of V_k , $k = 1, \dots, p$. After convergence, recover the factors by projecting the residual onto the estimated loading space, and map the loadings back to the original transformed coordinates:

$$\begin{aligned} \hat{f}_t &= (CJ)^{-1} \hat{\Lambda}' \left(\tilde{Y}_t - \sum_{k=1}^p \tilde{X}_{t-k} \text{Vec}(\hat{V}_k') \right), \\ \hat{\Lambda} &= (\mathcal{K}')^{-1} \hat{\Lambda}. \end{aligned}$$

3.4 Asymptotic Theory

We now study the large-sample properties of the estimator for the factor-adjusted VAR component. Starting from the model in (8), we work with its whitened representation, because the transformed coordinate space is equipped with the block-diagonal inner product matrix $\mathcal{H} = I_C \otimes H_J$, rather than the standard Euclidean one. Let $\mathcal{K} = I_C \otimes K_J$, where $K_J K_J' = H_J$, and define

$$\tilde{Y}_t = \mathcal{K}' Y_t, \quad \tilde{\Lambda} = \mathcal{K}' \Lambda, \quad \tilde{X}_t = \mathcal{K}' X_t, \quad \tilde{\varepsilon}_t = \mathcal{K}' \varepsilon_t.$$

Then the model can be rewritten as

$$\tilde{Y}_t = \tilde{W}_t \beta + \tilde{\Lambda} f_t + \tilde{\varepsilon}_t, \quad (9)$$

where $\tilde{W}_t = (\tilde{X}_{t-1}, \dots, \tilde{X}_{t-p})$, and

$$\beta = (\text{Vec}'(V_1'), \dots, \text{Vec}'(V_p'))'.$$

We derive a rate of convergence for the estimator of β and, under stronger conditions, its asymptotic normality under the regime

$$J, T \rightarrow \infty \quad \text{with } C \text{ fixed.}$$

Unlike Bai (2009), our regressors are generated by lagged dependent variables through the VAR structure, so the model does not satisfy the same exogeneity interpretation as a standard interactive fixed-effects regression. In particular, although the algebraic structure of the proof follows the Bai-type template, the dynamic nature of the regressors implies that certain weighted score terms involve internally generated regressors rather than externally given covariates. For this reason, the asymptotic theory requires additional conditions that are specific to the present factor-adjusted VAR setting.

A further difference from Bai (2009) is that the asymptotic roles of N and T_0 are reversed in our setting, because the PCA step is carried out along the time direction rather than the cross-sectional direction. As a result, the ratio conditions in the asymptotic theory are stated in terms of N/T_0 , rather than T_0/N . Throughout this subsection, a superscript 0 denotes the true value of a parameter.

An important feature of the proposed framework is that stationarity is required only for the idiosyncratic VAR component, not for the observed series as a whole. The common factor component may exhibit nonstationary behavior, including trend-like movements, provided that the empirical second-moment condition in Assumption B(i) holds. Thus, the model can accommodate multivariate density-valued time series with pervasive nonstationary common movements, while still allowing inference on directed dependence in the stable idiosyncratic component.

Because PCA is carried out along the time direction in our setting, the resulting projection and expansion arguments generate weighted score terms indexed by time averages of lagged regressors. In the Bai setting, the corresponding terms can be handled by exogeneity of the regressors, but that argument is no longer available here because the regressors are internally generated by the VAR dynamics. For this reason, Assumption C(vi) is introduced to control such weighted score terms, while Assumption E provides the additional limit and central limit conditions needed for asymptotic normality. These conditions are stronger and more model-specific than those used in the standard Bai setting, reflecting the fact that the present framework combines latent common factors with an idiosyncratic VAR structure rather than an exogenous regression design.

Regularity conditions.

Assumption A (Regression design and identification).

- (i) For the true $C \times C$ VAR coefficient matrices V_i^0 , $i = 1, \dots, p$, there exists a constant $\kappa \in (0, 1)$, independent of J and T , such that

$$\sum_{i=1}^p \|V_i^0\| \leq \kappa,$$

where $\|\cdot\|$ denotes the operator norm.

(ii) Let

$$\mathcal{F} = \{\Lambda \in \mathbb{R}^{CJ \times r} : (CJ)^{-1} \Lambda' \Lambda = I_r\}.$$

Then

$$\inf_{\tilde{\Lambda} \in \mathcal{F}} \lambda_{\min}(D(\tilde{\Lambda})) \geq c > 0,$$

where $\lambda_{\min}(D)$ denotes the minimum eigenvalue of D , and

$$\begin{aligned} D(\tilde{\Lambda}) &= \frac{1}{CJ(T-p)} \sum_{t=p+1}^T Z_t(\tilde{\Lambda})' Z_t(\tilde{\Lambda}), \\ Z_t(\tilde{\Lambda}) &= M_{\tilde{\Lambda}} \tilde{W}_t - \frac{1}{T-p} \sum_{s=p+1}^T a_{ts} M_{\tilde{\Lambda}} \tilde{W}_s, \\ a_{ts} &= f_t^{0'} \left(\frac{1}{T-p} F_0' F_0 \right)^{-1} f_s^0, \end{aligned}$$

with

$$M_{\tilde{\Lambda}} = I_{CJ} - (CJ)^{-1} \tilde{\Lambda} \tilde{\Lambda}', \quad F_0 = (f_{p+1}^0, \dots, f_T^0)'$$

Assumption B (Factors and loadings). Let r be fixed. Treat the common factors $\{f_t^0\}_{t=p+1}^T$ and the loading matrix

$$\tilde{\Lambda}_0 = (\tilde{\lambda}_{01}, \dots, \tilde{\lambda}_{0N})' \in \mathbb{R}^{N \times r}, \quad N = CJ,$$

as non-stochastic arrays.

(i) **Factor second-moment limit.**

$$\frac{1}{T_0} \sum_{t=p+1}^T f_t^0 f_t^{0'} \rightarrow \Sigma_f, \quad T_0 := T - p,$$

where Σ_f is positive definite. Moreover,

$$\sup_{p+1 \leq t \leq T} \|f_t^0\| \leq K.$$

(ii) **Loading second-moment limit.**

$$\frac{1}{N} \tilde{\Lambda}_0' \tilde{\Lambda}_0 \rightarrow \Sigma_\Lambda,$$

where Σ_Λ is positive definite. Moreover,

$$\sup_{1 \leq i \leq N} \|\tilde{\lambda}_{0i}\| \leq K.$$

Remark. Assumption B does not require the factor path to be stationary in the usual time-series sense. It only requires the empirical second-moment limit

$$\frac{1}{T_0} \sum_{t=p+1}^T f_t^0 f_t^{0'} \rightarrow \Sigma_f$$

to exist. In particular, under an infill interpretation, trend-like common components are allowed as long as this empirical second-moment limit exists.

Assumption C (Moments and dependence). Let $\{\tilde{\varepsilon}_t\}_{t=p+1}^T$ be the idiosyncratic error sequence in

$$\tilde{Y}_t = \tilde{W}_t \beta_0 + \tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t.$$

Assume:

(i) **Moments and scale.** There exists $\delta > 0$ such that

$$\sup_t \max_{1 \leq i \leq N} E|\tilde{\varepsilon}_{it}|^{4+\delta} \leq K < \infty.$$

(ii) **No serial correlation.** The sequence $\{\tilde{\varepsilon}_t\}$ is independent across t , and

$$E(\tilde{\varepsilon}_t) = 0, \quad E(\tilde{\varepsilon}_t \tilde{\varepsilon}_t') = \Sigma_{\varepsilon,t}.$$

(iii) **Uniformly bounded cross-sectional covariance.**

$$\sup_t \|\Sigma_{\varepsilon,t}\| \leq K < \infty.$$

(iv) **Weak cross-sectional dependence.** For each t ,

$$\max_{1 \leq i \leq N} \sum_{j=1}^N |\Sigma_{\varepsilon,t}(i, j)| \leq K < \infty.$$

(v) **Linear-form moment bound.** For any nonstochastic unit vectors $u, v \in \mathbb{R}^N$,

$$\sup_t E|u' \tilde{\varepsilon}_t|^{4+\delta} \leq K, \quad \sup_t E|(u' \tilde{\varepsilon}_t)(v' \tilde{\varepsilon}_t)|^2 \leq K.$$

(vi) **Uniform covariance summability for weighted score terms.** There exists $K < \infty$ such that, for any deterministic uniformly bounded array $\{c_{ts}\}_{t,s=p+1}^T$,

$$\frac{1}{NT_0} \sum_{t,u=p+1}^T \left| E \left[\left(\sum_{s=p+1}^T c_{ts} \tilde{W}_s' \tilde{\varepsilon}_t \right) \left(\sum_{v=p+1}^T c_{uv} \tilde{W}_v' \tilde{\varepsilon}_u \right) \right] \right| \leq K.$$

Remark. Assumption C(vi) is imposed only to control the weighted score term

$$\frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T c_{ts} \tilde{W}_s' \right\} \tilde{\varepsilon}_t,$$

which appears in the proof of Theorem 3.1. The difficulty is that the weighted regressor

$$b_t := \frac{1}{T_0} \sum_{s=p+1}^T c_{ts} \tilde{W}_s$$

contains both past and future values of \tilde{W}_s , so it is not measurable with respect to the information set at time $t - 1$. Therefore, the mean-independence condition in Assumption D is not sufficient by itself to control the second moment of $b_t' \tilde{\varepsilon}_t$. Assumption C(vi) is introduced as a convenient sufficient condition ensuring uniform summability of the covariance of such weighted score terms. It is stronger than what may be minimally necessary, but it avoids the need to impose more elaborate dependence conditions. This is precisely the point at which the present time-direction PCA argument differs from the standard Bai setting with exogenous regressors, where the corresponding weighted score term can be handled by exogeneity.

Assumption D (Sequential exogeneity). For each $t = p + 1, \dots, T$,

$$E(\tilde{\varepsilon}_t \mid \{\tilde{Y}_s : s < t\}) = 0.$$

Assumption E (Probability limits and a score CLT). Assumption E collects the additional probability-limit and central-limit conditions needed for asymptotic normality once the weighted score terms have been controlled. For the asymptotic normality result in Theorem 3.2, assume additionally that $N/T_0 \rightarrow 0$. Let $N := CJ$ and $T_0 := T - p$, and let $Z_t(\tilde{\Lambda})$ be defined in Assumption A(ii). Define

$$H_0 := \frac{1}{N} \tilde{\Lambda}'_0 \tilde{\Lambda}_0, \quad \bar{\Lambda}_0 := \tilde{\Lambda}_0 H_0^{-1/2},$$

so that $\bar{\Lambda}_0 \in \mathcal{F}$ and

$$P_{\bar{\Lambda}_0} = \frac{1}{N} \bar{\Lambda}_0 \bar{\Lambda}'_0 = \tilde{\Lambda}_0 (\tilde{\Lambda}'_0 \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}'_0 = \Pi_0.$$

Write

$$Z_t^0 := Z_t(\bar{\Lambda}_0), \quad \Sigma_{\varepsilon,t} := E(\tilde{\varepsilon}_t \tilde{\varepsilon}'_t).$$

Assume:

- (i) **Probability limits of $D(\bar{\Lambda}_0)$ and $\Omega(\bar{\Lambda}_0)$.** The following probability limits exist:

$$D(\bar{\Lambda}_0) := \frac{1}{NT_0} \sum_{t=p+1}^T Z_t^{0'} Z_t^0 \xrightarrow{p} D_0, \quad \Omega(\bar{\Lambda}_0) := \frac{1}{NT_0} \sum_{t=p+1}^T Z_t^{0'} \Sigma_{\varepsilon,t} Z_t^0 \xrightarrow{p} \Omega_0,$$

where D_0 is positive definite.

- (ii) **Central limit theorem for the score.**

$$\frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T Z_t^{0'} \tilde{\varepsilon}_t \xrightarrow{d} \mathcal{N}(0, \Omega_0).$$

We first derive a convergence rate under $N/T_0 \rightarrow \rho \geq 0$, and then establish asymptotic normality under the stronger condition $N/T_0 \rightarrow 0$.

Theorem 3.1 (Rate of convergence of $\hat{\beta}$). *Consider the model in (9),*

$$\tilde{Y}_t = \tilde{W}_t \beta_0 + \tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t, \quad t = p + 1, \dots, T,$$

and let $(\hat{\beta}, \hat{\Lambda})$ minimize

$$Q(\beta, \tilde{\Lambda}) = \sum_{t=p+1}^T (\tilde{Y}_t - \tilde{W}_t \beta)' M_{\tilde{\Lambda}} (\tilde{Y}_t - \tilde{W}_t \beta)$$

over $(\beta, \tilde{\Lambda}) \in \mathbb{R}^{pC^2} \times \mathcal{F}$, where

$$\mathcal{F} = \{\Lambda \in \mathbb{R}^{N \times r} : N^{-1} \Lambda' \Lambda = I_r\}, \quad M_{\tilde{\Lambda}} = I_N - N^{-1} \tilde{\Lambda} \tilde{\Lambda}', \quad N := CJ, \quad T_0 := T - p.$$

Suppose Assumptions A–D hold, with p , r , and C fixed. If $N \rightarrow \infty$, $T_0 \rightarrow \infty$, and

$$\frac{N}{T_0} \rightarrow \rho \geq 0,$$

then

$$\sqrt{NT_0} (\hat{\beta} - \beta_0) = O_p(1).$$

Theorem 3.2 (Asymptotic normality of $\hat{\beta}$). *Let $N := CJ$ and $T_0 := T - p$, and suppose Assumptions A–E hold, with r and C fixed. Assume additionally that*

$$\frac{N}{T_0} \rightarrow 0.$$

Then

$$\sqrt{NT_0}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, D_0^{-1}\Omega_0 D_0^{-1}).$$

Theorem 3.1 establishes a convergence rate, and hence consistency of the estimator, while Theorem 3.2 provides its asymptotic normality under the stronger condition $N/T_0 \rightarrow 0$. The main difficulty in the asymptotic analysis is Proposition 7.9, which derives a linear expansion for $\hat{\beta}$ in the presence of latent factors and internally generated VAR regressors. Although the overall proof follows the Bai-type projection-and-expansion strategy, the present setting is not a routine extension of Bai (2009). Because the regressors are lagged dependent variables and PCA is carried out along the time direction, the expansion produces a weighted score term involving time averages of lagged regressors. This term cannot be handled by the standard exogeneity argument used in the Bai setting. Assumption C(vi) is introduced precisely to control this weighted score contribution. The resulting expansion contains a leading score term and a bias term. The bias term is of order $O_p(\sqrt{N/T_0})$, so it affects the rate result in Theorem 3.1 when $N/T_0 \rightarrow \rho \geq 0$, but becomes asymptotically negligible under the stronger condition $N/T_0 \rightarrow 0$, which yields the asymptotic normality result in Theorem 3.2. Full proofs and auxiliary technical lemmas are provided in the online Appendix.

Remark on feasible inference. Theorem 3.2 implies that

$$\widehat{\text{Var}}(\hat{\beta}) = \frac{1}{NT_0} \hat{D}^{-1} \hat{\Omega} \hat{D}^{-1},$$

so coefficient-wise inference may be based on the corresponding standard errors, where

$$\hat{D} := \frac{1}{NT_0} \sum_{t=p+1}^T \hat{Z}'_t \hat{Z}_t, \quad \hat{Z}_t := Z_t(\hat{\Lambda}),$$

and

$$\hat{\Omega} := \frac{1}{NT_0} \sum_{t=p+1}^T \hat{Z}'_t \hat{\Sigma}_{\varepsilon,t} \hat{Z}_t.$$

Here \hat{Z}_t denotes the feasible analogue of $Z_t(\bar{\Lambda}_0)$, obtained by replacing $\bar{\Lambda}_0$ with $\hat{\Lambda}$ and a_{ts} with

$$\hat{a}_{ts} := \hat{f}'_t \left(\frac{1}{T_0} \hat{F}' \hat{F} \right)^{-1} \hat{f}_s, \quad \hat{F} := (\hat{f}_{p+1}, \dots, \hat{f}_T)'$$

$\hat{\Sigma}_{\varepsilon,t}$ denotes a feasible estimator of the contemporaneous covariance matrix of the idiosyncratic error. In finite samples, a fully unrestricted estimator of $\hat{\Omega}$ may be singular or numerically unstable when T_0 is not sufficiently large relative to the dimension of the projected score vectors. In the simulation and empirical analyses, we therefore estimate $\hat{\Sigma}_{\varepsilon,t}$ by a block-diagonal matrix with $J \times J$ blocks, one for each region. This block-diagonal approximation yields a stable and nonsingular feasible covariance estimator for coefficient-wise t -tests.

4 Simulation study

This section examines whether increasing the number of estimated factors mechanically eliminates the idiosyncratic VAR network in our procedure. This issue is important for interpreting the Ct-value application, because the detected network may weaken or disappear as the factor dimension is varied. The simulation study shows that, when genuine idiosyncratic VAR dependence is present, increasing the number of estimated factors does not mechanically erase the directed network. Rather, it mainly reduces spuriously detected edges induced by common factor movements and sharpens the remaining edge set.

4.1 Data-generating process

We generate data from a factor-augmented density-valued VAR model of the form

$$Y_t = (V_{\text{true}} \otimes I_J)Y_{t-1} + Lf_t + \varepsilon_t, \quad (10)$$

where $Y_t \in \mathbb{R}^{CJ}$, V_{true} is a $C \times C$ coefficient matrix, I_J is the $J \times J$ identity matrix, L is the loading matrix, and f_t is an r^* -dimensional factor process.

In the simulation, we set $T = 114$, $C = 20$, and $J = 15$, so that Y_t has dimension $CJ = 300$. The true number of common factors is fixed at $r^* = 5$. The factor process follows

$$f_t = Af_{t-1} + u_t, \quad (11)$$

where

$$A = \text{diag}(0, 0, 0.9, 0.9, 0.9),$$

and u_t is Gaussian with independent components and standard deviations

$$(1.0, 1.0, 0.3, 0.3, 0.3).$$

Thus, the first two factors are white noise, whereas the remaining factors are persistent.

To construct the loading matrix, let $Q = (q_1, \dots, q_{r^*})$ be a $C \times r^*$ orthonormal matrix obtained by applying a QR decomposition to a Gaussian random matrix, and let $U = (u_1, \dots, u_{r^*})$ be a $J \times r^*$ matrix whose columns are independently generated Gaussian vectors normalized to unit length. The loading matrix $L = (\ell_1, \dots, \ell_{r^*}) \in \mathbb{R}^{CJ \times r^*}$ is defined by

$$\ell_1 = (\mathbf{1}_C / \sqrt{C}) \otimes u_1, \quad \ell_k = q_k \otimes u_k, \quad k = 2, \dots, r^*,$$

where $\mathbf{1}_C$ denotes the C -vector of ones. Hence, the first factor has a global loading shared by all units, whereas the remaining factors generate structured cross-sectional heterogeneity.

The true VAR matrix V_{true} is constructed from

$$M = Q \text{diag}(1.00, 0.90, 0.80, 0.70, 0.60) Q'.$$

We restrict attention to off-diagonal entries in rows indexed by the source set $\{1, 2, 5\}$, rank these candidates by $|M_{ij}|$, and select the top 30 entries. For each selected pair (i, j) , we set

$$(V_{\text{true}})_{ij} = \alpha_V |M_{ij}|,$$

and set all remaining entries to zero. If the spectral radius of V_{true} exceeds 0.95, we rescale the entire matrix so that its spectral radius becomes 0.95. We consider three levels of VAR strength,

$\alpha_V \in \{0, 0.5, 1.0\}$. When $\alpha_V = 0$, there is no idiosyncratic VAR dependence. When $\alpha_V > 0$, the data contain genuine directed edges in the idiosyncratic VAR component.

The idiosyncratic disturbance is Gaussian,

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2 I_{CJ}), \quad \sigma_\varepsilon = 0.1.$$

For each design, we generate 100 Monte Carlo replications.

4.2 Estimation, edge selection, and results

For each simulated sample, we estimate the model using exactly the same procedure as in Section 3.3. We vary the number of estimated factors over

$$r = 0, 1, \dots, 8,$$

so that the estimated factor dimension may be smaller or larger than the true value $r^* = 5$.

When $r = 0$, we estimate a VAR without factors and compute block-diagonal robust t -statistics. More precisely, the covariance matrix of the error term is assumed to be block-diagonal across units, with each block given by a $J \times J$ covariance matrix.

When $r \geq 1$, we iteratively estimate the VAR coefficient matrix and the factor loading matrix by alternating between factor extraction from the residuals and projected least-squares estimation of the VAR component. Given the final estimator, we compute the robust t -statistics for the VAR coefficients based on the asymptotic covariance formula in Section 3.4. The asymptotic covariance matrix is estimated by

$$(NT_0)^{-1} \hat{D}^{-1} \hat{\Omega} \hat{D}^{-1},$$

where \hat{D} and $\hat{\Omega}$ are sample analogues of the corresponding population quantities in Assumption E. In estimating $\hat{\Omega}$, we assume that $\Sigma_{\varepsilon,t}$ has a block-diagonal structure,

$$\text{diag}(\varepsilon_{t,1} \varepsilon'_{t,1}, \dots, \varepsilon_{t,C} \varepsilon'_{t,C}),$$

where $\varepsilon_{t,c}$ denotes the J -dimensional idiosyncratic error vector for unit c at time t .

Directed edges are selected from positive off-diagonal entries of the estimated VAR matrix using one-sided t -tests with the Benjamini–Yekutieli false discovery rate control at level 10% (Benjamini and Yekutieli, 2001). For each replication and each value of r , we record the number of detected edges, the recall rate relative to the true edge set, the empirical false discovery proportion (FDP), and the precision. These four summaries allow us to assess whether increasing the estimated factor dimension mainly removes spurious edges induced by common factor movements, or instead erases genuine idiosyncratic VAR dependence.

Figure 1 reports four summaries of edge-selection performance as functions of the estimated number of factors r : the mean number of detected edges, the mean recall, the mean empirical FDP, and the mean precision. The results reveal a clear transition from many but noisy detected edges at small r to fewer but more credible edges once sufficiently many factors are included.

When $\alpha_V = 0$, recall is identically zero by construction, so any detected edge is spurious. In this case, the mean number of detected edges is large for small values of r , peaking around $r = 2$, while the empirical FDP is essentially one. Once r reaches about the true factor dimension, however, the mean number of detected edges falls to nearly zero. Thus, when the idiosyncratic VAR component is absent, apparent directed edges generated by common factor movements largely disappear after sufficient factor adjustment.

When $\alpha_V = 0.5$, the same reduction in the number of detected edges is accompanied not by the disappearance of the true signal, but by fewer false discoveries and higher precision. As r

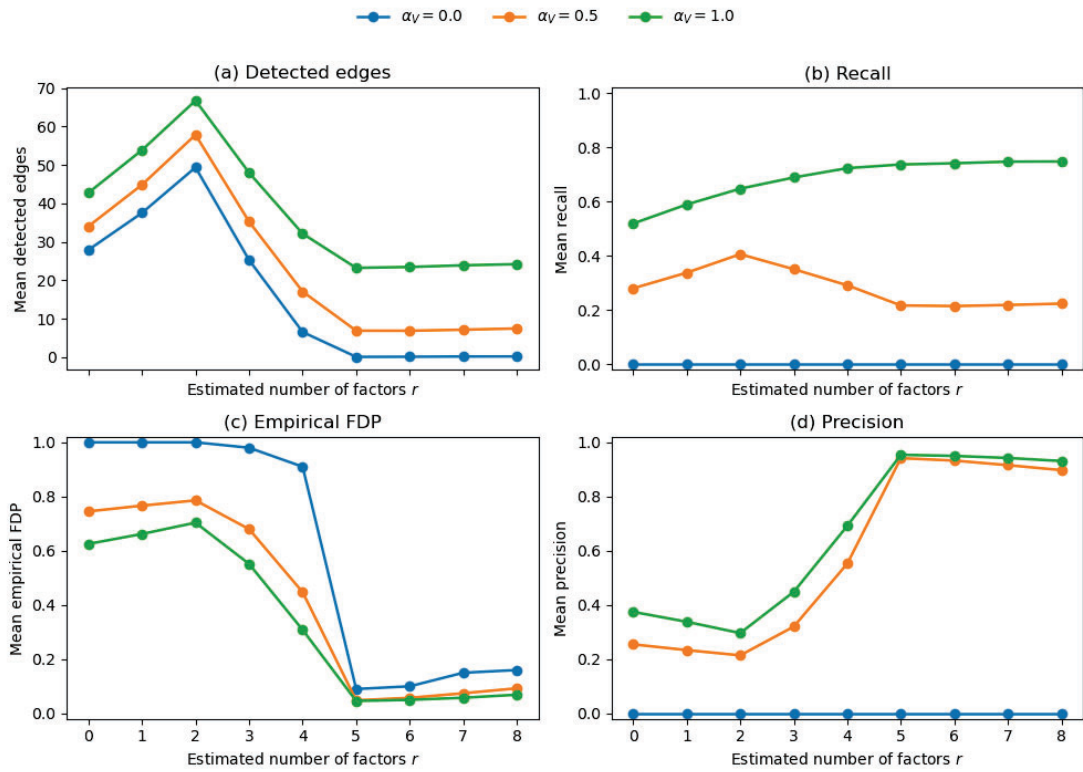


Figure 1: Simulation results as functions of the estimated number of factors r . Panel (a) reports the mean number of detected edges, panel (b) the mean recall, panel (c) the mean empirical false discovery proportion (FDP), and panel (d) the mean precision. The three curves correspond to $\alpha_V = 0, 0.5$, and 1.0 .

increases beyond 4, the mean number of detected edges drops sharply, but recall remains positive even for $r > r^*$. At the same time, the empirical FDP declines rapidly and precision rises to a high level. This shows that, under moderate idiosyncratic VAR dependence, increasing the number of estimated factors mainly removes false positives driven by common factor movements, while preserving a nontrivial fraction of genuine directed edges.

When $\alpha_V = 1.0$, the pattern is even clearer. Although the mean number of detected edges again declines as r increases, recall remains high and continues to improve slightly for large values of r . At the same time, the empirical FDP becomes small and precision becomes very high once r reaches about 5. Thus, when the true idiosyncratic VAR signal is strong, estimating more factors does not destroy the directed network; instead, it sharpens it by removing common-factor-driven spurious edges while retaining most of the genuine ones.

Overall, the simulation study shows that increasing the number of estimated factors does not mechanically eliminate genuine idiosyncratic VAR edges. What large values of r mainly do is reduce false discoveries and sharpen the selected network. Therefore, if detected edges disappear in the Ct-value application, this should not be viewed as a mechanical consequence of factor augmentation itself. Rather, it suggests that the apparent network at small r may be driven mainly by common factor movements, or that the idiosyncratic VAR component is weak or absent.

5 Empirical data analysis

The real-world dataset analyzed in this study consists of raw polymerase chain reaction (PCR) test results, including SARS-CoV-2 cycle threshold (Ct) values, obtained from a major clinical diagnostics laboratory in Brazil. The Ct value is the number of amplification cycles required for the viral signal to cross a detection threshold, so lower Ct values correspond to higher viral load. Temporal and regional changes in the Ct distribution can therefore be interpreted as changes in the distribution of viral load among tested individuals.

In total, the dataset comprises approximately 343,000 individual PCR test observations collected between 20 March 2020 and 22 May 2022. Although specimens were collected from more than 2,000 sites across the country, all laboratory analyses were performed at a centralized facility in São Paulo. A key feature of the dataset is therefore that cross-regional differences in the observed Ct distributions are less likely to reflect systematic differences in laboratory processing across locations. In addition, two PCR assays were used by the central laboratory (Seegene and Taq), and the laboratory confirmed that the resulting Ct values are comparable across assays.

We define the week index by

$$t = 1 + \lfloor (\text{date} - \text{start date})/7 \rfloor,$$

where the start date is 16 March 2020. We take Monday, 16 March 2020, as the reference date to define week boundaries, even though the first PCR observations enter from 20 March 2020. This yields $T = 114$ weekly observations in total. Figure 2 reports basic weekly summaries of the raw Ct observations: the weekly mean Ct value and the 5th and 95th percentiles computed from all observations nationwide, together with the weekly sample size. The 5th–95th percentile band of weekly Ct values is broadly stable over time, although it narrows around early 2022, driven mainly by an upward shift in the lower tail, even as the weekly sample size varies substantially.

Our empirical analysis follows the construction and modeling framework developed in Section 3. We first aggregate raw Ct observations into a weekly sequence of region-level density estimates, represented by B-spline mixture weights. We then apply the generalized logit transform to obtain the vector-valued series $Y_{t,c}$ and fit the factor-adjusted VAR model developed in Section 3. The

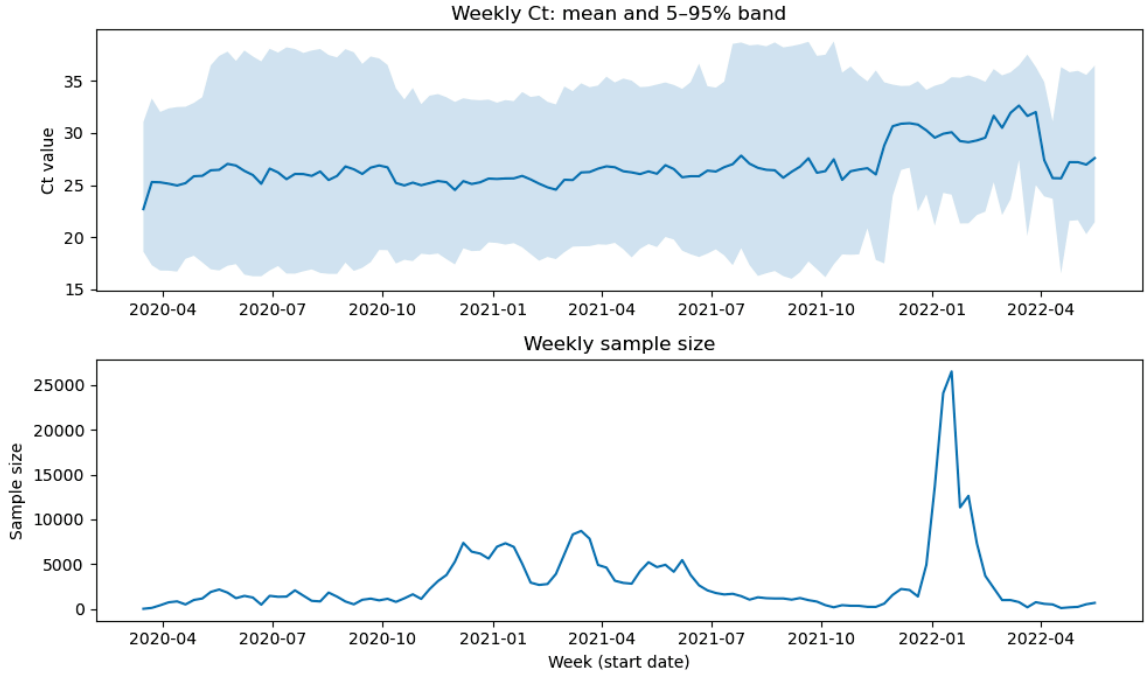


Figure 2: Weekly summaries of the raw Ct observations. The top panel reports the weekly mean Ct value together with the 5th and 95th percentiles, computed from nationwide observations. The bottom panel shows the weekly sample size. Both panels are indexed by the week start date.

resulting VAR coefficients are summarized as a directed network that captures lagged dependence in the evolution of Ct-value distributions across regions.

The main empirical issue is not merely whether one directed network can be drawn from the data, but under what conditions an idiosyncratic network becomes visible after strong common movements have been removed. In particular, both the sample period and the prior strength used in weekly density estimation affect how clearly the idiosyncratic VAR component can be separated from pervasive common factor movements. The empirical analysis below is therefore organized around this visibility question.

5.1 Construction of weekly Ct-value distributions

Each PCR observation is indexed by the date of sample collection and the reported municipality (city) and state. To connect city-level observations to the region-level model in Section 3, we construct $C = 20$ geographical regions by spatial clustering of Brazilian cities. We first apply spatial k -means clustering to city coordinates (latitude and longitude) with a deliberately large initial number of clusters ($K = 60$). We then iteratively merge clusters whose average weekly sample size falls below 30 tests per week, always merging the smallest cluster into its nearest neighboring cluster in terms of centroid distance. This procedure yields $C = 20$ geographically coherent regions with more balanced weekly observation counts, which helps avoid extremely sparse cells in the weekly density estimation. The resulting spatial partition is shown in Figure 3. Table 1 reports summary statistics of the weekly sample size $n_{t,c}$ for each region, labeled by its representative city.

Let $x_{t,c,i}$ denote the i th individual Ct value observed in region c during week t . For each region c and week t , we collect all Ct observations observed in that cell and estimate the weekly

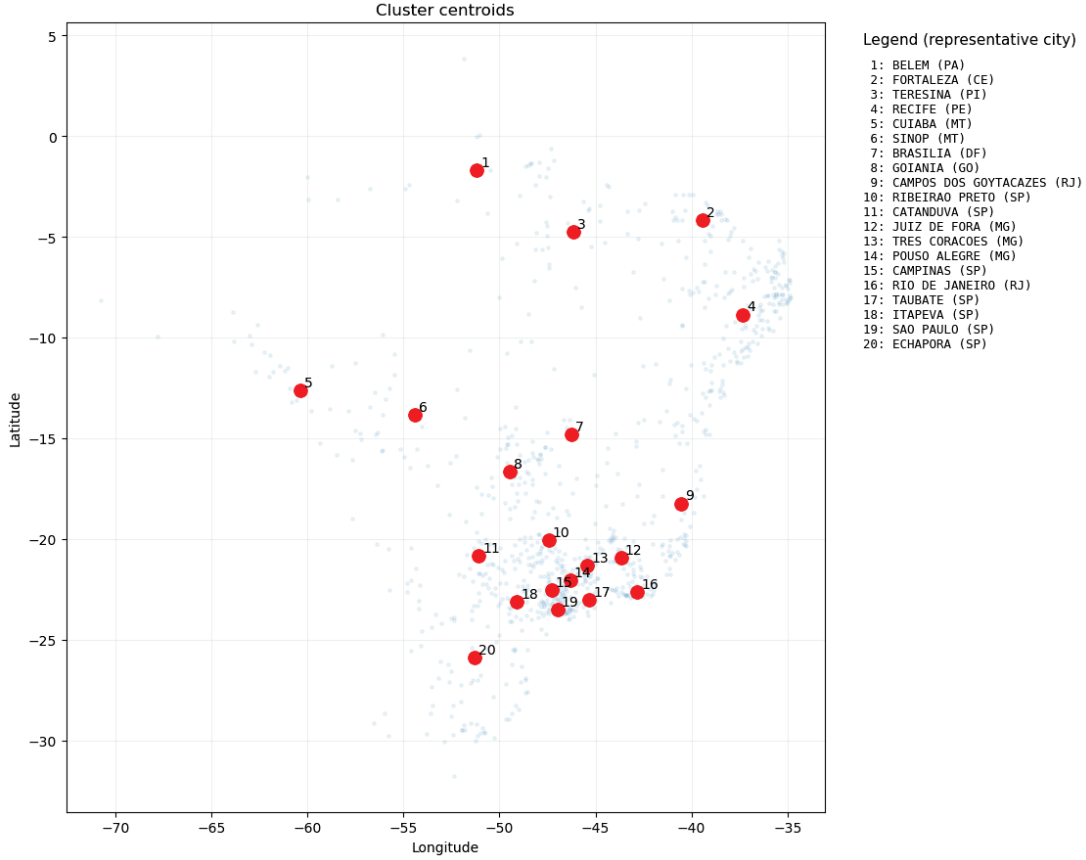


Figure 3: Map of the 20 regions in Brazil. Red dots indicate region centroids, and the numeric labels (ID 1–20) follow the north-to-south ordering. The legend lists each region ID and its representative city.

region-level density on the fixed support $[10, 40]$ by the B-spline mixture model in (1). In the empirical implementation, we use $J + 1$ normalized B-spline components and obtain the mixture weights $w_{t,c} \in \mathcal{S}_{J+1}$ by the EM algorithm with a Dirichlet prior, as described in Section 3.1. This regularization is important in practice because weekly sample sizes vary substantially across regions and are close to zero in the lower quartiles for several regions.

More specifically, for each week we use the nationwide weekly density as a common prior and estimate region-specific weekly densities as posterior-mean B-spline weights. The prior strength is controlled by the scalar parameter γ in (2). A larger γ pulls region-level densities more strongly toward the nationwide weekly pattern, whereas a smaller γ leaves more room for region-specific variation. Since this choice turns out to matter for the visibility of the idiosyncratic VAR network, we compare several values of γ in the empirical analysis below.

We transform the estimated weight vectors into \mathbb{R}^J via the generalized logit map in Section 3.2, $Y_{t,c} = \text{logit}_\delta(w_{t,c})$, and stack them as $Y_t = (Y'_{t,1}, \dots, Y'_{t,20})'$. All subsequent estimation uses the block-diagonal metric $\mathcal{H} = I_C \otimes H_J$ introduced in (5). Equivalently, we work with the whitened series

$$\tilde{Y}_t = \mathcal{K}' Y_t,$$

where $\mathcal{K} = I_C \otimes K_J$ and $\mathcal{H} = \mathcal{K}\mathcal{K}'$. Unless stated otherwise, we fix the tuning parameter $\delta = 1$ and

Table 1: Summary statistics of weekly sample size by representative area

ID	Rep. City	count	mean	std	min	25%	50%	75%	max
1	Belem	114	73.8	149.9	0	3.00	10.0	34.75	623
2	Fortaleza	114	43.9	67.6	0	0.25	17.0	58.00	396
3	Teresina	114	44.7	80.1	0	3.00	11.0	49.25	414
4	Recife	114	255.5	599.4	0	19.25	100.0	235.50	4640
5	Cuiaba	114	142.1	291.8	0	31.25	72.0	132.25	2372
6	Sinop	114	45.2	58.1	0	8.25	23.5	60.00	384
7	Brasilia	114	157.6	225.5	0	37.75	98.0	165.00	1462
8	Goiania	114	57.1	89.8	0	11.25	35.5	58.25	716
9	Campos dos Goytacazes	114	80.1	111.4	0	15.50	38.5	86.00	626
10	Ribeirao Preto	114	137.1	223.0	0	31.00	73.0	170.50	1528
11	Catanduva	114	110.9	236.0	0	12.00	33.0	96.75	1573
12	Juiz de Fora	114	94.6	131.6	0	15.00	34.5	119.50	675
13	Tres Coracoes	114	30.7	47.3	0	3.25	15.0	31.50	255
14	Pouso Alegre	114	143.3	364.1	0	8.25	28.0	133.75	2713
15	Campinas	114	271.0	336.4	1	62.25	141.0	358.75	1971
16	Rio de Janeiro	114	271.5	407.8	0	36.00	124.0	313.25	2988
17	Taubate	114	76.9	137.1	0	6.25	21.0	72.00	661
18	Itapeva	114	32.8	56.3	0	4.00	12.5	39.50	372
19	Sao Paulo	114	785.8	960.4	1	163.25	362.5	1127.75	5173
20	Echapora	114	150.6	281.9	0	12.25	30.0	127.00	1701

the spline specification ($J = 15, k = 3$) throughout the empirical analysis for numerical stability and comparability across regions and weeks.

5.2 Full-sample versus post-September 2020 results

We now examine how the detected cross-regional VAR network depends on the sample period and the prior strength used in the weekly density estimation. The main empirical question is whether the directed edges identified by our procedure reflect genuine idiosyncratic distributional dynamics, or whether they are masked by strong common factor movements and overshrinkage toward the nationwide weekly density.

To address this issue, we compare two sample periods. The first uses the full sample from 16 March 2020 onward. The second starts on 28 September 2020 and excludes the first six months of the pandemic. We focus on this later subsample because the early phase appears to be dominated by strong nationwide common movements, whereas the later phase is expected to contain richer regional heterogeneity. In addition, weekly sample sizes become more stable after the early phase, which improves the reliability of the region-level density estimates.

Figure 4 reports the number of detected directed edges as a function of the estimated number of factors r , for four values of the prior-strength parameter, $\gamma = 0.5, 1, 2, 5$. In each case, edges are selected from positive off-diagonal VAR coefficients using one-sided t -tests with Benjamini–Yekutieli false discovery rate control at the 5% level (Benjamini and Yekutieli, 2001).

The contrast between the two sample periods is sharp. In the full sample, the detected network is generally weak for all values of γ . Under the strong prior $\gamma = 5$, the number of detected edges increases temporarily around $r = 2$ and $r = 3$, but then falls back to nearly zero for $r \geq 5$. This pattern is consistent with the simulation evidence in Section 4: apparent directed edges induced

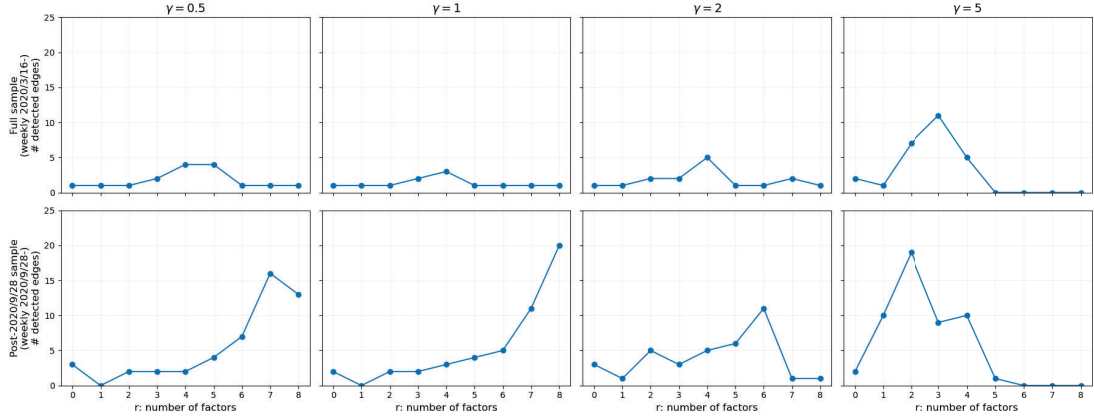


Figure 4: Number of detected directed edges as a function of the estimated number of factors r . The top row uses the full sample from 2020/3/16 onward, and the bottom row uses the subsample from 2020/9/28 onward. The four columns correspond to $\gamma = 0.5, 1, 2, 5$. Edges are selected from positive off-diagonal VAR coefficients using one-sided t -tests with Benjamini–Yekutieli false discovery rate control at the 5% level.

by strong common factor movements may remain visible when too few factors are included, but they disappear once enough factors are added. By contrast, in the post-2020/9/28 subsample, a substantial number of edges emerge under weak priors, especially for $\gamma = 0.5$ and $\gamma = 1$, when r is sufficiently large. This suggests that the early phase of the pandemic is more strongly dominated by common movements, whereas the later sample contains a more visible idiosyncratic VAR structure.

The prior-strength parameter also has a clear effect. Under weak priors ($\gamma = 0.5$ and $\gamma = 1$), the later subsample yields a substantial directed network for larger values of r . By contrast, stronger priors ($\gamma = 2$ and $\gamma = 5$) largely suppress the detected network even in the later subsample. This is consistent with the idea that a large γ overshrinks the region-level weekly densities toward the nationwide weekly pattern and thereby removes region-specific variation needed to identify the idiosyncratic VAR component.

Taken together, Figure 4 shows that two empirical choices are crucial for recovering an interpretable directed network from the Ct-value data: excluding the first six months of the sample and using a weak prior in the weekly density estimation. It also suggests that networks detected only at small values of r should be interpreted with caution, because they may still reflect insufficient removal of common movements. In the remainder of the empirical analysis, we therefore focus on the post-2020/9/28 subsample and use $\gamma = 1$ as our baseline specification. This choice is motivated by the preceding evidence: weak priors preserve region-specific variation, while sufficiently large values of r help remove strong common movements and make the idiosyncratic VAR network visible.

5.3 Directed network in the post-2020/9/28 weak-prior baseline

We next examine the directed network under the post-2020/9/28 subsample with the weak prior $\gamma = 1$, focusing on the large- r results for $r = 7$ and $r = 8$. As shown in Section 5.2, weak-prior specifications combined with sufficiently large values of r are the settings in which the idiosyncratic VAR network becomes visible after strong common movements are removed. Rather than singling out one specification as uniquely preferred, we compare two nearby large- r specifications and focus on the directed edges that remain selected in both of them.

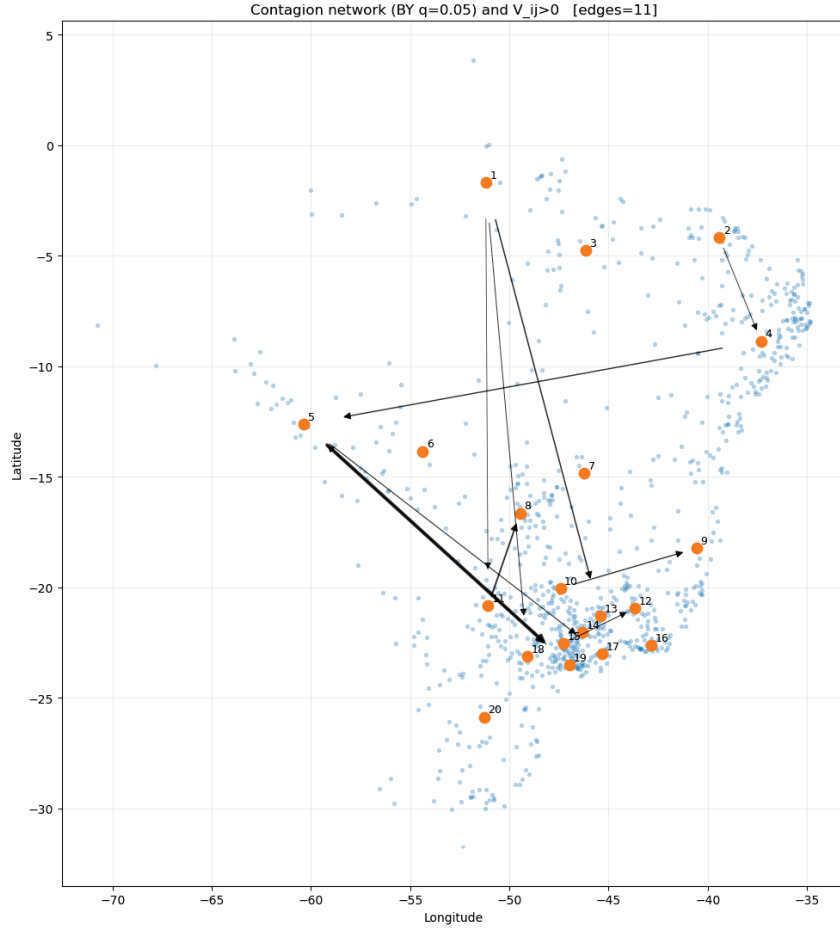


Figure 5: Directed network for the post-2020/9/28 subsample with $\gamma = 1$ and $r = 7$. All edges shown here remain selected in the $r = 8$ specification. Edges are selected from positive off-diagonal VAR coefficients using one-sided t -tests with Benjamini–Yekutieli false discovery rate control at the 5% level.

Figure 5 visualizes the network for $r = 7$, all of whose edges remain selected in the $r = 8$ specification, while Table 2 lists the edges common to both specifications. Each directed edge corresponds to a positive off-diagonal VAR coefficient selected by one-sided t -tests with Benjamini–Yekutieli false discovery rate control at the 5% level. An arrow $c' \rightarrow c$ indicates a Granger-predictive relation in the idiosyncratic VAR component (Granger, 1969; Lütkepohl, 2005): past distributional innovations in region c' help predict current innovations in region c after partialling out the common factor movements.

Two features stand out. First, the $r = 7$ network remains fully preserved in the broader $r = 8$ specification: all 11 edges selected at $r = 7$ remain selected at $r = 8$, with 9 additional edges. By contrast, the corresponding full-sample specifications yield only one selected edge in each case. This reinforces the conclusion of Section 5.2: once the first six months are excluded and the density prior is kept weak, the data contain a nontrivial idiosyncratic VAR network rather than a pattern driven solely by pervasive common movements.

Second, sender importance should be judged not only by the number of outgoing edges in a single specification, but also by whether those edges remain selected across nearby large- r specifications.

Table 2: Directed edges selected in both $r = 7$ and $r = 8$ under the post-2020/9/28 weak-prior baseline ($\gamma = 1$). Region labels follow the representative-city convention in Table 1.

Source ID	Target ID	Source rep. city	Target rep. city	Coef. ($r = 8$)	t -stat ($r = 8$)
5	19	Cuiaba	Sao Paulo	0.191	5.806
11	8	Catanduva	Goiania	0.203	4.069
1	13	Belem	Tres Coracoes	0.125	3.999
4	5	Recife	Cuiaba	0.115	3.534
10	9	Ribeirao Preto	Campos dos Goytacazes	0.171	3.805
19	5	Sao Paulo	Cuiaba	0.237	3.783
5	17	Cuiaba	Taubate	0.183	3.330
15	12	Campinas	Juiz de Fora	0.226	3.734
1	18	Belem	Itapeva	0.124	4.189
2	4	Fortaleza	Recife	0.154	3.703
1	11	Belem	Catanduva	0.080	3.691

Table 3: Outgoing-edge counts by source region under the post-2020/9/28 weak-prior baseline ($\gamma = 1$).

Source ID	Representative city	Outgoing at $r = 7$	Outgoing at $r = 8$
1	Belem	3	3
5	Cuiaba	2	2
2	Fortaleza	1	2
4	Recife	1	2
10	Ribeirao Preto	1	2
11	Catanduva	1	5
15	Campinas	1	1
19	Sao Paulo	1	2

The region represented by Belem has three outgoing edges at both $r = 7$ and $r = 8$, directed toward the regions represented by Tres Coracoes, Itapeva, and Catanduva. Although the representative city of region 1 is Belem, this northern region also contains Manaus. Thus, the persistent outgoing role of region 1 is broadly consistent with epidemiological accounts emphasizing the importance of the Manaus region during the spread of the Gamma (P.1) variant (Faria et al., 2021; Sabino et al., 2021; Banho et al., 2022).

Cuiaba also appears as a stable sender. Its outgoing edge count is two at both $r = 7$ and $r = 8$, and its links to Sao Paulo and Taubate remain selected in both specifications. In addition, one of these links is among the strongest selected edges in the table. By contrast, the region represented by Catanduva has the largest outgoing count at $r = 8$, but not at $r = 7$. This shows that the source region with the largest number of outgoing edges in a single specification need not be the most robust sender across nearby large- r specifications.

Taken together, Figure 5 and Tables 2–3 indicate that the post-2020/9/28 weak-prior baseline yields a network with multiple sender regions. Within this pattern, the northern region represented by Belem (including Manaus) and the inland region represented by Cuiaba stand out because their outgoing roles remain visible across both $r = 7$ and $r = 8$.

At the same time, the estimated network should be interpreted with caution. Our analysis does not model infections or case counts directly; instead, it studies lagged dependence in weekly Ct-value distributions. Hence, the detected edges should not be read as literal flows of infected

individuals across regions. Rather, they capture predictive dependence in how the shapes of the observed viral-load distributions evolve across regions. Since Ct distributions depend on who is tested and when samples are collected, the network may reflect variation in testing composition and timing as well as transmission dynamics. Even so, the estimated network provides evidence of geographically structured idiosyncratic dependence in the evolution of Ct-value distributions across Brazil.

5.4 Summary of empirical findings

The empirical analysis yields three main findings. First, when the full sample from March 2020 onward is used, the directed network implied by the idiosyncratic VAR component remains weak for all values of the prior-strength parameter considered here. Second, once the first six months are excluded and the sample starts on 28 September 2020, a substantial number of directed edges emerge under weak priors, especially for $\gamma = 0.5$ and $\gamma = 1$, when the number of factors is sufficiently large. Third, under the post-2020/9/28 weak-prior baseline with $\gamma = 1$, the large- r results reveal a network with multiple sender regions, among which the northern region represented by Belem (including Manaus) and the inland region represented by Cuiaba remain visible across both $r = 7$ and $r = 8$.

Taken together, these results indicate that the Ct-value data contain a visible idiosyncratic VAR structure only after the early pandemic phase is excluded and the density prior is kept weak enough to preserve region-specific variation. Thus, the disappearance of detected edges in some specifications should not be viewed as a mechanical failure of the factor-adjusted VAR framework, but rather as a consequence of strong common movements and excessive shrinkage in the density estimation step.

6 Conclusion

This paper proposed a density-valued VAR model with latent factors for multivariate time series of density functions observed across units. The new method combines B-spline mixture density estimation, a generalized logit transformation, an isometric inner-product structure for the transformed coordinates, latent factor decomposition, and cross-unit VAR dependence in a unified framework. This construction allows the separation between strong common movements and directed idiosyncratic distributional dynamics, while summarizing the latter as a Granger-predictive network.

On the theoretical side, we established large-sample properties of the factor-adjusted VAR estimator in the transformed Euclidean representation. Under a joint asymptotic regime in which the basis dimension and the time dimension diverge with the number of regions fixed, we derived its rate of convergence and, under stronger conditions, asymptotic normality for coefficient-wise inference. These results provide a formal basis for testing directed lagged dependence after removing latent common factors.

The simulation study showed that increasing the number of estimated factors does not mechanically eliminate a genuine idiosyncratic VAR network. When the true VAR component is absent, spuriously detected edges largely disappear once enough factors are included. When genuine idiosyncratic dependence is present, however, recall remains positive even for large values of the estimated factor dimension, while false discoveries decline sharply. This distinction is important for interpreting empirical results in settings with strong common movements.

In the empirical application to weekly regional distributions of SARS-CoV-2 Ct values in Brazil, the detected network depended crucially on both the sample period and the strength of the density prior. In the full sample, only a weak directed network was detected. By contrast, once the first

six months of the pandemic were excluded and the nationwide weekly prior was kept weak, a substantial idiosyncratic network became visible. Under the post-2020/9/28 weak-prior baseline, the large- r results revealed stable outgoing links from the northern region represented by Belem, which includes Manaus, and from Cuiabá toward southeastern metropolitan areas.

These findings show that the proposed framework can recover directed idiosyncratic dependence in multivariate time series of density functions after strong common movements are removed. More broadly, the paper provides a tractable approach to analyzing time series of distributions when common factors and directed idiosyncratic dependence are both present. Although the empirical application treats regions as the observational units, the same framework can also be applied when the units represent different variables, demographic groups, or other collections of density-valued observations over time. A natural next step is to extend the present VAR structure beyond scalar coefficients multiplying the transformed density coordinates as a whole, for example by allowing component-specific effects or more general cross-component interactions in the transformed space. Such extensions would broaden the class of directed distributional dynamics that can be represented within the same density-valued time-series framework.

7 Appendix

This Appendix collects auxiliary technical results and the proofs of Theorems 3.1 and 3.2 in the main text. The proofs follow the algebraic template of Bai (2009), adapted to the present density-valued setting with PCA carried out along the time direction.

Throughout the Appendix, for a matrix A , $\|A\|$ denotes its operator norm and $\|A\|_F$ its Frobenius norm. For a vector x , $\|x\|$ denotes its Euclidean norm.

7.1 Proofs

Proof of Theorem 3.1. Let $N := CJ$ and $T_0 := T - p$, and assume that $N \rightarrow \infty$, $T_0 \rightarrow \infty$, and $N/T_0 \rightarrow \rho \geq 0$.

By the projection-replacement argument established before the condition $N/T_0 \rightarrow 0$ is invoked in Corollary 7.13, we have

$$\begin{aligned} \sqrt{NT_0}(\hat{\beta} - \beta_0) &= D(\hat{\Lambda})^{-1} \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left[\tilde{W}'_t M_{\Pi_0} - \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\Pi_0} \right] \tilde{\varepsilon}_t \\ &\quad + \sqrt{\frac{N}{T_0}} \zeta_{NT_0} + o_p(1), \end{aligned}$$

where

$$a_{ts} = f_t^{0t} \left(F_0' F_0 / T_0 \right)^{-1} f_s^0, \quad F_0 = (f_{p+1}^0, \dots, f_T^0)',$$

and

$$\Pi_0 = \tilde{\Lambda}_0 (\tilde{\Lambda}_0' \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0'.$$

By Assumption A(ii),

$$\inf_{\tilde{\Lambda} \in \mathcal{F}} \lambda_{\min}\{D(\tilde{\Lambda})\} \geq c > 0$$

with probability approaching one. Since $\hat{\Lambda} \in \mathcal{F}$, it follows that

$$D(\hat{\Lambda})^{-1} = O_p(1).$$

Next consider the score term

$$S_{NT_0} := \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left[\tilde{W}_t' M_{\Pi_0} - \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}_s' \right\} M_{\Pi_0} \right] \tilde{\varepsilon}_t.$$

Since $\|M_{\Pi_0}\| \leq 1$, we have

$$\|S_{NT_0}\| \leq \left\| \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \tilde{W}_t' \tilde{\varepsilon}_t \right\| + \left\| \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}_s' \right\} \tilde{\varepsilon}_t \right\|.$$

The first term is $O_p(1)$ by Assumptions A–D and the same second-moment argument as in Step 1 of the proof of Lemma 7.12. For the second term, Assumption B(i) implies that the coefficients a_{ts} are uniformly bounded. Therefore, by Assumption C(vi),

$$\left\| \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}_s' \right\} \tilde{\varepsilon}_t \right\| = O_p(1).$$

Therefore

$$S_{NT_0} = O_p(1).$$

Moreover, $\zeta_{NT_0} = O_p(1)$ by Lemma 7.10. Since $N/T_0 \rightarrow \rho \geq 0$, we have

$$\sqrt{\frac{N}{T_0}} \zeta_{NT_0} = O_p(1).$$

Combining these bounds yields

$$\sqrt{NT_0} (\hat{\beta} - \beta_0) = O_p(1),$$

which completes the proof. \square

Proof of Theorem 3.2. Let $N := CJ$ and $T_0 := T - p$. Define

$$H_0 := \frac{1}{N} \tilde{\Lambda}_0' \tilde{\Lambda}_0, \quad \bar{\Lambda}_0 := \tilde{\Lambda}_0 H_0^{-1/2}, \quad \Pi_0 := \frac{1}{N} \bar{\Lambda}_0 \bar{\Lambda}_0' = \tilde{\Lambda}_0 (\tilde{\Lambda}_0' \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0'.$$

By Corollary 7.13, we have the linear expansion

$$\sqrt{NT_0} (\hat{\beta} - \beta_0) = D(\hat{\Lambda})^{-1} \left(\frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T Z_t(\bar{\Lambda}_0)' \tilde{\varepsilon}_t \right) + o_p(1), \quad (12)$$

where $Z_t(\bar{\Lambda}_0)$ is defined as in Assumption A(ii).

By Proposition 7.2(ii),

$$\|P_{\hat{\Lambda}} - \Pi_0\| \xrightarrow{p} 0,$$

and hence

$$\|M_{\hat{\Lambda}} - M_{\Pi_0}\| \xrightarrow{p} 0.$$

Hence, by the definition of $Z_t(\Lambda)$, we obtain

$$\|D(\hat{\Lambda}) - D(\bar{\Lambda}_0)\| = o_p(1).$$

Combined with Assumption E(i), this yields

$$D(\hat{\Lambda}) \xrightarrow{p} D_0.$$

Since $D_0 \succ 0$, we further obtain

$$D(\hat{\Lambda})^{-1} \xrightarrow{p} D_0^{-1}.$$

By Assumption E(ii),

$$\frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T Z_t(\bar{\Lambda}_0)' \tilde{\varepsilon}_t \xrightarrow{d} \mathcal{N}(0, \Omega_0).$$

Therefore, Slutsky's theorem (e.g., van der Vaart, 1998, Theorem 2.7) applied to (12) yields

$$\sqrt{NT_0}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, D_0^{-1}\Omega_0 D_0^{-1}).$$

This completes the proof. \square

7.2 Auxiliary results

Lemma 7.1. *Let*

$$\tilde{Y}_t = \tilde{W}_t \beta_0 + \tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t, \quad t = p+1, \dots, T,$$

where $\tilde{W}_t = (\tilde{X}_{t-1}, \dots, \tilde{X}_{t-p})$ and $\tilde{X}_t = I_C \otimes (\tilde{Y}_{t,1}, \dots, \tilde{Y}_{t,C})$. Let $N = CJ$ and, for any $\tilde{\Lambda} \in \mathcal{F} := \{\Lambda \in \mathbb{R}^{N \times r} : N^{-1}\Lambda'\Lambda = I_r\}$, define

$$P_{\tilde{\Lambda}} = N^{-1}\tilde{\Lambda}\tilde{\Lambda}', \quad M_{\tilde{\Lambda}} = I_N - P_{\tilde{\Lambda}}.$$

Assume Assumptions A(i), B, C, and D in the main text. Then, as $J, T \rightarrow \infty$ with C fixed,

(i)

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} \left\| \frac{1}{N(T-p)} \sum_{t=p+1}^T \tilde{W}_t' M_{\tilde{\Lambda}} \tilde{\varepsilon}_t \right\| = o_p(1).$$

(ii)

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} \left| \frac{1}{N(T-p)} \sum_{t=p+1}^T f_t^{0'} \tilde{\Lambda}_0' M_{\tilde{\Lambda}} \tilde{\varepsilon}_t \right| = o_p(1).$$

(iii)

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} \left| \frac{1}{N(T-p)} \sum_{t=p+1}^T \tilde{\varepsilon}_t' P_{\tilde{\Lambda}} \tilde{\varepsilon}_t \right| = o_p(1).$$

Proof of Lemma 7.1. Write $T_0 := T - p$. Fix r and let

$$\mathcal{F} = \{\tilde{\Lambda} \in \mathbb{R}^{N \times r} : N^{-1}\tilde{\Lambda}'\tilde{\Lambda} = I_r\}.$$

For $\tilde{\Lambda} \in \mathcal{F}$, define

$$P_{\tilde{\Lambda}} = N^{-1}\tilde{\Lambda}\tilde{\Lambda}', \quad M_{\tilde{\Lambda}} = I_N - P_{\tilde{\Lambda}}.$$

We first prove part (iii), which is then used in the proofs of parts (i) and (ii).

(iii). For $\tilde{\Lambda} \in \mathcal{F}$, write

$$P_{\tilde{\Lambda}} = UU', \quad U := N^{-1/2}\tilde{\Lambda} \in \mathbb{R}^{N \times r}, \quad U'U = I_r.$$

Then

$$S_3(\tilde{\Lambda}) = \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}'_t P_{\tilde{\Lambda}} \tilde{\varepsilon}_t = \frac{1}{N} \text{tr} \left[U' \left(\frac{1}{T_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t \tilde{\varepsilon}'_t \right) U \right].$$

Hence, for

$$A := \frac{1}{T_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t \tilde{\varepsilon}'_t,$$

we have

$$|S_3(\tilde{\Lambda})| \leq \frac{1}{N} \|U'AU\|_*,$$

where $\|\cdot\|_*$ denotes the nuclear norm. Since $U'AU$ is an $r \times r$ matrix,

$$\|U'AU\|_* \leq r \|U'AU\| \leq r \|A\|,$$

and therefore

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} |S_3(\tilde{\Lambda})| \leq \frac{r}{N} \left\| \frac{1}{T_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t \tilde{\varepsilon}'_t \right\|.$$

Now decompose

$$\frac{1}{T_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t \tilde{\varepsilon}'_t = \bar{\Sigma}_\varepsilon + \frac{1}{T_0} \sum_{t=p+1}^T (\tilde{\varepsilon}_t \tilde{\varepsilon}'_t - \Sigma_{\varepsilon,t}), \quad \bar{\Sigma}_\varepsilon := \frac{1}{T_0} \sum_{t=p+1}^T \Sigma_{\varepsilon,t}.$$

By Assumption C(iii),

$$\|\bar{\Sigma}_\varepsilon\| \leq \frac{1}{T_0} \sum_{t=p+1}^T \|\Sigma_{\varepsilon,t}\| \leq K.$$

Hence

$$\frac{r}{N} \|\bar{\Sigma}_\varepsilon\| = O(N^{-1}) = o(1).$$

For the centered part, define

$$C_t := \tilde{\varepsilon}_t \tilde{\varepsilon}'_t - \Sigma_{\varepsilon,t}, \quad t = p+1, \dots, T.$$

By Assumption C(ii), $\{C_t\}$ are independent across t and $E(C_t) = 0$. Using $\|A\| \leq \|A\|_F$,

$$\left\| \frac{1}{T_0} \sum_{t=p+1}^T C_t \right\| \leq \left\| \frac{1}{T_0} \sum_{t=p+1}^T C_t \right\|_F.$$

Moreover,

$$E \left\| \frac{1}{T_0} \sum_{t=p+1}^T C_t \right\|_F^2 = \frac{1}{T_0^2} \sum_{t=p+1}^T E \|C_t\|_F^2.$$

Now

$$\|C_t\|_F^2 \leq 2\|\tilde{\varepsilon}_t \tilde{\varepsilon}'_t\|_F^2 + 2\|\Sigma_{\varepsilon,t}\|_F^2.$$

Since

$$\|\tilde{\varepsilon}_t \tilde{\varepsilon}'_t\|_F^2 = (\tilde{\varepsilon}'_t \tilde{\varepsilon}_t)^2 = \left(\sum_{i=1}^N \tilde{\varepsilon}_{it}^2 \right)^2 = \sum_{i=1}^N \sum_{j=1}^N \tilde{\varepsilon}_{it}^2 \tilde{\varepsilon}_{jt}^2,$$

Assumption C(i) and Hölder's inequality imply

$$E\|\tilde{\varepsilon}_t \tilde{\varepsilon}'_t\|_F^2 \leq \sum_{i=1}^N \sum_{j=1}^N (E|\tilde{\varepsilon}_{it}|^4)^{1/2} (E|\tilde{\varepsilon}_{jt}|^4)^{1/2} \leq KN^2.$$

Also, since $\|\Sigma_{\varepsilon,t}\|_F \leq \sqrt{N} \|\Sigma_{\varepsilon,t}\|$, Assumption C(iii) gives

$$\|\Sigma_{\varepsilon,t}\|_F^2 \leq N\|\Sigma_{\varepsilon,t}\|^2 \leq KN.$$

Therefore

$$E\|C_t\|_F^2 \leq KN^2$$

uniformly in t , and hence

$$E \left\| \frac{1}{T_0} \sum_{t=p+1}^T C_t \right\|_F^2 \leq \frac{KN^2}{T_0}.$$

It follows that

$$\left\| \frac{1}{T_0} \sum_{t=p+1}^T C_t \right\| = O_p\left(\frac{N}{\sqrt{T_0}}\right).$$

Consequently,

$$\frac{r}{N} \left\| \frac{1}{T_0} \sum_{t=p+1}^T C_t \right\| = O_p(T_0^{-1/2}) = o_p(1).$$

Combining the deterministic and centered parts, we obtain

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} |S_3(\tilde{\Lambda})| \leq \frac{r}{N} \left\| \frac{1}{T_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t \tilde{\varepsilon}'_t \right\| = o_p(1).$$

(i). Decompose $S_1(\tilde{\Lambda})$ using $M_{\tilde{\Lambda}} = I_N - P_{\tilde{\Lambda}}$:

$$S_1(\tilde{\Lambda}) = \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t \tilde{\varepsilon}_t - \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t P_{\tilde{\Lambda}} \tilde{\varepsilon}_t =: A_1 - B_1(\tilde{\Lambda}).$$

Step 1: the non-projection term. By Assumptions C(ii) and D, the error $\tilde{\varepsilon}_t$ is mean independent of past outcomes, and hence of \tilde{W}_t , which is measurable with respect to $\{\tilde{Y}_s : s < t\}$. Therefore $E(A_1) = 0$, and

$$E\|A_1\|^2 = \frac{1}{N^2 T_0^2} \sum_{t=p+1}^T E \left[\tilde{\varepsilon}'_t \tilde{W}_t \tilde{W}'_t \tilde{\varepsilon}_t \right].$$

Conditioning on \tilde{W}_t and using $E(\tilde{\varepsilon}_t \tilde{\varepsilon}_t') = \Sigma_{\varepsilon,t}$,

$$E\left[\tilde{\varepsilon}_t' \tilde{W}_t \tilde{W}_t' \tilde{\varepsilon}_t \mid \tilde{W}_t\right] = \text{tr}(\Sigma_{\varepsilon,t} \tilde{W}_t \tilde{W}_t') \leq \|\Sigma_{\varepsilon,t}\| \|\tilde{W}_t\|_F^2 \leq K \|\tilde{W}_t\|_F^2.$$

Since \tilde{W}_t is $N \times (pC^2)$ and C, p are fixed, Assumption A(i) together with Assumptions B–D implies

$$E\|\tilde{W}_t\|_F^2 = O(N)$$

uniformly in t . Therefore

$$E\|A_1\|^2 \leq \frac{1}{N^2 T_0^2} \sum_{t=p+1}^T K \cdot O(N) = O\left(\frac{1}{NT_0}\right) \rightarrow 0,$$

so $A_1 = o_p(1)$.

Step 2: the projection term, uniformly in $\tilde{\Lambda}$. By Cauchy–Schwarz,

$$\|B_1(\tilde{\Lambda})\| \leq \left(\frac{1}{NT_0} \sum_{t=p+1}^T \|\tilde{W}_t\|_F^2\right)^{1/2} \left(\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t' P_{\tilde{\Lambda}} \tilde{\varepsilon}_t\right)^{1/2}.$$

The first factor is $O_p(1)$ because

$$\frac{1}{T_0} \sum_{t=p+1}^T \|\tilde{W}_t\|_F^2 = O_p(N),$$

and the second factor is $o_p(1)$ uniformly in $\tilde{\Lambda} \in \mathcal{F}$ by part (iii). Hence

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} \|B_1(\tilde{\Lambda})\| = o_p(1).$$

Combining Step 1 and Step 2,

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} \|S_1(\tilde{\Lambda})\| \leq \|A_1\| + \sup_{\tilde{\Lambda} \in \mathcal{F}} \|B_1(\tilde{\Lambda})\| = o_p(1).$$

(ii). Similarly, write

$$S_2(\tilde{\Lambda}) = \frac{1}{NT_0} \sum_{t=p+1}^T f_t^{0'} \tilde{\Lambda}'_0 \tilde{\varepsilon}_t - \frac{1}{NT_0} \sum_{t=p+1}^T f_t^{0'} \tilde{\Lambda}'_0 P_{\tilde{\Lambda}} \tilde{\varepsilon}_t =: A_2 - B_2(\tilde{\Lambda}).$$

Step 1: the non-projection term. Since $\tilde{\Lambda}_0$ and $\{f_t^0\}$ are non-stochastic arrays,

$$E(A_2) = 0$$

by Assumption C(ii), and by independence across t ,

$$E(A_2^2) = \frac{1}{N^2 T_0^2} \sum_{t=p+1}^T E\left[(f_t^{0'} \tilde{\Lambda}'_0 \tilde{\varepsilon}_t)^2\right].$$

Moreover,

$$E\left[(f_t^{0'} \tilde{\Lambda}'_0 \tilde{\varepsilon}_t)^2\right] = f_t^{0'} \tilde{\Lambda}'_0 \Sigma_{\varepsilon,t} \tilde{\Lambda}_0 f_t^0 \leq K \|\tilde{\Lambda}_0\|^2 \|f_t^0\|^2.$$

By Assumption B(ii),

$$\|\tilde{\Lambda}_0\|^2 = O(N),$$

and by Assumption B(i),

$$\frac{1}{T_0} \sum_{t=p+1}^T \|f_t^0\|^2 = O(1).$$

Hence

$$E(A_2^2) \leq \frac{K}{N^2 T_0^2} \|\tilde{\Lambda}_0\|^2 \sum_{t=p+1}^T \|f_t^0\|^2 = O\left(\frac{1}{NT_0}\right) = o(1),$$

so $A_2 = o_p(1)$.

Step 2: the projection term, uniformly in $\tilde{\Lambda}$. By Cauchy–Schwarz,

$$|B_2(\tilde{\Lambda})| \leq \left(\frac{1}{NT_0} \sum_{t=p+1}^T \|\tilde{\Lambda}_0 f_t^0\|^2 \right)^{1/2} \left(\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t' P_{\tilde{\Lambda}} \tilde{\varepsilon}_t \right)^{1/2}.$$

The first factor is $O_p(1)$, since

$$\|\tilde{\Lambda}_0 f_t^0\|^2 \leq \|\tilde{\Lambda}_0\|^2 \|f_t^0\|^2, \quad \|\tilde{\Lambda}_0\|^2 = O(N), \quad \frac{1}{T_0} \sum_{t=p+1}^T \|f_t^0\|^2 = O(1),$$

and the second factor is $o_p(1)$ uniformly in $\tilde{\Lambda}$ by part (iii). Hence

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} |B_2(\tilde{\Lambda})| = o_p(1),$$

and therefore

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} |S_2(\tilde{\Lambda})| \leq |A_2| + \sup_{\tilde{\Lambda} \in \mathcal{F}} |B_2(\tilde{\Lambda})| = o_p(1).$$

This completes the proof. \square

Proposition 7.2 (Consistency). *Consider the transformed model*

$$\tilde{Y}_t = \tilde{W}_t \beta_0 + \tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t, \quad t = p+1, \dots, T,$$

with $N := CJ$ and

$$\mathcal{F} := \{\Lambda \in \mathbb{R}^{N \times r} : N^{-1} \Lambda' \Lambda = I_r\}.$$

Let $(\hat{\beta}, \hat{\Lambda})$ minimize

$$Q(\beta, \tilde{\Lambda}) = \sum_{t=p+1}^T (\tilde{Y}_t - \tilde{W}_t \beta)' M_{\tilde{\Lambda}} (\tilde{Y}_t - \tilde{W}_t \beta), \quad M_{\tilde{\Lambda}} = I_N - N^{-1} \tilde{\Lambda} \tilde{\Lambda}'.$$

Under Assumptions A–D, as $J, T \rightarrow \infty$ with C fixed, the following hold:

(i) $\hat{\beta} \xrightarrow{p} \beta_0$.

(ii) $N^{-1} \tilde{\Lambda}_0' \hat{\Lambda}$ is invertible with probability approaching one, and

$$\|P_{\hat{\Lambda}} - \Pi_0\| \xrightarrow{p} 0,$$

where

$$P_{\hat{\Lambda}} = N^{-1} \hat{\Lambda} \hat{\Lambda}', \quad \Pi_0 := \tilde{\Lambda}_0 (\tilde{\Lambda}_0' \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0'.$$

Proof of Proposition 7.2. We divide the proof into four steps.

Step 1 (Decomposition). Let $N := CJ$ and $T_0 := T - p$. Define the normalized criterion

$$S_{NT_0}(\beta, \tilde{\Lambda}) := \frac{1}{NT_0} \sum_{t=p+1}^T (\tilde{Y}_t - \tilde{W}_t\beta)' M_{\tilde{\Lambda}} (\tilde{Y}_t - \tilde{W}_t\beta), \quad M_{\tilde{\Lambda}} := I_N - \frac{1}{N} \tilde{\Lambda} \tilde{\Lambda}', \quad \tilde{\Lambda} \in \mathcal{F}.$$

Without loss of generality, set $\beta_0 = 0$ by replacing \tilde{Y}_t with $\tilde{Y}_t - \tilde{W}_t\beta_0$. Then

$$\tilde{Y}_t = \tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t,$$

and hence

$$\begin{aligned} S_{NT_0}(\beta, \tilde{\Lambda}) &= \frac{1}{NT_0} \sum_{t=p+1}^T (\tilde{\Lambda}_0 f_t^0 - \tilde{W}_t\beta)' M_{\tilde{\Lambda}} (\tilde{\Lambda}_0 f_t^0 - \tilde{W}_t\beta) \\ &\quad + \frac{2}{NT_0} \sum_{t=p+1}^T (\tilde{\Lambda}_0 f_t^0 - \tilde{W}_t\beta)' M_{\tilde{\Lambda}} \tilde{\varepsilon}_t + \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t' M_{\tilde{\Lambda}} \tilde{\varepsilon}_t. \end{aligned}$$

Write

$$\tilde{\varepsilon}_t' M_{\tilde{\Lambda}} \tilde{\varepsilon}_t = \tilde{\varepsilon}_t' \tilde{\varepsilon}_t - \tilde{\varepsilon}_t' P_{\tilde{\Lambda}} \tilde{\varepsilon}_t, \quad P_{\tilde{\Lambda}} := \frac{1}{N} \tilde{\Lambda} \tilde{\Lambda}'.$$

Therefore,

$$S_{NT_0}(\beta, \tilde{\Lambda}) = \tilde{S}_{NT_0}(\beta, \tilde{\Lambda}) + \underbrace{\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t' \tilde{\varepsilon}_t}_{=: C_{NT_0}} + R_{NT_0}(\beta, \tilde{\Lambda}),$$

where

$$\begin{aligned} \tilde{S}_{NT_0}(\beta, \tilde{\Lambda}) &:= \frac{1}{NT_0} \sum_{t=p+1}^T (\tilde{\Lambda}_0 f_t^0 - \tilde{W}_t\beta)' M_{\tilde{\Lambda}} (\tilde{\Lambda}_0 f_t^0 - \tilde{W}_t\beta), \\ R_{NT_0}(\beta, \tilde{\Lambda}) &:= \frac{2}{NT_0} \sum_{t=p+1}^T (\tilde{\Lambda}_0 f_t^0 - \tilde{W}_t\beta)' M_{\tilde{\Lambda}} \tilde{\varepsilon}_t - \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t' P_{\tilde{\Lambda}} \tilde{\varepsilon}_t. \end{aligned}$$

Since C_{NT_0} does not depend on $(\beta, \tilde{\Lambda})$, it plays no role in minimization.

Uniform bound for the remainder. Expanding the cross term gives

$$(\tilde{\Lambda}_0 f_t^0 - \tilde{W}_t\beta)' M_{\tilde{\Lambda}} \tilde{\varepsilon}_t = f_t^{0'} \tilde{\Lambda}'_0 M_{\tilde{\Lambda}} \tilde{\varepsilon}_t - \beta' \tilde{W}_t' M_{\tilde{\Lambda}} \tilde{\varepsilon}_t.$$

Hence, for any $\beta \in \mathbb{R}^K$,

$$\begin{aligned} \sup_{\tilde{\Lambda} \in \mathcal{F}} |R_{NT_0}(\beta, \tilde{\Lambda})| &\leq 2 \sup_{\tilde{\Lambda} \in \mathcal{F}} \left| \frac{1}{NT_0} \sum_{t=p+1}^T f_t^{0'} \tilde{\Lambda}'_0 M_{\tilde{\Lambda}} \tilde{\varepsilon}_t \right| \\ &\quad + 2 \|\beta\| \sup_{\tilde{\Lambda} \in \mathcal{F}} \left\| \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}_t' M_{\tilde{\Lambda}} \tilde{\varepsilon}_t \right\| \\ &\quad + \sup_{\tilde{\Lambda} \in \mathcal{F}} \left| \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t' P_{\tilde{\Lambda}} \tilde{\varepsilon}_t \right|. \end{aligned}$$

By Lemma 7.1ii, Lemma 7.1i, and Lemma 7.1iii, respectively, each supremum term on the right-hand side is $o_p(1)$. Therefore,

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} |R_{NT_0}(\beta, \tilde{\Lambda})| = o_p(1) (1 + \|\beta\|).$$

Step 2 (Square completion). Fix $\tilde{\Lambda} \in \mathcal{F}$. Let $F_0 := (f_{p+1}^0, \dots, f_T^0)' \in \mathbb{R}^{T_0 \times r}$ and define

$$B := \left(\frac{1}{T_0} F_0' F_0 \right) \otimes \left(\frac{1}{N} I_N \right), \quad \eta(\tilde{\Lambda}) := \text{vec}(M_{\tilde{\Lambda}} \tilde{\Lambda}_0) \in \mathbb{R}^{Nr},$$

and

$$A(\tilde{\Lambda}) := \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}_t' M_{\tilde{\Lambda}} \tilde{W}_t \in \mathbb{R}^{K \times K}, \quad C(\tilde{\Lambda}) := -\frac{1}{NT_0} \sum_{t=p+1}^T (f_t^0 \otimes M_{\tilde{\Lambda}} \tilde{W}_t) \in \mathbb{R}^{Nr \times K},$$

where $K = \dim(\beta) = pC^2$.

Expanding $\tilde{S}_{NT_0}(\beta, \tilde{\Lambda})$ and using vec/Kronecker identities yields

$$\tilde{S}_{NT_0}(\beta, \tilde{\Lambda}) = \beta' A(\tilde{\Lambda}) \beta + \eta(\tilde{\Lambda})' B \eta(\tilde{\Lambda}) + 2 \beta' C(\tilde{\Lambda})' \eta(\tilde{\Lambda}).$$

Completing the square gives

$$\begin{aligned} \tilde{S}_{NT_0}(\beta, \tilde{\Lambda}) &= \beta' (A(\tilde{\Lambda}) - C(\tilde{\Lambda})' B^{-1} C(\tilde{\Lambda})) \beta \\ &\quad + \left(\eta(\tilde{\Lambda}) + B^{-1} C(\tilde{\Lambda}) \beta \right)' B \left(\eta(\tilde{\Lambda}) + B^{-1} C(\tilde{\Lambda}) \beta \right) \\ &=: \beta' D(\tilde{\Lambda}) \beta + \theta(\tilde{\Lambda}, \beta)' B \theta(\tilde{\Lambda}, \beta), \end{aligned}$$

where

$$D(\tilde{\Lambda}) := A(\tilde{\Lambda}) - C(\tilde{\Lambda})' B^{-1} C(\tilde{\Lambda}), \quad \theta(\tilde{\Lambda}, \beta) := \eta(\tilde{\Lambda}) + B^{-1} C(\tilde{\Lambda}) \beta.$$

Since $B \succ 0$ with probability approaching one by Assumption B(i), we have

$$\tilde{S}_{NT_0}(\beta, \tilde{\Lambda}) \geq \beta' D(\tilde{\Lambda}) \beta.$$

Let

$$a_{ts} := f_t^{0'} \left(\frac{1}{T_0} F_0' F_0 \right)^{-1} f_s^0$$

and define

$$Z_t(\tilde{\Lambda}) := M_{\tilde{\Lambda}} \tilde{W}_t - \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} M_{\tilde{\Lambda}} \tilde{W}_s.$$

Then the Frisch–Waugh–Lovell identity implies

$$D(\tilde{\Lambda}) = \frac{1}{NT_0} \sum_{t=p+1}^T Z_t(\tilde{\Lambda})' Z_t(\tilde{\Lambda}).$$

In particular, Assumption A(ii) implies that there exists $c_0 > 0$ such that

$$\inf_{\tilde{\Lambda} \in \mathcal{F}} \lambda_{\min}\{D(\tilde{\Lambda})\} \geq c_0 \quad \text{with probability approaching one.}$$

Step 3 (Consistency of $\hat{\beta}$). Let

$$H_0 := \frac{1}{N} \tilde{\Lambda}'_0 \tilde{\Lambda}_0, \quad \bar{\Lambda}_0 := \tilde{\Lambda}_0 H_0^{-1/2}.$$

Then $\bar{\Lambda}_0 \in \mathcal{F}$, and

$$P_{\bar{\Lambda}_0} = \frac{1}{N} \bar{\Lambda}_0 \bar{\Lambda}'_0 = \tilde{\Lambda}_0 (\tilde{\Lambda}'_0 \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}'_0 = \Pi_0,$$

so $M_{\bar{\Lambda}_0} \tilde{\Lambda}_0 = 0$.

By Step 2,

$$\tilde{S}_{NT_0}(\beta, \tilde{\Lambda}) \geq c_0 \|\beta\|^2$$

uniformly in $\tilde{\Lambda} \in \mathcal{F}$ with probability approaching one. Combining this with Step 1 gives

$$S_{NT_0}(\beta, \tilde{\Lambda}) \geq C_{NT_0} + c_0 \|\beta\|^2 - o_p(1)(1 + \|\beta\|)$$

uniformly in $\tilde{\Lambda} \in \mathcal{F}$. Since the quadratic term dominates the linear remainder for $\|\beta\|$ large enough, there exists $B < \infty$ such that

$$\inf_{\|\beta\| > B, \tilde{\Lambda} \in \mathcal{F}} S_{NT_0}(\beta, \tilde{\Lambda}) > C_{NT_0} + 1$$

with probability approaching one.

On the other hand, evaluating the criterion at $(0, \bar{\Lambda}_0)$, we have

$$S_{NT_0}(0, \bar{\Lambda}_0) = C_{NT_0} + R_{NT_0}(0, \bar{\Lambda}_0),$$

because $M_{\bar{\Lambda}_0} \tilde{\Lambda}_0 = 0$. By Step 1,

$$R_{NT_0}(0, \bar{\Lambda}_0) = o_p(1),$$

and hence

$$S_{NT_0}(0, \bar{\Lambda}_0) = C_{NT_0} + o_p(1).$$

Therefore, with probability approaching one, any minimizer $(\hat{\beta}, \hat{\Lambda})$ of $S_{NT_0}(\beta, \tilde{\Lambda})$ must satisfy

$$\|\hat{\beta}\| \leq B.$$

Restricting attention to the compact set $\{\beta : \|\beta\| \leq B\}$, Step 1 yields

$$\sup_{\|\beta\| \leq B, \tilde{\Lambda} \in \mathcal{F}} |R_{NT_0}(\beta, \tilde{\Lambda})| = o_p(1).$$

Hence

$$S_{NT_0}(\beta, \tilde{\Lambda}) = \tilde{S}_{NT_0}(\beta, \tilde{\Lambda}) + C_{NT_0} + o_p(1)$$

uniformly over $\|\beta\| \leq B$ and $\tilde{\Lambda} \in \mathcal{F}$. Since $(\hat{\beta}, \hat{\Lambda})$ is a minimizer,

$$\tilde{S}_{NT_0}(\hat{\beta}, \hat{\Lambda}) \leq \tilde{S}_{NT_0}(0, \bar{\Lambda}_0) + o_p(1).$$

Because $M_{\bar{\Lambda}_0} \tilde{\Lambda}_0 = 0$,

$$\tilde{S}_{NT_0}(0, \bar{\Lambda}_0) = 0,$$

and therefore

$$\tilde{S}_{NT_0}(\hat{\beta}, \hat{\Lambda}) = o_p(1). \tag{13}$$

Again by Step 2,

$$\tilde{S}_{NT_0}(\hat{\beta}, \hat{\Lambda}) \geq \hat{\beta}' D(\hat{\Lambda}) \hat{\beta} \geq c_0 \|\hat{\beta}\|^2 \quad \text{with probability approaching one.}$$

Combining this with (13) yields

$$\|\hat{\beta}\| = o_p(1),$$

that is,

$$\hat{\beta} \xrightarrow{p} 0 = \beta_0.$$

Step 4 (Consistency of the loading space). Expanding $\tilde{S}_{NT_0}(\beta, \tilde{\Lambda})$ yields

$$\begin{aligned}\tilde{S}_{NT_0}(\beta, \tilde{\Lambda}) &= \frac{1}{NT_0} \underbrace{\sum_{t=p+1}^T f_t^{0'} \tilde{\Lambda}'_0 M_{\tilde{\Lambda}} \tilde{\Lambda}_0 f_t^0}_{=:c(\tilde{\Lambda})} - 2\beta' \underbrace{\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\tilde{\Lambda}} \tilde{\Lambda}_0 f_t^0}_{=:b(\tilde{\Lambda})} \\ &\quad + \beta' \underbrace{\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\tilde{\Lambda}} \tilde{W}_t}_{=:A(\tilde{\Lambda})} \beta.\end{aligned}$$

Evaluating at $(\hat{\beta}, \hat{\Lambda})$ and using (13), we obtain

$$c(\hat{\Lambda}) = \tilde{S}_{NT_0}(\hat{\beta}, \hat{\Lambda}) + 2\hat{\beta}' b(\hat{\Lambda}) - \hat{\beta}' A(\hat{\Lambda}) \hat{\beta} = o_p(1),$$

where we used $\|\hat{\beta}\| = o_p(1)$ from Step 3 together with $b(\hat{\Lambda}) = O_p(1)$ and $A(\hat{\Lambda}) = O_p(1)$.

Write $F_0 = (f_{p+1}^0, \dots, f_T^0)' \in \mathbb{R}^{T_0 \times r}$. Then

$$c(\hat{\Lambda}) = \frac{1}{N} \text{tr} \left(\tilde{\Lambda}'_0 M_{\hat{\Lambda}} \tilde{\Lambda}_0 \cdot \frac{1}{T_0} F_0' F_0 \right) = o_p(1).$$

By Assumption B(i), $T_0^{-1} F_0' F_0 \rightarrow \Sigma_f \succ 0$, and therefore

$$\frac{1}{N} \tilde{\Lambda}'_0 M_{\hat{\Lambda}} \tilde{\Lambda}_0 = o_p(1). \quad (14)$$

Define

$$H_0 := \frac{1}{N} \tilde{\Lambda}'_0 \tilde{\Lambda}_0, \quad \bar{\Lambda}_0 := \tilde{\Lambda}_0 H_0^{-1/2}.$$

By Assumption B(ii), $H_0 \rightarrow \Sigma_\Lambda \succ 0$, so H_0 is invertible with probability approaching one and

$$\frac{1}{N} \bar{\Lambda}'_0 \bar{\Lambda}_0 = I_r.$$

Hence

$$P_{\bar{\Lambda}_0} := \frac{1}{N} \bar{\Lambda}_0 \bar{\Lambda}'_0$$

is an orthogonal projector of rank r , and

$$P_{\bar{\Lambda}_0} = \tilde{\Lambda}_0 (\tilde{\Lambda}'_0 \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}'_0 = \Pi_0.$$

From (14),

$$\frac{1}{N} \tilde{\Lambda}'_0 M_{\hat{\Lambda}} \tilde{\Lambda}_0 = H_0^{-1/2} \left(\frac{1}{N} \tilde{\Lambda}'_0 M_{\hat{\Lambda}} \tilde{\Lambda}_0 \right) H_0^{-1/2} = o_p(1).$$

Since both $P_{\hat{\Lambda}}$ and $P_{\bar{\Lambda}_0}$ are orthogonal projectors of rank r ,

$$\|P_{\hat{\Lambda}} - P_{\bar{\Lambda}_0}\|_F^2 = 2 \text{tr}(M_{\hat{\Lambda}} P_{\bar{\Lambda}_0}).$$

Using $P_{\bar{\Lambda}_0} = (1/N) \bar{\Lambda}_0 \bar{\Lambda}'_0$, we obtain

$$\text{tr}(M_{\hat{\Lambda}} P_{\bar{\Lambda}_0}) = \frac{1}{N} \text{tr}(\tilde{\Lambda}'_0 M_{\hat{\Lambda}} \tilde{\Lambda}_0) = o_p(1),$$

and therefore

$$\|P_{\hat{\Lambda}} - P_{\tilde{\Lambda}_0}\|_F^2 = o_p(1).$$

Since $\|A\| \leq \|A\|_F$ for any matrix A , we conclude that

$$\|P_{\hat{\Lambda}} - \Pi_0\| = \|P_{\hat{\Lambda}} - P_{\tilde{\Lambda}_0}\| = o_p(1).$$

Finally, since

$$\frac{1}{N} \tilde{\Lambda}'_0 P_{\hat{\Lambda}} \tilde{\Lambda}_0 = \frac{1}{N} \tilde{\Lambda}'_0 \tilde{\Lambda}_0 - \frac{1}{N} \tilde{\Lambda}'_0 M_{\hat{\Lambda}} \tilde{\Lambda}_0 = H_0 + o_p(1),$$

and H_0 is invertible with probability approaching one, it follows that

$$\frac{1}{N} \tilde{\Lambda}'_0 \hat{\Lambda}$$

is invertible with probability approaching one. This completes the proof. \square

Proposition 7.3 (Loading PCA). *Let*

$$\tilde{Y}_t = \tilde{W}_t \beta_0 + \tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t, \quad t = p+1, \dots, T,$$

with $N := CJ$ and $T_0 := T - p$. Let $\hat{\beta}$ be the estimator of β_0 defined in Proposition 7.2, and define

$$u_t := \tilde{Y}_t - \tilde{W}_t \hat{\beta}, \quad \hat{S} := \frac{1}{NT_0} \sum_{t=p+1}^T u_t u_t'.$$

Let $\hat{\Lambda} \in \mathbb{R}^{N \times r}$ collect the eigenvectors associated with the r largest eigenvalues of \hat{S} , normalized such that

$$\frac{1}{N} \hat{\Lambda}' \hat{\Lambda} = I_r.$$

Let \hat{V}_{NT_0} denote the corresponding diagonal matrix of eigenvalues, defined by

$$\hat{S} \hat{\Lambda} = \hat{\Lambda} \hat{V}_{NT_0}. \quad (15)$$

Under Assumptions A–D, as $J, T \rightarrow \infty$ with C fixed, the following hold:

(i) **Eigenvalues.** \hat{V}_{NT_0} is invertible with probability approaching one, and

$$\hat{V}_{NT_0} \xrightarrow{p} V,$$

where V is the $r \times r$ diagonal matrix consisting of the r eigenvalues of

$$\Sigma_f^{1/2} \Sigma_\Lambda \Sigma_f^{1/2} \quad (\text{equivalently, of } \Sigma_\Lambda^{1/2} \Sigma_f \Sigma_\Lambda^{1/2}).$$

(ii) **Loading space and explicit rotation.** Define

$$H := \left(\frac{1}{T_0} F_0' F_0 \right) \left(\frac{1}{N} \tilde{\Lambda}'_0 \hat{\Lambda} \right) \hat{V}_{NT_0}^{-1}, \quad F_0 := (f_{p+1}^0, \dots, f_T^0)'$$

Then H is invertible with probability approaching one, and

$$\frac{1}{N} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F^2 = O_p(\|\hat{\beta} - \beta_0\|^2) + O_p\left(\frac{1}{\min\{N, T_0\}}\right).$$

In particular, if $\hat{\beta} \xrightarrow{p} \beta_0$, then

$$\frac{1}{N} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F^2 = o_p(1), \quad \text{and hence} \quad \|P_{\hat{\Lambda}} - \Pi_0\| \xrightarrow{p} 0,$$

where

$$P_{\hat{\Lambda}} := N^{-1} \hat{\Lambda} \hat{\Lambda}', \quad \Pi_0 := \tilde{\Lambda}_0 (\tilde{\Lambda}'_0 \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}'_0.$$

Proof of Proposition 7.3. Let $T_0 := T - p$ and $N := CJ$. Define

$$u_t := \tilde{Y}_t - \tilde{W}_t \hat{\beta} = \tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t - \tilde{W}_t (\hat{\beta} - \beta_0), \quad t = p+1, \dots, T.$$

Write

$$s_t := \tilde{\Lambda}_0 f_t^0, \quad \varepsilon_t := \tilde{\varepsilon}_t, \quad r_t := -\tilde{W}_t (\hat{\beta} - \beta_0),$$

so that $u_t = s_t + \varepsilon_t + r_t$. Recall

$$\hat{S} = \frac{1}{NT_0} \sum_{t=p+1}^T u_t u_t', \quad \hat{S} \hat{\Lambda} = \hat{\Lambda} \hat{V}_{NT_0}, \quad \frac{1}{N} \hat{\Lambda}' \hat{\Lambda} = I_r.$$

Step 1 (Signal decomposition and a spectral bound). Define the signal matrix

$$S_f := \frac{1}{NT_0} \sum_{t=p+1}^T s_t s_t' = \frac{1}{N} \tilde{\Lambda}_0 \left(\frac{1}{T_0} F_0' F_0 \right) \tilde{\Lambda}_0', \quad F_0 := (f_{p+1}^0, \dots, f_T^0)'. \quad (16)$$

Then

$$\hat{S} = S_f + R,$$

where

$$R = \frac{1}{NT_0} \sum_{t=p+1}^T \left(s_t \varepsilon_t' + \varepsilon_t s_t' + s_t r_t' + r_t s_t' + \varepsilon_t r_t' + r_t \varepsilon_t' + \varepsilon_t \varepsilon_t' + r_t r_t' \right).$$

Claim. Under Assumptions A–D,

$$\|R\| = O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\sqrt{T_0}}\right) + O_p\left(\frac{1}{N}\right). \quad (17)$$

Proof of the claim. We bound each block in operator norm.

(a) Terms involving r_t . By Cauchy–Schwarz and $\|ab'\| \leq \|a\| \|b\|$,

$$\left\| \frac{1}{NT_0} \sum s_t r_t' \right\| \leq \left(\frac{1}{NT_0} \sum \|s_t\|^2 \right)^{1/2} \left(\frac{1}{NT_0} \sum \|r_t\|^2 \right)^{1/2}.$$

By (16),

$$\frac{1}{NT_0} \sum \|s_t\|^2 = \frac{1}{N} \text{tr} \left(\tilde{\Lambda}_0' \tilde{\Lambda}_0 \cdot \frac{1}{T_0} F_0' F_0 \right) = O(1)$$

using Assumption B(i)–(ii). Moreover,

$$\frac{1}{NT_0} \sum \|r_t\|^2 \leq \|\hat{\beta} - \beta_0\|^2 \cdot \frac{1}{NT_0} \sum \|\tilde{W}_t\|_F^2 = O_p(1) \|\hat{\beta} - \beta_0\|^2,$$

because $K = \dim(\beta) = pC^2$ is fixed and, by Assumption A(i) together with Assumptions B and C, $E\|\tilde{W}_t\|_F^2 = O(N)$ uniformly in t . Hence

$$\left\| \frac{1}{NT_0} \sum s_t r_t' \right\| = O_p(\|\hat{\beta} - \beta_0\|).$$

The same bound holds for $\frac{1}{NT_0} \sum r_t s_t'$, $\frac{1}{NT_0} \sum \varepsilon_t r_t'$, and $\frac{1}{NT_0} \sum r_t \varepsilon_t'$. Finally,

$$\left\| \frac{1}{NT_0} \sum r_t r_t' \right\| \leq \frac{1}{NT_0} \sum \|r_t\|^2 = O_p(\|\hat{\beta} - \beta_0\|^2).$$

(b) Cross terms $s_t \varepsilon'_t$ and $\varepsilon_t s'_t$. Since $s_t = \tilde{\Lambda}_0 f_t^0$ is deterministic and $E(\varepsilon_t) = 0$ by Assumption C(ii), we have $E(s_t \varepsilon'_t) = 0$. Moreover, the summands are independent over t by Assumption C(ii). Using $\|A\| \leq \|A\|_F$,

$$E \left\| \frac{1}{NT_0} \sum s_t \varepsilon'_t \right\|_F^2 = \frac{1}{N^2 T_0^2} \sum E(\|s_t\|^2 \|\varepsilon_t\|^2).$$

Moreover,

$$\|s_t\|^2 = \|\tilde{\Lambda}_0 f_t^0\|^2 \leq \|\tilde{\Lambda}_0\|^2 \|f_t^0\|^2 = O(N)$$

uniformly in t , by Assumption B(i)–(ii), and $E\|\varepsilon_t\|^2 = O(N)$ by Assumption C(i). Therefore

$$E \left\| \frac{1}{NT_0} \sum s_t \varepsilon'_t \right\|_F^2 = O\left(\frac{1}{T_0}\right),$$

and hence

$$\left\| \frac{1}{NT_0} \sum s_t \varepsilon'_t \right\| = O_p\left(\frac{1}{\sqrt{T_0}}\right).$$

The same bound holds for $\frac{1}{NT_0} \sum \varepsilon_t s'_t$.

(c) Pure error term. By Assumption C(ii),

$$E\left(\frac{1}{NT_0} \sum \varepsilon_t \varepsilon'_t\right) = \frac{1}{N} \cdot \frac{1}{T_0} \sum_{t=p+1}^T \Sigma_{\varepsilon,t},$$

so by Assumption C(iii),

$$\left\| E\left(\frac{1}{NT_0} \sum \varepsilon_t \varepsilon'_t\right) \right\| \leq \frac{K}{N}.$$

Moreover,

$$E \left\| \frac{1}{NT_0} \sum (\varepsilon_t \varepsilon'_t - E(\varepsilon_t \varepsilon'_t)) \right\|_F^2 = O\left(\frac{1}{T_0}\right),$$

so the centered part is $O_p(T_0^{-1/2})$ in operator norm. Hence

$$\left\| \frac{1}{NT_0} \sum \varepsilon_t \varepsilon'_t \right\| = O_p\left(\frac{1}{\sqrt{T_0}}\right) + O\left(\frac{1}{N}\right).$$

Collecting (a)–(c) proves (17). □

Step 2 (Eigenvalues). Let $\mu_1(\cdot) \geq \dots \geq \mu_N(\cdot)$ denote ordered eigenvalues. By Weyl's inequality (e.g., [Horn and Johnson, 2012](#), Theorem 4.3.1),

$$\max_{1 \leq j \leq r} |\mu_j(\hat{S}) - \mu_j(S_f)| \leq \|\hat{S} - S_f\| = \|R\|.$$

Thus, by (17),

$$\max_{1 \leq j \leq r} |\mu_j(\hat{S}) - \mu_j(S_f)| = O_p(\|\hat{\beta} - \beta_0\|) + O_p(T_0^{-1/2}) + O_p(N^{-1}). \quad (18)$$

The r nonzero eigenvalues of S_f are the eigenvalues of the $r \times r$ matrix

$$\left(\frac{1}{T_0} F_0' F_0\right)^{1/2} \left(\frac{1}{N} \tilde{\Lambda}'_0 \tilde{\Lambda}_0\right) \left(\frac{1}{T_0} F_0' F_0\right)^{1/2}.$$

By Assumption B(i)–(ii),

$$\frac{1}{T_0} F_0' F_0 \rightarrow \Sigma_f \succ 0, \quad \frac{1}{N} \tilde{\Lambda}_0' \tilde{\Lambda}_0 \rightarrow \Sigma_\Lambda \succ 0,$$

so the r nonzero eigenvalues of S_f converge to those of

$$\Sigma_f^{1/2} \Sigma_\Lambda \Sigma_f^{1/2}.$$

Denote by V the diagonal matrix collecting these limiting eigenvalues. Since Proposition 7.2(i) gives $\hat{\beta} \xrightarrow{p} \beta_0$, combining this with (18) yields

$$\hat{V}_{NT_0} \xrightarrow{p} V.$$

In particular, since $V \succ 0$, \hat{V}_{NT_0} is invertible with probability approaching one.

Step 3 (Rotation and loading error bound). From $\hat{S} = S_f + R$ and $\hat{S}\hat{\Lambda} = \hat{\Lambda}\hat{V}_{NT_0}$,

$$S_f \hat{\Lambda} + R \hat{\Lambda} = \hat{\Lambda} \hat{V}_{NT_0}.$$

Using (16),

$$\hat{\Lambda} \hat{V}_{NT_0} - \tilde{\Lambda}_0 \left(\frac{1}{T_0} F_0' F_0 \right) \left(\frac{1}{N} \tilde{\Lambda}_0' \hat{\Lambda} \right) = R \hat{\Lambda}. \quad (19)$$

Define

$$H := \left(\frac{1}{T_0} F_0' F_0 \right) \left(\frac{1}{N} \tilde{\Lambda}_0' \hat{\Lambda} \right) \hat{V}_{NT_0}^{-1}.$$

Right-multiplying (19) by $\hat{V}_{NT_0}^{-1}$ yields

$$\hat{\Lambda} - \tilde{\Lambda}_0 H = R \hat{\Lambda} \hat{V}_{NT_0}^{-1}. \quad (20)$$

Taking Frobenius norms and using $\|ABC\|_F \leq \|A\| \|B\|_F \|C\|$,

$$\frac{1}{\sqrt{N}} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F \leq \|R\| \cdot \frac{1}{\sqrt{N}} \|\hat{\Lambda}\|_F \cdot \|\hat{V}_{NT_0}^{-1}\|.$$

Since $N^{-1} \hat{\Lambda}' \hat{\Lambda} = I_r$, we have $\|\hat{\Lambda}\|_F = \sqrt{Nr}$, and hence

$$\frac{1}{\sqrt{N}} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F \leq \sqrt{r} \|R\| \|\hat{V}_{NT_0}^{-1}\|.$$

By Step 2, $\|\hat{V}_{NT_0}^{-1}\| = O_p(1)$. Combining this with (17) gives

$$\frac{1}{\sqrt{N}} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F = O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\sqrt{T_0}}\right) + O_p\left(\frac{1}{N}\right).$$

Squaring both sides yields

$$\frac{1}{N} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F^2 = O_p(\|\hat{\beta} - \beta_0\|^2) + O_p\left(\frac{1}{T_0}\right) + O_p\left(\frac{1}{N^2}\right) = O_p(\|\hat{\beta} - \beta_0\|^2) + O_p\left(\frac{1}{\min\{N, T_0\}}\right).$$

Step 4 (Invertibility of H and projection consistency). By Proposition 7.2(i), $\hat{\beta} \xrightarrow{p} \beta_0$. Hence, by (17), $\|\hat{S} - S_f\| = o_p(1)$.

By Assumption B(i), $T_0^{-1}F_0'F_0$ is invertible with probability approaching one, and by Step 2, \hat{V}_{NT_0} is invertible with probability approaching one. It therefore remains to show that $N^{-1}\tilde{\Lambda}'_0\hat{\Lambda}$ is invertible with probability approaching one.

The r nonzero eigenvalues of S_f are separated from the remaining eigenvalues, since the r th eigenvalue converges to a strictly positive limit while the $(r+1)$ th eigenvalue is zero. Hence, by the Davis–Kahan eigenspace perturbation theorem (see Davis and Kahan, 1970), the principal angles between the column spaces of $\hat{\Lambda}$ and $\tilde{\Lambda}_0$ converge to zero in probability. In particular, $N^{-1}\tilde{\Lambda}'_0\hat{\Lambda}$ is invertible with probability approaching one, and therefore H is invertible with probability approaching one.

Since H is nonsingular and

$$\frac{1}{N}\|\hat{\Lambda} - \tilde{\Lambda}_0H\|_F^2 = o_p(1),$$

the column space of $\hat{\Lambda}$ converges to that of $\tilde{\Lambda}_0$. Equivalently,

$$\|P_{\hat{\Lambda}} - \Pi_0\| \xrightarrow{p} 0, \quad P_{\hat{\Lambda}} := \frac{1}{N}\hat{\Lambda}\hat{\Lambda}', \quad \Pi_0 := \tilde{\Lambda}_0(\tilde{\Lambda}'_0\tilde{\Lambda}_0)^{-1}\tilde{\Lambda}'_0.$$

This completes the proof. \square

Lemma 7.4. *Under Assumptions A–D, there exists a finite constant $M < \infty$ such that*

$$E\left\|T_0^{-1/2} \sum_{t=p+1}^T \frac{1}{N} \tilde{\Lambda}'_0 \left(\tilde{\varepsilon}_t \tilde{\varepsilon}'_t - \Sigma_{\varepsilon,t} \right) \tilde{\Lambda}_0\right\|_F^2 \leq M, \quad \Sigma_{\varepsilon,t} := E(\tilde{\varepsilon}_t \tilde{\varepsilon}'_t).$$

Proof of Lemma 7.4. Define the $r \times r$ random matrix

$$Z_t := \frac{1}{N} \tilde{\Lambda}'_0 \left(\tilde{\varepsilon}_t \tilde{\varepsilon}'_t - \Sigma_{\varepsilon,t} \right) \tilde{\Lambda}_0, \quad t = p+1, \dots, T.$$

Let $a_k := N^{-1/2}\tilde{\Lambda}_{0,\cdot k}$ be the k th normalized loading vector. Then the (k, ℓ) element of Z_t is

$$(Z_t)_{k\ell} = (a'_k \tilde{\varepsilon}_t)(a'_\ell \tilde{\varepsilon}_t) - E[(a'_k \tilde{\varepsilon}_t)(a'_\ell \tilde{\varepsilon}_t)].$$

By Assumption C(ii), $\{Z_t\}$ are independent across t , and $E(Z_t) = 0$. Hence,

$$E\left\|T_0^{-1/2} \sum_{t=p+1}^T Z_t\right\|_F^2 = \frac{1}{T_0} \sum_{t=p+1}^T E\|Z_t\|_F^2.$$

Since r is fixed and $\|Z_t\|_F^2 = \sum_{k,\ell \leq r} (Z_t)_{k\ell}^2$, it suffices to bound $E|(Z_t)_{k\ell}|^2$ uniformly in t . Write

$$a_k = \|a_k\|u_k, \quad a_\ell = \|a_\ell\|u_\ell,$$

where u_k, u_ℓ are deterministic unit vectors whenever $a_k \neq 0, a_\ell \neq 0$. By Assumption B(ii),

$$\|a_k\|^2 = \frac{1}{N}\|\tilde{\Lambda}_{0,\cdot k}\|^2 = O(1), \quad \|a_\ell\|^2 = \frac{1}{N}\|\tilde{\Lambda}_{0,\cdot \ell}\|^2 = O(1).$$

Therefore, using $\mathbb{V}\mathcal{D}\setminus(X) \leq E(X^2)$,

$$\begin{aligned} E|(Z_t)_{k\ell}|^2 &\leq E\left[(a'_k \tilde{\varepsilon}_t)^2 (a'_\ell \tilde{\varepsilon}_t)^2\right] \\ &= \|a_k\|^2 \|a_\ell\|^2 E\left[(u'_k \tilde{\varepsilon}_t)^2 (u'_\ell \tilde{\varepsilon}_t)^2\right] \leq K, \end{aligned}$$

where the last inequality follows from Assumption C(v) and the uniform boundedness of $\|a_k\|, \|a_\ell\|$.

Consequently,

$$\sup_t E\|Z_t\|_F^2 \leq \sum_{k,\ell \leq r} \sup_t E|(Z_t)_{k\ell}|^2 \leq r^2 K < \infty.$$

Hence

$$E\left\|T_0^{-1/2} \sum_{t=p+1}^T Z_t\right\|_F^2 \leq \sup_t E\|Z_t\|_F^2 \leq r^2 K =: M.$$

□

Lemma 7.5. *Under Assumptions A–D, let*

$$H := \left(\frac{1}{T_0} F_0' F_0\right) \left(\frac{1}{N} \tilde{\Lambda}'_0 \hat{\Lambda}\right) \hat{V}_{NT_0}^{-1}, \quad F_0 := (f_{p+1}^0, \dots, f_T^0)'$$

Then H is nonsingular with probability approaching one, and the following hold:

(i)

$$\left\| \frac{1}{N} \tilde{\Lambda}'_0 (\hat{\Lambda} - \tilde{\Lambda}_0 H) \right\|_F = O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T_0}\}}\right).$$

(ii)

$$\left\| \frac{1}{N} (\hat{\Lambda} - \tilde{\Lambda}_0 H)' (\hat{\Lambda} - \tilde{\Lambda}_0 H) \right\|_F = O_p(\|\hat{\beta} - \beta_0\|^2) + O_p\left(\frac{1}{\min\{N, T_0\}}\right).$$

(iii) For any fixed $t \in \{p+1, \dots, T\}$,

$$\left\| \frac{1}{N} \tilde{Y}'_t (\hat{\Lambda} - \tilde{\Lambda}_0 H) \right\| = O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T_0}\}}\right).$$

(iv)

$$\left\| \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} (\hat{\Lambda} - \tilde{\Lambda}_0 H) \right\|_F = O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T_0}\}}\right).$$

Proof of Lemma 7.5. By Proposition 7.3(ii), H is nonsingular with probability approaching one and

$$\frac{1}{N} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F^2 = O_p(\|\hat{\beta} - \beta_0\|^2) + O_p\left(\frac{1}{\min\{N, T_0\}}\right). \quad (21)$$

In particular,

$$\frac{1}{\sqrt{N}} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F = O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T_0}\}}\right). \quad (22)$$

(ii). This is exactly (21), since r is fixed.

(i). By Cauchy–Schwarz,

$$\left\| \frac{1}{N} \tilde{\Lambda}'_0 (\hat{\Lambda} - \tilde{\Lambda}_0 H) \right\|_F \leq \frac{1}{N} \|\tilde{\Lambda}_0\|_F \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F.$$

Under Assumption B(ii),

$$N^{-1} \|\tilde{\Lambda}_0\|_F^2 = \text{tr}\left(N^{-1} \tilde{\Lambda}'_0 \tilde{\Lambda}_0\right) = O(1),$$

so $\|\tilde{\Lambda}_0\|_F = O_p(\sqrt{N})$. Combining this with (22) yields

$$\left\| \frac{1}{N} \tilde{\Lambda}'_0 (\hat{\Lambda} - \tilde{\Lambda}_0 H) \right\|_F = O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T_0}\}}\right).$$

(iii). Fix t . Decompose $\tilde{Y}_t = \tilde{W}_t \beta_0 + \tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t$ and write

$$\frac{1}{N} \tilde{Y}'_t (\hat{\Lambda} - \tilde{\Lambda}_0 H) = \frac{1}{N} (\tilde{W}_t \beta_0)' (\hat{\Lambda} - \tilde{\Lambda}_0 H) + \frac{1}{N} (\tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t)' (\hat{\Lambda} - \tilde{\Lambda}_0 H).$$

By Cauchy–Schwarz,

$$\left\| \frac{1}{N} (\tilde{W}_t \beta_0)' (\hat{\Lambda} - \tilde{\Lambda}_0 H) \right\| \leq \frac{1}{N} \|\tilde{W}_t\|_F \|\beta_0\| \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F.$$

For fixed t , Assumption A(i), together with Assumptions B and C, implies $\|\tilde{W}_t\|_F = O_p(\sqrt{N})$, and (22) yields the claimed rate for this term.

For the second term, again by Cauchy–Schwarz,

$$\left\| \frac{1}{N} (\tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t)' (\hat{\Lambda} - \tilde{\Lambda}_0 H) \right\| \leq \frac{1}{N} \|\tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t\| \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F.$$

For fixed t , Assumptions B(i) and C(i)–(iii) imply

$$\|\tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t\| = O_p(\sqrt{N}),$$

and (22) again yields the claimed rate. Combining the two bounds proves (iii).

(iv). Using $\|M_{\hat{\Lambda}}\| \leq 1$ and Cauchy–Schwarz,

$$\left\| \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} (\hat{\Lambda} - \tilde{\Lambda}_0 H) \right\|_F \leq \left(\frac{1}{NT_0} \sum_{t=p+1}^T \|\tilde{W}_t\|_F^2 \right)^{1/2} \left(\frac{1}{N} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F^2 \right)^{1/2}.$$

Assumption A(i), together with Assumptions B and C, implies

$$\frac{1}{NT_0} \sum_{t=p+1}^T \|\tilde{W}_t\|_F^2 = O_p(1),$$

and (21) gives

$$\left(\frac{1}{N} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F^2 \right)^{1/2} = O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T_0}\}}\right),$$

which yields (iv). □

Lemma 7.6. *Under Assumptions A–D, the loading error bounds in Lemma 7.5 imply that the following hold:*

(i)

$$\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}'_t (\hat{\Lambda} - \tilde{\Lambda}_0 H) = o_p(1) \quad (\text{as a } 1 \times r \text{ row vector}).$$

(ii)

$$\frac{1}{NT_0} \sum_{t=p+1}^T f_t^{0'} \tilde{\Lambda}'_0 (\hat{\Lambda} - \tilde{\Lambda}_0 H) \tilde{\varepsilon}_t = o_p(1) \quad (\text{as a scalar}).$$

(iii)

$$\left\| \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t (\hat{\Lambda} - \tilde{\Lambda}_0 H)' \tilde{\Lambda}_0 f_t^0 \right\| = o_p(1) \quad (\text{Euclidean norm in } \mathbb{R}^N).$$

Proof of Lemma 7.6. We prove parts (i)–(iii) in turn. By Proposition 7.2(i), $\hat{\beta} \xrightarrow{p} \beta_0$. Hence Lemma 7.5(ii) implies

$$\frac{1}{N} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F^2 = o_p(1).$$

(i). By Cauchy–Schwarz,

$$\left\| \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t' (\hat{\Lambda} - \tilde{\Lambda}_0 H) \right\| \leq \left(\frac{1}{NT_0} \sum_{t=p+1}^T \|\tilde{\varepsilon}_t\|^2 \right)^{1/2} \left(\frac{1}{N} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F^2 \right)^{1/2}.$$

By Assumption C(i),

$$\frac{1}{NT_0} \sum_{t=p+1}^T \|\tilde{\varepsilon}_t\|^2 = O_p(1),$$

and by Lemma 7.5(ii),

$$\frac{1}{N} \|\hat{\Lambda} - \tilde{\Lambda}_0 H\|_F^2 = o_p(1).$$

Hence the product is $o_p(1)$.

(ii). Write each summand as

$$f_t^{0'} \tilde{\Lambda}'_0 (\hat{\Lambda} - \tilde{\Lambda}_0 H) \tilde{\varepsilon}_t = \{(\hat{\Lambda} - \tilde{\Lambda}_0 H)' \tilde{\Lambda}_0 f_t^0\}' \tilde{\varepsilon}_t.$$

By Cauchy–Schwarz,

$$\left| \frac{1}{NT_0} \sum_{t=p+1}^T \{(\hat{\Lambda} - \tilde{\Lambda}_0 H)' \tilde{\Lambda}_0 f_t^0\}' \tilde{\varepsilon}_t \right| \leq \left(\frac{1}{NT_0} \sum_{t=p+1}^T \|\tilde{\varepsilon}_t\|^2 \right)^{1/2} \left(\frac{1}{NT_0} \sum_{t=p+1}^T \|(\hat{\Lambda} - \tilde{\Lambda}_0 H)' \tilde{\Lambda}_0 f_t^0\|^2 \right)^{1/2}.$$

For the second factor, use $\|A'x\| \leq \|A\| \|x\|$ and $\|\tilde{\Lambda}_0 f_t^0\| \leq \|\tilde{\Lambda}_0\| \|f_t^0\|$:

$$\|(\hat{\Lambda} - \tilde{\Lambda}_0 H)' \tilde{\Lambda}_0 f_t^0\| \leq \|\hat{\Lambda} - \tilde{\Lambda}_0 H\| \|\tilde{\Lambda}_0\| \|f_t^0\|.$$

Under Assumption B(ii), $\|\tilde{\Lambda}_0\| = O(\sqrt{N})$, and Lemma 7.5(ii) implies

$$\|\hat{\Lambda} - \tilde{\Lambda}_0 H\| = o_p(\sqrt{N}),$$

since r is fixed and $\|\cdot\| \leq \|\cdot\|_F$. Hence

$$\frac{1}{NT_0} \sum_{t=p+1}^T \|(\hat{\Lambda} - \tilde{\Lambda}_0 H)' \tilde{\Lambda}_0 f_t^0\|^2 \leq \frac{\|\hat{\Lambda} - \tilde{\Lambda}_0 H\|^2 \|\tilde{\Lambda}_0\|^2}{N} \cdot \left(\frac{1}{T_0} \sum_{t=p+1}^T \|f_t^0\|^2 \right) = o_p(1),$$

where Assumption B(i) gives $T_0^{-1} \sum_t \|f_t^0\|^2 = O(1)$. Combining this with Assumption C(i) proves (ii).

(iii). By Cauchy–Schwarz,

$$\left\| \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t (\hat{\Lambda} - \tilde{\Lambda}_0 H)' \tilde{\Lambda}_0 f_t^0 \right\| \leq \left(\frac{1}{NT_0} \sum_{t=p+1}^T \|\tilde{\varepsilon}_t\|^2 \right)^{1/2} \left(\frac{1}{NT_0} \sum_{t=p+1}^T \|(\hat{\Lambda} - \tilde{\Lambda}_0 H)' \tilde{\Lambda}_0 f_t^0\|^2 \right)^{1/2}.$$

The first factor is $O_p(1)$ by Assumption C(i), and the second factor is $o_p(1)$ by the bound established in the proof of part (ii). Hence (iii) follows. \square

Lemma 7.7. *Under Assumptions C(i)–(ii) and r fixed,*

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} \left| \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t' P_{\tilde{\Lambda}} \tilde{\varepsilon}_t - \frac{1}{NT_0} \sum_{t=p+1}^T E[\tilde{\varepsilon}_t' P_{\tilde{\Lambda}} \tilde{\varepsilon}_t] \right| = o_p(1), \quad P_{\tilde{\Lambda}} := N^{-1} \tilde{\Lambda} \tilde{\Lambda}'.$$

Proof of Lemma 7.7. For $\tilde{\Lambda} \in \mathcal{F}$, write $P_{\tilde{\Lambda}} = UU'$ with

$$U := N^{-1/2} \tilde{\Lambda} \in \mathbb{R}^{N \times r}, \quad U'U = I_r.$$

Define

$$A := \frac{1}{T_0} \sum_{t=p+1}^T \left(\tilde{\varepsilon}_t \tilde{\varepsilon}_t' - E[\tilde{\varepsilon}_t \tilde{\varepsilon}_t'] \right).$$

Then

$$\frac{1}{NT_0} \sum_{t=p+1}^T \left(\tilde{\varepsilon}_t' P_{\tilde{\Lambda}} \tilde{\varepsilon}_t - E[\tilde{\varepsilon}_t' P_{\tilde{\Lambda}} \tilde{\varepsilon}_t] \right) = \frac{1}{N} \text{tr}(P_{\tilde{\Lambda}} A) = \frac{1}{N} \text{tr}(U'AU).$$

Hence,

$$\sup_{\tilde{\Lambda} \in \mathcal{F}} \frac{1}{N} |\text{tr}(U'AU)| \leq \sup_{U'U=I_r} \frac{1}{N} \|U'AU\|_F \|I_r\|_F \leq \frac{\sqrt{r}}{N} \|A\|_F,$$

where we used $\|U'AU\|_F \leq \|A\|_F$ for $U'U = I_r$.

Let

$$B_t := \tilde{\varepsilon}_t \tilde{\varepsilon}_t' - E[\tilde{\varepsilon}_t \tilde{\varepsilon}_t'].$$

Then $A = T_0^{-1} \sum_{t=p+1}^T B_t$, $E(B_t) = 0$, and by Assumption C(ii), $\{B_t\}$ are independent across t . Therefore,

$$E\|A\|_F^2 = \frac{1}{T_0^2} \sum_{t=p+1}^T E\|B_t\|_F^2.$$

Moreover,

$$E\|B_t\|_F^2 \leq E\|\tilde{\varepsilon}_t \tilde{\varepsilon}_t'\|_F^2 = \sum_{i=1}^N \sum_{j=1}^N E(\tilde{\varepsilon}_{it}^2 \tilde{\varepsilon}_{jt}^2).$$

By Hölder's inequality and Assumption C(i),

$$E(\tilde{\varepsilon}_{it}^2 \tilde{\varepsilon}_{jt}^2) \leq (E|\tilde{\varepsilon}_{it}|^4)^{1/2} (E|\tilde{\varepsilon}_{jt}|^4)^{1/2} \leq K,$$

uniformly in t . Hence $E\|B_t\|_F^2 \leq KN^2$, and thus

$$E\|A\|_F^2 \leq \frac{1}{T_0^2} \sum_{t=p+1}^T KN^2 = \frac{K}{T_0} N^2.$$

It follows that $\|A\|_F = O_p(N/\sqrt{T_0})$, and therefore

$$\sup_{\hat{\Lambda} \in \mathcal{F}} \frac{1}{N} |\text{tr}(U'AU)| \leq \frac{\sqrt{r}}{N} O_p\left(\frac{N}{\sqrt{T_0}}\right) = O_p\left(\frac{1}{\sqrt{T_0}}\right) = o_p(1).$$

This proves the lemma. \square

Lemma 7.8 (Projected regressor average bound). *Let $\{b_t\}_{t=p+1}^T$ be a sequence of N -dimensional non-stochastic vectors such that*

$$\sup_{p+1 \leq t \leq T} \|b_t\| = O(1).$$

Define

$$\Gamma_T(b) := \frac{1}{T_0} \sum_{t=p+1}^T b_t \tilde{W}_t' M_{\hat{\Lambda}}, \quad T_0 := T - p.$$

Under Assumptions A–D,

$$\frac{1}{\sqrt{N}} \|\Gamma_T(b)\| = O_p(1).$$

Proof. By the Cauchy–Schwarz inequality,

$$\|\Gamma_T(b)\| \leq \left(\frac{1}{T_0} \sum_{t=p+1}^T \|b_t\|^2 \right)^{1/2} \left(\frac{1}{T_0} \sum_{t=p+1}^T \|M_{\hat{\Lambda}} \tilde{W}_t\|_F^2 \right)^{1/2}.$$

Since $\sup_t \|b_t\| = O(1)$, the first factor is $O(1)$. Also,

$$\|M_{\hat{\Lambda}} \tilde{W}_t\|_F \leq \|\tilde{W}_t\|_F$$

because $M_{\hat{\Lambda}}$ is an orthogonal projection matrix and hence $\|M_{\hat{\Lambda}}\| \leq 1$. Therefore,

$$\frac{1}{N} \frac{1}{T_0} \sum_{t=p+1}^T \|M_{\hat{\Lambda}} \tilde{W}_t\|_F^2 \leq \frac{1}{N} \frac{1}{T_0} \sum_{t=p+1}^T \|\tilde{W}_t\|_F^2.$$

By Assumption A(i) and Assumptions B–D, the right-hand side is $O_p(1)$. Hence

$$\frac{1}{\sqrt{N}} \left(\frac{1}{T_0} \sum_{t=p+1}^T \|M_{\hat{\Lambda}} \tilde{W}_t\|_F^2 \right)^{1/2} = O_p(1).$$

Combining the two bounds yields

$$\frac{1}{\sqrt{N}} \|\Gamma_T(b)\| = O_p(1),$$

as claimed. \square

Proposition 7.9 (Linear expansion of $\hat{\beta}$). *Under Assumptions A–D and $N/T_0 \rightarrow \rho \geq 0$, we have*

$$\begin{aligned} \sqrt{NT_0}(\hat{\beta} - \beta_0) &= D(\hat{\Lambda})^{-1} \left[\frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \tilde{\varepsilon}_t - \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\hat{\Lambda}} \tilde{\varepsilon}_t \right] \\ &\quad + \sqrt{\frac{N}{T_0}} \zeta_{NT_0} + o_p(1), \end{aligned} \quad (23)$$

where $N = CJ$, $T_0 = T - p$, and

$$M_{\hat{\Lambda}} = I_N - N^{-1} \hat{\Lambda} \hat{\Lambda}'.$$

Define

$$a_{ts} := f_t^{0'} \left(\frac{1}{T_0} F_0' F_0 \right)^{-1} f_s^0, \quad F_0 := (f_{p+1}^0, \dots, f_T^0)'$$

For any $\tilde{\Lambda} \in \mathcal{F} := \{\Lambda \in \mathbb{R}^{N \times r} : (1/N)\Lambda' \Lambda = I_r\}$, define

$$Z_t(\tilde{\Lambda}) := M_{\tilde{\Lambda}} \tilde{W}_t - \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} M_{\tilde{\Lambda}} \tilde{W}_s, \quad D(\tilde{\Lambda}) := \frac{1}{NT_0} \sum_{t=p+1}^T Z_t(\tilde{\Lambda})' Z_t(\tilde{\Lambda}),$$

as in Assumption A(ii). In (23), $D(\hat{\Lambda})$ denotes $D(\tilde{\Lambda})$ evaluated at $\tilde{\Lambda} = \hat{\Lambda}$, and ζ_{NT_0} collects the remaining bias terms in the expansion.

Proof of Proposition 7.9. Write $T_0 := T - p$ and

$$M_{\hat{\Lambda}} = I_N - N^{-1} \hat{\Lambda} \hat{\Lambda}'.$$

The first-order condition of $Q(\beta, \tilde{\Lambda})$ with respect to β , evaluated at $(\hat{\beta}, \hat{\Lambda})$, is

$$0 = \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} (\tilde{Y}_t - \tilde{W}_t \hat{\beta}).$$

Using

$$\tilde{Y}_t = \tilde{W}_t \beta_0 + \tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t,$$

we obtain

$$\left(\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \tilde{W}_t \right) (\hat{\beta} - \beta_0) = \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \tilde{\Lambda}_0 f_t^0 + \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \tilde{\varepsilon}_t. \quad (24)$$

Step 1 (Rewriting the factor term). Fix H as in Proposition 7.3(ii), namely,

$$H = \left(\frac{1}{T_0} F_0' F_0 \right) \left(\frac{1}{N} \tilde{\Lambda}_0' \hat{\Lambda} \right) \hat{V}_{NT_0}^{-1}, \quad F_0 := (f_{p+1}^0, \dots, f_T^0)'$$

Define

$$R := \tilde{\Lambda}_0 - \hat{\Lambda} H^{-1}.$$

Since $M_{\hat{\Lambda}} \hat{\Lambda} = 0$, we have

$$M_{\hat{\Lambda}} \tilde{\Lambda}_0 f_t^0 = M_{\hat{\Lambda}} (\tilde{\Lambda}_0 - \hat{\Lambda} H^{-1}) f_t^0 = M_{\hat{\Lambda}} R f_t^0.$$

Therefore,

$$\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \tilde{\Lambda}_0 f_t^0 = \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} R f_t^0. \quad (25)$$

Moreover, Proposition 7.3(ii) gives

$$\frac{1}{N} \|R\|_F^2 = \frac{1}{N} \|\tilde{\Lambda}_0 - \hat{\Lambda} H^{-1}\|_F^2 = O_p(\|\hat{\beta} - \beta_0\|^2) + O_p\left(\frac{1}{\min\{N, T_0\}}\right). \quad (26)$$

In particular,

$$\|R\|_F = O_p(\sqrt{N} \|\hat{\beta} - \beta_0\|) + O_p\left(\sqrt{\frac{N}{\min\{N, T_0\}}}\right) = o_p(\sqrt{N}),$$

where the last relation uses Proposition 7.2(i).

Step 2 (Expansion of R). Define the estimated factor matrix

$$\hat{F} := (\hat{f}_{p+1}, \dots, \hat{f}_T)', \quad \hat{f}_t = \frac{1}{N} \hat{\Lambda}' (\tilde{Y}_t - \tilde{W}_t \hat{\beta}), \quad t = p+1, \dots, T.$$

Start from the loading eigen-equation

$$\hat{S} \hat{\Lambda} = \hat{\Lambda} \hat{V}_{NT_0}, \quad \hat{S} := \frac{1}{NT_0} \sum_{t=p+1}^T u_t u_t', \quad u_t := \tilde{Y}_t - \tilde{W}_t \hat{\beta}.$$

Using

$$u_t = \tilde{\Lambda}_0 f_t^0 + \tilde{\varepsilon}_t - \tilde{W}_t (\hat{\beta} - \beta_0),$$

expand $u_t u_t'$ and collect terms to obtain

$$\hat{\Lambda} \hat{V}_{NT_0} - \tilde{\Lambda}_0 \left(\frac{1}{T_0} F_0' F_0 \right) \left(\frac{1}{N} \tilde{\Lambda}_0' \hat{\Lambda} \right) = \sum_{m=1}^8 \mathcal{J}_m, \quad (27)$$

where

$$\begin{aligned}
\mathcal{J}_1 &:= \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}_t(\hat{\beta} - \beta_0)(\hat{\beta} - \beta_0)' \tilde{W}_t' \hat{\Lambda}, \\
\mathcal{J}_2 &:= -\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}_t(\hat{\beta} - \beta_0) f_t^{0'} \tilde{\Lambda}_0' \hat{\Lambda}, \\
\mathcal{J}_3 &:= -\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}_t(\hat{\beta} - \beta_0) \tilde{\varepsilon}_t' \hat{\Lambda}, \\
\mathcal{J}_4 &:= -\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\Lambda}_0 f_t^0 (\hat{\beta} - \beta_0)' \tilde{W}_t' \hat{\Lambda}, \\
\mathcal{J}_5 &:= -\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t (\hat{\beta} - \beta_0)' \tilde{W}_t' \hat{\Lambda}, \\
\mathcal{J}_6 &:= \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\Lambda}_0 f_t^0 \tilde{\varepsilon}_t' \hat{\Lambda}, \\
\mathcal{J}_7 &:= \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t f_t^{0'} \tilde{\Lambda}_0' \hat{\Lambda}, \\
\mathcal{J}_8 &:= \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t \tilde{\varepsilon}_t' \hat{\Lambda}.
\end{aligned}$$

Right-multiplying (27) by

$$\left(\frac{1}{N} \tilde{\Lambda}_0' \hat{\Lambda} \right)^{-1} \left(\frac{1}{T_0} F_0' F_0 \right)^{-1},$$

and using

$$H = \left(\frac{1}{T_0} F_0' F_0 \right) \left(\frac{1}{N} \tilde{\Lambda}_0' \hat{\Lambda} \right) \hat{V}_{NT_0}^{-1}, \quad H^{-1} = \hat{V}_{NT_0} \left(\frac{1}{N} \tilde{\Lambda}_0' \hat{\Lambda} \right)^{-1} \left(\frac{1}{T_0} F_0' F_0 \right)^{-1},$$

we obtain

$$R = \tilde{\Lambda}_0 - \hat{\Lambda} H^{-1} = \sum_{m=1}^8 \mathcal{I}_m, \tag{28}$$

where

$$\mathcal{I}_m := -\mathcal{J}_m \left(\frac{1}{N} \tilde{\Lambda}_0' \hat{\Lambda} \right)^{-1} \left(\frac{1}{T_0} F_0' F_0 \right)^{-1}, \quad m = 1, \dots, 8.$$

By Proposition 7.3, Lemmas 7.4, 7.5, 7.6, 7.7, and 7.8, together with Assumptions A–D, the terms $\mathcal{I}_1, \dots, \mathcal{I}_8$ satisfy the bounds required in Step 3: the terms $\mathcal{I}_2, \mathcal{I}_7$, and the mean part of \mathcal{I}_8 generate the leading contributions, while the remaining terms are of smaller order and are absorbed into

$$o_p(1) \|\hat{\beta} - \beta_0\| \quad \text{or} \quad o_p\left(\frac{1}{\sqrt{NT_0}}\right),$$

by the bounds stated above.

Step 3 (Expansion of the projected factor term). Here each \mathcal{I}_m is an $N \times r$ matrix collecting the contribution of the m th term in the decomposition of R in (28). Accordingly, J_m denotes the contribution of \mathcal{I}_m to the projected factor term

$$\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} R f_t^0.$$

Substituting (28) into (25) gives

$$\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \tilde{\Lambda}_0 f_t^0 = J_1 + \cdots + J_8,$$

where

$$J_m := \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \mathcal{I}_m f_t^0, \quad m = 1, \dots, 8.$$

The leading contributions are J_2 , J_7 , and the mean part of J_8 . The remaining terms J_1 , J_3 , J_4 , J_5 , J_6 , together with the centered part of J_8 , are absorbed into

$$o_p(1) \|\hat{\beta} - \beta_0\| \quad \text{or} \quad o_p\left(\frac{1}{\sqrt{NT_0}}\right).$$

For J_2 , substituting the definition of \mathcal{I}_2 yields

$$J_2 = \frac{1}{NT_0} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\hat{\Lambda}} \tilde{W}_t (\hat{\beta} - \beta_0), \quad (29)$$

where

$$a_{ts} := f_t^{0'} \left(\frac{1}{T_0} F_0' F_0 \right)^{-1} f_s^0.$$

For J_7 , substituting the definition of \mathcal{I}_7 yields

$$J_7 = -\frac{1}{NT_0} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\hat{\Lambda}} \tilde{\varepsilon}_t. \quad (30)$$

Next define the average idiosyncratic covariance

$$\Omega := \frac{1}{T_0} \sum_{t=p+1}^T \Omega_t, \quad \Omega_t := E(\tilde{\varepsilon}_t \tilde{\varepsilon}_t').$$

Decompose

$$J_8 = \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{\varepsilon}_t \tilde{\varepsilon}_t' \hat{\Lambda} = \frac{1}{NT_0} \sum_{t=p+1}^T \Omega_t \hat{\Lambda} + \frac{1}{NT_0} \sum_{t=p+1}^T (\tilde{\varepsilon}_t \tilde{\varepsilon}_t' - \Omega_t) \hat{\Lambda},$$

and write accordingly

$$J_8 = J_{8a} + J_{8b},$$

where J_{8a} is the contribution of the average covariance term and J_{8b} is the contribution of the centered remainder.

Define

$$J_{8a} = \frac{1}{\sqrt{NT_0}} \sqrt{\frac{N}{T_0}} \tilde{\zeta}_{NT_0}, \quad (31)$$

where

$$\tilde{\zeta}_{NT_0} := -\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \Omega \hat{\Lambda} \left(\frac{1}{N} \tilde{\Lambda}'_0 \hat{\Lambda} \right)^{-1} \left(\frac{1}{T_0} F'_0 F_0 \right)^{-1} f_t^0.$$

Therefore,

$$\begin{aligned} \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \tilde{\Lambda}_0 f_t^0 &= \frac{1}{NT_0} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\hat{\Lambda}} \tilde{W}_t (\hat{\beta} - \beta_0) \\ &\quad - \frac{1}{NT_0} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\hat{\Lambda}} \tilde{\varepsilon}_t + \frac{1}{\sqrt{NT_0}} \sqrt{\frac{N}{T_0}} \tilde{\zeta}_{NT_0} \\ &\quad + o_p\left(\frac{1}{\sqrt{NT_0}}\right) + o_p(1) \|\hat{\beta} - \beta_0\|. \end{aligned} \quad (32)$$

Step 4 (Conclusion). Substituting (32) into (24) and multiplying both sides by $\sqrt{NT_0}$ give

$$\begin{aligned} &\left[\frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \tilde{W}_t - \frac{1}{NT_0} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\hat{\Lambda}} \tilde{W}_t \right] \sqrt{NT_0} (\hat{\beta} - \beta_0) \\ &= \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \tilde{\varepsilon}_t - \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\hat{\Lambda}} \tilde{\varepsilon}_t \\ &\quad + \sqrt{\frac{N}{T_0}} \tilde{\zeta}_{NT_0} + o_p(1). \end{aligned}$$

By the definition of $Z_t(\hat{\Lambda})$ and $D(\hat{\Lambda})$, a direct expansion shows that the bracketed term on the left-hand side is equal to

$$D(\hat{\Lambda}) + o_p(1).$$

Hence

$$\begin{aligned} [D(\hat{\Lambda}) + o_p(1)] \sqrt{NT_0} (\hat{\beta} - \beta_0) &= \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \tilde{\varepsilon}_t - \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\hat{\Lambda}} \tilde{\varepsilon}_t \\ &\quad + \sqrt{\frac{N}{T_0}} \tilde{\zeta}_{NT_0} + o_p(1). \end{aligned}$$

Left-multiplying by $D(\hat{\Lambda})^{-1}$ and using

$$D(\hat{\Lambda})^{-1} [D(\hat{\Lambda}) + o_p(1)] = I + o_p(1),$$

we obtain (23) with

$$\zeta_{NT_0} := D(\hat{\Lambda})^{-1} \tilde{\zeta}_{NT_0}.$$

□

Lemma 7.10. *Under Assumptions A–D, we have*

$$\zeta_{NT_0} = O_p(1),$$

where ζ_{NT_0} is the bias term appearing in Proposition 7.9.

Proof of Lemma 7.10. Recall that

$$\zeta_{NT_0} := -D(\hat{\Lambda})^{-1} \frac{1}{NT_0} \sum_{t=p+1}^T \tilde{W}'_t M_{\hat{\Lambda}} \Omega \hat{\Lambda} \left(\frac{1}{N} \tilde{\Lambda}'_0 \hat{\Lambda} \right)^{-1} \left(\frac{1}{T_0} F'_0 F_0 \right)^{-1} f_t^0,$$

where

$$\Omega := \frac{1}{T_0} \sum_{t=p+1}^T \Omega_t, \quad \Omega_t := E(\tilde{\varepsilon}_t \tilde{\varepsilon}'_t).$$

Write

$$B_{NT_0} := \left(\frac{1}{N} \tilde{\Lambda}'_0 \hat{\Lambda} \right)^{-1}, \quad A_{NT_0} := \left(\frac{1}{T_0} F'_0 F_0 \right)^{-1}.$$

Then, by the Cauchy–Schwarz inequality,

$$\|\zeta_{NT_0}\| \leq \|D(\hat{\Lambda})^{-1}\| \frac{1}{N} \left(\frac{1}{T_0} \sum_{t=p+1}^T \|\tilde{W}_t\|_F^2 \right)^{1/2} \left(\frac{1}{T_0} \sum_{t=p+1}^T \|M_{\hat{\Lambda}} \Omega \hat{\Lambda} B_{NT_0} A_{NT_0} f_t^0\|^2 \right)^{1/2}.$$

Now

$$\|D(\hat{\Lambda})^{-1}\| = O_p(1)$$

by Assumption A(ii), and

$$\|M_{\hat{\Lambda}}\| \leq 1, \quad \frac{1}{N} \frac{1}{T_0} \sum_{t=p+1}^T \|\tilde{W}_t\|_F^2 = O_p(1)$$

by Assumption A(i) together with Assumptions B–D. Also,

$$\|\Omega\| = \left\| \frac{1}{T_0} \sum_{t=p+1}^T \Omega_t \right\| \leq \frac{1}{T_0} \sum_{t=p+1}^T \|\Omega_t\| = O(1)$$

by Assumption C(iii).

Next, since $(1/N)\hat{\Lambda}'\hat{\Lambda} = I_r$, we have

$$\|\hat{\Lambda}\|_F = \sqrt{Nr},$$

so

$$\frac{1}{\sqrt{N}} \|\hat{\Lambda}\|_F = O(1).$$

By Proposition 7.3(ii) and Assumption B(ii),

$$\frac{1}{N} \tilde{\Lambda}'_0 \hat{\Lambda} = \left(\frac{1}{N} \tilde{\Lambda}'_0 \tilde{\Lambda}_0 \right) H^{-1} + o_p(1),$$

and hence

$$\left(\frac{1}{N} \tilde{\Lambda}'_0 \hat{\Lambda} \right)^{-1} = O_p(1),$$

because $(1/N)\tilde{\Lambda}'_0\tilde{\Lambda}_0$ converges to a positive definite matrix and $H^{-1} = O_p(1)$. Moreover, by Assumption B(i),

$$\left(\frac{1}{T_0}F'_0F_0\right)^{-1} = O(1), \quad \frac{1}{T_0} \sum_{t=p+1}^T \|f_t^0\|^2 = O(1).$$

Therefore,

$$\begin{aligned} \frac{1}{T_0} \sum_{t=p+1}^T \|M_{\hat{\Lambda}}\Omega\hat{\Lambda}B_{NT_0}A_{NT_0}f_t^0\|^2 &\leq \|M_{\hat{\Lambda}}\|^2\|\Omega\|^2\|\hat{\Lambda}\|_F^2\|B_{NT_0}\|^2\|A_{NT_0}\|^2\frac{1}{T_0} \sum_{t=p+1}^T \|f_t^0\|^2 \\ &= O_p(N). \end{aligned}$$

Combining the above bounds yields

$$\|\zeta_{NT_0}\| \leq O_p(1) \cdot \frac{1}{N} \cdot O_p(\sqrt{N}) \cdot O_p(\sqrt{N}) = O_p(1).$$

Thus

$$\zeta_{NT_0} = O_p(1).$$

□

Lemma 7.11. *Under Assumptions A–D, let H be the rotation matrix defined in Proposition 7.3(ii). Then the following hold:*

(i)

$$HH' = \left(\frac{1}{N}\tilde{\Lambda}'_0\tilde{\Lambda}_0\right)^{-1} + O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\sqrt{\min\{N, T_0\}}}\right).$$

(ii)

$$\|P_{\hat{\Lambda}} - \Pi_0\|^2 = O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\sqrt{T_0}}\right),$$

where

$$P_{\hat{\Lambda}} := N^{-1}\hat{\Lambda}\hat{\Lambda}', \quad \Pi_0 := \tilde{\Lambda}_0(\tilde{\Lambda}'_0\tilde{\Lambda}_0)^{-1}\tilde{\Lambda}'_0.$$

Proof of Lemma 7.11. Let

$$\Delta_{\Lambda} := \hat{\Lambda} - \tilde{\Lambda}_0H.$$

By Proposition 7.3(ii),

$$\frac{1}{N}\|\Delta_{\Lambda}\|_F^2 = O_p(\|\hat{\beta} - \beta_0\|^2) + O_p\left(\frac{1}{\min\{N, T_0\}}\right), \quad \frac{1}{N}\hat{\Lambda}'\hat{\Lambda} = I_r.$$

Hence

$$I_r = \frac{1}{N}\hat{\Lambda}'\hat{\Lambda} = \frac{1}{N}(H'\tilde{\Lambda}'_0 + \Delta'_{\Lambda})(\tilde{\Lambda}_0H + \Delta_{\Lambda}).$$

Expanding gives

$$I_r = H'\left(\frac{1}{N}\tilde{\Lambda}'_0\tilde{\Lambda}_0\right)H + \frac{1}{N}H'\tilde{\Lambda}'_0\Delta_{\Lambda} + \frac{1}{N}\Delta'_{\Lambda}\tilde{\Lambda}_0H + \frac{1}{N}\Delta'_{\Lambda}\Delta_{\Lambda}.$$

By Cauchy–Schwarz and the rate for Δ_{Λ} , the last three terms are

$$O_p(\|\hat{\beta} - \beta_0\|) + O_p\left(\frac{1}{\sqrt{\min\{N, T_0\}}}\right).$$

Therefore

$$H' \left(\frac{1}{N} \tilde{\Lambda}'_0 \tilde{\Lambda}_0 \right) H = I_r + O_p(\|\hat{\beta} - \beta_0\|) + O_p \left(\frac{1}{\sqrt{\min\{N, T_0\}}} \right),$$

which implies

$$HH' = \left(\frac{1}{N} \tilde{\Lambda}'_0 \tilde{\Lambda}_0 \right)^{-1} + O_p(\|\hat{\beta} - \beta_0\|) + O_p \left(\frac{1}{\sqrt{\min\{N, T_0\}}} \right).$$

This proves (i).

Next, write

$$P_{\hat{\Lambda}} = \frac{1}{N} \hat{\Lambda} \hat{\Lambda}', \quad \Pi_0 = \tilde{\Lambda}_0 (\tilde{\Lambda}'_0 \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}'_0.$$

Using $\hat{\Lambda} = \tilde{\Lambda}_0 H + \Delta_\Lambda$, we have

$$\begin{aligned} P_{\hat{\Lambda}} &= \frac{1}{N} (\tilde{\Lambda}_0 H + \Delta_\Lambda) (\tilde{\Lambda}_0 H + \Delta_\Lambda)' \\ &= \tilde{\Lambda}_0 H H' \tilde{\Lambda}'_0 / N + \frac{1}{N} \tilde{\Lambda}_0 H \Delta_\Lambda' + \frac{1}{N} \Delta_\Lambda H' \tilde{\Lambda}'_0 + \frac{1}{N} \Delta_\Lambda \Delta_\Lambda'. \end{aligned}$$

By part (i), the first term equals $\Pi_0 + O_p(\|\hat{\beta} - \beta_0\|) + O_p \left(\frac{1}{\sqrt{\min\{N, T_0\}}} \right)$, and the remaining three terms are of the same or smaller order by Cauchy-Schwarz. Hence

$$\|P_{\hat{\Lambda}} - \Pi_0\|^2 = O_p(\|\hat{\beta} - \beta_0\|) + O_p \left(\frac{1}{\sqrt{T_0}} \right).$$

This proves (ii). □

Lemma 7.12. *Under Assumptions A–D, define*

$$a_{ts} := f_t^{0'} \left(\frac{1}{T_0} F_0' F_0 \right)^{-1} f_s^0, \quad F_0 := (f_{p+1}^0, \dots, f_T^0)'$$

Assume $N/T_0 \rightarrow 0$. Then

$$\begin{aligned} & \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left[\tilde{W}'_t M_{\hat{\Lambda}} - \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\hat{\Lambda}} \right] \tilde{\varepsilon}_t \\ &= \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left[\tilde{W}'_t M_{\Pi_0} - \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\Pi_0} \right] \tilde{\varepsilon}_t + o_p(1). \end{aligned} \quad (33)$$

Proof of Lemma 7.12. Let

$$\Delta_M := M_{\hat{\Lambda}} - M_{\Pi_0} = \Pi_0 - P_{\hat{\Lambda}}.$$

Then the left-hand side minus the right-hand side of (33) is

$$\begin{aligned} R_{NT_0} &= \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left[\tilde{W}'_t \Delta_M - \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} \Delta_M \right] \tilde{\varepsilon}_t \\ &=: R_{1,NT_0} - R_{2,NT_0}, \end{aligned}$$

where

$$R_{1,NT_0} := \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \tilde{W}'_t \Delta_M \tilde{\varepsilon}_t, \quad R_{2,NT_0} := \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} \Delta_M \tilde{\varepsilon}_t.$$

By Cauchy–Schwarz,

$$\|R_{1,NT_0}\| \leq \|\Delta_M\| \left\| \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \tilde{W}'_t \tilde{\varepsilon}_t \right\|,$$

and

$$\|R_{2,NT_0}\| \leq \|\Delta_M\| \left\| \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} \tilde{\varepsilon}_t \right\|.$$

By Lemma 7.11(ii),

$$\|\Delta_M\|^2 = \|P_{\hat{\Lambda}} - \Pi_0\|^2 = O_p(\|\hat{\beta} - \beta_0\|) + O_p(T_0^{-1}).$$

Since $N/T_0 \rightarrow 0$, Proposition 7.2(i) implies

$$\|\Delta_M\| = o_p(1).$$

It therefore remains to show that the two bracketed terms are $O_p(1)$.

Step 1: the unweighted score term. Define the deterministic array $\{c_{ts}^{(1)}\}_{t,s=p+1}^T$ by

$$c_{ts}^{(1)} := \mathbf{1}\{t = s\}.$$

Then $c_{ts}^{(1)}$ is uniformly bounded, and

$$\sum_{s=p+1}^T c_{ts}^{(1)} \tilde{W}'_s \tilde{\varepsilon}_t = \tilde{W}'_t \tilde{\varepsilon}_t.$$

Hence Assumption C(vi) yields

$$\frac{1}{NT_0} \sum_{t,u=p+1}^T \left| E \left[(\tilde{W}'_t \tilde{\varepsilon}_t) (\tilde{W}'_u \tilde{\varepsilon}_u) \right] \right| \leq K.$$

Therefore

$$E \left\| \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \tilde{W}'_t \tilde{\varepsilon}_t \right\|^2 \leq K,$$

and thus

$$\left\| \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \tilde{W}'_t \tilde{\varepsilon}_t \right\| = O_p(1).$$

Step 2: the weighted score term. Define the deterministic array $\{c_{ts}^{(2)}\}_{t,s=p+1}^T$ by

$$c_{ts}^{(2)} := \frac{1}{T_0} a_{ts}.$$

By Assumption B(i), the coefficients a_{ts} are uniformly bounded, so $\{c_{ts}^{(2)}\}$ is also uniformly bounded. Moreover,

$$\sum_{s=p+1}^T c_{ts}^{(2)} \tilde{W}'_s \tilde{\varepsilon}_t = \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} \tilde{\varepsilon}_t.$$

Applying Assumption C(vi) again, we obtain

$$\frac{1}{NT_0} \sum_{t,u=p+1}^T \left| E \left[\left(\frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \tilde{\varepsilon}_t \right) \left(\frac{1}{T_0} \sum_{v=p+1}^T a_{uv} \tilde{W}'_v \tilde{\varepsilon}_u \right) \right] \right| \leq K.$$

Therefore

$$E \left\| \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} \tilde{\varepsilon}_t \right\|^2 \leq K,$$

and hence

$$\left\| \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} \tilde{\varepsilon}_t \right\| = O_p(1).$$

Combining Steps 1 and 2, we conclude that

$$\|R_{1,NT_0}\| + \|R_{2,NT_0}\| = o_p(1),$$

and therefore

$$R_{NT_0} = o_p(1).$$

This proves (33). □

Corollary 7.13. *Under Assumptions A–D and $N/T_0 \rightarrow 0$, we have*

$$\sqrt{NT_0} (\hat{\beta} - \beta_0) = D(\hat{\Lambda})^{-1} \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left[\tilde{W}'_t M_{\Pi_0} - \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\Pi_0} \right] \tilde{\varepsilon}_t + o_p(1),$$

where

$$a_{ts} = f_t^{0t} \left(\frac{1}{T_0} F_0' F_0 \right)^{-1} f_s^0, \quad F_0 = (f_{p+1}^0, \dots, f_T^0)',$$

and $D(\cdot)$ is defined in Assumption A(ii).

Proof of Corollary 7.13. By Proposition 7.9,

$$\begin{aligned} \sqrt{NT_0} (\hat{\beta} - \beta_0) &= D(\hat{\Lambda})^{-1} \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left[\tilde{W}'_t M_{\hat{\Lambda}} - \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\hat{\Lambda}} \right] \tilde{\varepsilon}_t \\ &\quad + \sqrt{\frac{N}{T_0}} \zeta_{NT_0} + o_p(1). \end{aligned}$$

Under $N/T_0 \rightarrow 0$, the bias term is negligible, and by Lemma 7.12, the score term with $M_{\hat{\Lambda}}$ can be replaced by the one with M_{Π_0} up to $o_p(1)$. Therefore

$$\sqrt{NT_0} (\hat{\beta} - \beta_0) = D(\hat{\Lambda})^{-1} \frac{1}{\sqrt{NT_0}} \sum_{t=p+1}^T \left[\tilde{W}'_t M_{\Pi_0} - \left\{ \frac{1}{T_0} \sum_{s=p+1}^T a_{ts} \tilde{W}'_s \right\} M_{\Pi_0} \right] \tilde{\varepsilon}_t + o_p(1),$$

which proves the claim. □

Acknowledgments

The authors are grateful to Takuya Ishihara of Tohoku University and Daisuke Kurisu of the University of Tokyo for valuable comments. Ishihara provided helpful suggestions on the analysis of density-valued data, and Kurisu offered valuable comments on defining the inner product in an isometric way.

Funding

This work was supported by a JSPS KAKENHI Grant-in-Aid for Scientific Research (A), Grant Number 26H01952.

Disclosure Statement

The authors report there are no competing interests to declare.

Data and Code Availability Statement

Code and replication materials will be made available upon acceptance. Access to the raw Ct-value data is subject to the terms of the data provider.

References

- Baggio, S., L’Huillier, A. G., Yerly, S., Bellon, M., Wagner, N., Rohr, M., Huttner, B., Blanchard-Rohner, G., Loevy, N., Kaiser, L., and Gervaix, A. (2021). Severe acute respiratory syndrome coronavirus 2 (sars-cov-2) viral load in the upper respiratory tract of children and adults with early acute coronavirus disease 2019 (covid-19). *Clinical Infectious Diseases*, 73(1):148–150.
- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.
- Banho, C. A., Sacchetto, L., Campos, G. R. F., et al. (2022). Impact of sars-cov-2 gamma lineage introduction and covid-19 vaccination on the epidemiological landscape of a brazilian city. *Communications Medicine*, 2:41.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188.
- Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559.
- Davis, C. and Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. *SIAM Journal on Numerical Analysis*, 7(1):1–46.

- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38.
- Diebold, F. X. and Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.
- Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, 137(2):334–353.
- Faria, N. R., Mellan, T. A., Whittaker, C., et al. (2021). Genomics and epidemiology of the p.1 sars-cov-2 lineage in manaus, brazil. *Science*, 372(6544):815–821.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471):830–840.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.
- Harrison, R. E., Hamada, A., Haswell, N., Groves, A., Vihta, K.-D., Cella, K., Garner, S., Walker, A. S., and Seale, A. C. (2023). Cycle threshold values as indication of increasing sars-cov-2 new variants, england, 2020–2022. *Emerging Infectious Diseases*, 29(10):2024–2031.
- Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press, Cambridge, 2 edition.
- Hron, K., Menafoglio, A., Templ, M., Hrušová, K., and Filzmoser, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. *Computational Statistics & Data Analysis*, 94:330–350.
- Jones, T. C., Biele, G., Mühlemann, B., Veith, T., Schneider, J., Beheim-Schwarzbach, J., Bleicker, T., Tesch, J., Schmidt, M. L., Sander, L. E., et al. (2021). Estimating infectiousness throughout SARS-CoV-2 infection course. *Science*, 373:eabi5273.
- Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press, Cambridge.
- Kutta, T., Jach, A., Haddad, M. F. C., Kokoszka, P., and Wang, H. (2025). Detection and localization of changes in a panel of densities. *Journal of Multivariate Analysis*, 205:105374.
- Lin, Y., Yang, B., Cobey, S., Lau, E. H., Adam, D. C., Wong, J. Y., Bond, H. S., Cheung, J. K., Ho, F., Gao, H., et al. (2022). Incorporating temporal distribution of population-level viral load enables real-time estimation of COVID-19 transmission. *Nature Communications*, 13:1155.
- Liu, Y., Yan, L.-M., Wan, L., Xiang, T.-X., Le, A., Liu, J.-M., Peiris, M., Poon, L. L. M., and Zhang, W. (2020). Viral dynamics in mild and severe cases of covid-19. *The Lancet Infectious Diseases*, 20(6):656–657.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

- Matsuda, Y. and Iwafuchi, R. (2025). Density-valued arma models by spline mixtures. *Journal of Time Series Analysis*.
- McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. Wiley.
- Panaretos, V. M. and Zemel, Y. (2020). *An Invitation to Statistics in Wasserstein Space*. Springer Series in Statistics. Springer, Cham.
- Petersen, A. and Müller, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *The Annals of Statistics*, 44(1):183–218.
- Petersen, A., Zhang, C., and Kokoszka, P. (2022). Modeling probability density functions as data objects. *Econometrics and Statistics*, 21:159–178.
- Rao, S. N., Manissero, D., Steele, V. R., and Pareja, J. (2020). A systematic review of the clinical utility of cycle threshold values in the context of covid-19. *Infectious Diseases and Therapy*, 9(3):573–586.
- Sabino, E. C., Buss, L. F., Carvalho, M. P. S., et al. (2021). Resurgence of covid-19 in manaus, brazil, despite high seroprevalence. *The Lancet*, 397(10273):452–455.
- Srivastava, A., Klassen, E., Joshi, S., and Jermyn, I. (2011). Shape analysis of elastic curves in euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1415–1428.
- Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.
- van den Boogaart, K. G., Egozcue, J. J., and Pawlowsky-Glahn, V. (2010). Bayes linear spaces. *SORT-Statistics and Operations Research Transactions*, 34(2):201–222.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- Zhang, C., Kokoszka, P., and Petersen, A. (2022). Wasserstein autoregressive models for density time series. *Journal of Time Series Analysis*, 43(1):30–52.