Department of Economics Working Paper Series

Robust Inference for Diffusion-Index Forecasts with Cross-Sectionally Dependent Data

by

Min Seong Kim
University of Connecticut

Working Paper 2021-04
March 2021

365 Fairfield Way, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
http://www.econ.uconn.edu/

Robust Inference for Diffusion-Index Forecasts with Cross-Sectionally Dependent Data

Min Seong Kim*
University of Connecticut

Abstract

In this paper, we propose the time-series average of spatial HAC estimators for the variance of the estimated common factors under the approximate factor structure. Based on this, we provide the confidence interval for the conditional mean of the diffusion-index forecasting model with cross-sectional heteroskedasticity and dependence of the idiosyncratic errors. We establish the asymptotics under very mild conditions, and no prior information about the dependence structure is needed to implement our procedure. We employ a bootstrap to select the bandwidth parameter. Simulation studies show that our procedure performs well in finite samples. We apply the proposed confidence interval to the problem of forecasting the unemployment rate using data by Ludvigson and Ng (2010).

JEL Classification Numbers: C12, C31, C38

Keywords: Approximate factor structure, Bandwidth selection, Diffusion index forecast, Robust inference, Spatial HAC estimator

*Email: min_seong.kim@uconn.edu.

1 Introduction

Since Stock and Watson (2002), factor model based forecasting has become widely used in macroeconomic time series and empirical finance. See, for example, Stock and Watson (2005), Ludvigson and Ng (2007, 2009), Christiansen, Eriksen, and Møller (2014), Jurado, Ludvigson, and Ng (2015), and Giglio, Kelly, and Pruitt (2016). This approach summarizes the information from a large number of time-series predictors with a few common factors, and includes the estimated common factors in a regression to forecast the object of interest.

Let

$$y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h}, \tag{1}$$

where h is the forecasting horizon and W_t is an a-vector of observed regressors. F_t is a p-vector of latent common factors that capture the co-movement of n candidate predictors $\{X_{it}, i = 1, \cdots, n\}$. We model X_{it} with the factor structure

$$X_{it} = \lambda_i' F_t + e_{it}, \quad i = 1, ..., n; \; t = 1, ..., T, \tag{2}$$

where λ_i is a vector of loadings and e_{it} is an idiosyncratic error. In this paper, p is assumed to be known. Stock and Watson (2002) refer to this factor-augmented regression as the diffusion-index forecasting model. Since F_t is latent, the forecast of y_{t+h} based on (1) employs a two-stage approach. In the first stage, we estimate F_t, denoted by $\tilde{F}_t$, using the principal component method based on (2). We normalize $\tilde{F} = \left(\tilde{F}_1, ..., \tilde{F}_T\right)'$ such that $\tilde{F}'\tilde{F}/T = \mathbb{I}_p$, so it consists of the product of$\sqrt{T}$ and the eigenvectors that correspond to the p largest eigenvalues of $XX'/(nT)$ in decreasing order, where $X = (X_1, ..., X_T)'$ and $X_t = (X_{1t}, ..., X_{nt})'$. $\Lambda = (\lambda_1, ..., \lambda_n)'$ is estimated by the least squares

$$\tilde{\Lambda} = X'\tilde{F}\left(\tilde{F}'\tilde{F}\right)^{-1} = X'\tilde{F}/T. \tag{3}$$

In the second stage, we regress y_{t+h} on $\hat{z}_t = \left(\tilde{F}_t', W_t'\right)'$, $t = 1, ..., T-h$, to obtain the LS coefficients

$$\hat{\delta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \left(\sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t'\right)^{-1} \sum_{t=1}^{T-h} \hat{z}_t y_{t+h}. \tag{4}$$

Let $z_t = (F_t', W_t')'$. Based on the estimated common factors and LS coefficients, we estimate the conditional mean of y_{T+h},

$$y_{T+h|T} = E\left(y_{T+h} | z_T, z_{T-1}, ...\right) = \alpha' F_T + \beta' W_T := \delta' z_T, \tag{5}$$

with

$$\hat{y}_{T+h|T} = \hat{\alpha}'\tilde{F}_T + \hat{\beta}' W_T = \hat{\delta}'\hat{z}_T. \tag{6}$$

As is well known, $y_{T+h|T}$ is the optimal forecast of y_{T+h} in terms of the mean squared forecast error if F_T is observed and $E\left(\varepsilon_{T+h} | z_T, z_{T-1}, ...\right) = 0$.

When e_{it} are cross-sectionally correlated, (2) has an approximate factor structure as in Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986, 1993). One of the challenges of this approach under the approximate factor structure is to obtain a valid confidence interval for $y_{T+h|T}$. Since the limiting distribution of $\hat{y}_{T+h|T}$ involves the estimation uncertainty in $\tilde{F}_T$, we have to estimate not only the variance of the regression coefficients but also that of $\tilde{F}_T$ in order to construct the confidence interval. However, it is difficult to correctly estimate the latter when there exists cross-sectional dependence of the idiosyncratic errors. To address this problem, Bai and Ng (2006) provide the cross-sectional heteroskedasticity and autocorrelation consistent (CS-HAC) variance estimator, and Gonçalves and Perron (2020) propose the cross-sectional dependence (CSD) bootstrap method. Assuming that the covariance structure of $\{e_{it}\}_{i=1}^{n}$ is time invariant, these two methods utilize the fact that cross-section units are repeatedly observed over time.

As an alternative to the CS-HAC estimation method of Bai and Ng, in the paper we consider the time-series average of the spatial HAC estimators, which we refer to as the AV-SHAC estimator, to estimate the variance of $\tilde{F}_T$. The spatial HAC estimator was first proposed by Conley (1999) and has been studied further in the linear regression and GMM contexts. See, for example, Pinkse, Slade, and Brett (2002), Conley and Molinari (2007), Kelejian and Prucha (2007), and Kim and Sun (2011). We extend the spatial HAC estimation approach to the factor model by using the time-series average. We follow Kelejian and Prucha (2007) by modeling the cross-sectional dependence using a linear representation in which the coefficients of iid innovations are not assumed to be known and are not parameterized. The advantage of this approach is that we need not rely on a mixing-type condition to establish the asymptotics, which, according to Bai and Ng (2006), is difficult to justify in the cross-sectional dimension. The extension of spatial HAC estimation to the factor model is empirically relevant in that approximate factor model applications have become very popular in macroeconomics and empirical finance. The extension is technically nontrivial as well. We have to examine the effect of estimation errors in the factor model on the asymptotics, and there are several issues to be addressed in regard to its implementation as discussed in this paper.

To examine the asymptotics, we decompose the difference of our AV-SHAC estimator from the true variance as a sum of three terms. The first term is due to estimation errors in the factor model, and the second and third terms represent the bias and variation of the infeasible estimator in which model parameter values are assumed to be known. We find that the variation of the infeasible estimator decreases faster than the effect of the estimation errors, which implies that the optimal rate of convergence is achieved by balancing the bias and the effect of the estimation errors. This is in sharp contrast to the asymptotics of the usual HAC estimators (e.g., Andrews, 1991; Kim and Sun, 2011) in which the trade-off is between the bias and the variation of the infeasible estimator. The practical implication of this result is that an explicit formula for the asymptotic MSE is not available for the AV-SHAC estimator, so we cannot establish the bandwidth selection procedure based on it. We address this practical issue by proposing a bandwidth selection procedure based on the cluster wild bootstrap. In this approach, each cluster contains all the units in one time period in order to replicate cross-sectional dependence of the original data. We select the bandwidth that maximizes the bootstrap version of the AV-SHAC estimator under the constraint of the rate condition for consistency. Simulation studies show that the proposed bandwidth selection procedure performs well in finite samples.

Since our estimator is constructed based on the spatial HAC estimator, we need a distance measure that characterizes the dependence structure of the data. That is, the covariance of two potential predictors is assumed to be a decreasing function of the distance between them. A typical approach in this regard is to find a relevant auxiliary variable that captures the decaying pattern of dependence and use that variable as the distance measure. The choice of the auxiliary variable tends to depend on the type of application, for example, the transportation cost (Conley and Ligon, 2002), the geographic distance (Pinkse, Slade, and Brett, 2002), or the similarity of the input and output structures (Chen and Conley, 2001; Conley and Dupor, 2003). An alternative approach is to define the distance in such a way that it reflects the dependence structure directly. For example, we define a distance $d_{ij}^{\mathrm{D}} = |1/Corr\left(e_{it}, e_{jt}\right)| - 1$ in our simulation and empirical application which by definition reflects the degree of dependence very well. Using the assumption that the covariance structure is time invariant, we can approximate this quantity using time-series observations. A crucial advantage of this approach is that we need no prior information for its implementation. Constructing the distance based on the correlation coefficient has been considered in spatial panel data models. See, for example, Mantegna (1998), Fernandez (2011), and Cui, Sarafidis, and Yamagata (2020).

While this paper uses our AV-SHAC estimator for the estimation of $Var\left(\hat{y}_{T+h|T}\right)$, we note that it is possible to use it in different contexts. For example, Ludvigson and Ng (2010) and Gonçalves and Perron (2014) study the asymptotic bias in the factor augmented model when $\sqrt{T}/n \to c \neq 0$, and they show that the bias is a function of the asymptotic variance of $\tilde{F}_t$. It would be a natural extension of this paper to employ the AV-SHAC estimator to correct this bias under the approximate factor structure.

The rest of the paper is organized as follows. Section 2 introduces the proposed variance estimator and associated confidence interval for diffusion-index forecasts. We study their asymptotic properties in Section 3. In Section 4, we discuss our bootstrap based bandwidth selection procedure. Section 5 presents simulation results and an empirical illustration in which we apply the proposed confidence interval to the problem of forecasting the unemployment rate using data by Ludvigson and Ng (2010). The last section concludes. The appendix consists of two parts. In the first part, we introduce a diagnostic test for the existence of cross-sectional dependence by comparing the spatial HAC estimator with Bai's (2003) heteroskedasticity robust (HR) variance estimator. The second part provides the proofs of our theoretical results.

2 Variance estimator and confidence interval

In this section, we review the asymptotics of factor models, and introduce the AV-SHAC estimator and associated confidence interval for $y_{T+h|T}$. We follow Bai (2003) and Bai and Ng (2006) by making the following assumptions.

Assumption F1 *(i)* $E\|F_t\|^4 \leq M$ *and* $T^{-1}\sum_t^T F_tF_t' \to^p \Sigma_F$, *where* Σ_F *is nonrandom and positive definite. (ii)* $E\|\lambda_i\|^4 \leq M$ *and* $n^{-1}\sum_{i=1}^n \lambda_i\lambda_i' \to^p \Sigma_\Lambda$, *where* Σ_Λ *is nonrandom and positive definite.*

Assumption F2 *(i)* $Ee_{it} = 0$ *and* $Ee_{it}^8 \leq M$ *for all* i *and* t. *(ii) Let* $\sigma_{ij,ts} = E(e_{it}e_{js})$ *such that* $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ *for all* t *and* s *and* $|\sigma_{ij,ts}| \leq \tau_{ts}$ *for all* i *and* j. *For all* T *and* n, *and for every* $t \leq T$ *and every* $i \leq n$, $\sum_{j=1}^n \bar{\sigma}_{ij} \leq M$, $\sum_{s=1}^T \tau_{ts} \leq M$, *and* $(nT)^{-1}\sum_{i,j,t,s}|\sigma_{ij,ts}| \leq M$. *(iii)* $E\left|n^{-1/2}\sum_{i=1}^n [e_{is}e_{it} - E(e_{is}e_{it})]\right|^4 \leq M$ *for all* t *and* s.

Assumption F3 *For all* n *and* T *the following hold:*
(i) For each t, $E\left\|(nT)^{-1/2}\sum_{i=1}^n\sum_{t=1}^T F_t[e_{it}e_{is} - E(e_{it}e_{is})]\right\|^2 \leq M$.
(ii) $E\left\|(nT)^{-1/2}\sum_{i=1}^n\sum_{t=1}^T F_t\lambda_i'e_{it}\right\|^2 \leq M$.
(iii) For each t, $n^{-1/2}\sum_{i=1}^n \lambda_i e_{it} \to^d N(0,\Gamma_t)$, *where*

$$\Gamma_t = \lim_{n\to\infty} Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \lambda_i e_{it}\right) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n E(\lambda_i\lambda_j'e_{it}e_{jt})$$

Assumption F4 *The variables* $\{\lambda_i\},\{F_t\}$, *and* $\{e_{it}\}$ *are mutually independent.*

Assumption F5 *(i)* $E\|z_t\|^4 \leq M$, $E(\varepsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, \ldots) = 0$ *for every* $h > 0$, *and* z_t *and* ε_t *are independent of* e_{is} *for all* i *and* s. *(ii)* $T^{-1}\sum_{t=1}^T z_tz_t' \to^p \Sigma_{zz}$, *and* Σ_{zz} *is positive definite. (iii)* $T^{-1/2}\sum_{t=1}^T z_t\varepsilon_{t+h} \to^d N(0,\Sigma_{zz,\varepsilon})$, *where* $\Sigma_{zz,\varepsilon} = p\lim\, T^{-1}\sum_{t=1}^T \varepsilon_{t+h}^2 z_tz_t'$ *and* $\Sigma_{zz,\varepsilon}$ *is positive definite.*

Assumption F6 $n, T \to \infty$, $\sqrt{T}/n \to 0$, *and* $\sqrt{n}/T \to 0$.

Assumption F1 provides moment conditions for the factors and loadings. Assumption F2 states the moment conditions, and the serial and cross-sectional heteroskedasticity and weak dependence of the idiosyncratic errors. Assumption F3(iii) is a central limit theorem for the moment process, which is satisfied under various conditions. The independence condition in Assumption F4 is standard in the literature. Assumption F5(iii) states that $z_t\varepsilon_{t+h}$ is serially uncorrelated, which is true when $y_{T+h|T}$ is defined as the conditional mean given past information.

Gonçalves, Perron, and Djogbenou (2015) consider the case where serial correlation is present in $z_t\varepsilon_{t+h}$. Assumption F6 provides the rate condition for establishing the asymptotics of $\hat{\delta}$ and $\tilde{F}_t$.

Let $H = \tilde{V}^{-1}\left(\tilde{F}'F/T\right)(\Lambda'\Lambda/n)$, where $\tilde{V}$ is the $(p \times p)$ diagonal matrix containing the p largest eigenvalues of $XX'/(nT)$ in decreasing order. It is well known that F and Λ are not separately identifiable, and that $\tilde{F}_t$ and $\tilde{\lambda}_i$ are the estimators of HF_t and $H^{-1}\lambda_i$, respectively. We define $V = p\lim \tilde{V}$, $H_0 = p\lim H$, and $Q = p\lim \tilde{F}'F/T = H_0^{-1}$. Bai and Ng (2006) show that, under the assumptions,

$$\frac{\hat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{Var\left(\hat{y}_{T+h|T}\right)}} \to^d N(0,1), \tag{7}$$

where

$$Var\left(\hat{y}_{T+h|T}\right) = \frac{1}{T}\hat{z}_T' Avar\left(\hat{\delta}\right)\hat{z}_T + \frac{1}{n}\hat{\alpha}' Avar\left(\tilde{F}_T\right)\hat{\alpha}, \tag{8}$$

$$Avar\left(\hat{\delta}\right) = \Upsilon'^{-1}\Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}^{-1}\Upsilon^{-1}, \text{ and } Avar\left(\tilde{F}_T\right) = V^{-1}Q\Gamma_T Q' V^{-1},$$

with $\Upsilon = diag\left(V^{-1}Q\Sigma_\Lambda, \mathbb{I}_a\right)$ and $\delta = \left(\alpha' H^{-1}, \beta'\right)'$. The variance of $\hat{y}_{T+h|T}$ in (8) consists of two terms. The first is associated with the variance of the regression coefficients, and the second is associated with the variance of the preliminary common factor estimator. Thus we need to estimate these two quantities in order to construct the confidence interval based on (7). Estimating $Avar\left(\hat{\delta}\right)$ is straightforward under Assumption F5, since it is consistently estimated with the sample counterpart

$$\widehat{Avar}\left(\hat{\delta}\right) = \left(\frac{1}{T}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\right)^{-1}\frac{1}{T}\sum_{t=1}^{T-h}\hat{\varepsilon}_{t+h}^2\hat{z}_t\hat{z}_t'\left(\frac{1}{T}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\right)^{-1}.$$

In contrast, it is challenging to estimate $Avar\left(\tilde{F}_T\right)$ under the approximate factor structure. To formulate the estimation problem, let's define $F_t^H = H_0F_t$, $\lambda_i^H = H_0^{-1}\lambda_i$ and $\eta_{it}^H = \lambda_i^H e_{it}$. Since $Q = H_0^{-1}$, we can write

$$\begin{aligned} Avar\left(\tilde{F}_t\right) &= V^{-1}H_0^{-1}\Gamma_t\left(H_0^{-1}\right)'V^{-1} \\ &= V^{-1}\Gamma_t^H V^{-1}, \end{aligned}$$

where

$$\Gamma_t^H = \lim_{n\to\infty}\Gamma_{t,n}^H \text{ and } \Gamma_{t,n}^H = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n E(\eta_{it}^H\left(\eta_{it}^H\right)'). \tag{9}$$

In the absence of cross-sectional correlation, Γ_t^H is consistently estimated with the HR variance estimator given by

$$\tilde{\Gamma}_t^{HR} = \frac{1}{n}\sum_{i=1}^n\tilde{\eta}_{it}\tilde{\eta}_{it}', \tag{10}$$

where $\tilde{\eta}_{it} = \tilde{\lambda}_i \tilde{e}_{it}$, $\tilde{e}_{it} = X_{it} - \tilde{\lambda}_i' \tilde{F}_t$, and $\tilde{\lambda}_i$ is the vector of estimated factor loadings from (3). In the presence of cross-sectional dependence, $\tilde{\Gamma}_t^{HR}$ is evidently inconsistent, so the associated confidence interval is invalid. This problem would be more serious when T/n is large, because, as implied in (8), the variance of $\hat{y}_{T+h|T}$ is mainly determined by the variance of $\tilde{F}_T$ in this case.

To address the issue, Bai and Ng (2006) propose the CS-HAC estimator, which allows for cross-sectional dependence. That estimator is defined as

$$\tilde{\Gamma}^{BN} = \frac{1}{n_{sub}} \sum_{i=1}^{n_{sub}} \sum_{j=1}^{n_{sub}} \tilde{\lambda}_i \tilde{\lambda}_j' \frac{1}{T} \sum_{t=1}^{T} \tilde{e}_{it} \tilde{e}_{jt},$$

where $n_{sub}/\min[n, T] \to 0$. They show that $\tilde{\Gamma}^{BN}$ is consistent under Assumption H1 below.

Assumption H1 *$E(e_{it}e_{jt}) = \sigma_{ij}$ for all i, j, t.*

This assumption states that the covariance structure of $\{e_{it}\}$ remains invariant over time, which results in

$$\Gamma_t^H = \Gamma^H \text{ for all } t, \text{ where } \Gamma^H = \lim_{n,T\to\infty} \Gamma_{nT}^H \text{ with } \Gamma_{nT}^H = \frac{1}{T} \sum_{t=1}^{T} \Gamma_{t,n}^H. \tag{11}$$

Relying on the same assumption, we propose the AV-SHAC estimator, which is the time-series average of the spatial HAC estimators. Our estimator is defined as

$$\tilde{\Gamma} = \frac{1}{T} \sum_{t=1}^{T} \tilde{\Gamma}_t \text{ with } \tilde{\Gamma}_t = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right) \tilde{\eta}_{it} \tilde{\eta}_{jt}', \tag{12}$$

where $K(\cdot)$ is a real-valued kernel function and d_n is the bandwidth parameter. d_{ij} is a distance between units i and j that reflects the strength of the covariance between η_{it}^H and η_{jt}^H such that $\left|Cov\left(\eta_{it}^H, \eta_{jt}^H\right)\right|$ is a decreasing function of d_{ij}. We make the following assumption about d_{ij} to establish the asymptotics.

Assumption H2 *(i) $d_{ij} \geq 0$, $d_{ii} = 0$, and $d_{ij} = d_{ji}$; (ii) d_{ij} is time invariant.*

The assumption indicates that the AV-SHAC estimator does not require d_{ij} to satisfy the triangle inequality, $d_{ij} \leq d_{ik} + d_{kj}$, while the other properties of distance are assumed to hold. This is in sharp contrast to the standard spatial HAC estimation literature (e.g., Conley, 1999; Kim and Sun, 2011), which relies on this inequality to establish the asymptotics. Assumption H2(ii) is implied by Assumption H1.

A typical approach for the construction of d_{ij} in the literature is to find a relevant auxiliary variable that characterizes the decaying pattern of dependence in the data and use that variable as the distance. The variable tends to be different for different applications, for example, the transportation cost (Conley and Ligon, 2000), the geographic distance (Pinkse, Slade, and Brett, 2002), or the similarity of the input and output structures (Chen and Conley, 2001; Conley and Dupor, 2003). A problem with this approach in our setting is that the diffusion-index forecasts often employ macroeconomic and financial data as candidate predictors, and such variables are unlikely to be available in practice. For example, Stock and Watson's (2002) dataset contains 215 macroeconomic time series in 8 different categories, including real output and income, exchange rates, interest rates, price index, etc. It would be almost impossible to find a variable that can be used for the distance between, for instance, the USD-JPY exchange rate and the CPI in service goods.

An alternative approach is to define the distance in such a way that it reflects the dependence structure directly. For example, in Section 5 we define $d_{ij} = |1/Corr\left(e_{it}, e_{jt}\right)| - 1$ in the simulation and empirical illustration, which by definition captures the degree of dependence very well. This approach gives us a valid distance under Assumption H2 that does not require d_{ij} to satisfy the triangle inequality. $Corr\left(e_{it}, e_{jt}\right)$ is unobserved, but it is easy to approximate using time-series observations under Assumptions H1 and H2(ii). A crucial advantage of this approach is that no prior information is required for its implementation.

It is important to note that nontrivial information about the dependence structure is needed for the use of $\tilde{\Gamma}^{BN}$. Researchers should be able to select n_{sub} observations ($n_{sub}/\min[n, T] \to 0$) which replicate the overall cross-sectional dependence structure so that $T^{-1}\sum_{s=1}^{T} Var\left(n_{sub}^{-1/2}\sum_{i=1}^{n_{sub}}\eta_{is}^{H}\right)$ becomes a good copy of Γ^{H}. Our simulation studies show that the performance of $\tilde{\Gamma}^{BN}$ depends strongly on this selection. To the best of our knowledge, however, there is no practical guidance in this regard in the literature.

Using $\tilde{\Gamma}$, we estimate $Avar\left(\tilde{F}_t\right)$ with

$$\widehat{Avar}\left(\tilde{F}_t\right) = \tilde{V}^{-1}\tilde{\Gamma}\tilde{V}^{-1}. \tag{13}$$

We propose the confidence interval for $y_{T+h|T}$ at the $100\,(1-\alpha)\,\%$ level with

$$CI\left(y_{T+h|T}\right) = \left[\hat{y}_{T+h|T} + \mathfrak{q}_{\alpha/2}\sqrt{\widehat{Var}\left(\hat{y}_{T+h|T}\right)}, \hat{y}_{T+h|T} + \mathfrak{q}_{1-\alpha/2}\sqrt{\widehat{Var}\left(\hat{y}_{T+h|T}\right)}\right], \tag{14}$$

where $\mathfrak{q}_{\alpha}$ denotes the α quantile of the standard normal distribution and

$$\widehat{Var}\left(\hat{y}_{T+h|T}\right) = \frac{1}{T}\hat{z}_T'\widehat{Avar}\left(\hat{\delta}\right)\hat{z}_T + \frac{1}{n}\hat{\alpha}'\widehat{Avar}\left(\tilde{F}_T\right)\hat{\alpha}. \tag{15}$$

If we assume that $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$, we can also construct the forecasting interval for y_{T+h} as

$$FI\left(y_{T+h}\right) = \left[\hat{y}_{T+h|T} + \mathfrak{q}_{\alpha/2}\sqrt{\widehat{Var}\left(\hat{y}_{T+h|T}\right) + \hat{\sigma}_\varepsilon^2}, \hat{y}_{T+h|T} + \mathfrak{q}_{1-\alpha/2}\sqrt{\widehat{Var}\left(\hat{y}_{T+h|T}\right) + \hat{\sigma}_\varepsilon^2}\right],$$

where

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^{T-h}\hat{\varepsilon}_{t+h}^2 \text{ with } \hat{\varepsilon}_{t+h} = y_{t+h} - \hat{y}_{t+h|T}.$$

3 Asymptotic properties

In this section, we examine the asymptotics of $\tilde{\Gamma}$ and associated confidence interval for $y_{T+h|T}$. We first employ the following linear-array process to model η_{it}^H. Let

$$\eta_{it}^H = R_{it}\epsilon, \tag{16}$$

where

$$R_{it} = \begin{bmatrix} r_{it,1}^{(1)} & r_{it,2}^{(1)} & \cdots & r_{it,nTp}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ r_{it,1}^{(p)} & r_{it,2}^{(p)} & \cdots & r_{it,nTp}^{(p)} \end{bmatrix} \tag{17}$$

is a $p \times nTp$ nonstochastic matrix and $\epsilon = (\epsilon_1, \ldots, \epsilon_l, \ldots, \epsilon_{nTp})'$ is a vector of iid innovations. The elements of R_{it} are not assumed to be known and are not parameterized. The linear-array process includes widely used spatial/spatiotemporal parametric models as special cases, and is commonly used to characterize spatial dependence in the literature. See, for example, Kelejian and Prucha (2007), Kim and Sun (2011, 2013), Robinson (2011), and Pesaran and Tosetti (2011). A crucial advantage of using a linear array is that we do not need to introduce a mixing-type condition to establish the asymptotics, which, according to Bai and Ng (2006), is difficult to justify in the cross-sectional dimension.

Define the infeasible version of $\tilde{\Gamma}$ as

$$\tilde{\Gamma}^0 = \frac{1}{T}\sum_{t=1}^{T}\tilde{\Gamma}_t^0 \text{ with } \tilde{\Gamma}_t^0 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right)\eta_{it}^H\left(\eta_{jt}^H\right)', \tag{18}$$

which is obtained by substituting η_{it}^H for $\tilde{\eta}_{it}$ in $\tilde{\Gamma}$. Using $\tilde{\Gamma}^0$, we decompose $\tilde{\Gamma} - \Gamma_{nT}^H$ as a sum of three terms:

$$\tilde{\Gamma} - \Gamma_{nT}^H = \left(\tilde{\Gamma} - \tilde{\Gamma}^0\right) + \left(\tilde{\Gamma}^0 - E\tilde{\Gamma}^0\right) + \left(E\tilde{\Gamma}^0 - \Gamma_{nT}^H\right). \tag{19}$$

The first term represents the effect of the estimation errors in the factor model, and the second and third terms are due to the variance and bias of $\tilde{\Gamma}^0$, respectively. We make Assumptions H1–H7 to characterize the variance and bias of $\tilde{\Gamma}^0$, and to control the effect of the estimation errors based on Assumptions F1–F4 and F6.

Assumption H3 $\epsilon \sim (0, \mathbb{I}_{nTp})$, *with* $E\left(\epsilon_l^4\right) \leq M,\ l = 1, ..., nTp$, *for some constant* $M > 0$.

Assumption H3 provides the moment condition for ϵ_l. Under this assumption, Γ_{nT}^H in (11) can be rewritten as

$$\Gamma_{nT}^H = \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T} R_{it}R_{jt}'.$$

Assumption H4 *For all $c = 1, \cdots, p$, (i) $\lim_{n,T\to\infty} \sum_{i=1}^{n} \sum_{t=1}^{T} \left| r_{it,l}^{(c)} \right| < M$ for all t and l; (ii) $\lim_{n,T\to\infty} \sum_{l=1}^{nTp} \left| r_{it,l}^{(c)} \right| < M$ for all i and t.*

This assumption states the summability conditions for the coefficients of the linear process, which is related to the weak dependence of η_{it}^{H}. Note that $r_{it,l}^{(c)}$ represents the amount of change in the cth element of η_{it}^{H} as the result of a one-unit increase in ϵ_l. Thus Assumption H4(i) and H4(ii) state that the aggregate absolute response of $\left\{\eta_{it}^{H}\right\}$ to a single innovation ϵ_l and the sum of the absolute responses of η_{it}^{H} to innovations $\epsilon_l, ..., \epsilon_{nTp}$ are finite. We introduce these conditions to control the variance of $\tilde{\Gamma}^0$. An alternative approach would be to introduce some mixing and stationarity assumptions to obtain the fourth-order cumulant condition as in the time-series case (e.g., Andrews, 1991).

Let

$$\ell_{in} = \sum_{j=1}^{n} 1\left\{d_{ij} \leq d_n\right\} \text{ and } \ell_n = \frac{1}{n}\sum_{i=1}^{n} \ell_{in}. \tag{20}$$

ℓ_{in} is the number of unit i's pseudo-neighbors whose distance from i is within the bandwidth, and ℓ_n is the average number of pseudo-neighbors. It is obvious from (20) that ℓ_{in} and ℓ_n are increasing functions of d_n.

Assumption H5 *$\ell_{in} \leq c_\ell \ell_n$ for all $i = 1, ..., n$ with some constant c_ℓ.*

Assumption H5 allows units to have different numbers of pseudo-neighbors to a certain degree, but it excludes cases where only a few units have many correlated units while others have none or very few.

Let q be the Parzen characteristic exponent of $K(x)$. That is, q is the largest value of q_0 for which

$$K_{q_0} = \lim_{x\to 0} \frac{1 - K(x)}{|x|^{q_0}}$$

is finite. The following assumption is made to characterize the weak dependence of η_{it}^{H} with respect to d_{ij}.

Assumption H6 *There exists a finite constant M such that*

$$\lim_{n,T\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \left\| E\left[\eta_{it}^{H}\left(\eta_{jt}^{H}\right)'\right] \right\| d_{ij}^{q} < M, \tag{21}$$

where $\|A\|$ denotes the Euclidean norm of matrix A.

Assumption H6 provides the key condition that d_{ij} should satisfy, as it implies that d_{ij} captures the decaying pattern of the dependence structure, so that the covariance between η_{it}^{H} and η_{jt}^{H} decreases to zero quickly as d_{ij} grows. This assumption practically enables us to control the bias of $\tilde{\Gamma}$ caused by truncation and downweighting with the kernel function. That is, when d_{ij} is large, $K(d_{ij}/d_n)$ assigns a weight of zero or close to zero to $\tilde{\eta}_{it}\tilde{\eta}_{jt}'$ in (12), but it does not cause much bias under this assumption, because $E\left[\eta_{it}^H\left(\eta_{jt}^H\right)'\right]$ is also close to zero.

The results in this paper can be generalized to the case where d_{ij} is error contaminated. Following Kim and Sun (2011), we can show that our asymptotic results are still valid under the following conditions: (i) the measurement error is independent of $\{\epsilon_l\}$; (ii) it is of order $o(d_n)$ as d_n increases; and (iii) the summability condition in Assumption H6 holds with the error-contaminated distance measure. For simplicity, however, in this paper we do not consider measurement errors.

The asymptotic results for $\tilde{\Gamma}$ in Theorem 1 are based on the following assumption on the kernel function.

Assumption H7 *The kernel $K:\mathbb{R}\to[-1,1]$ satisfies $K(0)=1,\ K(x)=K(-x), K(x)=0$ for $|x|>1$.*

Assumption H7 is standard and is satisfied by kernels that are commonly used for the HAC estimator, including the Bartlett, Gaussian, Tukey–Hanning and Parzen kernels.

Theorem 1 states the consistency and convergence rate of $\tilde{\Gamma}$ based on the decomposition in (19).

Theorem 1 *Suppose that Assumptions F1–F4, F6 and H1–H7 hold, and that $d_n, \ell_n, n, T\to\infty$ such that $\ell_n/n, \ell_n/T\to 0$.*

(i) $E\tilde{\Gamma}^0-\Gamma_{nT}^H=O\left(\frac{1}{d_n^q}\right)$, *(ii)* $\tilde{\Gamma}^0-E\tilde{\Gamma}^0=O_P\left(\sqrt{\frac{\ell_n}{nT}}\right)$, *(iii)* $\tilde{\Gamma}-\tilde{\Gamma}^0=O_P\left(\sqrt{\frac{\ell_n}{T}}\right)+O_P\left(\frac{\ell_n}{n}\right)$.

(iv) Let $c_R\in[0,\infty)$ and $\tau_1,\tau_2,\tau_3\in(0,\infty)$. Then

$$\tilde{\Gamma}-\Gamma_{nT}^H=\begin{cases} O_P\left(\sqrt{\frac{\ell_n}{T}}\right) & \text{if } \frac{T}{n}\to c_R \text{ and } \frac{d_n^{2q}\ell_n}{T}\to\tau_1; \\ O_P\left(\frac{\ell_n}{n}\right) & \text{if } \frac{T}{n}\to\infty,\ \frac{\sqrt{T}\ell_n}{n}\to\tau_2 \text{ and } \frac{d_n^q n}{T}\to\tau_3 \end{cases}$$

The proofs are given in Appendix B. The first part of the Theorem indicates that the bias of $\tilde{\Gamma}^0$ converges to zero as $d_n\to\infty$. This is because the degree of truncation and downweighting with the kernel function decreases as d_n grows. The convergence of the variation of $\tilde{\Gamma}^0$ is provided in Theorem 1(ii). Since our estimator is based on the time-series average of the spatial HAC estimators, the amount of smoothing is determined by Tn/ℓ_n and the variation decreases to zero as $nT/\ell_n\to\infty$. Theorem 1(iii) implies that the difference between $\tilde{\Gamma}$ and $\tilde{\Gamma}^0$, which is caused by estimation errors in $\{\tilde{\eta}_{it}\}$, vanishes as $\ell_n/n, \ell_n/T\to 0$. The optimal rate of convergence is summarized in Theorem 1(iv).

As ℓ_n is an increasing function of d_n, the Theorem implies that both the variance and the effect of the estimation errors increase as $d_n \to \infty$. Comparing the rates in (ii) and (iii) shows that the variation of $\tilde{\Gamma}^0$ is of smaller order, so the optimal rate of convergence for $\tilde{\Gamma}$ is achieved with the sequence of d_n that balances the bias of $\tilde{\Gamma}^0$ and $\tilde{\Gamma} - \tilde{\Gamma}^0$. This result is in sharp contrast to Andrews (1991) and Kim and Sun (2011, 2013), where the effect of the estimation errors is dominated by the variance of the infeasible estimator. The difference is due to the fact that (1) $\tilde{\Gamma}$ improves the rate for the variance by employing the time-series average of $\tilde{\Gamma}_t$, and (2) the estimation errors in $\tilde{\eta}_{it}$ involve $\tilde{F}_t$ and $\tilde{\lambda}_i$, which do not achieve $\sqrt{nT}$ convergence differently from the standard GMM/LS estimators in panel models.

Note that the Theorem establishes the consistency and the convergence rate but does not establish the asymptotic MSE. Alternatively, we may follow Kim and Sun (2011), who make a set of more restrictive assumptions in order to derive the asymptotic bias and variance of the infeasible estimator in the cross-sectional setting. They introduce a certain version of the stationarity condition in the cross-sectional dimension (Assumption 6) and assume that the effect of units in the boundary is asymptotically negligible. They also require the distance measure to satisfy the triangle inequality. However, these conditions may not be suitable in the diffusion-index forecast model, where macroeconomic and financial data are often employed as candidate predictors. By showing the validity of $\tilde{\Gamma}$ without such restrictive assumptions, our procedure becomes applicable to a variety of factor model applications. The practical usefulness of having explicit formulas for the bias and variance is to establish the MSE optimal bandwidth selection based on them (see Andrews (1991) and Kim and Sun (2011, 2013)). However, this does not apply to our model, because, as discussed above, the variance of $\tilde{\Gamma}^0$ is dominated by the effect of the estimation errors in $\tilde{\eta}_{it}$ so the asymptotic MSE of $\tilde{\Gamma}$ does not consist of the bias and variance of $\tilde{\Gamma}^0$.

As emphasized in the literature since Newey and West (1987), positive semi-definiteness of $\tilde{\Gamma}$ is highly desirable. In the time-series context, the HAC estimator is a weighted average of periodogram, and we can obtain a psd estimator by choosing a kernel for which the Fourier transform is nonnegative. However, this is not the case for $\tilde{\Gamma}$, which is the average of the spatial HAC estimators. In this regard, Kelejian and Prucha (2007) introduce a class of kernels that generate a psd spatial HAC estimator. An example is

$$K_v(x) = \begin{cases} (1-x)^v, & 0 \leq x \leq 1 \\ 0, & x > 1 \end{cases}.$$

This class of kernels ensures that $\tilde{\Gamma}$ is psd if $v \geq (p+1)/2$ and the Euclidean distance is employed. Another solution to this problem is to modify $\tilde{\Gamma}$ to render a psd estimator. This method is suggested by Politis (2011). Consider the eigen-decomposition of $\tilde{\Gamma}$, that is, $\tilde{\Gamma} = \Psi \Xi \Psi'$, where $\Xi = diag(\mu_1, ..., \mu_p)$ is a diagonal matrix containing the eigenvalues of $\tilde{\Gamma}$ and Ψ is the matrix of orthonormal eigenvectors. We can define the modified psd estimator as

$$\tilde{\Gamma}^{+}=\Psi\Xi^{+}\Psi', \tag{22}$$

where $\Xi^{+}=diag\left(\mu_1^{+},...,\mu_p^{+}\right)$, $\mu_\iota^{+}=\max\left(\mu_\iota,c_\mu\right)$, and c_μ is a small nonnegative number. It is easy to show that $\tilde{\Gamma}^{+}$ has the same convergence rate as $\tilde{\Gamma}$. We use this method (with $c_\mu=10^{-6}$) in the simulation and empirical illustration.

Corollary 1 below follows directly from Theorem 1.

Corollary 1 *Under the conditions in Theorem 1 and Assumptions F5 and F6, we have*

$$\frac{\hat{y}_{T+h|T}-y_{T+h|T}}{\sqrt{\widehat{Var}\left(\hat{y}_{T+h|T}\right)}}\to^{d} N\left(0,1\right).$$

4 Bandwidth selection procedure

An important issue in implementing the AV-SHAC estimator is to select the bandwidth parameter properly. While the inference based on Corollary 1 is valid in the asymptotic sense, the theory does not provide practical guidance on the choice of d_n, which the finite sample performance depends strongly on. This is a particularly challenging problem in this setting, because, as discussed in the previous section, the conventional MSE optimal bandwidth parameter is not available.

To address this, we consider a bandwidth selection procedure based on the cluster wild bootstrap. Let $D_n=\left\{d_n^{(1)},...,d_n^{(M)}\right\}$ be the set of reasonable d_n for a given sample size. The procedure involves the following steps.

Step 1 For $t=1,...,T$, let $\tilde{e}_t=\left(\tilde{e}_{1t},...,\tilde{e}_{nt}\right)'$ denote the n-vector of residuals in time period t. Generate B bootstrap samples $\left\{X_{k,t}^{*},\ t=1,...,T\right\}_{k=1}^{B}$ based on

$$\underbrace{X_{k,t}^{*}}_{n\times 1}=\tilde{\Lambda}\tilde{F}_t+e_{k,t}^{*}.$$

The vector of bootstrap errors $e_{k,t}^{*}=\left(e_{k,1t}^{*},...,e_{k,nt}^{*}\right)'$ is generated from the following process:

$$e_{k,t}^{*}=\tilde{e}_t\xi_{k,t},\ \text{where}\ \xi_{k,t}\overset{\text{iid}}{\sim}\left(0,1\right).$$

Step 2 Using the principal component method, estimate the bootstrap factors and bootstrap loadings in order to obtain $\left\{\tilde{F}_{k,t}^{*}\right\}_{t=1}^{T}$ and $\left\{\tilde{\lambda}_{k,i}^{*}\right\}_{i=1}^{n}$, and construct the bootstrap version of the AV-SHAC estimator with $d_n^{(1)}$:

$$\tilde{\Gamma}_k^{*}\left(d_n^{(1)}\right)=\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}K\left(\frac{d_{ij}}{d_n^{(1)}}\right)\tilde{\eta}_{k,it}^{*}\left(\tilde{\eta}_{k,jt}^{*}\right)',$$

where $\tilde{\eta}^*_{k,it} = \tilde{\lambda}^*_{k,i}\tilde{e}^*_{k,it}$ and $\tilde{e}^*_{k,it} = X^*_{k,it} - \left(\tilde{\lambda}^*_{k,i}\right)' \tilde{F}^*_{k,t}$.

Step 3 Compute

$$T^*\left(d_n^{(1)}\right) = \frac{1}{B}\sum_{k=1}^{B} T_k^*\left(d_n^{(1)}\right), \text{ where } T_k^*\left(d_n^{(1)}\right) = \hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_k^*\left(d_n^{(1)}\right)\tilde{V}^{-1}\hat{\alpha}.$$

Step 4 Repeat Steps 1 and 3 for each $d_n \in D_n$. Let $\ell_n(d_n)$ denote the value of ℓ_n with bandwidth d_n, $\pi \in (0,1)$, and $c_\pi > 0$. Our bandwidth selection $d_n^\dagger$ solves

$$d_n^\dagger = \arg\max_{d_n \in D_n} T^*(d_n) \text{ s.t. } \ell_n(d_n) \leq c_\pi \min\{n, T\}^\pi . \tag{23}$$

As we can see in Step 1, we use the cluster wild bootstrap to generate bootstrap samples $\left\{X^*_{k,t}\right\}$. Each cluster contains all the units in one time period, and the external random variable $\xi_{k,t}$ is common to all units in t, which enables replication of the cross-sectional dependence of the original samples. Thus $T^*\left(d_n^{(m)}\right)$ is expected to be a good approximation to $\hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}\tilde{V}^{-1}\hat{\alpha}$ with $d_n = d_n^{(m)}$. We employ the Rademacher random variable for $\xi_{k,t}$ in our simulation and empirical application.

Recall that the proposed confidence interval for $y_{T+h|T}$ at the $100(1-\alpha)\%$ level is

$$CI\left(y_{T+h|T}\right) = \left[\hat{y}_{T+h|T} + \mathfrak{q}_{\alpha/2}\sqrt{\widehat{Var}\left(\hat{y}_{T+h|T}\right)}, \hat{y}_{T+h|T} + \mathfrak{q}_{1-\alpha/2}\sqrt{\widehat{Var}\left(\hat{y}_{T+h|T}\right)}\right],$$

where

$$\widehat{Var}\left(\hat{y}_{T+h|T}\right) = \frac{1}{T}\hat{z}_T'\widehat{Avar}\left(\hat{\delta}\right)\hat{z}_T + \frac{1}{n}\hat{\alpha}'\widehat{Avar}\left(\tilde{F}_T\right)\hat{\alpha} \text{ and } \widehat{Avar}\left(\tilde{F}_t\right) = \tilde{V}^{-1}\tilde{\Gamma}\tilde{V}^{-1}.$$

Our bootstrap criterion function in (23) is designed to choose a $d_n^\dagger$ that improves the coverage of $CI\left(y_{T+h|T}\right)$, because the coverage rates can be much lower than its confidence level under the approximate factor structure. The constraint that $\ell_n(d_n) \leq c_\pi \min\{n,T\}^\pi$ with $\pi \in (0,1)$ and $c_\pi > 0$ is given to impose the rate condition for consistency in Theorem 1. In our simulation and empirical illustration in Section 5, we set $c_\pi = 1$ and $\pi = 2/3$.

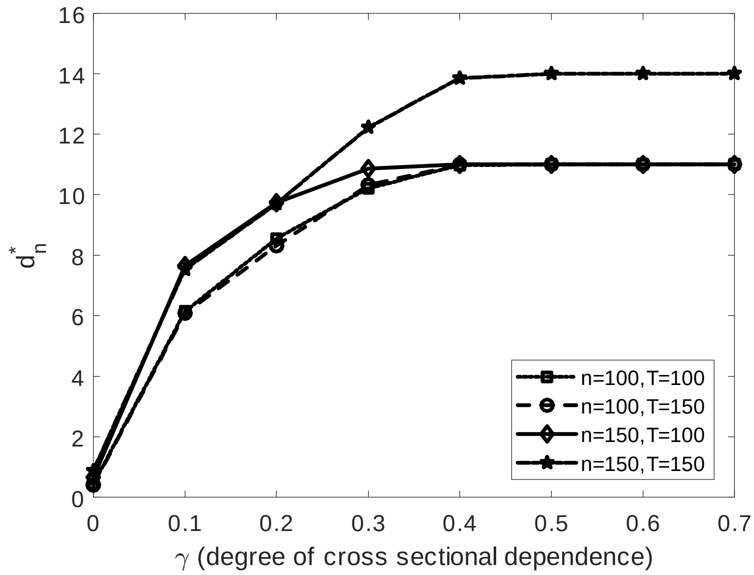

Figure 1: Bandwidth selection for different degrees of cross-sectional dependence

Figure 1 illustrates the relationship between the bootstrap bandwidth selection and cross-sectional dependence. The data in our simulation were generated based on DGP1, where γ is the parameter that determines the degree of cross-sectional dependence. The DGP is explained in detail in Section 5.1. The figure shows that when the degree of dependence increases, our procedure tends to yield a larger bandwidth, which reduces the bias of $\tilde{\Gamma}$. We can also see that given γ, the procedure tends to choose a smaller bandwidth if n decreases, and this enables $\tilde{\Gamma}$ to control its variance.

Different bootstrap methods could be used in Step 1 as long as they replicate the cross-sectional dependence. For example, we could use the CSD bootstrap by Gonçalves and Perron (2020) that employs the thresholding technique. In the current setting, our cluster wild bootstrap seems to be easier to use than the CSD bootstrap, because the latter requires selecting the thresholding parameter. If information about the location of each unit is available, we could also consider a parametric bootstrap based on the Cliff–Ord type spatial regression or the parametric spatial covariance model (e.g., the Matérn function). However, it is unlikely that such information is available with macroeconomic and financial data that the factor model usually employs in the diffusion-index forecasting context. This paper does not examine the theoretical properties of our cluster wild bootstrap, and we leave that for future research.

5 Simulation and empirical illustration

5.1 Simulation

In this section, we report the results of simulation studies that we conducted to investigate the finite sample properties of the proposed inference procedure. For comparative purposes, we also consider existing methods which are based on $\tilde{\Gamma}_T^{HR}$ and $\tilde{\Gamma}^{BN}$. The following simulation design is employed.

$$
\begin{aligned}
& y_{t+h} = \beta_0 + \alpha_1 F_{1t} + \alpha_2 F_{2t} + \varepsilon_{t+h};\ \varepsilon_{t+h} \sim N(0,1) \\
& X_{it} = \lambda_i' F_t + e_{it}, \\
& \lambda_i \sim^{iid} U(0,1), \\
& F_{jt} = \theta_j F_{jt-1} + \sqrt{1-\theta_j^2} u_{jt},\ \theta_j = 0.3^j,\ j = 1,2; \\
& u_{jt} \sim^{iid} N\left(0, \sigma_u^2(j)\right),\ \sigma_u^2(j) = U[.5, 1.5].
\end{aligned}
$$

We set $\beta_0 = \alpha_1 = \alpha_2 = 1$ and $h = 4$. The number of common factors, p, is 2, and is assumed to be known. The number of replications is 3000. We consider two DGPs to generate cross-sectional dependence in idiosyncratic errors.

The first DGP follows Bai and Ng (2013) and Gonçalves and Perron (2020).

$$
\begin{aligned}
& \underline{\text{DGP1}} \\
& e_t = \bar{\Omega}(\gamma)^{1/2} v_t,\ e_t = (e_{1t}, ..., e_{nt})',\ v_t = (v_{1t}, ..., v_{nt})', \\
& v_{it} \sim^{iid} N\left(0, \sigma_v^2(i)\right),\ \sigma_v^2(i) = U[.5, 1.5],
\end{aligned}
$$

where v_{it} is independent of u_{jt}. $\bar{\Omega}(\gamma)^{1/2}$ is the Choleski decomposition of $n \times n$ Toeplitz matrix in which the ιth main diagonal is γ^ι if $\iota \leq 10$ and zero otherwise. Thus we can generate cross-sectional dependence by choosing a nonzero γ.

The second DGP generates cross-sectionally dependent data using a popular spatio-temporal parametric model. The design is based on an $L_n \times L_n$ square integer lattice ($L_n = 10, 12$), where unit i is located at a lattice point (i_1, i_2).

$$
\begin{aligned}
& \underline{\text{DGP2}} \\
& e_t = 0.3 e_{t-1} + (I_n + \varrho M) v_t,\ e_0 = (0, ..., 0)', \\
& v_t = (v_{1t}, ..., v_{nt})',\ v_{it} \sim^{iid} N(0,1),
\end{aligned}
$$

where v_{it} is independent of u_{jt}. $M = [m_{ij}]_{i,j=1}^n$ is an $n \times n$ spatial weight matrix such that for units i, j,

$$
m_{ij} = \begin{cases} 1 & \text{if } d_{ij} = 1 \\ 0 & \text{if } d_{ij} \neq 1 \end{cases},
$$

where $d_{ij} = \sqrt{(i_1 - j_1)^2 + (i_2 - j_2)^2}$. Thus units i and j are dependent on each other if the distance d_{ij} is 1.

To construct $\tilde{\Gamma}$, we employ two different distance measures. The first one, denoted by d_{ij}^{T}, is the true distance, that is, $d_{ij}^{\mathrm{T}} = |i-j|$ for DGP1 and $d_{ij}^{\mathrm{T}} = \sqrt{(i_1-j_1)^2 + (i_2-j_2)^2}$ for DGP2. The second distance measure, denoted by d_{ij}^{D}, is defined as

$$d_{ij}^{\mathrm{D}} = \frac{1}{|\rho_{ij}|} - 1,$$

where $\rho_{ij} = Corr\,(e_{it}, e_{jt})$. By definition, this reflects the degree of dependence between i and j very well. Strictly speaking, d_{ij}^{D} is not a valid distance, since it does not satisfy the triangle inequality. However, as discussed in Section 2, the triangle inequality is not required for our method. Though d_{ij}^{D} is unknown in practice, we can approximate it under Assumptions H1 and H2(ii) by utilizing information obtained from the time dimension:

$$\tilde{d}_{ij}^{\mathrm{D}} = \min\left\{\frac{1}{|\tilde{\rho}_{ij}|}, 100\right\} - 1,$$

where $\tilde{\rho}_{ij} = \sum_{t=1}^{T} \tilde{e}_{it}\tilde{e}_{jt} / \sqrt{\sum_{t=1}^{T} \tilde{e}_{it}^2 \sum_{t=1}^{T} \tilde{e}_{jt}^2}$. An important advantage of using $\tilde{d}_{ij}^{\mathrm{D}}$ is that no prior information about the dependence structure is needed for its construction. For the selection of d_n, we use the bootstrap based selection procedure proposed in Section 4. The Parzen kernel is employed.

$\tilde{\Gamma}^{BN}$ is constructed in two different ways. The first one, denoted by CS-HAC$_{\mathrm{T}}$, supposes that the true covariance structure is known. For DGP1, we randomly select n_{sub} $\left(= \min\left\{\sqrt{n}, \sqrt{T}\right\}\right)$ consecutive units g times to obtain $\tilde{\Gamma}_{(1)}^{BN}, ..., \tilde{\Gamma}_{(g)}^{BN}$ and then take their average. For DGP2, we randomly select g blocks of units. The block sizes are 3×3 units ($n_{sub} = 9$) when $\min\left\{\sqrt{n}, \sqrt{T}\right\} = 10$, and 3×4 units ($n_{sub} = 12$) when $\min\left\{\sqrt{n}, \sqrt{T}\right\} = 12$. The second approach, denoted by CS-HAC$_{\mathrm{R}}$, selects n_{sub} units randomly, so the selected observations do not maintain the dependence of the data at all. We set $g = \left\lfloor \min\left(\sqrt{n}, \sqrt{T}\right) \right\rfloor$, where $\lfloor x \rfloor$ represents the largest integer that does not exceed x.

Table 1 presents the empirical coverage probabilities (ECPs) of the 95% confidence interval for $y_{T+h|T}$ and the forecasting interval for y_{T+h} under DGP1. A few patterns emerge. First, while the ECP of HR for y_{T+h} is generally close to the nominal coverage probability regardless of the existence and strength of cross-sectional dependence, its ECP for $y_{T+h|T}$ tends to be very sensitive to cross-sectional dependence. For example, when $\gamma = 0.4$, $n = 100$, and $T = 50$, the ECPs of HR for y_{T+h} and $y_{T+h|T}$ are 0.931 and 0.873, respectively. When $\gamma = 0.7$, $n = 100$, and $T = 50$, the ECP for $y_{T+h|T}$ decreases to 0.774, which implies that the distortion of the ECPs becomes more serious as the strength of the dependence increases. We also find that the performance of HR in the presence of cross-sectional dependence becomes worse when T/n is large. See, for example, that the ECP for $y_{T+h|T}$ decreases further to 0.664 when $\gamma = 0.7$, $n = 100$, and $T = 200$. This is well expected from (8), which shows that the variance of $\hat{y}_{T+h|T}$ is mainly determined by $Avar\left(\tilde{F}_T\right)$ when T/n is large.

Second, our method improves upon coverage rates regardless of whether the true distance d_{ij}^{T} or the data-driven distance $\tilde{d}_{ij}^{\mathrm{D}}$ is employed. The former tends to perform slightly better in general. For example, when $\gamma = 0.7$, $n = 150$, and $T = 200$, the ECPs of AV-SHAC with d_{ij}^{T} and $\tilde{d}_{ij}^{\mathrm{D}}$ for $y_{T+h|T}$ are 0.907 and 0.886, respectively, which are comparable to the ECP of CS-HAC$_{\mathrm{T}}$ ($= 0.902$) and are substantially superior to the ones based on HR ($= 0.701$) and CS-HAC$_{\mathrm{R}}$ ($= 0.688$). In the absence of cross-sectional dependence, that is, $\gamma = 0$, AV-SHAC performs as well as HR, which is constructed under the zero covariance assumption.

The finding that AV-SHAC performs well with $\tilde{d}_{ij}^{\mathrm{D}}$ gives us an important implication from an empirical point of view. A typical approach for the construction of d_{ij} is to find a relevant auxiliary variable that characterizes the dependence structure of the data. However, such a variable is unlikely to be available in the diffusion-index forecast which often uses macroeconomic and financial data as candidate predictors. Our simulation studies show that even without observing such a variable, we can still use our method with $\tilde{d}_{ij}^{\mathrm{D}}$, which is directly obtained from time-series observations.

Third, the performance of CS-HAC depends strongly on how the n_{sub} observations are selected. CS-HAC$_{\mathrm{T}}$, in which $\tilde{\Gamma}^{BN}$ is constructed with consecutive units, improves the accuracy of the confidence intervals for $y_{T+h|T}$ substantially. For example, when $\gamma = 0.7$, $n = 150$, and $T = 200$, its ECP is 0.902, while the ECP of HR is only 0.701. However, the results are remarkably different if $\tilde{\Gamma}^{BN}$ is constructed with randomly selected observations that do not maintain the dependence of the data. When $\gamma = 0.7$, $n = 150$, and $T = 200$, CS-HAC$_{\mathrm{R}}$ has an ECP of 0.688, so its performance is similar to HR.

We also compare the performance based on the mean absolute error to examine the precision of each estimator. The results of comparison are similar to the ones based on the ECP, and we omit the table to save the space.

Table 2 reports the ECPs when the data are generated from DGP2. The results are similar to those presented in Table 1. All the estimators yield accurate confidence intervals in the absence of cross-sectional dependence. However, when $\varrho = 0.2, 0.4$, HR and CS-HAC$_{\mathrm{R}}$ are substantially inferior to AV-SHAC and CS-HAC$_{\mathrm{T}}$.

We conduct additional simulations to examine the sensitivity of the finite sample performance to the choices of kernel function and c_{μ}, which is a threshold parameter for the modified psd variance estimator in (22). We find that our procedure is robust to those choices.

5.2 Empirical Illustration

In this section, we report the results of applying the proposed confidence interval to the problem of forecasting the unemployment rate. Our forecasting exercise is based on the dataset used by Ludvigson and Ng (2010), which contains 131 monthly macroeconomic time series for 1964:1–2007:12. The dataset is available at Ludvigson's webpage (https://www.sydneyludvigson.com/).

We consider one-period ahead ($h = 1$) forecasts of the monthly growth in the unemployment rate using the following model:

$$\Delta UER_{t+1} = \beta_0 + \beta_1 \Delta UER_t + \alpha_1 F_{1,t} + \alpha_2 F_{2,t} + \varepsilon_{t+1}, \tag{24}$$

where $\Delta UER_{t+1} = \log\left(UER_{t+1}/UER_t\right)$ and UER_t is the unemployment rate in time period t. Thus the model employs two common factors and uses ΔUER_t as the observed predictor.

The forecasting exercise begins by estimating factors with the data from 1964:1 to 1988:12 (300 months). We then estimate the coefficients in (24) by regressing ΔUER_{t+1} on $\left(1, \Delta UER_t, \tilde{F}_{1,t}, \tilde{F}_{2,t}\right)'$ for $t = 1964{:}1$ to 1988:11 and obtain the estimate of the conditional mean of $\Delta UER_{1989:1}$ given by

$$\widehat{\Delta UER}_{1989:1} = \hat{\beta}_0 + \hat{\beta}_1 \Delta UER_{1988:12} + \hat{\alpha}_1 \tilde{F}_{1,1988:12} + \hat{\alpha}_2 \tilde{F}_{2,1988:12}.$$

Finally, we construct the 95% confidence interval for the conditional mean of $\Delta UER_{1989:1}$. This is constructed in two different ways. The first is the one that we propose, in which we employ $\tilde{\Gamma}$ with $\tilde{d}_{ij}^{\mathrm{D}}$ and use the bootstrap based bandwidth selection method. The second approach is based on $\tilde{\Gamma}_T^{HR}$, which does not account for cross-sectional dependence. Using the same procedure, we obtain $\widehat{\Delta UER}_{1989:2}$ and the associated confidence intervals based on data from 1964:2 to 1989:1. The procedure is repeated until the forecast is made for $\Delta UER_{2007:12}$.

The result is summarized in the table below.

95% Confidence intervals for the conditional mean of ΔUER_{T+1}

	HR	AV-SHAC ($\tilde{d}_{ij}^{\mathrm{D}}$)
Average (1989:1–2007:12)	$[-0.0495,\ 0.0586]$	$[-0.0546,\ 0.0636]$

The table reports the averages of the confidence intervals for the conditional mean of ΔUER_{T+1} over the period 1989:1–2007:12. By taking cross-sectional dependence of the data into account, our procedure produces confidence intervals which are about 9.3% wider, on average, than the conventional confidence intervals based on $\tilde{\Gamma}_T^{HR}$.

Figure 2 presents the diffusion-index forecast of ΔUER_{T+1} and the associated confidence intervals in each time period between 2006:1 and 2007:12. Differences of two confidence intervals show how important it is to account for cross-sectional dependence. We can see that our procedure produces wider confidence intervals in most of the time periods.

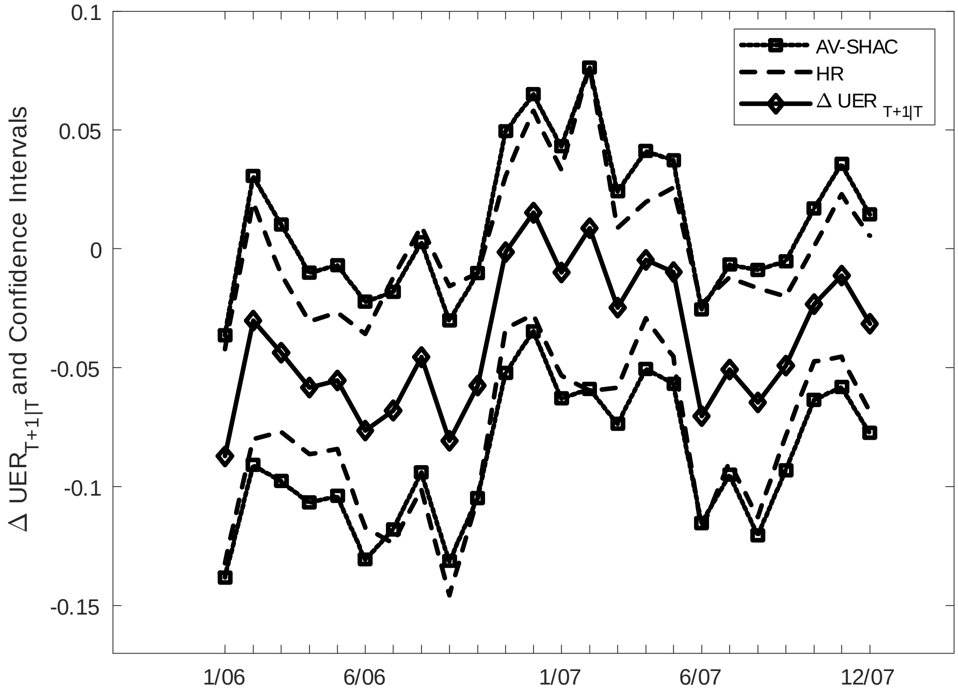

Figure 2: Diffusion-index forecast and confidence interval for the growth rate of unemployment rate.

6 Conclusion

In this paper, we propose the time-series average of the spatial HAC estimators for the variance of the estimated common factors under the approximate factor structure. We then provide the confidence interval for the conditional mean of the diffusion-index forecasts, which is robust to the cross-sectional heteroskedasticity and dependence of unknown forms in idiosyncratic errors. We establish the asymptotics under very mild conditions. Since the performance of our procedure depends strongly on the choice of bandwidth, we provide a bandwidth selection procedure using the cluster wild bootstrap. A crucial advantage of our procedure is that no prior information about the dependence structure is required for its implementation.

Table 1: Empirical coverage probabilities of 95% confidence intervals using different variance estimators: DGP1

Method		AV-SHAC (d_{ij}^{T})		AV-SHAC ($\tilde{d}_{ij}^{\mathrm{D}}$)		CS-HAC$_{\mathrm{T}}$		CS-HAC$_{\mathrm{R}}$		HR	
n	T	$\hat{y}_{T+h\|T}$	$\hat{y}_{T+h}$	$\hat{y}_{T+h\|T}$	$\hat{y}_{T+h}$	$\hat{y}_{T+h\|T}$	$\hat{y}_{T+h}$	$\hat{y}_{T+h\|T}$	$\hat{y}_{T+h}$	$\hat{y}_{T+h\|T}$	$\hat{y}_{T+h}$
						$\gamma = 0.0$					
100	50	0.935	0.923	0.935	0.923	0.932	0.923	0.932	0.923	0.936	0.923
100	100	0.944	0.944	0.944	0.944	0.933	0.943	0.934	0.944	0.942	0.943
100	200	0.936	0.946	0.936	0.946	0.930	0.945	0.927	0.945	0.936	0.946
150	50	0.940	0.931	0.940	0.931	0.938	0.931	0.939	0.931	0.943	0.931
150	100	0.941	0.941	0.941	0.941	0.942	0.938	0.941	0.938	0.941	0.943
150	200	0.942	0.948	0.942	0.948	0.935	0.948	0.935	0.948	0.942	0.948
						$\gamma = 0.4$					
100	50	0.905	0.934	0.882	0.931	0.904	0.934	0.867	0.931	0.873	0.931
100	100	0.915	0.942	0.888	0.941	0.909	0.942	0.848	0.938	0.857	0.939
100	200	0.921	0.948	0.898	0.947	0.918	0.947	0.838	0.946	0.846	0.947
150	50	0.922	0.938	0.898	0.936	0.923	0.939	0.888	0.935	0.889	0.935
150	100	0.931	0.943	0.905	0.942	0.929	0.943	0.875	0.940	0.880	0.941
150	200	0.935	0.947	0.913	0.947	0.927	0.947	0.854	0.944	0.863	0.945
						$\gamma = 0.7$					
100	50	0.862	0.935	0.837	0.934	0.869	0.936	0.771	0.929	0.774	0.929
100	100	0.874	0.945	0.838	0.944	0.873	0.945	0.706	0.940	0.712	0.940
100	200	0.859	0.950	0.837	0.949	0.861	0.950	0.653	0.943	0.664	0.944
150	50	0.863	0.933	0.851	0.932	0.875	0.934	0.788	0.930	0.792	0.930
150	100	0.899	0.947	0.874	0.946	0.901	0.947	0.757	0.941	0.762	0.941
150	200	0.907	0.954	0.886	0.953	0.902	0.954	0.688	0.947	0.701	0.948

Table 2: Empirical coverage probabilities of 95% confidence intervals using different variance estimators: DGP2

Method n	T	AV-SHAC (d_{ij}^{T}) $\hat{y}_{T+h\|T}$	$\hat{y}_{T+h}$	AV-SHAC ($\tilde{d}_{ij}^{\mathrm{D}}$) $\hat{y}_{T+h\|T}$	$\hat{y}_{T+h}$	CS-HAC$_{\mathrm{T}}$ $\hat{y}_{T+h\|T}$	$\hat{y}_{T+h}$	CS-HAC$_{\mathrm{R}}$ $\hat{y}_{T+h\|T}$	$\hat{y}_{T+h}$	HR $\hat{y}_{T+h\|T}$	$\hat{y}_{T+h}$
					$\varrho = 0.0$						
100	100	0.935	0.948	0.935	0.948	0.920	0.948	0.926	0.948	0.938	0.948
100	150	0.939	0.948	0.939	0.948	0.923	0.948	0.928	0.948	0.937	0.948
100	200	0.946	0.948	0.946	0.948	0.927	0.947	0.935	0.948	0.945	0.948
144	100	0.935	0.942	0.935	0.942	0.925	0.942	0.931	0.942	0.935	0.942
144	150	0.945	0.937	0.945	0.937	0.938	0.937	0.939	0.937	0.942	0.938
144	200	0.950	0.948	0.950	0.948	0.940	0.948	0.940	0.948	0.948	0.949
					$\varrho = 0.2$						
100	100	0.901	0.945	0.869	0.944	0.887	0.945	0.821	0.943	0.835	0.943
100	150	0.891	0.952	0.864	0.951	0.870	0.951	0.796	0.948	0.809	0.949
100	200	0.899	0.951	0.879	0.950	0.881	0.949	0.801	0.945	0.817	0.945
144	100	0.913	0.941	0.875	0.941	0.889	0.941	0.834	0.939	0.840	0.940
144	150	0.907	0.951	0.879	0.950	0.901	0.951	0.818	0.948	0.830	0.948
144	200	0.910	0.949	0.890	0.948	0.908	0.949	0.817	0.947	0.829	0.947
					$\varrho = 0.4$						
100	100	0.870	0.944	0.836	0.943	0.849	0.944	0.737	0.939	0.749	0.940
100	150	0.861	0.960	0.835	0.959	0.831	0.959	0.704	0.954	0.717	0.955
100	200	0.859	0.953	0.840	0.953	0.832	0.952	0.900	0.947	0.713	0.947
144	100	0.893	0.945	0.863	0.942	0.867	0.943	0.765	0.939	0.774	0.939
144	150	0.887	0.956	0.865	0.954	0.880	0.955	0.738	0.948	0.749	0.948
144	200	0.883	0.957	0.865	0.956	0.874	0.957	0.722	0.952	0.729	0.952

Appendix A: Testing cross-sectional dependence

In this appendix, we consider a diagnostic test for the existence of cross-sectional dependence in case where each candidate predictor can be mapped onto an integer lattice. This test is useful for choosing an appropriate estimator of Γ_T^H when researchers use spatial variables as candidate predictors.

If there is no cross-sectional dependence, it may be preferable to use $\tilde{\Gamma}_T^{HR}$, since that is generally more efficient. Otherwise, an estimator that is robust under the approximate factor structure should be employed. We consider the following null and alternative hypotheses:

$$\mathbb{H}_0 : \hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_T^{HR}\tilde{V}^{-1}\hat{\alpha} \text{ is consistent for } \alpha' Avar\left(\tilde{F}_T\right)\alpha,$$
$$\mathbb{H}_1 : \hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_T^{HR}\tilde{V}^{-1}\hat{\alpha} \text{ underestimates } \alpha' Avar\left(\tilde{F}_T\right)\alpha.$$

Define

$$\mathcal{T}_t = \frac{\hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_t\tilde{V}^{-1}\hat{\alpha}}{\hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_t^{HR}\tilde{V}^{-1}\hat{\alpha}}. \tag{A.1}$$

Note that $\mathcal{T}_t$ is based on $\tilde{\Gamma}_t$ with $\tilde{\Gamma}_t^{HR}$ and does not employ the AV-SHAC estimator. We derive the asymptotically equivalent distribution of $\mathcal{T}_t$ using the theory of fixed-b asymptotics in which ℓ_n/n is assumed to be fixed with the sample size. Under this asymptotics, the numerator $\hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_t\tilde{V}^{-1}\hat{\alpha}$ becomes asymptotically equal in distribution to a random variable that is proportional to $a' Avar\left(\tilde{F}_t\right)\alpha$, while the denominator, $\hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_t^{HR}\tilde{V}^{-1}\hat{\alpha}$, is consistent only in the absence of cross-sectional dependence.

To establish the fixed-b asymptotics of $\mathcal{T}_t$, we follow Conley (1999) and Bester, Conley, Hansen, and Vogelsang (2016) in assuming that each unit is mapped onto a two-dimensional integer lattice. The extension of this to a higher-dimensional lattice is straightforward. In a lattice setting, a variable is indexed based on location, and we let $(i_1, i_2) \in [1, 2, ..., L_n] \otimes [1, 2, ..., M_n]$ denote the location of unit i. Let

$$\eta^H_{(i_1,i_2),t} = \begin{cases} \eta^H_{it}, & \text{if a unit is present at } (i_1, i_2) \\ 0, & \text{otherwise.} \end{cases}$$

d_{ij} is now a distance between two locations, (i_1, i_2) and (j_2, j_2).

Assumption A1 $\ell_n/n \to b^o \in (0, 1]$, *and* $\ell_n/T = o(1)$.

Under Assumption A1, the bias of $\tilde{\Gamma}_t$ vanishes, but the variation does not disappear even as $n, T \to \infty$, so $\tilde{\Gamma}_t$ converges in distribution to a random matrix. The second part of this assumption is made to control the effect of the estimation errors in the factor model.

We follow Kim and Sun (2013) in making Assumptions A2 and A3, which present conditions on the distance measure and the kernel functions.

Assumption A2 *Let $d_{(i_1,i_2),(j_1,j_2)}$ denote the distance between the two units located at (i_1,i_2) and (j_2,j_2). Then*

$$\frac{d_{(i_1,i_2),(j_1,j_2)}}{d_n} = d\left(\frac{|i_1-j_1|}{d_n}, \frac{|i_2-j_2|}{d_n}\right)$$

and $d(\cdot,\cdot)$ is continuously differentiable.

Assumption A2 implies that $d_{ij} = d_{(i_1,i_2),(j_1,j_2)}$ is a function of $|i_1-j_1|$ and $|i_2-j_2|$ and homogeneous, and p-norm distances satisfy this condition. Let $b=(b_1,b_2)$ with $b_1 = d_n/L_n$ and $b_2 = d_n/M_n$, and let

$$\begin{aligned}\mathbb{K}_b\left(\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right),\left(\frac{j_1}{L_n},\frac{j_2}{M_n}\right)\right) &= \mathbb{K}\left(\left(\frac{i_1}{b_1L_n},\frac{i_2}{b_2M_n}\right),\left(\frac{j_1}{b_1L_n},\frac{j_2}{b_2M_n}\right)\right) \\ &= K\left(\frac{d_{(i_1,i_2),(j_1,j_2)}}{d_n}\right).\end{aligned} \tag{A.2}$$

Based on (A.2), we can rewrite the spatial HAC estimator as

$$\tilde{\Gamma}_t = \frac{1}{n}\sum_{i_1,j_1=1}^{L_n}\sum_{i_2,j_2=1}^{M_n}\mathbb{K}_b\left(\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right),\left(\frac{j_1}{L_n},\frac{j_2}{M_n}\right)\right)\tilde{\eta}_{(i_1,i_2)}\tilde{\eta}'_{(j_1,j_2)}.$$

Assumption A3 *(i) Assumption H7 holds. (ii) For all $x_1,x_2\in\mathbb{R}$, there is a constant $c_K<\infty$ such that $|K(x_1)-K(x_2)|\le c_K|x_1-x_2|$. (iii) $\mathbb{K}_b((v_1,v_2),(w_1,w_2))$ is continuous and continuously differentiable almost everywhere on $[0,1]^2\times[0,1]^2$.*

Assumption A3 accommodates all kernels commonly used in HAC estimation, but it excludes the rectangular kernel. Under this assumption, we have the Fourier series representation

$$\mathbb{K}_b((v_1,v_2),(w_1,w_2)) = \lim_{\mathcal{L}\to\infty}\sum_{\iota=1}^{\mathcal{L}}\kappa_\iota\psi_{b,\iota}(v_1,v_2)\,\psi_{b,\iota}(w_1,w_2), \tag{A.3}$$

where $\{\psi_{b,\iota}(v_1,v_2)\,\psi_{b,\iota}(w_1,w_2)\}$ is a sequence of continuously differentiable functions on $\mathbb{L}^2\left([0,1]^2\times[0,1]^2\right)$. By default, we set $\psi_{b,1}(\cdot,\cdot)$ to be a constant function. The convergence in (A.3) is absolute and uniform in $(v_1,v_2)\in[0,1]^2$ and $(w_1,w_2)\in[0,1]^2$. We state the high level assumptions below.

Assumption A4 *Let $\xi_i\sim^{iid}N(0,\mathbb{I}_p)$. For a given t, the following holds:*

$$\begin{aligned}&P\left(\left[\frac{1}{\sqrt{n}}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\psi_{b,\iota}\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right)\eta^H_{(i_1,i_2)}\right]<c \text{ for } \iota=1,2,...,\mathcal{L}\right)\\ &= P\left(\left[\frac{J_t}{\sqrt{n}}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\psi_{b,\iota}\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right)\xi_{(i_1,i_2)}\right]<c \text{ for } \iota=1,2,...,\mathcal{L}\right)+o_P(1)\end{aligned}$$

as $n\to\infty$ for every fixed $\mathcal{L}$, where $c\in\mathbb{R}^p, b\in(0,1]\times(0,1]$, and J_t is the matrix square root of Γ^H_t, that is, $J_tJ_t'=\Gamma^H_t$.

Assumption A4 implies that $n^{-1/2}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\psi_{b,\iota}\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right)\eta_{(i_1,i_2)}^{H}$ is asymptotically equivalent in distribution to a normal random variable with mean zero and variance Γ_t^H. We make this assumption to approximate the distribution of $\mathcal{T}_t$ under the fixed-b asymptotics. This condition is satisfied if a CLT holds jointly over $\iota=1,...,\mathcal{L}$. Primitive conditions for Assumption A4 are provided in Sun and Kim (2015).

Assumption A5 *Let $\Sigma_{\Lambda^H}=H^{-1}\Sigma_\Lambda H^{-1}$. For all $(r_1,r_2)\in[0,1]^2$,*

$$\frac{1}{n}\sum_{i_1=1}^{[r_1L_n]}\sum_{i_2=1}^{[r_2M_n]}\lambda_{(i_1,i_2)}^{H}\left(\lambda_{(i_1,i_2)}^{H}\right)'\to^p r_1r_2\Sigma_{\Lambda^H}.$$

Assumption A6 *For each t, $E\left\|(nT)^{-1/2}\sum_{i=1}^{n}\sum_{s=1}^{T}\left(F_s^He_{is}e_{it}-E\left(F_s^He_{is}e_{it}\right)\right)\right\|\leq M$.*

Proposition A1 provides the asymptotically equivalent distribution of $\tilde{\Gamma}_t$ under the fixed-b asymptotics.

Proposition A1 *Suppose that Assumptions F1–F4, F6 and A1–A6 hold. For a given t, we have*

$$\tilde{\Gamma}_t\sim^a J_t\tilde{\Gamma}^aJ_t'$$

and

$$\tilde{\Gamma}^a=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\left(\xi_i-\bar{\xi}\right)\left(\xi_j-\bar{\xi}\right)'$$

with $\bar{\xi}=n^{-1}\sum_{i=1}^{n}\xi_i$. '$\sim^a$' denotes asymptotic equivalence in distribution as $n,T\to\infty$.

The proof is in the supplementary appendix. This proposition states that, under the fixed-b asymptotics, $\tilde{\Gamma}_t$ is asymptotically equivalent in distribution to a random matrix which is proportional to Γ_t^H in the matrix sense. We develop our cross-sectional dependence test based on this result.

As $\hat{\alpha}$ and $\tilde{V}$ are consistent for α and V, Proposition A1 implies that $\hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_t\tilde{V}^{-1}\hat{\alpha}$ is asymptotically equivalent to $\alpha'V^{-1}J_t\tilde{\Gamma}^aJ_t'V^{-1}\alpha$. Note that $\underbrace{\alpha'V^{-1}J_t}_{1\times p}\underbrace{\xi_i}_{p\times1}=^d\sqrt{\alpha'Avar\left(\tilde{F}_t\right)\alpha}\zeta_i$ with $\zeta_i\sim^{iid}N(0,1)$. Thus, we have

$$\begin{aligned}\alpha'V^{-1}J_t\tilde{\Gamma}^aJ_t'V^{-1}\alpha&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\left(\alpha'V^{-1}J_t\left(\xi_i-\bar{\xi}\right)\right)\left(\alpha'V^{-1}J_t\left(\xi_j-\bar{\xi}\right)\right)'\\&=^d\alpha'Avar\left(\tilde{F}_t\right)\alpha\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\left(\zeta_i-\bar{\zeta}\right)\left(\zeta_j-\bar{\zeta}\right).\end{aligned}\tag{A.4}$$

If $\hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_t^{HR}\tilde{V}^{-1}\hat{\alpha}$ is consistent for $\alpha' Avar\left(\tilde{F}_t\right)\alpha$, we have

$$\mathcal{T}_t \sim^a \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right)\left(\zeta_i-\bar{\zeta}\right)\left(\zeta_j-\bar{\zeta}\right) := \Psi_{bn}. \tag{A.5}$$

We summarize the result in the theorem below.

Theorem A1 *Let Assumptions F1–F4, F6 and A1–A6 hold. Then, for a given* t, *if* $\hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_t^{HR}\tilde{V}^{-1}\hat{\alpha}$ *is consistent for* $\alpha' Avar\left(\tilde{F}_t\right)\alpha$ *we have*

$$P\left(\mathcal{T}_t < c\right) - P\left(\Psi_{bn} < c\right) = o\left(1\right) \text{ as } n, T \to \infty$$

for fixed b.

The proof is omitted, because it is directly implied by Proposition A1 and the consistency of $\tilde{\Gamma}_t^{HR}$ in the absence of cross-sectional dependence. This theorem enables us to use $\mathcal{T}_T$ as the test statistic and the $(1-\alpha)$ quantile of Ψ_{bn} as the critical value for the cross-sectional dependence test at the significance level α. If $\hat{\alpha}'\tilde{V}^{-1}\tilde{\Gamma}_T^{HR}\tilde{V}^{-1}\hat{\alpha}$ underestimates $\alpha' Avar\left(\tilde{F}_T\right)\alpha$, the rejection probability exceeds the nominal level α. The distribution of Ψ_{bn} is easy to simulate based on (A.5). Some simulation results are provided in the supplementary appendix.

Appendix B: Proofs

Lemma B1 *Let* $\delta_{nT} = \min\left\{\sqrt{n}, \sqrt{T}\right\}$. *Under Assumption F1–F4, we have the following:*
(i) $\tilde{\lambda}_i - \lambda_i^H = T^{-1}\sum_{s=1}^{T} F_s^H e_{is} + O_P\left(\delta_{nT}^{-2}\right)$
(ii) $T^{-1}\sum_{t=1}^{T} e_{it}\left(\tilde{F}_t - F_t^H\right) = O_P\left(\delta_{nT}^{-2}\right)$
(iii) $T^{-1}\sum_{t=1}^{T}\left\|\tilde{F}_t - F_t^H\right\|^2 = O_P\left(\delta_{nT}^{-2}\right)$
(iv) $T^{-1}\sum_{t=1}^{T}\left(\tilde{F}_t - F_t^H\right)' F_t = O_P\left(\delta_{nT}^{-2}\right)$

The proofs of all four parts of this lemma are given in Bai (2003).

Proof of Theorem 1 Since $\tilde{\Gamma}-\Gamma = \left(E\tilde{\Gamma}^0-\Gamma\right)+\left(\tilde{\Gamma}^0-E\tilde{\Gamma}^0\right)+\left(\tilde{\Gamma}-\tilde{\Gamma}^0\right)$, we can establish the asymptotics of $\tilde{\Gamma}$ by examining each term on the right-hand side. Note that the equation $\tilde{\Gamma}-\Gamma = o_P\left(1\right)$ holds if and only if $a\prime\tilde{\Gamma}a - a'\Gamma a$ for any $a \in \mathbb{R}^p$. Therefore, without loss of generality, we assume that $\tilde{\Gamma}$ is a scalar, that is, $p=1$.
(i) $E\tilde{\Gamma}^0 - \Gamma_{nT}^H = O\left(\frac{1}{d_n^q}\right)$

We have

$$\begin{aligned} E\tilde{\Gamma}^0 - \Gamma_{nT}^H &= \frac{1}{d_n^q}\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T} E\left(\eta_{it}^H\left(\eta_{jt}^H\right)'\right) d_{ij}^q\left[\frac{K\left(\frac{d_{ij}}{d_n}\right)-1}{\left(\frac{d_{ij}}{d_n}\right)^q}\right] \\ &= \frac{1}{d_n^q}\left(\Gamma^{(q)}K_q + o\left(1\right)\right) = O\left(\frac{1}{d_n^q}\right) \end{aligned}$$

as $n, T, d_n \to \infty$.

(ii) $\tilde{\Gamma}^0 - E\tilde{\Gamma}^0 = O_P\left(\sqrt{\frac{\ell_n}{nT}}\right)$

We note that

$$
\begin{aligned}
& E\left(\tilde{\Gamma}^0 - E\tilde{\Gamma}^0\right)^2 \\
&= \frac{1}{n^2T^2}\sum_{i,j=1}^{n}\sum_{a,b=1}^{n}\sum_{t,s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ab}}{d_n}\right)\left[E\left(\eta_{it}^H\eta_{jt}^H\eta_{as}^H\eta_{bs}^H\right) - E\left(\eta_{it}^H\eta_{jt}^H\right)E\left(\eta_{as}^H\eta_{bs}^H\right)\right] \\
&= \frac{1}{n^2T^2}\sum_{i,j=1}^{n}\sum_{a,b=1}^{n}\sum_{t,s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ab}}{d_n}\right)\left[\left\{E\left(\eta_{it}^H\eta_{jt}^H\eta_{as}^H\eta_{bs}^H\right) - E\left(\eta_{it}^H\eta_{jt}^H\right)E\left(\eta_{as}^H\eta_{bs}^H\right)\right.\right. \\
&\quad \left. - E\left(\eta_{it}^H\eta_{as}^H\right)E\left(\eta_{jt}^H\eta_{bs}^H\right) - E\left(\eta_{it}^H\eta_{bs}^H\right)E\left(\eta_{as}^H\eta_{jt}^H\right)\right\} + E\left(\eta_{it}^H\eta_{as}^H\right)E\left(\eta_{jt}^H\eta_{bs}^H\right) \\
&\quad \left. + E\left(\eta_{it}^H\eta_{bs}^H\right)E\left(\eta_{as}^H\eta_{jt}^H\right)\right] \\
&= A_1 + A_2 + A_3.
\end{aligned}
$$

For A_1, we use the linear representation of η_{it}^H in (16) to obtain

$$
\begin{aligned}
& E\left(\eta_{it}^H\eta_{jt}^H\eta_{as}^H\eta_{bs}^H\right) - E\left(\eta_{it}^H\eta_{jt}^H\right)E\left(\eta_{as}^H\eta_{bs}^H\right) - E\left(\eta_{it}^H\eta_{as}^H\right)E\left(\eta_{jt}^H\eta_{bs}^H\right) - E\left(\eta_{it}^H\eta_{bs}^H\right)E\left(\eta_{as}^H\eta_{jt}^H\right) \\
&= \sum_{l=1}^{nTp} r_{it,l}r_{jt,l}r_{as,l}r_{bs,l}\left(E\epsilon_{lt}^4 - 3\right).
\end{aligned}
$$

Thus under Assumption H4,

$$
\begin{aligned}
nT\left|A_1\right| &\leq \frac{1}{nT}\sum_{i,j=1}^{n}\sum_{a,b=1}^{n}\sum_{t,s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ab}}{d_n}\right)\sum_{l=1}^{nTp}\left|r_{it,l}r_{jt,l}r_{as,l}r_{bs,l}\right|\left|E\epsilon_{lt}^4 - 3\right| \\
&\leq \frac{|M-3|}{nT}\sum_{l=1}^{nTp}\underbrace{\left(\sum_{t=1}^{T}\sum_{i=1}^{n}\left|r_{it,l}\right|\right)}_{\leq M}\underbrace{\left(\sum_{j=1}^{n}\left|r_{jt,l}\right|\right)}_{\leq M}\underbrace{\left(\sum_{s=1}^{T}\sum_{a=1}^{n}\left|r_{as,l}\right|\right)}_{\leq M}\underbrace{\left(\sum_{b=1}^{n}\left|r_{bs,l}\right|\right)}_{\leq M} \\
&= O(1). \qquad \text{(B.6)}
\end{aligned}
$$

For A_2,

$$
\begin{aligned}
\frac{nT}{\ell_n}\left|A_2\right| &\leq \frac{1}{\ell_n nT}\sum_{t=1}^{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{a=1}^{n}\sum_{j\in\{d_{ij}\leq d_n\}}\sum_{b\in\{d_{ab}\leq d_n\}}\left|E\left(\eta_{it}^H\eta_{as}^H\right)\right|\left|E\left(\eta_{jt}^H\eta_{bs}^H\right)\right| \\
&\leq \frac{1}{\ell_n nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\sum_{l=1}^{nTp}\left|r_{it,l}\right|\right)\left(\sum_{s=1}^{T}\sum_{a=1}^{n}\left|r_{as,l}\right|\right)\left(\sum_{k=1}^{nTp}\left|r_{jt,k}\right|\right)\left(\sum_{b=1}^{n}\left|r_{bs,k}\right|\right) \\
&= O(1). \qquad \text{(B.7)}
\end{aligned}
$$

Using the same argument, we can show that

$$
\frac{nT}{\ell_n}\left|A_3\right| = O(1). \qquad \text{(B.8)}
$$

Combining (B.6) through (B.8), we have $E\left(\tilde{\Gamma}^0 - E\tilde{\Gamma}^0\right)^2 = O\left(\frac{1}{nT}\right) + O\left(\frac{\ell_n}{nT}\right)$, which implies that

$$\tilde{\Gamma}^0 - E\tilde{\Gamma}^0 = O_P\left(\sqrt{\frac{\ell_n}{nT}}\right)$$

as $\ell_n, n, T \to \infty$ such that $\ell_n/nT \to 0$.

(iii) $\tilde{\Gamma} - \tilde{\Gamma}^0 = O_P\left(\sqrt{\frac{\ell_n}{T}}\right) + O_P\left(\frac{\ell_n}{n}\right)$

We can write

$$\begin{aligned}\tilde{\Gamma} - \tilde{\Gamma}^0 &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\left(\frac{d_{ij}}{d_n}\right)\left(\tilde{\lambda}_i\tilde{\lambda}_j - \lambda_i^H\lambda_j^H\right)\tilde{e}_{it}\tilde{e}_{jt} \\ &+ \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left(\tilde{e}_{it}\tilde{e}_{jt} - e_{it}e_{jt}\right) \\ &:= B_1 + B_2.\end{aligned}$$

Since $\tilde{e}_{it} = e_{it} - \left\{\tilde{F}_t\left(\tilde{\lambda}_i - \lambda_i^H\right) + \left(\tilde{F}_t - F_t^H\right)\lambda_i^H\right\}$, we have

$$\begin{aligned}B_1 &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i - \lambda_i^H\right)\left(\tilde{\lambda}_j - \lambda_j^H\right) + 2\left(\tilde{\lambda}_i - \lambda_i^H\right)\lambda_j^H\right]e_{it}e_{jt} \\ &- \frac{2}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i - \lambda_i^H\right)\left(\tilde{\lambda}_j - \lambda_j^H\right) + 2\left(\tilde{\lambda}_i - \lambda_i^H\right)\lambda_j^H\right]e_{it} \\ &\times\left\{\tilde{F}_t\left(\tilde{\lambda}_j - \lambda_j^H\right) + \left(\tilde{F}_t - F_t^H\right)\lambda_j^H\right\} \\ &+ \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i - \lambda_i^H\right)\left(\tilde{\lambda}_j - \lambda_j^H\right) + 2\left(\tilde{\lambda}_i - \lambda_i^H\right)\lambda_j^H\right] \\ &\times\left\{\tilde{F}_t\left(\tilde{\lambda}_i - \lambda_i^H\right) + \left(\tilde{F}_t - F_t^H\right)\lambda_i^H\right\}\left\{\tilde{F}_t\left(\tilde{\lambda}_j - \lambda_j^H\right) + \left(\tilde{F}_t - F_t^H\right)\lambda_j^H\right\} \\ &:= B_{11} + B_{12} + B_{13}. \qquad\qquad (B.9)\end{aligned}$$

For B_{11},

$$\begin{aligned}B_{11} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)\left(\tilde{\lambda}_i - \lambda_i^H\right)\left(\tilde{\lambda}_j - \lambda_j^H\right)e_{it}e_{jt} \\ &+ \frac{2}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)\left(\tilde{\lambda}_i - \lambda_i^H\right)\lambda_j^H e_{it}e_{jt} \\ &:= B_{11}^{(1)} + B_{11}^{(2)}\end{aligned}$$

For $B_{11}^{(1)}$, it is easy to show that $\frac{T}{\ell_n}\left|B_{11}^{(1)}\right| = O_P(1)$. For $B_{11}^{(2)}$,

$$\begin{aligned}
\sqrt{\frac{T}{\ell_n}}\left|B_{11}^{(2)}\right| &\leq \frac{2}{T}\sum_{t=1}^{T}\left\{\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^H e_{is}\right)^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}e_{it}^4\right)^{1/4}\right]\right.\\
&\quad\left.\times\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{\ell_n}}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\lambda_j^H e_{jt}\right)^2\right]^{1/2}\right\} + o_P(1)\\
&= O_P(1).
\end{aligned}$$

Therefore,

$$B_{11} = O_P\left(\sqrt{\frac{\ell_n}{T}}\right). \tag{B.10}$$

B_{12} can be rewritten as

$$\begin{aligned}
B_{12} &= -\frac{2}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i-\lambda_i^H\right)\left(\tilde{\lambda}_j-\lambda_j^H\right)^2 e_{it}\tilde{F}_t\right.\\
&\quad-\left(\tilde{\lambda}_i-\lambda_i^H\right)\left(\tilde{\lambda}_j-\lambda_j^H\right)e_{it}\lambda_j^H\left(\tilde{F}_t-F_t^H\right)-2\left(\tilde{\lambda}_i-\lambda_i^H\right)\lambda_j^H e_{it}\tilde{F}_t\left(\tilde{\lambda}_j-\lambda_j^H\right)\\
&\quad\left.-2\left(\tilde{\lambda}_i-\lambda_i^H\right)\lambda_j^H e_{it}\left(\tilde{F}_t-F_t^H\right)\lambda_j^H\right]\\
&:= B_{12}^{(1)}+B_{12}^{(2)}+B_{12}^{(3)}+B_{12}^{(4)}.
\end{aligned}$$

For $B_{12}^{(1)}$,

$$\begin{aligned}
&\frac{T^{3/2}}{\ell_n}\left|B_{12}^{(1)}\right|\\
&\leq 2\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^H e_{is}\right)^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}\tilde{F}_t\right)^4\right)^{1/4}\\
&\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^H e_{j\tau}\right)^2\right)^2\right)^{1/2} + o_P(1)\\
&= O_P(1),
\end{aligned} \tag{B.11}$$

because

$$\frac{1}{T}\sum_{t=1}^{T}e_{it}\tilde{F}_t = \frac{1}{T}\sum_{t=1}^{T}e_{it}\left(\tilde{F}_t-F_t^H\right)+O_P\left(\frac{1}{\sqrt{T}}\right)$$

and

$$\frac{1}{T}\sum_{t=1}^{T}e_{it}\left(\tilde{F}_t-F_t^H\right) = O_P\left(\frac{1}{\delta_{nT}^2}\right) \tag{B.12}$$

by Lemma B1(ii). Similarly, we can show that

$$\frac{T}{\ell_n}\left|B_{12}^{(2)}\right| = \frac{T}{\ell_n}\left|B_{12}^{(3)}\right| = \frac{T}{\ell_n}\left|B_{12}^{(4)}\right| = O_P(1).$$

Thus we have

$$B_{12} = O_P\left(\frac{\ell_n}{T}\right). \tag{B.13}$$

For B_{13},

$$\begin{aligned}
B_{13} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)\Big[&\left(\tilde{\lambda}_i - \lambda_i^H\right)^2\left(\tilde{\lambda}_j - \lambda_j^H\right)^2 \tilde{F}_t^2 + 2\left(\tilde{\lambda}_i - \lambda_i^H\right)^2\left(\tilde{\lambda}_j - \lambda_j^H\right)\lambda_j^H \tilde{F}_t^2 \\
&+ 2\left(\tilde{\lambda}_i - \lambda_i^H\right)^2\left(\tilde{\lambda}_j - \lambda_j^H\right)\lambda_j^H \tilde{F}_t\left(\tilde{F}_t - F_t^H\right) + 4\left(\tilde{\lambda}_i - \lambda_i^H\right)^2\left(\lambda_j^H\right)^2 \tilde{F}_t\left(\tilde{F}_t - F_t^H\right) \\
&+ \left(\tilde{\lambda}_i - \lambda_i^H\right)\lambda_i^H\left(\tilde{\lambda}_j - \lambda_j^H\right)\lambda_j^H\left(\tilde{F}_t - F_t^H\right)^2 + 2\left(\tilde{\lambda}_i - \lambda_i^H\right)\lambda_i^H\left(\lambda_j^H\right)^2\left(\tilde{F}_t - F_t^H\right)^2\Big] \\
:= B_{13}^{(1)} + B_{13}^{(2)} + B_{13}^{(3)}
\end{aligned}$$

For $B_{13}^{(1)}$, it is easy to show that $\frac{T^{3/2}}{\ell_n}\left|B_{13}^{(1)}\right| = O_P(1)$. For $B_{13}^{(2)}$,

$$\begin{aligned}
\frac{T\delta_{nT}^2}{\ell_n}\left|B_{13}^{(2)}\right| \leq\ & \frac{2}{\sqrt{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^H e_{is}\right)^4\right)^{1/2}\left|\frac{\delta_{nT}^2}{T}\sum_{t=1}^{T}\tilde{F}_t\left(\tilde{F}_t - F_t^H\right)\right| \\
&\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^H e_{j\tau}\right)^2\right)\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\lambda_j^H\right)^2\right)\right)^{1/2} \\
&+ 4\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^H e_{is}\right)^4\right)^{1/2}\left|\frac{\delta_{nT}^2}{T}\sum_{t=1}^{T}\tilde{F}_t\left(\tilde{F}_t - F_t^H\right)\right| \\
&\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\lambda_j^H\right)^2\right)^2\right)^{1/2} \\
=\ & O_P(1),
\end{aligned}$$

where

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\left(\tilde{F}_t - F_t^H\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\tilde{F}_t - F_t^H\right)^2 + \frac{1}{T}\sum_{t=1}^{T}F_t^H\left(\tilde{F}_t - F_t^H\right) = O_P\left(\frac{1}{\delta_{nT}^2}\right)$$

by Lemmas B1(iii) and (iv).

Similarly, we can show that $\frac{T}{\ell_n}\left|B_{13}^{(3)}\right| = O_P(1)$. Thus we have

$$B_{13} = O_P\left(\frac{\ell_n}{T}\right), \tag{B.14}$$

which, together with (B.10) and (B.13), implies that

$$B_1 = O_P\left(\sqrt{\frac{\ell_n}{T}}\right).$$

We can write B_2 as follows:

$$\begin{aligned}
B_2 &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left[\left(X_{it}-\tilde{\lambda}_i\tilde{F}_t\right)\left(X_{jt}-\tilde{\lambda}_j\tilde{F}_t\right)-e_{it}e_{jt}\right] \\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left[\left(\tilde{\lambda}_i\tilde{F}_t-\lambda_i^H F_t^H\right)\left(\tilde{\lambda}_j\tilde{F}_t-\lambda_j^H F_t^H\right)\right. \\
&\left. -2e_{it}\left(\tilde{\lambda}_j\tilde{F}_t-\lambda_j^H F_t^H\right)\right] \\
&:= B_{21}+B_{22},
\end{aligned}$$

where

$$\begin{aligned}
B_{21} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left(\tilde{\lambda}_i\tilde{F}_t-\lambda_i^H F_t^H\right)\left(\tilde{\lambda}_j\tilde{F}_t-\lambda_j^H F_t^H\right), \\
B_{22} &= -\frac{2}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H e_{it}\left(\tilde{\lambda}_j\tilde{F}_t-\lambda_j^H F_t^H\right).
\end{aligned}$$

B_{21} can be rewritten as

$$\begin{aligned}
B_{21} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left[\left(\tilde{\lambda}_i-\lambda_i^H\right)\tilde{F}_t+\lambda_i^H\left(\tilde{F}_t-F_t^H\right)\right] \\
&\times\left[\left(\tilde{\lambda}_j-\lambda_j^H\right)\tilde{F}_t+\lambda_j^H\left(\tilde{F}_t-F_t^H\right)\right] \\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left[\left(\tilde{\lambda}_i-\lambda_i^H\right)\left(\tilde{\lambda}_j-\lambda_j^H\right)\tilde{F}_t^2\right. \\
&\left. +2\left(\tilde{\lambda}_i-\lambda_i^H\right)\lambda_j^H\tilde{F}_t\left(\tilde{F}_t-F_t^H\right)+\lambda_i^H\lambda_j^H\left(\tilde{F}_t-F_t^H\right)^2\right] \\
&= B_{21}^{(1)}+B_{21}^{(2)}+B_{21}^{(3)}.
\end{aligned}$$

We can show that

$$\frac{T}{\ell_n}\left|B_{21}^{(1)}\right| = \frac{T}{\ell_n}\left|B_{21}^{(2)}\right| = O_P(1) \text{ and } \frac{\delta_{nT}^2}{\ell_n}\left|B_{21}^{(3)}\right| = O_P(1).$$

Thus

$$B_{21} = O_P\left(\frac{\ell_n}{\delta_{nT}^2}\right)$$

For B_{22}, we have

$$
\begin{aligned}
B_{22} &= -\frac{2}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H e_{it}\lambda_j^H \left[\left(\tilde{\lambda}_j-\lambda_j^H\right)\tilde{F}_t+\lambda_j^H\left(\tilde{F}_t-F_t^H\right)\right] \\
&= B_{22}^{(1)}+B_{22}^{(2)}.
\end{aligned}
$$

For $B_{22}^{(1)}$,

$$
\begin{aligned}
&\sqrt{\frac{T}{\ell_n}}\left|B_{22}^{(1)}\right| \\
&\leq 2\left(\frac{1}{n}\sum_{j=1}^{n}\left(\lambda_j^H\right)^2\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^H e_{j\tau}\right)^2\right)^{1/2}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\right| \\
&\times\left(\frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{\sqrt{\ell_n}}\sum_{i=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H e_{it}\right)^2\right)^{1/2} + o_P(1) \\
&= O_P(1).
\end{aligned}
$$

We also have

$$
\frac{\delta_{nT}^2}{\ell_n}\left|B_{22}^{(2)}\right| = O_P(1)
$$

and

$$
B_{22} = O_P\left(\sqrt{\frac{\ell_n}{T}}\right) + \underbrace{O_P\left(\frac{\ell_n}{\delta_{nT}^2}\right)}_{=O_P\left(\frac{\ell_n}{n}\right)+O_P\left(\frac{\ell_n}{T}\right)} = O_P\left(\sqrt{\frac{\ell_n}{T}}\right) + O_P\left(\frac{\ell_n}{n}\right).
$$

Since $B_{21} = O_P\left(\frac{\ell_n}{\delta_{nT}^2}\right)$, we have $B_2 = O_P\left(\sqrt{\frac{\ell_n}{T}}\right) + O_P\left(\frac{\ell_n}{n}\right)$ and

$$
\tilde{\Gamma}-\tilde{\Gamma}^0 = O_P\left(\sqrt{\frac{\ell_n}{T}}\right) + O_P\left(\frac{\ell_n}{n}\right).
$$

This completes the proof.

Supplementary Appendix

Simulation for the cross-sectional dependence test

We conduct simulation studies to examine the finite sample properties of our cross-sectional dependence test. DGP1 and DGP2 in Section 5.1 are employed to generate the data. Since this test is based on the spatial HAC estimator $\tilde{\Gamma}_T$ and does not employ AV-SHAC, we consider the MSE criterion and parametric plug-in method proposed by Kim and Sun (2011) to select the bandwidth. We first approximate the MSE of $\alpha' V^{-1}\tilde{\Gamma}_T V^{-1}\alpha$ with

$$AMSE = \frac{1}{d_n^{2q}} K_q^2 \left(\alpha' V^{-1}\Gamma_T^{(q)} V^{-1}\alpha\right)^2 + \frac{\ell_n}{n} 2\bar{\mathcal{K}}_T \left(\alpha' V^{-1}\Gamma_T^{H} V^{-1}\alpha\right)^2, \tag{S.1}$$

where $\Gamma_T^{(q)} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n E\left[\eta_{iT}^H \left(\eta_{jT}^H\right)'\right] d_{ij}^q$. To avoid the effect of an unduly small value of d_n, our bandwidth selection is based on the following modified MSE criterion,

$$d_n^* = \max\left\{\arg\min_{d_n} AMSE, \underline{d}\right\}, \tag{S.2}$$

where $\underline{d}$ is the prespecified minimum value of the bandwidth.

For the plug-in method, we employ the SAR(1) model

$$\eta_{a,iT}^H = \phi_a \mathcal{W} \eta_{a,iT}^H + u_{iT}, \tag{S.3}$$

where $\eta_{a,iT}^H$ is the ath component of η_{iT}^H and $u_{iT} \sim^{iid} (0,1)$, and we estimate ϕ_a by the QML method, which is given by

$$\hat{\phi}_a = \arg\max_{\phi} \log L\left(\tilde{\eta}_{a,iT} | \phi_a\right) \tag{S.4}$$

with

$$\log L\left(\tilde{\eta}_{a,iT} | \phi_a\right) = -\frac{n}{2}\log\left(\tilde{\eta}_{a,iT} - \phi_a \mathcal{W}\tilde{\eta}_{a,iT}\right)'\left(\tilde{\eta}_{a,iT} - \phi_a \mathcal{W}\tilde{\eta}_{a,iT}\right) - \log\left|\mathbb{I}_n - \phi_a \mathcal{W}\right| + const.$$

$\mathcal{W}$ is a contiguity matrix in which we treat units i and j as neighbors if $d_{ij} \leq 1$. For this matrix, row standardization is applied and all the diagonal elements are zero. See Kim and Sun (2011) for details of this plug-in method.

Table S1 below reports the empirical rejection probabilities (ERPs) of our cross-sectional dependence test. The table shows that the test works very well. The empirical sizes of $\mathcal{T}_T$ are always close to the nominal level $\alpha = 0.05$. In the presence of cross-sectional dependence, the ERP becomes larger than α and grows as the strength of the dependence and/or n, T increase. We set $\underline{d} = 20$ for DGP1, and $\underline{d} = 4$ for DGP2. Additional simulations (not reported here to save space) show that the test tends to lose power when $\underline{d}$ becomes larger.

Table S1: Empirical rejection probabilities of the cross-sectional dependence test ($\alpha = 0.05$)

DGP1				DGP2			
n	T	γ	ERP	n	T	ϱ	ERP
		0.0	0.045			0.0	0.065
100	50	0.3	0.303	100	50	0.1	0.308
		0.5	0.656			0.3	0.690
		0.7	0.925			0.5	0.790
		0.0	0.050			0.0	0.058
100	100	0.3	0.328	100	100	0.1	0.323
		0.5	0.686			0.3	0.710
		0.7	0.939			0.5	0.800
		0.0	0.060			0.0	0.060
100	200	0.3	0.331	100	200	0.1	0.333
		0.5	0.682			0.3	0.718
		0.7	0.944			0.5	0.817
		0.0	0.044			0.0	0.060
150	50	0.3	0.406	144	50	0.1	0.396
		0.5	0.821			0.3	0.783
		0.7	0.986			0.5	0.869
		0.0	0.050			0.0	0.057
150	100	0.3	0.413	144	100	0.1	0.421
		0.5	0.827			0.3	0.798
		0.7	0.990			0.5	0.890
		0.0	0.050			0.0	0.049
150	200	0.3	0.427	144	200	0.1	0.445
		0.5	0.845			0.3	0.813
		0.7	0.993			0.5	0.894

Proof of Proposition A1

Lemma S2 *Under Assumption T5,*

$$\frac{1}{n}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\psi_b\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right)\lambda^H_{(i_1,i_2)}\left(\lambda^H_{(i_1,i_2)}\right)' \to^p \frac{1}{n}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\psi_b\left(\frac{i}{L_n},\frac{i_2}{M_n}\right)\Sigma_{\Lambda^H}.$$

The proof of this lemma is included in the proof of Lemma 1 in Kim and Sun (2013).

Proof of Proposition A1 Let $\Gamma_t^0 = \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{d_{ij}}{d_n}\right) \lambda_i^H \left(\lambda_j^H\right)' e_{it} e_{jt}$ denote the infeasible spatial HAC estimator in time period t. Then

$$\begin{aligned}
\tilde{\Gamma}_t - \Gamma_t^0 &= \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{d_{ij}}{d_n}\right)\left[\tilde{\lambda}_i \tilde{\lambda}_j \tilde{e}_{it}\tilde{e}_{jt} - \lambda_i^H \lambda_j^H e_{it} e_{jt}\right] \\
&= \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i \tilde{\lambda}_j - \lambda_i^H \lambda_j^H\right)\tilde{e}_{it}\tilde{e}_{jt} + \lambda_i^H \lambda_j^H \left(\tilde{e}_{it}\tilde{e}_{jt} - e_{it}e_{jt}\right)\right] \\
&:= C_1 + C_2,
\end{aligned}$$

where

$$\begin{aligned}
C_1 &= \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{d_{ij}}{d_n}\right)\left(\tilde{\lambda}_i \tilde{\lambda}_j - \lambda_i^H \lambda_j^H\right)\tilde{e}_{it}\tilde{e}_{jt}, \\
C_2 &= \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H \lambda_j^H \left(\tilde{e}_{it}\tilde{e}_{jt} - e_{it}e_{jt}\right).
\end{aligned}$$

For C_1, we have

$$\begin{aligned}
C_1 = &\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i - \lambda_i^H\right)\left(\tilde{\lambda}_j - \lambda_j^H\right)' + 2\left(\tilde{\lambda}_i - \lambda_i^H\right)\left(\lambda_j^H\right)'\right] e_{it}e_{jt} \\
&- \frac{2}{n}\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i - \lambda_i^H\right)\left(\tilde{\lambda}_j - \lambda_j^H\right) + 2\left(\tilde{\lambda}_i - \lambda_i^H\right)\lambda_j^H\right] e_{it} \\
&\times \left\{\tilde{F}_t\left(\tilde{\lambda}_j - \lambda_j^H\right) + \left(\tilde{F}_t - F_t^H\right)\lambda_j^H\right\} \\
&+ \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i - \lambda_i^H\right)\left(\tilde{\lambda}_j - \lambda_j^H\right) + 2\left(\tilde{\lambda}_i - \lambda_i^H\right)\lambda_j^H\right] \\
&\times \left\{\tilde{F}_t\left(\tilde{\lambda}_i - \lambda_i^H\right) + \left(\tilde{F}_t - F_t^H\right)\lambda_i^H\right\}\left\{\tilde{F}_t\left(\tilde{\lambda}_j - \lambda_j^H\right) + \left(\tilde{F}_t - F_t^H\right)\lambda_j^H\right\} \\
&:= C_{11} + C_{12} + C_{13}.
\end{aligned}$$

For C_{11},

$$\begin{aligned}
C_{11} = &\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{d_{ij}}{d_n}\right)\left(\tilde{\lambda}_i - \lambda_i^H\right)\left(\tilde{\lambda}_j - \lambda_j^H\right) e_{it}e_{jt} \\
&+ \frac{2}{n}\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{d_{ij}}{d_n}\right)\left(\tilde{\lambda}_i - \lambda_i^H\right)\lambda_j^H e_{it}e_{jt} \\
&:= C_{11}^{(1)} + C_{11}^{(2)}.
\end{aligned}$$

For $C_{11}^{(1)}$

$$\left|C_{11}^{(1)}\right| \leq \frac{\ell_n}{T}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^H e_{is}\right)^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}e_{it}^4\right)^{1/4}$$
$$\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^H e_{j\tau}\right)^2\right)\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}e_{jt}^2\right)\right)^{1/2}+o_P(1)$$
$$= O_P\left(\frac{\ell_n}{T}\right). \tag{S.5}$$

For $C_{11}^{(2)}$,

$$\begin{aligned}C_{11}^{(2)} &= \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\left(\tilde{\lambda}_i-\lambda_i^H\right)\lambda_j^H e_{it}e_{jt}\\ &= \frac{2}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)F_s^H e_{is}e_{it}\eta_{jt}^H+o_P(1)\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)F_s^H e_{is}e_{it}\eta_{jt}^H\\ &= \frac{\sqrt{\ell_n}}{nT}\sum_{i=1}^{n}\sum_{s=1}^{T}\left(F_s^H e_{is}e_{it}-E\left(F_s^H e_{is}e_{it}\right)\right)\frac{1}{\sqrt{\ell_n}}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\eta_{jt}^H\\ &+\frac{\sqrt{\ell_n}}{nT}\sum_{i=1}^{n}\sum_{s=1}^{T}E\left(F_s^H e_{is}e_{it}\right)\frac{1}{\sqrt{\ell_n}}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\eta_{jt}^H\\ &= c_1+c_2.\end{aligned}$$

For c_1,

$$\begin{aligned}|c_1| &\leq \sqrt{\frac{\ell_n}{nT}}\left|\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{s=1}^{T}\left(F_s^H e_{is}e_{it}-E\left(F_s^H e_{is}e_{it}\right)\right)\right|\sup_a\left|\frac{1}{\sqrt{\ell_n}}\sum_{j=1}^{n}K\left(\frac{d_{aj}}{d_n}\right)\eta_{jt}^H\right|\\ &= O_P\left(\sqrt{\frac{\ell_n}{nT}}\right)\end{aligned}$$

under Assumption T6.

For c_2,

$$\begin{aligned}|c_2| &\leq \frac{\sqrt{\ell_n}}{nT}\sum_{i=1}^{n}\sum_{s=1}^{T}\left|E\left(F_s^H\right)\right|\left|E\left(e_{is}e_{it}\right)\right|\left|\frac{1}{\sqrt{\ell_n}}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\eta_{jt}^H\right| \\ &\leq \frac{\sqrt{\ell_n}}{T}\frac{M^2}{n}\sum_{i=1}^{n}\left|\frac{1}{\sqrt{\ell_n}}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\eta_{jt}^H\right| \\ &= O_P\left(\frac{\sqrt{\ell_n}}{T}\right).\end{aligned}$$

Thus we have

$$C_{11} = o_P\left(1\right), \tag{S.6}$$

as $\ell_n/T \to 0$ and $d_n, \ell_n, n, T \to \infty$.

C_{12} can be rewritten as

$$\begin{aligned}C_{12} = &-\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\Big[\left(\tilde{\lambda}_i-\lambda_i^H\right)\left(\tilde{\lambda}_j-\lambda_j^H\right)^2 e_{it}\tilde{F}_t \\ &+\left(\tilde{\lambda}_i-\lambda_i^H\right)\left(\tilde{\lambda}_j-\lambda_j^H\right)e_{it}\lambda_j^H\left(\tilde{F}_t-F_t^H\right)+2\left(\tilde{\lambda}_i-\lambda_i^H\right)\lambda_j^H e_{iT}\tilde{F}_t\left(\tilde{\lambda}_j-\lambda_j^H\right) \\ &+2\left(\tilde{\lambda}_i-\lambda_i^H\right)\left(\lambda_j^H\right)^2 e_{it}\left(\tilde{F}_t-F_t^H\right)\Big] \\ &:= C_{12}^{(1)}+C_{12}^{(2)}+C_{12}^{(3)}+C_{12}^{(4)}.\end{aligned}$$

For $C_{12}^{(1)}$,

$$\begin{aligned}\left|C_{12}^{(1)}\right| &\leq 2\left|\tilde{F}_t\right|\frac{\ell_n}{T\sqrt{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s e_{is}\right)^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}e_{it}^4\right)^{1/4} \\ &\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau e_{j\tau}\right)^2\right)^2\right)^{1/2}+o_P\left(1\right) \\ &= o_P\left(1\right).\end{aligned} \tag{S.7}$$

For $C_{12}^{(2)}$,

$$\begin{aligned}\left|C_{12}^{(2)}\right| &\leq 2\frac{\ell_n}{T\sqrt{n}}\left|\sqrt{n}\left(\tilde{F}_t-F_t^H\right)\right|\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^H e_{is}\right)^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}e_{it}^4\right)^{1/4} \\ &\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^H e_{j\tau}\right)^2\right)\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\lambda_j^H\right)^2\right)\right)^{1/2} \\ &= o_P\left(1\right)\end{aligned} \tag{S.8}$$

For $C_{12}^{(3)}$,

$$
\begin{aligned}
&\left|C_{12}^{(3)}\right| \\
&\leq 4\left|\tilde{F}_t\right| \frac{\ell_n}{T}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^H e_{is}\right)^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}e_{it}^4\right)^{1/4} \\
&\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^H e_{j\tau}\right)^2\right)\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\lambda_j^H\right)^2\right)\right)^{1/2}+o_P(1) \\
&= o_P(1).
\end{aligned}
\tag{S.9}
$$

For $C_{12}^{(4)}$,

$$
\begin{aligned}
\left|C_{12}^{(4)}\right| &\leq 4\frac{\ell_n}{\sqrt{nT}}\left|\sqrt{n}\left(\tilde{F}_t-F_t^H\right)\right|\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^H e_{is}\right)^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}e_{it}^4\right)^{1/4} \\
&\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\lambda_j^H\right)^2\right)^2\right)^{1/2}+o_P(1) \\
&= o_P(1).
\end{aligned}
\tag{S.10}
$$

From (S.7), (S.8), (S.9), and (S.10), we have

$$
C_{12} = o_P(1),
$$

as $\ell_n/T, \ell_n/\sqrt{nT} \to 0$ and $d_n, \ell_n, n, T \to \infty$.

For C_{13},

$$
\begin{aligned}
C_{13} &= \frac{\tilde{F}_t^2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i-\lambda_i^H\right)^2\left(\tilde{\lambda}_j-\lambda_j^H\right)^2+2\left(\tilde{\lambda}_i-\lambda_i^H\right)^2\left(\tilde{\lambda}_j-\lambda_j^H\right)\lambda_j^H\right] \\
&+\frac{2\tilde{F}_t}{n}\left(\tilde{F}_t-F_t^H\right)\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i-\lambda_i^H\right)^2\left(\tilde{\lambda}_j-\lambda_j^H\right)\lambda_j^H+2\left(\tilde{\lambda}_i-\lambda_i^H\right)^2\left(\lambda_j^H\right)^2\right] \\
&+\frac{\left(\tilde{F}_t-F_t^H\right)^2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\left[\left(\tilde{\lambda}_i-\lambda_i^H\right)\left(\tilde{\lambda}_j-\lambda_j^H\right)\lambda_i^H\lambda_j^H+2\left(\tilde{\lambda}_i-\lambda_i^H\right)\lambda_i^H\left(\lambda_j^H\right)^2\right] \\
&= A_{13}^{(1)}+A_{13}^{(2)}+A_{13}^{(3)}
\end{aligned}
$$

For $C_{13}^{(1)}$,

$$
\begin{aligned}
\left|C_{13}^{(1)}\right| \leq & \frac{\ell_n}{T^2}\tilde{F}_t^2\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^H e_{is}\right)^4\right)^{1/2} \\
& \times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^H e_{j\tau}\right)^2\right)^2\right)^{1/2} \\
& + 2\frac{\ell_n}{T\sqrt{T}}\tilde{F}_t^2\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^H e_{is}\right)^4\right)^{1/2} \\
& \times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^H e_{j\tau}\right)^2\right)\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\lambda_j^H\right)^2\right)\right)^{1/2} \\
& = o_P(1)
\end{aligned}
$$

Based on similar procedures, it is easy to show that

$$C_{13}^{(2)} = C_{13}^{(3)} = o_P(1).$$

Thus we have

$$C_1 = o_P(1). \tag{S.11}$$

For C_2,

$$
\begin{aligned}
C_2 &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left[\left(X_{it}-\tilde{\lambda}_i\tilde{F}_t\right)\left(X_{jt}-\tilde{\lambda}_j\tilde{F}_t\right)-e_{it}e_{jt}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left[\left(\tilde{\lambda}_i\tilde{F}_t-\lambda_i^H F_t^H\right)\left(\tilde{\lambda}_j\tilde{F}_t-\lambda_j^H F_t^H\right)-2e_{it}\left(\tilde{\lambda}_j\tilde{F}_t-\lambda_j^H F_t^H\right)\right] \\
&:= C_{21}+C_{22},
\end{aligned}
$$

where

$$
\begin{aligned}
C_{21} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left(\tilde{\lambda}_i\tilde{F}_t-\lambda_i^H F_t^H\right)\left(\tilde{\lambda}_j\tilde{F}_t-\lambda_j^H F_t^H\right), \\
C_{22} &= -\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H e_{it}\left(\tilde{\lambda}_j\tilde{F}_t-\lambda_j^H F_t^H\right).
\end{aligned}
$$

C_{21} can be rewritten as

$$C_{21} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left[\left(\tilde{\lambda}_i - \lambda_i^H\right)\tilde{F}_t + \lambda_i^H\left(\tilde{F}_t - F_t^H\right)\right]$$
$$\times\left[\left(\tilde{\lambda}_j - \lambda_j^H\right)\tilde{F}_t + \lambda_j^H\left(\tilde{F}_t - F_t^H\right)\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\lambda_j^H\left[\left(\tilde{\lambda}_i - \lambda_i^H\right)\left(\tilde{\lambda}_j - \lambda_j^H\right)\tilde{F}_t^2\right.$$
$$\left. +2\left(\tilde{\lambda}_i - \lambda_i^H\right)\lambda_j^H\tilde{F}_t\left(\tilde{F}_t - F_t^H\right) + \lambda_i^H\lambda_j^H\left(\tilde{F}_t - F_t^H\right)^2\right]$$
$$= C_{21}^{(1)} + C_{21}^{(2)} + C_{21}^{(3)}.$$

For $C_{21}^{(1)}$,

$$\left|C_{21}^{(1)}\right|$$
$$\leq \tilde{F}_t^2\frac{\ell_n}{T}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^He_{is}\right)^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\lambda_i^H\right)^4\right)^{1/4}$$
$$\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^He_{j\tau}\right)^2\right)\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\lambda_j^H\right)^2\right)\right)^{1/2} + o_P(1)$$
$$= o_P(1).$$

For $C_{21}^{(2)}$,

$$\left|C_{21}^{(2)}\right| \leq \frac{2\ell_n}{\sqrt{nT}}\left|\tilde{F}_t\right|\left|\sqrt{n}\left(\tilde{F}_t - F_t^H\right)\right|\left(\frac{1}{n}\sum_{i=1}^{n}\left(\lambda_i^H\right)^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^He_{is}\right)^2\right)^{1/4}$$
$$\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\lambda_j^H\right)^2\right)^2\right)^{1/2} + o_P(1)$$
$$= o_P(1),$$

as $\ell_n/\sqrt{nT} \to 0$.

For $C_{21}^{(3)}$,

$$\left|C_{21}^{(3)}\right|$$
$$\leq \frac{\ell_n}{n}\left(\sqrt{n}\left(\tilde{F}_t - F_t^H\right)\right)^2\left(\frac{1}{n}\sum_{i=1}^{n}\left(\lambda_i^H\right)^4\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\{d_{ij}\leq d_n\}}\left(\lambda_j^H\right)^2\right)^2\right)^{1/2}$$
$$= O_P\left(\frac{\ell_n}{n}\right).$$

For C_{22},

$$
\begin{aligned}
C_{22} &= -\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H e_{it}\lambda_j^H\left[\left(\tilde{\lambda}_j-\lambda_j^H\right)\tilde{F}_t+\lambda_j^H\left(\tilde{F}_t-F_t^H\right)\right] \\
&= C_{22}^{(1)}+C_{22}^{(2)}.
\end{aligned}
$$

For $C_{22}^{(1)}$,

$$
\begin{aligned}
&C_{22}^{(1)} \\
&\le \frac{2\ell_n}{\sqrt{nT}}\left|\tilde{F}_t\right|\left(\frac{1}{n}\sum_{i=1}^{n}\left(\lambda_j^H\right)^2\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^H e_{j\tau}\right)^2\right)^{1/2} \\
&\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{\ell_n}}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\eta_{it}^H\right)^2\right)^{1/2}+o_P(1) \\
&\le \frac{2\ell_n}{\sqrt{nT}}\left|\tilde{F}_t\right|\left(\frac{1}{n}\sum_{i=1}^{n}\left(\lambda_j^H\right)^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}F_\tau^H e_{j\tau}\right)^4\right)^{1/4} \\
&\times\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{\ell_n}}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\eta_{it}^H\right)^2\right)^{1/2}+o_P(1) \\
&= o_P(1),
\end{aligned}
$$

as $\ell_n/\sqrt{nT}\to 0$.

For $C_{22}^{(2)}$,

$$
\begin{aligned}
&\left|C_{22}^{(2)}\right| \\
&\le \frac{2\ell_n}{n}\left|\sqrt{n}\left(\tilde{F}_t-F_t^H\right)\right|\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\lambda_i^H e_{it}\frac{1}{\ell_n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\left(\lambda_j^H\right)^2\right| \\
&\le \frac{2\ell_n}{n}\left|\sqrt{n}\left(\tilde{F}_t-F_t^H\right)\right|\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\lambda_i^H e_{it}\right|\left|\sup_c\frac{1}{\ell_n}\sum_{j=1}^{n}K\left(\frac{d_{cj}}{d_n}\right)\left(\lambda_j^H\right)^2\right| \\
&= O_P\left(\frac{\ell_n}{n}\right).
\end{aligned}
$$

Under the rate condition $\ell_n/n\to b^o$ with $\ell_n/T=o(1)$, we have

$$
\tilde{\Gamma}_t=\tilde{\Gamma}_t^0+C_{21}^{(3)}+C_{22}^{(2)}+o_P(1).
$$

Using matrix notation, we have

$$\tilde{\Gamma}_t^0 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right)\eta_{it}^H\left(\eta_{jt}^H\right)',$$

$$C_1 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right)\lambda_i^H\left(\lambda_i^H\right)'\left(\tilde{F}_t - F_t^H\right)\left(\tilde{F}_t - F_t^H\right)'\lambda_j^H\left(\lambda_j^H\right)',$$

$$C_2 = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right)\left[\eta_{it}^H\left(\tilde{F}_t - F_t^H\right)'\lambda_j^H\left(\lambda_j^H\right)' + \lambda_i^H\left(\lambda_i^H\right)'\left(\tilde{F}_t - F_t^H\right)\left(\eta_{it}^H\right)'\right].$$

By the Fourier series representation, $\tilde{\Gamma}_t$ can be written as

$$\begin{aligned}
\tilde{\Gamma}_t &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right)\left[\eta_{it}^H - \lambda_i^H\left(\lambda_i^H\right)'\left(\tilde{F}_t - F_t^H\right)\right]\left[\eta_{jt}^H - \lambda_j^H\left(\lambda_j^H\right)'\left(\tilde{F}_t - F_t^H\right)\right]' + o_P(1) \\
&= \sum_{\iota=1}^{\infty}\kappa_\iota\left(\frac{1}{\sqrt{n}}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\psi_b\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right)\left[\eta_{(i_1,i_2),t}^H - \lambda_{(i_1,i_2)}^H\left(\lambda_{(i_1,i_2)}^H\right)'\left(\tilde{F}_t - F_t^H\right)\right]\right) \\
&\times\left(\frac{1}{\sqrt{n}}\sum_{j_1=1}^{L_n}\sum_{j_2=1}^{M_n}\psi_b\left(\frac{j}{L_n},\frac{j_2}{M_n}\right)\left[\eta_{(j_1,j_2),t}^H - \lambda_{(j_1,j_2)}^H\left(\lambda_{(j_1,j_2)}^H\right)'\left(\tilde{F}_t - F_t^H\right)\right]\right)' + o_P(1) \\
&= \sum_{\iota=1}^{\infty}\kappa_\iota\frac{1}{\sqrt{n}}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\psi_b\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right)\left(\eta_{(i_1,i_2),t}^H - \lambda_{(i_1,i_2)}^H\left(\lambda_{(i_1,i_2)}^H\right)'V^{-1}\left(\frac{\tilde{F}'F^H}{T}\right)\frac{1}{n}\sum_{a_1=1}^{L_n}\sum_{a_2=1}^{M_n}\eta_{(a_1,a_2),t}^H\right) \\
&\times\frac{1}{\sqrt{n}}\sum_{j_1=1}^{L_n}\sum_{j_2=1}^{M_n}\psi_b\left(\frac{j_1}{L_n},\frac{j_2}{M_n}\right)\left(\eta_{(j_1,j_2),t}^H - \lambda_{(j_1,j_2)}^H\left(\lambda_{(j_1,j_2)}^H\right)'V^{-1}\left(\frac{\tilde{F}'F^H}{T}\right)\frac{1}{n}\sum_{a_1=1}^{L_n}\sum_{a_2=1}^{M_n}\eta_{(a_1,a_2),t}^H\right)' \\
&+ o_P(1) \qquad \text{(S.12)}
\end{aligned}$$

According to Lemma 1, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\psi_b\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right)\lambda_{(i_1,i_2)}^H\left(\lambda_{(i_1,i_2)}^H\right)'V^{-1}\left(\frac{\tilde{F}'F^H}{T}\right)\frac{1}{n}\sum_{a_1=1}^{L_n}\sum_{a_2=1}^{M_n}\eta_{(a_1,a_2),t}^H \\
&= \frac{1}{n}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\psi_b\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right)\left(\Sigma_{\Lambda^H} + o_P(1)\right)V^{-1}\left(\frac{\tilde{F}'F^H}{T}\right)\frac{1}{\sqrt{n}}\sum_{a_1=1}^{L_n}\sum_{a_2=1}^{M_n}\eta_{(a_1,a_2),t}^H \\
&= \frac{1}{\sqrt{n}}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\psi_b\left(\frac{i_1}{L_n},\frac{i_2}{M_n}\right)\frac{1}{n}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\eta_{(i_{1t},i_{2t}),t}^H + o_P(1), \qquad \text{(S.13)}
\end{aligned}$$

where we use $\Sigma_{\Lambda^H} V^{-1} \left(\tilde{F}' F^H / T\right) \to^p \mathbb{I}_p$. Substituting this result in (S.12), we have

$$\begin{aligned}
\tilde{\Gamma}_t &= \sum_{\iota=1}^{\infty} \kappa_\iota \frac{1}{\sqrt{n}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \psi_b \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}\right) \left(\eta^H_{(i_1,i_2),t} - \frac{1}{n} \sum_{a_1=1}^{L_n} \sum_{a_2=1}^{M_n} \eta^H_{(a_1,a_2),t}\right) \\
&\times \frac{1}{\sqrt{n}} \sum_{j_1=1}^{L_n} \sum_{j_2=1}^{M_n} \psi_b \left(\frac{j_1}{L_n}, \frac{j_2}{M_n}\right) \left(\eta^H_{(j_1,j_2),t} - \frac{1}{n} \sum_{a_1=1}^{L_n} \sum_{a_2=1}^{M_n} \eta^H_{(a_1,a_2),t}\right)' + o_P(1) \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right) \left(\eta^H_{it} - \frac{1}{n} \sum_{a=1}^{n} \eta^H_{at}\right) \left(\eta^H_{jt} - \frac{1}{n} \sum_{a=1}^{n} \eta^H_{at}\right)' + o_P(1) \\
&\overset{a}{\sim} J_t \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{d_{ij}}{d_n}\right) \left(\xi_i - \bar{\xi}\right) \left(\xi_j - \bar{\xi}\right)' J_t' + o_P(1),
\end{aligned} \tag{S.14}$$

where the asymptotic equivalence is a direct application of Proposition 2 in Kim and Sun (2013). This completes the proof.

References

[1] Andrews, D. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59(3), 817–858.

[2] Bai, J. (2003) Inferential theory for factor models of large dimensions, *Econometrica,* 71, 135–172.

[3] Bai, J., and Ng, S. (2006) Confidence intervals for diffusion index forecasts and inference with factor-augmented regressions, *Econometrica,* 74, 1133–1150.

[4] Bester, A., Conley, T., Hansen, C., and Vogelsang, T. (2016) Fixed-b asymptotics for spatially dependent robust nonparametric covariance matrix estimators, *Econometric Theory*, 32(1), 154–186.

[5] Chamberlain, G., and Rothschild, M. (1983) Arbitrage factor structure, and mean-variance analysis of large asset markets, *Econometrica,* 51, 1281–1304.

[6] Chen, X., and Conley, T. (2001) A new semiparametric spatial model for panel time series, *Journal of Econometrics,* 105(1), 59–83.

[7] Christiansen, C., Eriksen, J. N., and Møller, S. V. (2014) Forecasting US recession: The role of sentiment, *Journal of Banking and Finance*, 49, 459–468.

[8] Conley, T. (1999) GMM estimation with cross sectional dependence, *Journal of Econometrics,* 92(1), 1–45.

[9] Conley, T., and Dupor, B. (2003) A spatial analysis of sectoral complementarity, *Journal of Political Economics*, 111(2), 311–352.

[10] Conley, T., and Ligon, E. (2002) Economic distance and cross-country spillovers, *Journal of Economic Growth*, 7(2), 157–187.

[11] Conley, T., and Molinari, F. (2007) Spatial correlation robust inference with errors in location or distance, *Journal of Econometrics* 140(1), 76–96.

[12] Connor, G., and Korajczyk, R. A. (1986) Performance measurement with the arbitrage pricing theory, *Journal of Financial Economics,* 15, 373–394.

[13] Connor, G., and Korajczyk, R. A. (1993) A test for the number of factors in an approximate factor model, *Journal of Finance,* 48(4), 1263–1291.

[14] Giglio, S., Kelly, B., and Pruitt, S. (2016) Systemic risk and the macroeconomy: An empirical evaluation, *Journal of Financial Economics*, 119, 457–471.

[15] Gonçalves, S., and Perron, B. (2014) Bootstrapping factor-augmented regression models, *Journal of Econometrics,* 182, 156–173.

[16] Gonçalves, S., and Perron, B. (2020) Bootstrapping factor models with cross sectional dependence, *Journal of Econometrics,* 218, 476–495.

[17] Gonçalves, S., Perron, B., and Djogbenou, A. (2015) Bootstrap prediction intervals for factor models, *Journal of Time Series Analysis*, 36 (3), 481–502.

[18] Jurado, K., Ludvigson, S. C., and Ng, S. (2015) Measuring uncertainty, *American Economic Review*, 105(3), 1177–1216.

[19] Kelejian, H., and Prucha, I. (2007) HAC estimation in a spatial framework, *Journal of Econometrics*, 140(1), 131–154.

[20] Kim, M. S., and Sun, Y. (2011) Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix, *Journal of Econometrics*, 160, 349–371

[21] Kim, M. S., and Sun, Y. (2013) Heteroskedasticity and spatiotemporal dependence robust inference for linear panel models with fixed effects, *Journal of Econometrics*, 177, 85–108.

[22] Ludvigson, S. C., and Ng, S. (2007) The empirical risk-return relation: a factor analysis approach, *Journal of Financial Economics*, 83, 171–222.

[23] Ludvigson, S. C., and Ng, S. (2009) Macro factors in bond risk premia, *Review of Financial Studies*, 22(12), 5027–5067.

[24] Ludvigson, S. C., Ng, S. (2010) A factor analysis of bond risk premia, *Handbook of Empirical Economics and Finance*, 313–372.

[25] Newey, W., and West, K. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703–708.

[26] Pesaran, H., and Tosetti, E. (2011) Large panels with common factors and spatial correlation, *Journal of Econometrics*, 161(2), 182–202.

[27] Pinkse, J., Slade, M. E., and Brett, C. (2002) Spatial price competition: A semiparametric approach, *Econometrica*, 70(3), 1111–1153.

[28] Politis, D. (2011) Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory,* 27, 703–744.

[29] Robinson, P. M. (2011) Asymptotic theory for nonparametric regression with spatial data, *Journal of Econometrics*, 165(1), 5–19.

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

[31] Stock, J. H., and Watson, M. W. (2005) Understanding changes in international business cycle dynamics, *Journal of European Economic Association*, 3(5), 968–1006.

[32] Sun, Y., and Kim, M. S. (2015) Asymptotic F test in a GMM framework with cross sectional dependence, *Review of Economics and Statistics*, 97(1), 210–223.