INFERENCE WITHOUT SMOOTHING FOR LARGE PANELS WITH CROSS-SECTIONAL AND TEMPORAL DEPENDENCE

JAVIER HIDALGO AND MARCIA SCHAFGANS

Abstract. This paper addresses inference in large panel data models in the presence of both cross-sectional and temporal dependence of unknown form. We are interested in making inferences that do not rely on the choice of any smoothing parameter as is the case with the often employed "HAC" estimator for the covariance matrix. To that end, we propose a cluster estimator for the asymptotic covariance of the estimators and valid bootstrap schemes that do not require the selection of a bandwidth or smoothing parameter and accommodate the nonparametric nature of both temporal and cross-sectional dependence. Our approach is based on the observation that the spectral representation of the fixed effect panel data model is such that the errors become approximately temporally uncorrelated. Our proposed bootstrap schemes can be viewed as wild bootstraps in the frequency domain. We present some Monte Carlo simulations to shed some light on the small sample performance of our inferential procedure.
JEL classification: C12, C13, C23

Keywords: Large panel data models. Cross-sectional strong-dependence. Central Limit Theorems. Clustering. Discrete Fourier Transformation. Nonparametric bootstrap algorithms.

Affiliation: Economics Department, London School of Economics.

[0] *Corresponding author:* Marcia Schafgans. Economics Department, London School of Economics, Houghton Street, London WC2A 2AE, UK. Tel.: +44 207 955 7487; fax: +44 207 955 6592.
E-mail addresses f.j.hidalgo@lse.ac.uk (J. Hidalgo), m.schafgans@lse.ac.uk.(M. Schafgans)

1. INTRODUCTION

Nowadays we often encounter panel data sets where both the number of individuals, n, and the time dimension, T, are large or increase without limit. Phillips and Moon (1999) and Pesaran and Yamagata (2008) provide some theoretical results for the parameter estimators in large panel data models, that is where both n and T tend to infinity. These works were done under the assumption of no dependence among the cross-sectional units. Yet, it is well recognized that the latter assumption is not very realistic, and there has been a surge of work on how to provide valid inferences when this type of dependence is present. The issues are closely related to Zellner's (1962) $SURE$ (Seemingly Unrelated Regression) model, be it that here both dimensions are allowed to increase without limit.

Once one accepts the possibility that the errors of the model may exhibit cross-sectional and/or temporal dependence, a key component to make valid inferences is the consistent estimation of the asymptotic covariance matrix of the estimators. For that purpose, we might proceed by explicitly assuming some specific dependence structure on the error term. In our context, this route appears to be quite cumbersome mainly for two reasons. First, it is quite difficult to specify an appropriate model in the presence of cross-sectional dependence as there are ample generic models capable to justify such a dependence. Some examples are the Spatial Autoregressive (SAR) model of Cliff and Ord (1973), which has its origins in Whittle (1954), Andrews' (2005) proposal to capture common shocks (e.g., macroeconomic, technological, legal/institutional) across observations and Pesaran's (2006) factor model. Second, in many settings it may be quite unrealistic to assume that the temporal dependence is the same for all individuals, so finding a correct specification may be infeasible as n increases with no limit. Inferential properties based on parameter estimates that use a specific (wrong) structure, moreover, may be worse than the least squares estimates (LSE). The latter observation was first documented in Engle (1974) and latter examined in Nicholls and Pagan (1977), who illustrated the adverse consequences of imposing incorrect temporal dependence assumptions on inference, say when the practitioner assumes an $AR(1)$ model instead of the true underlying $AR(2)$ specification.

As the task of finding an appropriate model for the dependence can be very daunting, one of our main aims in this paper is then to provide inference in panel data not only when the error term (potentially) exhibits both temporal and cross-sectional dependence, but also more importantly doing so without relying on any parametric functional form for such a dependence. Under these circumstances, one standard methodology is based on the HAC estimator, whose implementation requires the choice of one (or more) bandwidth parameter(s).[1] While this approach is often invoked and used in the context of time series regression models, application of spatial HAC estimators is less common. The use of HAC estimators in spatial econometrics was advocated by Conley (1999) and Kelejian and Prucha (2007) studied its use in Cliff–Ord type spatial models. Recently, a HAC estimator accounting for both the temporal and the spatial correlation been considered by Kim

[1] In a time series regression model context several proposals, both in the time and frequency domain, have been employed and bootstrap applications commonly approximate the long term covariance by using a long AR polynomial (sieve method). Other methods include the use of orthogonal polynomials, see, e.g., Sun (2013) and Phillips (2005), instead of the use of Fourier sequences. All of them have in common that they require the choice of a bandwidth parameter and/or base function. Lazarus et al. (2018) provide an interesting simulation study.

and Sun (2013). The implementation requires not only the selection of a bandwidth parameter but, also more importantly, an associated measure of distance between the cross-sectional units. This explicitly assumes that there is some type of ordering among the individuals or cross-sectional units which, in contrast with the time dimension, is not unambiguous. Even if one accepts the existence of such an ordering, it is likely that various economics and/or geographical distance measures, each requiring their own bandwidth, may be required to encapsulate the order. For instance, simply relying on the geographic "as the crow flies" distance measure for ordering is questionable as one cannot expect that two cross-sectional units located in the Rockies would behave the same as if they were in the Midwest. Clearly, a distance measure which captures the topography and other economic measures may be required. In addition, if we recognize that the temporal dependence may not be the same for all individuals, even the selection of a bandwidth parameter to account for the temporal dependence may become infeasible. Any cross-validation algorithm used to determine the bandwidth parameter for temporal dependence may then need to be performed for each individual.

To deal with the potential caveats of HAC estimators, we shall propose a cluster based estimator which is able to take into account both types of dependence and permits the temporal dependence to vary across individuals, see Condition $C1$ and its discussion, extending the work of Arellano (1987) and Driscoll and Kraay (1998) in a substantial way. While Driscoll and Kraay (1998) employed a cluster type of estimator to account for the cross-sectional dependence, they relied on the HAC methodology to accommodate the temporal dependence subjecting it to the drawback mentioned before. We avoid the use of the HAC methodology altogether. In addition, we provide a new CLT that accounts for an unknown and general temporal spatial dependence structure that permits strong spatial dependence. Our approach allows for more general dependence structure than permitted by Kim and Sun (2013) and Driscoll and Kraay (1998). Our new results can therefore be regarded as providing primitive conditions that guarantee Kim and Sun's and Driscoll and Kraay's assumption of the existence of a suitable CLT.

Our approach is based on the observation that the spectral representation of the fixed effect panel data model (2.1) is such that the errors become approximately temporally uncorrelated whilst heteroskedastic. It is this observation that enables us to conduct inference without any smoothing. To provide finite sample improvements for inference based on our cluster estimator, we present and examine bootstrap schemes which also do not require the choice of any bandwidth parameter, contrary to the sieve or moving block bootstrap (henceforth denoted MBB). Two bootstrap algorithms are presented, one where we assume homogeneous temporal dependence, which we shall denote as the naïve bootstrap, and a second one, denoted the wild bootstrap, where we allow for heterogeneous temporal dependence. Our bootstrap schemes can be viewed as wild bootstraps in the frequency domain which are shown to have good finite sample properties.

We compare our proposal to other methods that also do not require any ordering of the cross-sectional units. In particular, we consider Driscoll and Kraay's HAC estimator and the fixed-b asymptotic framework advocated by Vogelsang (2012). We also consider the MBB bootstrap applied to the vector containing all the individual observations at each point in time as proposed by Gonçalves (2011).

While our estimator does permit more general spatio temporal dependence and does not require any smoothing parameters, in line with Robinson (1998), the approach examined in Section 2 precludes the presence of conditional heteroskedasticity. In Section 4, we examine how we can relax this by introducing a multiplicative error structure, $v_{pt} = \sigma_1(w_p)\sigma_2(\varrho_t)\, u_{pt}$, where w_p and ϱ_t can be functions of the fixed effects and/or variables which are correlated with the included regressors. It is worth noting that we do not need to observe these variables, as is the case when w_p, say, is the fixed effect. That is, we can allow for "groupwise" heteroskedasticity and applications in development economics are commonplace, see Deaton (1996) and Greene (2018). Of particular interest, here, is the realization that our cluster based inference is robust to the presence of heteroskedasticity that is only cross-sectional in nature (i.e., where $\sigma_2(\varrho_t)$ is constant). In the presence of a non-constant $\sigma_2(\varrho_t)$, we propose a simple way to robustify our cluster based inference. Whereas more general forms of heteroskedasticity, where $v_{pt} = \sigma(x_{pt})u_{pt}$, can be permitted, its implementation would require the use of nonparametric methods which would require the selection of a bandwidth parameter to estimate the heteroskedasticity function. We shall indicate how we should proceed if this were the case. Finally, a benefit of our estimator is that it permits the temporal dependence to vary across individuals, which is more realistic. It is important to point out that MBB would not be valid in these settings as it depends on some type of temporal homogeneity or even stationarity.

The remainder of the paper is organized as follows: In Section 2 we discuss the regularity conditions for our model and describe the main results. In Section 3 we introduce our bandwidth parameter free bootstrap schemes and we demonstrate their validity. Section 4 discusses a generalization of our model that permits (conditional) heteroskedasticity. Section 5 presents a Monte Carlo simulation experiment to shed some light on the finite sample performance of our cluster estimator and its comparison to others and we illustrate the finite sample benefits of our bootstrap schemes. In Section 6 we summarize. The proofs of our main results are given in Appendix A, which employs a series of lemmas given in Appendix B.

2. THE REGULARITY CONDITIONS AND MAIN RESULTS

We shall begin by considering the panel data model

$$y_{pt} = \beta' x_{pt} + \eta_p + \alpha_t + u_{pt}, \quad p = 1, ..., n, \qquad t = 1, ..., T, \tag{2.1}$$

where β is a $k \times 1$ vector of unknown parameters, x_{pt} is a $k \times 1$ vector of covariates, α_t and η_p represent respectively the time and individual fixed effects and $\{u_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$, are sequences of zero mean errors with heterogeneous variance $E\left(u_{pt}^2\right) = \sigma_p^2$, $p \in N^+$. We allow for general (unknown) temporal and cross-sectional dependence structures of the sequence $\{u_{pt}\}_{t\in\mathbb{Z}}$, $p \in N^+$, detailed in Condition $C1$ and $\{x_{pt}\}_{t\in\mathbb{Z}}$, $p \in N^+$, detailed in Condition $C2$. Further details are provided in our discussion of these conditions below. For simplicity, we shall assume that the sequences $\{x_{pt}\}_{t\in\mathbb{Z}}$, $p \in N^+$, are mutually independent of the error term $\{u_{pt}\}_{t\in\mathbb{Z}}$, $p \in N^+$, whilst allowing for dependence of the covariates with the fixed effects η_p and/or α_t.[2] In Section 4, we shall relax this condition allowing for heteroskedasticity. A straightforward extension that

[2] In fact, all that is needed is that the first conditional moment of the error is zero and the second conditional moment is equal to the unconditional one.

allows for lagged endogenous variables $\{y_{p,t-\ell}\}_{\ell=1}^{k_1}$, as in Hidalgo and Schafgans (2017), requires the use of the instrumental variable estimator, where $\{x_{p,t-\ell}\}_{\ell=1}^{k_1}$ provide natural instruments for $\{y_{p,t-\ell}\}_{\ell=1}^{k_1}$. We have avoided this generalization as it would detract from the main contribution of the paper and it will only add some extra technicalities and/or considerations which are well known and understood when $n = 1$.

Our first aim in the paper is to perform inference on the slope parameters β in the presence of a very general and unknown spatio-temporal dependence structure. To that end, we first need to extend a Central Limit Theorem provided in Phillips and Moon (1999), see also Hahn and Kuersteiner (2002). The reason for this is that in their work the sequences of random variables, say $\{\psi_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$, are assumed to be independent, that is $\{\psi_{pt}\}_{t\in\mathbb{Z}}$ and $\{\psi_{qt}\}_{t\in\mathbb{Z}}$ are mutually independent for any $p \neq q$, which is ruled out in our context as we permit cross-sectional dependence. Moreover, as we shall allow for "*strong-dependence*" in our error and regressor sequences, we cannot use results and arguments based on any type of "*strong-mixing*" conditions, so that results in Jenish and Prucha (2009, 2012) cannot be invoked in our framework either. Our theorem also extends the results provided in Hidalgo and Schafgans (2017) by allowing the errors u_{pt} to exhibit temporal dependence. A second aim of the paper is to extend the work of Driscoll and Kraay (1998) by examining, in the presence of individual and temporal fixed effects, a cluster estimator of the asymptotic covariance of the estimator of the slope parameters that does not require the ordering of the observations (in the cross-sectional dimension) or the selection of a bandwidth parameter.

The fixed effect model and the estimator for the slope parameters we consider are well known. Denoting for any generic sequence $\{\varsigma_{pt}\}_{t=1}^{T}$, $p = 1, ..., n$, the transformation

$$\begin{aligned} \widetilde{\varsigma}_{pt} &= \varsigma_{pt} - \overline{\varsigma}_{\cdot t} - \overline{\varsigma}_{p\cdot} + \overline{\overline{\varsigma}}_{\cdot\cdot}; \quad (2.2)\\ \overline{\varsigma}_{\cdot t} &= \frac{1}{n}\sum_{p=1}^{n}\varsigma_{pt};\ \overline{\varsigma}_{p\cdot} = \frac{1}{T}\sum_{t=1}^{T}\varsigma_{pt};\ \text{and}\ \overline{\overline{\varsigma}}_{\cdot\cdot} = \frac{1}{nT}\sum_{t=1}^{T}\sum_{p=1}^{n}\varsigma_{pt}, \end{aligned}$$

the estimator of β is obtained by performing least squares on the transformed model (where the individual and time effects are removed)

$$\widetilde{y}_{pt} = \beta'\widetilde{x}_{pt} + \widetilde{u}_{pt}, \ \ p = 1, ..., n \ \text{ and } t = 1, ..., T, \quad (2.3)$$

so that $\widehat{\beta}$ is defined as

$$\widehat{\beta} = \left(\sum_{p=1}^{n}\sum_{t=1}^{T}\widetilde{x}_{pt}\widetilde{x}_{pt}'\right)^{-1}\left(\sum_{p=1}^{n}\sum_{t=1}^{T}\widetilde{x}_{pt}\widetilde{y}_{pt}\right). \quad (2.4)$$

It is obvious that we can take $Ex_{pt} = 0$ as $\widetilde{x}_{pt}$ is invariant to additive constants, say μ_t or ν_p, to x_{pt}.

In this paper, we shall focus on an equivalent frequency domain formulation of (2.1) and (2.3). It is the application of the Discrete Fourier Transform (DFT) to our model, as will become clear shortly, that plays an important role in describing and motivating the cluster estimator of the asymptotic covariance matrix of $\widehat{\beta}$, or equivalently $\tilde{\beta}$ given in (2.7) below, and the bootstrap schemes described in Section 3.

For this purpose, we denote the DFT for generic sequences $\{\varsigma_{pt}\}_{t=1}^{T}$, $p \geq 1$, by

$$\mathcal{J}_{\varsigma,p}(\lambda_j) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} \varsigma_{pt} e^{-it\lambda_j}, \quad j = 1, ..., \widetilde{T} = [T/2], \;\; \lambda_j = \frac{2\pi j}{T} \tag{2.5}$$

and $\mathcal{J}_{\varsigma,p}(\lambda_j) = \mathcal{J}_{\varsigma,p}(-\lambda_{T-j})$, $j = \widetilde{T}+1, .., T$. We can then rewrite (2.3) as

$$\mathcal{J}_{\widetilde{y},p}(\lambda_j) = \beta' \mathcal{J}_{\widetilde{x},p}(\lambda_j) + \mathcal{J}_{\widetilde{u},p}(\lambda_j), \;\; p = 1, ..., n; \;\; j = 1, ..., T-1. \tag{2.6}$$

Given that our sequences $\{\widetilde{\varsigma}_{pt}\}_{t=1}^{T}$, $p \geq 1$, are centred around their sample means, we can leave out the frequency λ_j for $j = T$ (and 0) as $J_{\widetilde{\varsigma},p}(0) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} \widetilde{\varsigma}_{pt} = 0$. The interesting property of $\mathcal{J}_{\widetilde{u},p}(\lambda_j)$, $j = 1, ..., T-1$, that allows us to formulate our new cluster estimator that accounts for both types of dependence, is that it is serially uncorrelated over the Fourier frequencies for large T, whilst possibly heteroskedastic. Based on the frequency domain formulation of our model (2.6), we can also compute our estimator of β as

$$\widetilde{\beta} = \left(\sum_{p=1}^{n} \sum_{j=1}^{T-1} \mathcal{J}_{\widetilde{x},p}(\lambda_j) \mathcal{J}'_{\widetilde{x},p}(-\lambda_j) \right)^{-1} \left(\sum_{p=1}^{n} \sum_{j=1}^{T-1} \mathcal{J}_{\widetilde{x},p}(\lambda_j) \mathcal{J}_{\widetilde{y},p}(-\lambda_j) \right). \tag{2.7}$$

We introduce our regularity conditions next. To that end, and in what follows, we denote for any generic sequence $\{v_{pt}\}_{t \in \mathbb{Z}}$, $p \in N$,

$$\varphi_v(p,q) = Cov(v_{pt}; v_{qt}), \text{ for any } p, q \geq 1.$$

Condition C1: $\{u_{pt}\}_{t \in \mathbb{Z}}$, $p \in \mathbb{N}^+$, *are zero mean sequences of random variables such that*

(i) $$u_{pt} = \sum_{k=0}^{\infty} d_k(p) \xi_{p,t-k}, \qquad \sum_{k=0}^{\infty} k d_k < \infty, \quad d_k =: \sup_p |d_k(p)|,$$

where $E(\xi_{pt} \mid \mathcal{V}_{p,t-1}) = 0$*;* $E(\xi_{pt}^2 \mid \mathcal{V}_{p,t-1}) = \sigma_{\xi,p}^2$ *and finite fourth moments, with* $\mathcal{V}_{p,t}$ *denoting the* $\sigma-$*algebra generated by* $\{\xi_{ps}, \, s \leq t\}$.

(ii) *For all* $t \in \mathbb{Z}$ *and* $p \in \mathbb{N}^+$,

$$\xi_{pt} = \sum_{\ell=1}^{\infty} a_\ell(p) \varepsilon_{\ell t}, \qquad \sup_{p \in \mathbb{N}^+} \sum_{\ell=1}^{\infty} |a_\ell(p)|^2 < \infty, \;\; \sup_{\ell \geq 1} \sum_{p=1}^{n} |a_\ell(p)|^2 < \infty,$$

where the sequences $\{\varepsilon_{\ell t}\}_{t \in \mathbb{Z}}$, $\ell \in \mathbb{N}^+$, *are zero mean independent identically distributed (iid) random variables.*

(iii) *The fourth cumulant of* $\{u_{pt}\}_{t \in \mathbb{Z}}$, $p \in \mathbb{N}^+$, *satisfies*

$$\lim_{T \nearrow \infty} \sup_{p \in \mathbb{N}^+} \sum_{t_1, t_2, t_3 = 1}^{T} |Cum(u_{pt_1}; u_{pt_2}; u_{pt_3}; u_{p0})| < \infty.$$

Condition C2: $\{x_{pt}\}_{t \in \mathbb{Z}}$, $p \in \mathbb{N}^+$, *are sequences of random variables such that:*

(i) $$x_{pt} = \sum_{k=0}^{\infty} c_k(p) \chi_{p,t-k}, \qquad \sum_{k=0}^{\infty} k c_k < \infty, \quad c_k =: \sup_p \|c_k(p)\|,$$

where $\|B\|$ *denotes the norm of the matrix* B *and* $E(\chi_{pt} \mid \Upsilon_{p,t-1}) = 0$*;* $Cov(\chi_{pt} \mid \Upsilon_{p,t-1}) = \Sigma_{\chi,p}$ *and* $E\|\chi_{pt}\|^4 < \infty$, *with* $\Upsilon_{p,t}$ *denoting the* $\sigma-$*algebra generated by* $\{\chi_{ps}, \, s \leq t\}$.

(ii) *The sequences of random variables* $\{\chi_{pt}\}_{t \in \mathbb{Z}}$, $p \in \mathbb{N}^+$, *are such that*

$$\chi_{pt} = \sum_{\ell=1}^{\infty} b_\ell(p) \eta_{\ell t}, \qquad \sup_{p \in \mathbb{N}^+} \sum_{\ell=1}^{\infty} |b_\ell(p)|^2 < \infty, \quad \sup_{\ell \geq 1} \sum_{p=1}^{n} |b_\ell(p)|^2 < \infty,$$

where the sequences $\{\eta_{\ell t}\}_{t\in\mathbb{Z}}$, $\ell \in \mathbb{N}^+$, *are zero mean iid random variables.*

(iii) *Denoting* $\Sigma_{x,p} = E\left(x_{pt}x_{pt}'\right)$, *we have*

$$0 < \Sigma_x = \lim_{n\to\infty} \frac{1}{n} \sum_{p=1}^{n} \Sigma_{x,p} \tag{2.8}$$

and the fourth cumulant of $\{x_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$, *satisfies*

$$\lim_{T\to\infty} \sup_{p\in\mathbb{N}^+} \sum_{t_1,t_2,t_3=1}^{T} \left| Cum\left(x_{pt_1,a}; x_{pt_2,b}; x_{pt_3,c}; x_{p0,d}\right)\right| < \infty,\ a,b,c,d = 1,...,k,$$

where $x_{pt,a}$ *denotes the* $a-th$ *element of* x_{pt}.

Condition C3: *For all* $p \in \mathbb{N}^+$, *the sequences* $\{u_{pt}\}_{t\in\mathbb{Z}}$ *and* $\{x_{pt}\}_{t\in\mathbb{Z}}$ *are mutually independent and*

$$0 < \max_{1\le p\le n} \sum_{q=1}^{n} \|\varphi(p,q)\| < \infty, \tag{2.9}$$

where $\varphi(p,q) := \varphi_u(p,q)\,\varphi_x(p,q)$.

We now comment on our conditions. Conditions $C1$ and $C2$ indicate that $\{u_{pt}\}_{t\in\mathbb{Z}}$ and $\{x_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$, are linear processes that permit the usual SAR (or more generally $SARMA$) model. Indeed, by definition of the SAR model, with W a spatial weight matrix we have

$$\begin{aligned} u &= (I-\omega W)^{-1}\varepsilon \\ &= (I+\Xi)\,\varepsilon, \quad \Xi = \left(\psi_q(p)\right)_{p,q=1}^{n}, \end{aligned}$$

so that $u_p = \sum_{q=0}^{n} \psi_q(p)\,\varepsilon_q$, which implies that the SAR model satisfies Condition $C1$. Unlike the SAR model, Condition $C1$ does permit the sequence $\sum_{p=1}^{n} |a_\ell(p)|$ to grow with n. One can allow the weights $a_\ell(p)$ to depend on the sample size "n" as is often done in SAR models with weight matrices W row-normalized, but it does not add anything significant. Our conditions, therefore, appear to be weaker than those typically assumed when cross-sectional dependence is allowed while being similar to those of Lee and Robinson (2013). As the sequences may exhibit long memory spatial dependence, the condition of strong mixing for the spatial dependence in Jenish and Prucha (2012) is ruled out. This appears to be the case as Ibragimov and Rozanov (1978) showed; if the sequence $\left\{\gamma_{u,pq}(j)\right\}_{j\in\mathbb{Z}}$ is not summable, the process $\{u_{pt}\}_{t\in\mathbb{Z}}, p \in \mathbb{N}^+$, cannot be *strong-mixing*. The long memory dependence also rules out that the process is Near Epoch Dependent with size greater than 1/2, which appears to be a necessary condition for standard asymptotic CLT results.

Conditions $C1$ and $C2$ do rule out long memory temporal dependence on the sequences $\{x_{pt}\}_{t\in\mathbb{Z}}$ and $\{u_{pt}\}_{t\in\mathbb{Z}}$ for each p. Even though there are several results available allowing their temporal dependence to exhibit long memory, see Robinson and Hidalgo (1997) or Hidalgo (2003), we have decided to assume the temporal dependence of the regressors and errors to be weakly dependent to simplify the arguments. It is worth pointing out that our Conditions $C1\,(\mathbf{i})$ and $C2\,(\mathbf{i})$ can be relaxed to some extent to allow some type of mixing condition such as L^4-Near Epoch dependence with size greater than or equal to 2. The latter condition is often invoked when we allow the errors to have a nonlinear type of dependence structure or if (2.1) were replaced by a nonlinear panel data model

$$y_{pt} = g(x_{pt};\beta) + \eta_p + \alpha_t + u_{pt}, \quad p = 1,...,n, \quad t = 1,...,T.$$

In fact, we expect the conclusions of our results to hold under such a mixing condition as it has been shown in numerous situations. Conditions $C1$ and $C2$ do permit heterogeneity in its second moments as $E\left(\xi_{pt}^2 \mid \mathcal{V}_{p,t-1}\right) = \sigma_{\xi,p}^2$ and $Cov\left(\chi_{pt} \mid \Upsilon_{p,t-1}\right) = \Sigma_{\chi,p}$. This follows from our conditions because $E\left(\xi_{pt}^2 \mid \mathcal{V}_{p,t-1}\right) = \sum_{\ell=1}^{\infty} \left|a_\ell(p)\right|^2$ clearly depends on p. Furthermore, we allow for some trending behaviour of the sequences $\{x_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$, as we allow the mean of x_{pt} to depend on time.

An important consequence of Conditions $C1$ and $C2$ is that they guarantee that the covariance structure of the sequences $\{u_{pt}\}_{t\in\mathbb{Z}}$ and $\{x_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$, is multiplicative. For instance, Condition $C1$ implies that, for all $p, q \in \mathbb{N}^+$,

$$\begin{aligned} E\left(u_{pt}u_{qs}\right) &= E\left(\sum_{k=0}^{\infty} d_k(p)\,\xi_{p,t-k} \sum_{\ell=0}^{\infty} d_\ell(q)\,\xi_{q,s-\ell}\right) \\ &= E\left(\xi_{p1}\xi_{q1}\right)\begin{cases} \sum_{\ell=0}^{\infty} d_{t-s+\ell}(p)\,d_\ell(q) & t>s \\ \sum_{\ell=0}^{\infty} d_\ell(p)\,d_{s-t+\ell}(q) & t\le s \end{cases} \\ &= \varphi_u(p,q)\,\gamma_{u;pq}(t-s)\,. \end{aligned} \tag{2.10}$$

Following the spatio-temporal literature, see Cressie and Huang (1999), we can denote this covariance structure as *separable*. Of course, there are *nonseparable* covariance structures, see Gneiting (2002) and tests for separability are available, see Fuentes (2006) or Matsuda and Yajima (2004). Notice that in the absence of cross-sectional dependence, $E\left(\xi_{p1}\xi_{q1}\right) = \sigma_{\xi p}^2 \mathbf{1}(p=q)$ and $E\left(u_{pt}u_{qs}\right) = \sigma_{\xi p}^2 \gamma_{u;pp}(t-s)\,\mathbf{1}(p=q)$. Here, and in what follows, $\mathbf{1}(A)$ denotes the indicator function.

Remark 1. *The condition* $\sup_{p\in\mathbb{N}^+}\sum_{\ell=0}^{\infty}\left|a_\ell(p)\right|^2 < \infty$ *guarantees that for any reordering of the sequence* $\left\{\left|a_\ell(p)\right|^2\right\}_{\ell\in\mathbb{N}^+}$*, say* $\left\{\left|a_{\ell(\tau)}(p)\right|^2\right\}_{\ell(\tau)\in\mathbb{N}^+}$*, we have that* $a_{\ell(\tau)}(p) = O\left(\ell(\tau)^{-\zeta}\right)$ *for some* $\zeta > 1/2$*. Similarly the requirement* $\sup_{\ell\ge1}\sum_{p=1}^{n}\left|a_\ell(p)\right|^2 < \infty$ *will mean that* $a_\ell(p) = O\left(p^{-\zeta}\right)$ *for some* $\zeta > 1/2$ *uniformly in* $\ell \ge 1$*. Similar arguments follow for* $\left\{\left|b_\ell(p)\right|^2\right\}_{\ell\in\mathbb{N}^+}$, $p \ge 1$.

Condition $C3$ assumes that the sequences $\{x_{pt}\}_{t\in\mathbb{Z}}$ and $\{u_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$ are independent, although we envisage that it can be relaxed to require only conditional independence in first and second moments. To simplify the arguments somewhat, we have preferred to keep the condition as it stands. Even though we allow long memory spatial dependence of the individual sequences, the absolute summability requirement in (2.9) limits the combined cross-sectional dependence, that is the dependence of the sequence $\{z_{pt} = u_{pt}x_{pt}\}_{t\in\mathbb{Z}}$, $p \in N^+$, is "weakly spatially dependent", see also Hidalgo and Schafgans (2017). We have adopted the convention that $\gamma_{u;pp}(t-s) = E\left(u_{pt}u_{ps}\right)/\varphi_u(p,p)$. Importantly, as we assume that the errors and regressors are uncorrelated, the spectral density matrix of the sequences $\{z_{pt} =: u_{pt}x_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$ is given by the convolution of the spectral density matrix of $\{x_{pt}\}_{t\in\mathbb{Z}}$ and spectral density function of $\{u_{pt}\}_{t\in\mathbb{Z}}$, that is

$$f_p(\lambda) =: \int_{-\pi}^{\pi} f_{u,p}(\upsilon)\, f_{x,p}(\lambda-\upsilon)\, d\upsilon, \;\; p \in \mathbb{N}^+, \tag{2.11}$$

where Conditions $C1$ and $C2$ imply that $f_p(\lambda)$ is twice continuous differentiable. By Fuller's (1996) Theorem 3.4.1, or Corollary 3.4.1.2, the Fourier coefficients of $f_p(\lambda)$ are given by $\gamma_p(j) =$

$\gamma_{x,p}(j)\gamma_{u,p}(j)$, $p\in\mathbb{N}^{+}$, so that

$$\sup_{p,q=1,..,n}\sum_{\ell=-\infty}^{\infty}\left\|\gamma_{pq}(\ell)\right\|<\infty;\quad Cov\left(z_{pt};z_{qs}\right)=\gamma_{pq}(t-s)\varphi(p,q).$$

With the convention that $\gamma_{u,pq}(0)=\gamma_{x,pq}(0)=1$, $Cov\left(z_{pt},z_{qt}\right)=\varphi(p,q)=:\varphi_{u}(p,q)\varphi_{x}(p,q)$ as defined in Condition $C3$.

Remark 2. *It is worth noticing that (2.9) ensures that* $\varphi(p,q)=O\left(q^{-1-\delta}\right)$ *or* $\varphi(p,q)=O\left(p^{-1-\delta}\right)$ *for some* $\delta>0$, *so that*

$$\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{p,q=1}^{n}\varphi(p,q)<\infty.$$

The latter displayed expression can be regarded as a type of weak dependence in the cross-sectional dimension, see also Robinson (2011) *or Lee and Robinson* (2013)*. In addition, the ergodicity in second mean, that is*

$$\frac{1}{n^{2}}\sum_{p,q=1}^{n}\left(\varphi_{u}(p,q)+\varphi_{x}(p,q)\right)=o(1),$$

implies that $\varphi_{u}(p,q)=O\left(q^{-\varsigma_{u}}\right)$ *and* $\varphi_{x}(p,q)=O\left(q^{-\varsigma_{x}}\right)$ *such that* $\varsigma_{u}+\varsigma_{x}=1+\delta>0$.

Conditions $C1-C3$, therefore, imply that the "*average*" long-run variance of the sequences $\left\{z_{pt}=:u_{pt}x_{pt}\right\}_{t\in\mathbb{Z}}$, $p\in\mathbb{N}^{+}$, is given by

$$\Phi=:2\pi\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{p,q=1}^{n}f_{pq}(0)\varphi(p,q)<\infty \tag{2.12}$$

$$2\pi f_{pq}(0)=\sum_{\ell=-\infty}^{\infty}\gamma_{pq}(\ell).$$

Observe that standard algebra yields that

$$\begin{aligned}\Phi &=: \lim_{n\rightarrow\infty}\lim_{T\rightarrow\infty}\frac{1}{nT}E\left\{\left(\sum_{p=1}^{n}\sum_{t=1}^{T}x_{pt}u_{pt}\right)\left(\sum_{p=1}^{n}\sum_{t=1}^{T}x_{pt}^{\prime}u_{pt}\right)\right\}\\ &= \lim_{n\rightarrow\infty}\lim_{T\rightarrow\infty}\frac{1}{nT}\sum_{p,q=1}^{n}\sum_{t,s=1}^{T}E\left(x_{pt}x_{qs}^{\prime}\right)E\left(u_{pt}u_{qs}\right), \qquad (2.13)\end{aligned}$$

or, using its spectral domain formulation,

$$\begin{aligned}\Phi &= \lim_{n\rightarrow\infty}\lim_{T\rightarrow\infty}\frac{1}{nT}E\left\{\left(\sum_{j=1}^{T-1}\sum_{p=1}^{n}\mathcal{J}_{x,p}(\lambda_{j})\mathcal{J}_{u,p}(-\lambda_{j})\right)\left(\sum_{j=1}^{T-1}\sum_{p=1}^{n}\mathcal{J}_{x,p}^{\prime}(-\lambda_{j})\mathcal{J}_{u,p}(\lambda_{j})\right)\right\}\\ &= \lim_{n\rightarrow\infty}\lim_{T\rightarrow\infty}\frac{1}{nT}\sum_{j=1}^{T-1}\sum_{p,q=1}^{n}E\left(\mathcal{J}_{x,p}(\lambda_{j})\mathcal{J}_{x,q}^{\prime}(-\lambda_{j})\right)E\left(\mathcal{J}_{u,p}(-\lambda_{j})\mathcal{J}_{u,q}(\lambda_{j})\right). \qquad (2.14)\end{aligned}$$

Finally, we denote

$$V=\Sigma_{x}^{-1}\Phi\Sigma_{x}^{-1}, \tag{2.15}$$

where $\Sigma_{x}>0$ was defined in Condition $C2$.

The following theorem presents our result establishing the CLT for our slope parameter estimates in the presence of general temporal and cross-sectional dependence.

Theorem 1. *Under Conditions $C1-C3$, we have that as $n,T\to\infty$,*

$$(nT)^{1/2}\left(\widetilde{\beta}-\beta\right)\overset{d}{\to}\mathcal{N}\left(0,V\right).$$

Proof. The proof of this result, based on either the time or frequency domain formulation, will be given in Appendix A. All other proofs are relegated to this appendix as well. □

Remark 3. *While the result could be shown to hold with finite n, a setting considered by Robinson (1998), the presence of the time fixed effect would require special attention since the dependence structure of u_{pt} and $u_{pt}-n^{-1}\sum_{p=1}^{n}u_{pt}$ are not quite the same when n is finite.*

With V defined in (2.15), Theorem 1 indicates that to make inferences on β, we need to provide a consistent estimator of Φ. A first glance at (2.13) or (2.14) suggests that this might be complicated or computationally burdensome due to the general spatio-temporal dependence structure of the data. As we pointed out in the introduction, the standard approach to deal with such dependence, that is a HAC type of estimator, has various potential drawbacks in the presence of cross-sectional dependence. While choosing a bandwidth parameter associated with the cross-sectional dependence requires or induces an artificial and/or nontrivial ordering, the presence of individual heterogeneous temporal dependence (as assumed in Conditions $C1$ and $C2$) would even render any cross validation method used to choose the temporal bandwidth parameter intractable.

While Kim and Sun's (2013) approach is subject to both these criticisms, Driscoll and Kraay (1998) avoid the need to specify an ordering of individuals by introducing an HAC estimator of cross-sectional averages, so that one can consider their estimator as a hybrid between an HAC and a cluster one: they employ the HAC methodology to deal with the temporal dependence whereas they employ a cluster type of estimator to account for the cross-sectional dependence. We advocate to use an approach that does not require any ordering and/or selection of a bandwidth parameter and also permits a more general spatio-temporal dependence than allowed by either Driscoll and Kraay (1998) or Kim and Sun (2013) and permits the cross-sectional dependence to be "long-memory" which latter work ruled out. Moreover our approach permits the temporal dependence to be heterogeneous across individuals, which is more realistic.

Our approach can be regarded as a natural extension of the earlier work by Robinson (1998) on inference without smoothing in a time series regression model context. In his case, abstracting from cross-sectional dependence

$$\Phi=:\lim_{n\to\infty}\frac{2\pi}{n}\sum_{p=1}^{n}f_{pp}\left(0\right).$$

Applying his estimator to our model, would yield the estimator

$$\frac{2\pi}{n}\sum_{p=1}^{n}\frac{1}{T}\sum_{j=1}^{T}\mathcal{I}_{u,p}\left(\lambda_{j}\right)\mathcal{I}_{x,p}\left(-\lambda_{j}\right)=\frac{1}{n}\sum_{p=1}^{n}\sum_{\ell=-T+1}^{T-1}\widehat{\gamma}_{x,p}\left(\ell\right)\widehat{\gamma}_{u,p}\left(\ell\right),\tag{2.16}$$

where $\widehat{\gamma}_{x,p}\left(j\right)$ and $\widehat{\gamma}_{u,p}\left(j\right)$ are respectively the standard sample moment estimators of $\gamma_{x,p}\left(j\right)$ and $\gamma_{u,p}\left(j\right)$ and $I_{u,p}\left(\lambda\right)=T^{-1}\left(\sum_{t=1}^{T}u_{pt}e^{it\lambda}\right)\left(\sum_{t=1}^{T}u_{pt}e^{-it\lambda}\right)'$ with $I_{x,p}\left(\lambda\right)$ similarly defined. When cross-sectional dependence is allowed, the latter arguments suggest that (2.16) is not a consistent (cluster) estimator of Φ. The reason for this (see also the proof of Proposition 1) is that

$$\frac{1}{n}\sum_{p=1}^{n}\sum_{\ell=-T+1}^{T-1}\gamma_{x,p}\left(\ell\right)\gamma_{u,p}\left(\ell\right)\nrightarrow\Phi$$

as expected since the first moment of (2.16) does not capture the cross-sectional dependence. The purpose of the next section is therefore to provide a consistent "cluster" estimator of Φ that accounts for the presence of cross-sectional dependence.

2.1. **Cluster estimator of Φ.**

We shall present a simple cluster estimator of Φ using the "frequency" domain methodology. Obviously, there is a time domain analogue, which we shall briefly describe at the end of the section. Our cluster estimator appears to be the first one which permits both time and cross-sectional dependence and gives a formal justification of its statistical properties. Our estimator therefore becomes an extension of previous cluster estimators in the literature such as that in Arellano (1987) (where only temporal dependence is present) or Bester, Conley and Hansen (2011) (where only cross-sectional dependence is present).

Our main motivation to propose a cluster estimator using the frequency domain methodology comes from the well known observation that for all $j\neq k$, $J_{u,p}\left(\lambda_j\right)$ and $J_{u,q}\left(\lambda_k\right)$ can be considered as being uncorrelated although possibly heteroskedastic. This observation was employed in the landmark paper by Hannan (1963) on adaptive estimation in a time series regression model. The fact that we may therefore consider $\mathcal{J}_{\widetilde{x},p}\left(\lambda_j\right)\mathcal{J}_{u,p}\left(-\lambda_j\right)$ as a sequence of uncorrelated and heteroskedastic random variables in j, although not in p, suggests that, in a spirit similar to White's (1980) estimator, we may estimate Φ by

$$\breve{\Phi}=\frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}\left(\lambda_j\right)\mathcal{J}_{\widehat{u},p}\left(-\lambda_j\right)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}'_{\widetilde{x},p}\left(-\lambda_j\right)\mathcal{J}_{\widehat{u},p}\left(\lambda_j\right)\right)\right\}. \tag{2.17}$$

Based on the DFT formulation, we denote the estimator of Σ_x by

$$\widetilde{\Sigma}_x=\frac{1}{nT}\sum_{j=1}^{T-1}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}\left(\lambda_j\right)\mathcal{J}'_{\widetilde{x},p}\left(-\lambda_j\right).$$

The following proposition establishes the consistency of our cluster estimator for the "average" long-run variance of the sequences $\left\{z_{pt}=:u_{pt}x_{pt}\right\}_{t\in\mathbb{Z}}$, $p\in N^{+}$.

Proposition 1. *Under the conditions of Theorem 1, we have that*

$$\begin{aligned}&(\mathbf{a}) \qquad \breve{\Phi}-\Phi=o_p\left(1\right)\\&(\mathbf{b}) \qquad \widetilde{\Sigma}_x-\Sigma_x=o_p\left(1\right).\end{aligned}$$

Denoting $\hat{V}=:\widetilde{\Sigma}_x^{-1}\breve{\Phi}\widetilde{\Sigma}_x^{-1}$, we now obtain the following corollary.

Corollary 1. *Under the conditions of Theorem 1, we have that*

$$\left(nT\right)^{1/2}\hat{V}^{-1/2}\left(\widetilde{\beta}-\beta\right)\overset{d}{\to}\mathcal{N}\left(0,I\right).$$

Proof. The proof is standard from Theorem 1 and Proposition 1, and is therefore omitted. □

We now describe the time domain analogue estimator of Φ. For that purpose, using $\sum_{t=1}^{T} e^{it\lambda_\ell} = 0$ if $1 \le \ell \le T-1$, we have after standard algebra that

$$\breve{\Phi} = \frac{1}{n} \sum_{p,q=1}^{n} \sum_{|\ell|=0}^{T-1} \widehat{\gamma}_{x,pq}(\ell) \widehat{\gamma}_{u,pq}(\ell),$$

where due to (2.10)

$$\begin{aligned} \widehat{\gamma}_{x,pq}(\ell) &= \frac{1}{T} \sum_{t=1}^{T-|\ell|} \widetilde{x}_{pt} \widetilde{x}'_{q,t+\ell}; \\ \widehat{\gamma}_{u,pq}(\ell) &= \frac{1}{T} \sum_{t=1}^{T-|\ell|} \widehat{u}_{pt} \widehat{u}_{q,t+\ell} \mathbf{1}(\ell > 0) + \frac{1}{T} \sum_{t=1}^{T-|\ell|} \widehat{u}_{qt} \widehat{u}_{p,t+\ell} \mathbf{1}(\ell < 0), \end{aligned}$$

and $\widehat{u}_{pt} = \widetilde{y}_{pt} - \widetilde{\beta}' \widetilde{x}_{pt}$, $p = 1, ..., n$; $t = 1, ..., T$.

3. BOOTSTRAP SCHEMES

Our motivation to introduce bootstrap schemes emanates from findings in our Monte Carlo experiment, which suggest that the asymptotic distribution of $(nT)^{1/2} \hat{V}^{-1/2} \left(\widetilde{\beta} - \beta\right)$ does not appear to provide a good approximation of its finite sample distribution. In such situations, the use of the bootstrap has been advocated as it has been shown to improve the finite sample performance. The general spatio-temporal dependence inherent in our model suggests that a valid bootstrap mechanism may not to be easy to implement since one of the basic requirements for its validity is that it has to preserve the covariance structure of the data/model. Drawing analogies from the time series literature, one might be tempted to use the block bootstrap (BB) principle as it is not clear how the sieve bootstrap can be implemented under cross-sectional dependence in the absence of a clear ordering of the data. Applying a BB in both dimensions, however, would also be sensitive to the particular ordering chosen by the practitioner and be subject to the absence of weak stationarity, where the dependence structure of say $(x_{p_1,t}, ..., x_{p_1+m,t})'$ and $(x_{p_1+1,t}, ..., x_{p_1+m+1,t})'$ need not be identical.

Avoiding the need to establish a particular ordering of the cross sectional units, Gonçalves (2011) proposes to apply a moving block bootstrap (MBB) to the vector containing all individual observations for each t, that is it only applies a BB in the time dimension. The MBB, however, does require the choice of the block size and is known to be sensitive to its choice in finite samples. In the absence of temporal dependence, the block size equals one, and the approach is similar to Hidalgo and Schafgans (2017).

Here we propose valid bootstrap schemes with the interesting feature that they are computationally simple (there is no need to estimate, either by parametric or nonparametric methods, the time and/or cross-sectional dependence of the error term) and do not require the choice of any bandwidth parameter for its implementation, thereby avoiding any level of arbitrariness.

Both bootstrap schemes considered are in the frequency domain. We recall the DFT for generic sequences $\{\varsigma_{pt}\}_{t=1}^{T}$, $p \ge 1$, as $\mathcal{J}_{\varsigma,p}(\lambda_j) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} \varsigma_{pt} e^{-it\lambda_j}$, $j = 1, ..., \widetilde{T} = [T/2]$, $\lambda_j = \frac{2\pi j}{T}$, and define its periodogram

$$\mathcal{I}_{\varsigma,p}(\lambda_j) = |\mathcal{J}_{\varsigma,p}(\lambda_j)|^2 \quad j = 1, ..., \widetilde{T} = [T/2], \quad p = 1, ..., n.$$

The first scheme, labelled the naïve bootstrap, imposes Condition $C4$ under which the time dependence is assumed to be homogeneous across individuals. We relax this assumption in the second scheme, labelled the wild bootstrap, in line with Conditions $C1$ and $C2$.

For homogeneous temporal dependence, therefore, we impose

Condition C4: Homogeneous time dependence: $d_k(p)$ and $c_k(p)$ defined in Conditions $C1$ and $C2$ do not vary over p.

Denoting $\sigma_u^2(p) = Eu_{pt}^2$ and $f_{u,p}(\lambda)$ the spectral density function of the sequence $\{u_{pt}\}_{t\in\mathbb{Z}}$, for any $p = 1, 2, ...,$ Condition $C4$ ensures that

$$\frac{f_{u,p}(\lambda)}{\sigma_u^2(p)} =: g_u(\lambda), \quad p = 1, 2, 3, ... \tag{3.1}$$

That is, the spectral density function normalized by the variance does not depend on p. This enables us to use the average periodogram of standardized residuals in constructing the valid bootstrap model. What really matters here is that the "correlation" structure is the same.

The naïve bootstrap, involves resampling from the residuals of the model[3] and involves the following simple *3 STEPS:*

***STEP 1*:** Obtain the residuals

$$\widehat{u}_{pt} = \widetilde{y}_{pt} - \widetilde{\beta}'\widetilde{x}_{pt}, \;\; p = 1, ..., n; \; t = 1, ..., T,$$

compute $\widetilde{\sigma}_{\widehat{u}}^2(p) = T^{-1}\sum_{t=1}^{T}\widehat{u}_{pt}^2$, and obtain the standardized residuals

$$\check{u}_{pt} = \widehat{u}_{pt}/\widetilde{\sigma}_{\widehat{u}}(p).$$

***STEP 2*:** Denoting $\hat{U}_t = \{\hat{u}_{pt}\}_{p=1}^{n}$, do standard random sampling from the empirical distribution of the residuals $\{\hat{U}_t\}_{t=1}^{T}$. That is, we assign probability T^{-1} to each $n\times 1$ vector $\hat{U}_t$. Denote the bootstrap sample by $\{U_t^*\}_{t=1}^{T}$, where $U_t^* = \left\{u_{pt}^*\right\}_{p=1}^{n}$. Compute the bootstrap analogue of (2.3) as

$$\mathcal{J}_{y*,p}(\lambda_j) = \widetilde{\beta}'\mathcal{J}_{\widetilde{x},p}(\lambda_j) + \left(\frac{1}{n}\sum_{q=1}^{n}\mathcal{I}_{\check{u},q}(\lambda_j)\right)^{1/2}\mathcal{J}_{u^*,p}(\lambda_j)$$

for $p = 1, ..., n$ and $j = 1, ..., T-1$.

***STEP 3*:** Compute the corresponding bootstrap analogue of (2.7) as

$$\widetilde{\beta}^* = \left(\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{J}_{\widetilde{x},p}(\lambda_j)\mathcal{J}_{\widetilde{x},p}'(-\lambda_j)\right)^{-1}\left(\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{J}_{\widetilde{x},p}(\lambda_j)\mathcal{J}_{\widetilde{y}*,p}(-\lambda_j)\right), \tag{3.2}$$

with $\mathcal{J}_{\widetilde{y}*,p}(\lambda_j) = \mathcal{J}_{y*,p}(\lambda_j) - \frac{1}{n}\sum_{q=1}^{n}\mathcal{J}_{y*,q}(\lambda_j)$.

Remark 4. *Since $\overline{\widehat{u}}_p = 0$ there is no need the recentre in Step 1. The standardization of the residuals (the variance is not the same for all individuals) is used in Step 2 to impose the appropriate dependence structure on our bootstrap regression. As the bootstrap is done on the vector containing all individual observations for each t there is no need for standardization otherwise.*

[3]See also Remark 6.

Remark 5. *We use the average periodogram of the standardized residuals to impose the appropriate dependence structure on our bootstrap regression in Step 2. When the time dependence is homogeneous among the cross-sectional units* $\frac{1}{n}\sum_{q=1}^{n}\mathcal{I}_{\check{u},q}(\lambda_j)=\sigma_u^{-2}(p)f_{u,p}(\lambda_j)(1+o_p(1))=:g_u(\lambda_j)(1+o_p(1))$, *see also (3.1). That is, if the temporal dependence were given by an AR(1) model, the right side becomes the spectral density function of an AR(1) sequence, where the innovation sequence has variance equal to 1. In addition, as we bootstrap from* $\hat{u}_{pt}$*, which are the residuals, we ensure that the variance of* u^*_{pt} *is that of* u_{pt}*.*

Remark 6. *Alternatively, we could have used random sampling from the normalized DFT of the residuals as considered by Hidalgo (2003). In that case, denoting* $T_{\widehat{u}}(\lambda_j)=\left\{\mathcal{J}_{\widehat{u},p}(\lambda_j)/|\mathcal{J}_{\widehat{u},p}(\lambda_j)|\right\}_{p=1}^{n}$, $T_{u^*,p}(\lambda_j)$ *form independent draws from the empirical distribution of* $\tilde{T}_{\widehat{u}}(\lambda_j)=(T_{\widehat{u}}(\lambda_j)-\bar{T}_{\widehat{u}})/\hat{\sigma}_{\mathcal{T}}$ *where* $\bar{T}_{\widehat{u}}=[T/2]^{-1}\sum_{j=1}^{T/2}T_{\widehat{u}}(\lambda_j)$ *and* $\hat{\sigma}^2_{\mathcal{T}}=[T/2]^{-1}\sum_{j=1}^{T/2}\left(\mathcal{T}_{\widehat{u}}(\lambda_j)-\bar{\mathcal{T}}\right)^2$. *The bootstrap analogue of (2.3) would then be obtained using* $\mathcal{J}_{y*,p}(\lambda_j)=\widetilde{\beta}'\mathcal{J}_{\widetilde{x},p}(\lambda_j)+\left(\frac{1}{n}\sum_{q=1}^{n}\mathcal{I}_{\check{u},q}(\lambda_j)\right)^{1/2}\widetilde{\sigma}_{\widehat{u}}(p)\,\mathcal{T}_{u^*,p}(\lambda_j)$. *Our scheme uses Step 2, which has better finite sample properties as observed in Hidalgo (2003).*

The key feature of this naïve bootstrap, is that there is no need to choose any bandwidth parameter for its implementation. Under Condition $C4$, uniformly in $j=1,...,T-1$, we have that

$$\begin{aligned}\mathcal{I}_{\check{u},p}(\lambda_j) &= \widetilde{\sigma}_{\widehat{u}}^{-2}(p)\left\{\mathcal{I}_{u,p}(\lambda_j)+\left(\widetilde{\beta}-\beta\right)^2\mathcal{I}_{x,p}(\lambda_j)+\left(\widetilde{\beta}-\beta\right)\mathcal{J}_{x,p}(\lambda_j)\,\mathcal{J}_{u,p}(-\lambda_j)\right\}\\ &= \sigma_u^{-2}(p)\,\mathcal{I}_{u,p}(\lambda_j)(1+o_p(1)),\end{aligned}$$

and

$$\begin{aligned}&E\mathcal{I}_{u,p}(\lambda_j)=f_{u,p}(\lambda_j)(1+o(1))\\ &E^*\left(\mathcal{J}_{u^*,p}(\lambda_j)\,\mathcal{J}_{u^*,p}(-\lambda_\ell)\right)=0,\quad \text{if } j\neq\ell,\ \sigma_u^2(p)\text{ otherwise}\\ &\widetilde{\sigma}^2_{\widehat{u}}(p)=\sigma^2_u(p)(1+o_p(1)).\end{aligned}$$

The last displayed expressions suggest that, under Condition $C4$, we can consider $\left(\frac{1}{n}\sum_{q=1}^{n}\mathcal{I}_{\check{u},q}(\lambda_j)\right)^{1/2}\mathcal{J}_{u^*,p}(\lambda_j)$ as some type of wild bootstrap in the frequency domain because under homogeneous time dependence

$$\begin{aligned}E^*\left|\left(\frac{1}{n}\sum_{q=1}^{n}\mathcal{I}_{\check{u},q}(\lambda_j)\right)^{1/2}\mathcal{J}_{u^*,p}(\lambda_j)\right|^2 &= \left(\sigma_u^{-2}(p)f_{u,p}(\lambda_j)\right)\cdot\sigma_u^2(p)\,.\,(1+o_p(1))\\ &= f_{u,p}(\lambda_j)(1+o_p(1)).\end{aligned}$$

The following theorem is used to establish the validity of our naïve bootstrap scheme.

Theorem 2. *(Naïve Bootstrap) Under Conditions $C1$–$C4$, we have that in probability,*

$$(nT)^{1/2}\left(\widetilde{\beta}^*-\widetilde{\beta}\right)\xrightarrow{d^*}\mathcal{N}(0,V).$$

With the bootstrap cluster estimator of the asymptotic covariance, given by,

$$\check{\Phi}^*=\frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}(\lambda_j)\,\mathcal{J}_{\hat{u}^*,p}(-\lambda_j)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}'_{\widetilde{x},p}(-\lambda_j)\,\mathcal{J}_{\hat{u}^*,p}(\lambda_j)\right)\right\}\qquad(3.3)$$

the next proposition establishes the consistency of the bootstrap cluster estimator.

Proposition 2. *(Naïve Bootstrap) Under the assumptions of Theorem 2, we have*

$$\breve{\Phi}^{*}-\breve{\Phi}=o_{p^{*}}(1).$$

The previous results can be extended to incorporate the more realistic situation where the temporal dynamics might differ by individual, as allowed by Conditions $C1$ and $C2$. This bootstrap, labelled the wild bootstrap, merges ideas from Hidalgo (2003) and Chan and Ogden (2009). As the DFT residuals are heterogeneous whilst independent over the Fourier frequencies, it applies the wild-type bootstrap approach to the increasing dimensional vector $\{\mathcal{J}_{\hat{u},p}(\lambda_j)\}_{p=1}^{n}$. It requires a modification of the above bootstrap, which primarily involves replacing *STEP 2.* For completeness we provide all steps:

***STEP 1′*:** Obtain the residuals

$$\widehat{u}_{pt}=\widetilde{y}_{pt}-\widetilde{\beta}'\widetilde{x}_{pt},\ \ p=1,...,n;\ t=1,...,T.$$

***STEP 2′*:** Denote $\{\eta_j\}_{j=1}^{\widetilde{T}}$ a sequence of independent identically distributed random variables with mean zero and unit variance. We then compute the bootstrap analogue of (2.3) as

$$\mathcal{J}_{y*,p}(\lambda_j)=\widetilde{\beta}'\mathcal{J}_{\widetilde{x},p}(\lambda_j)+\mathcal{J}_{\hat{u},p}(\lambda_j)\eta_j,\qquad \begin{cases} p=1,...,n\\ j=1,...,T-1,\end{cases}$$

where $\mathcal{J}_{y*,p}(\lambda_j)=\overline{\mathcal{J}_{y*,p}(\lambda_{T-j})}$ and $\eta_j=\eta_{T-j}$, for $j=\widetilde{T}+1,...,T-1$.

***STEP 3′*:** Compute the corresponding bootstrap analogue of (2.7) as

$$\widetilde{\beta}^{*}=\left(\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{J}_{\widetilde{x},p}(\lambda_j)\mathcal{J}_{\widetilde{x},p}'(-\lambda_j)\right)^{-1}\left(\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{J}_{\widetilde{x},p}(\lambda_j)\mathcal{J}_{\widetilde{y}*,p}(-\lambda_j)\right),$$

with $\mathcal{J}_{\widetilde{y}*,p}(\lambda_j)=\mathcal{J}_{y*,p}(\lambda_j)-\frac{1}{n}\sum_{q=1}^{n}\mathcal{J}_{y*,q}(\lambda_j)$.

Remark 7. *For a discussion regarding the requirement that $\eta_j=\eta_{T-j}$ for $j=\widetilde{T}+1,...,T-1.$, we refer to Hidalgo* (2003)*.*

The validity of the wild bootstrap scheme follows from the following proposition.

Proposition 3. *(Wild Bootstrap) Under Conditions $C1$–$C3$, in probability,*

$$(nT)^{1/2}\left(\widetilde{\beta}^{*}-\widetilde{\beta}\right)\xrightarrow{d^{*}}\mathcal{N}(0,V)$$

and $\breve{\Phi}^{*}-\breve{\Phi}=o_{p^{*}}(1)$.

We conclude with the stating the validity of the standardized bootstrap statistic

Corollary 2. *Under Conditions $C1$–$C3$, we have that in probability,*

$$(nT)^{1/2}\hat{V}^{*-1/2}\left(\widetilde{\beta}^{*}-\widetilde{\beta}\right)\xrightarrow{d^{*}}\mathcal{N}(0,I),$$

where $\hat{V}^{*}=\widetilde{\Sigma}_x^{-1}\breve{\Phi}^{*}\widetilde{\Sigma}_x^{-1}$.

Proof. The proof is standard after Theorem 2 and Propositions 1, 2 and 3. □

4. (CONDITIONAL) HETEROSKEDASTICITY

In this section, we extend our model to permit general forms of heteroskedasticity. Specifically, we begin by considering

$$y_{pt} = \beta' \acute{x}_{pt} + \eta_p + \alpha_t + v_{pt}, \quad p = 1, ..., n, \quad t = 1, ..., T, \tag{4.1}$$

where

$$v_{pt} =: \sigma_1 (w_p) \sigma_2 (\varrho_t) u_{pt}. \tag{4.2}$$

The error $\{u_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$satisfies the same regularity conditions given in Condition $C1$, exhibiting general spatial and temporal dependence. The sequences $\{w_p\}_{p\in\mathbb{N}}$ and $\{\varrho_t\}_{t\in\mathbb{Z}}$, which can even be functions of the fixed effects, are not required to be mutually independent of the regressors $\{\acute{x}_{pt}\}_{t\in\mathbb{Z}}$, $p \in N^+$. Without loss of generality we will normalize $\sigma^2_{u,p} = 1$ in Condition $C1$; $\sigma^2_{u,p}$ is not separately identified from $\sigma^2_1(w_p)$ and $\sigma^2_2(\varrho_t)$. The error $\{u_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$ is assumed to be independent of the regressors $\{\acute{x}_{pt}\}_{t\in\mathbb{Z}}$, $p \in N^+$, $\{w_p\}_{p\in\mathbb{N}^+}$ and $\{\varrho_t\}_{t\in\mathbb{Z}}$, see also footnote 2. Here the errors v_{pt} permit conditional heteroskedasticity. This is an extension of the so-called *groupwise heteroskedasticity*, where observations belonging to different groups have distinct variances, see for instance Greene (2018). This type of heteroskedasticity is not uncommon in applications such as in development economics, where it has been suggested that observations within a villages or strata would have the same (conditional) variance while differences over villages or strata exist (Deaton, 1996), that is the variance depends on some specific village variable(s).

Before we modify our Condition $C3$ to ensure we can permit this generalization, it is useful to introduce some notation. We shall denote $\{\ddot{x}_{pt}\}_{t\in\mathbb{Z}}, p \in \mathbb{N}^+$, the sequence that applies the usual transformation to remove the fixed effects, see (2.2), to the sequence $\{\acute{x}_{pt}\}_{t\in\mathbb{Z}}, p \in \mathbb{N}^+$, such that

$$\ddot{x}_{pt} = \acute{x}_{pt} - \frac{1}{n}\sum_{q=1}^{n} \acute{x}_{qt} - \frac{1}{T}\sum_{s=1}^{T} \acute{x}_{ps} + \frac{1}{nT}\sum_{q=1}^{n}\sum_{s=1}^{T} \acute{x}_{qs}.$$

Observe that as it happens with $\tilde{x}_{pt}$, we can take $E(\ddot{x}_{pt}) = 0$. Our new Condition $C3'$ is given next.

Condition C3′: *For all $p \in \mathbb{N}^+$, the sequence $\{u_{pt}\}_{t\in\mathbb{Z}}$ is independent of $\{\acute{x}_{pt}\}_{t\in\mathbb{Z}}$, $\{w_p\}_{p\in\mathbb{N}}$ and $\{\varrho_t\}_{t\in\mathbb{Z}}$ and*

$$0 < \max_{1\le p\le n} \sum_{q=1}^{n} \|\varphi(p,q)\| < \infty, \tag{4.3}$$

where $\varphi(p,q) := \varphi_u(p,q)\,\varphi_{\ddot{x}}(p,q)$ and

$$\varphi_{\ddot{x}}(p,q) = Cov\left(\sigma_1(w_p)\,\ddot{x}_{pt}; \sigma_1(w_q)\,\ddot{x}'_{qt}\right), \textit{ for any } p, q \ge 1.$$

The requirement given in (4.3) limits the combined cross-sectional dependence in v_{pt} and $\ddot{x}_{pt}$ ($\acute{x}_{pt}$) needed to ensure the existence of a consistent estimator of the "*average*" long-run variance of the sequences $\{z_{pt} =: v_{pt}\ddot{x}_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$ in this framework. This is an obvious extension of our previous Condition $C3$, since our expression for Φ under our generalization becomes

$$\begin{aligned}\Phi &=: \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}E\left\{\left(\sum_{p=1}^{n}\sum_{t=1}^{T}\ddot{x}_{pt}v_{pt}\right)\left(\sum_{p=1}^{n}\sum_{t=1}^{T}\ddot{x}_{pt}'v_{pt}\right)\right\}\\ &= \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}\sum_{p,q=1}^{n}\sum_{t,s=1}^{T}E\left(\left\{\sigma_2(\varrho_t)\sigma_2(\varrho_s)\right\}\left\{\sigma_1(w_p)\ddot{x}_{pt}\right\}\left\{\sigma_1(w_q)\ddot{x}_{qs}'\right\}\right)E(u_{pt}u_{qs}) \quad (4.4)\end{aligned}$$

We shall now give some examples. We can allow

$$\textbf{(i)}\quad \acute{x}_{pt} =: x_{pt} + h_1(w_p) + h_2(\varrho_t) \text{ and } \textbf{(ii)}\quad \acute{x}_{pt} =: h(w_p;\varrho_t)x_{pt}, \quad (4.5)$$

for given h_1, h_2 and h, where x_{pt} satisfies Condition $C2$. It is clear that under the additive structure in (**i**), the transformed variables that account for the fixed effects, recall (2.2), satisfy

$$\widetilde{\acute{x}}_{pt} =: \ddot{x}_{pt} \equiv \widetilde{x}_{pt},\ p = 1,...,n;\ t = 1,...,T,$$

which renders this, potentially, the most straightforward setting. In this case we have

$$\begin{aligned}\Phi &=: \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}E\left\{\left(\sum_{p=1}^{n}\sum_{t=1}^{T}\ddot{x}_{pt}v_{pt}\right)\left(\sum_{p=1}^{n}\sum_{t=1}^{T}\ddot{x}_{pt}'v_{pt}\right)\right\}\\ &= \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}E\left\{\left(\sum_{p=1}^{n}\sum_{t=1}^{T}x_{pt}\sigma_1(w_p)\sigma_2(\varrho_t)u_{pt}\right)\left(\sum_{p=1}^{n}\sum_{t=1}^{T}x_{pt}'\sigma_1(w_p)\sigma_2(\varrho_t)u_{pt}\right)\right\}\\ &= \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}\sum_{p,q=1}^{n}\sum_{t,s=1}^{T}E\left(\ddot{\mathrm{x}}_{pt}\ddot{\mathrm{x}}_{qs}'\right)E(u_{pt}u_{qs}),\end{aligned}$$

where $\ddot{\mathrm{x}}_{pt} = x_{pt}\sigma_1(w_p)\sigma_2(\varrho_t)$. The behaviour of the second moments of $\ddot{\mathrm{x}}_{pt}$ are essentially those of x_{pt} because

$$Cov\left(\ddot{\mathrm{x}}_{pt},\ddot{\mathrm{x}}_{qs}\right) = E\left(\sigma_1(w_p)\sigma_1(w_q)\right)E\left(\sigma_2(\varrho_t)\sigma_2(\varrho_s)\right)Cov(x_{pt},x_{qs}).$$

Remark 8. *In the last displayed assumption we have only assumed that*

$$E\left(x_{pt} \mid w_p,\varrho_t\right) = 0; \quad E\left(x_{pt}x_{qs} \mid w_p, w; \varrho_t,\varrho_s\right) = E(x_{pt}x_{qs}) =: Cov(x_{pt},x_{qs})$$

so some type of dependence between x_{pt} *and* (w_p,ϱ_t) *is still allowed.*

With the multiplicative structure in (**ii**), it is basically the same since

$$\begin{aligned}\Phi &=: \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}E\left\{\left(\sum_{p=1}^{n}\sum_{t=1}^{T}\ddot{x}_{pt}v_{pt}\right)\left(\sum_{p=1}^{n}\sum_{t=1}^{T}\ddot{x}_{pt}'v_{pt}\right)\right\}\\ &= \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}E\left\{\left(\sum_{p=1}^{n}\sum_{t=1}^{T}x_{pt}h(w_p;\varrho_t)\sigma_1(w_p)\sigma_2(\varrho_t)u_{pt}\right)\left(\sum_{p=1}^{n}\sum_{t=1}^{T}x_{pt}'h(w_p;\varrho_t)\sigma_1(w_p)\sigma_2(\varrho_t)u_{pt}\right)\right\}\\ &= \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}\sum_{p,q=1}^{n}\sum_{t,s=1}^{T}E\left(\ddot{\mathrm{x}}_{pt}\ddot{\mathrm{x}}_{qs}'\right)E(u_{pt}u_{qs}),\end{aligned}$$

where now $\ddot{\mathrm{x}}_{pt} = h(w_p;\varrho_t)\sigma_1(w_p)\sigma_2(\varrho_t)x_{pt}$,and $\left|Cov\left(\ddot{\mathrm{x}}_{pt},\ddot{\mathrm{x}}_{qs}\right)\right| \le K\left|Cov(x_{pt},x_{qs})\right|$ using Markov inequality. The same caveats mentioned in the last remark apply in this case.

We now turn to the consistent estimator of the "*average*" long-run variance of the sequences $\{z_{pt} =: v_{pt}\ddot{x}_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$ in this framework (recognizing that we have established the necessary regularity conditions for its existence). Following a rescaling of our regressors,

$$\dot{x}_{pt} =: \ddot{x}_{pt}\sigma_1(w_p)\sigma_2(\varrho_t),$$

for given $\sigma_1(w_p)\sigma_2(\varrho_t)$, our estimator for Φ, see also (2.17), becomes

$$\breve{\Phi} = \frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\tilde{\breve{x}},p}(\lambda_j)\mathcal{J}_{\hat{u},p}(-\lambda_j)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}'_{\tilde{\breve{x}},p}(-\lambda_j)\mathcal{J}_{\hat{u},p}(\lambda_j)\right)\right\}, \tag{4.6}$$

where $\hat{u}_{pt} := \hat{v}_{pt}/(\hat{\sigma}_1(w_p)\hat{\sigma}_2(\varrho_t))$ and $\tilde{\breve{x}}_{pt} = \hat{\sigma}_1(w_p)\hat{\sigma}_2(\varrho_t)\tilde{\dot{x}}_{pt}$. Implementation of this estimator only requires a consistent estimator of $\sigma_2^2(\varrho_t)$ (up to unknown scale of proportionality), and a natural estimator we can use is

$$\widehat{\sigma}_2^2(\varrho_t) = \frac{1}{n}\sum_{p=1}^{n}\hat{v}_{pt}^2.$$

The estimator for $\sigma_1^2(w_p)$, $\hat{\sigma}_1^2(w_p)$, indeed cancels out when considering the product $\mathcal{J}_{\tilde{\breve{x}},p}(\lambda_j)\mathcal{J}_{\hat{u},p}(-\lambda_j)$, as

$$\begin{aligned}\mathcal{J}_{\tilde{\breve{x}},p}(\lambda_j)\mathcal{J}_{\widehat{u},p}(-\lambda_j) &= \frac{1}{T^{1/2}}\sum_{t=1}^{T}\tilde{\dot{x}}_{pt}\hat{\sigma}_1(w_p)\hat{\sigma}_2(\varrho_t)e^{-it\lambda_j}\sum_{t=1}^{T}\frac{\widehat{v}_{pt}}{\widehat{\sigma}_2(\varrho_t)\hat{\sigma}_1(w_p)}e^{-it\lambda_j}\\ &= \frac{1}{T^{1/2}}\sum_{t=1}^{T}\tilde{\dot{x}}_{pt}\widehat{\sigma}_2(\varrho_t)e^{-it\lambda_j}\sum_{t=1}^{T}\frac{\widehat{v}_{pt}}{\widehat{\sigma}_2(\varrho_t)}e^{-it\lambda_j}\end{aligned}$$

Moreover, this result shows that when $\sigma_2(\varrho_t)$ is a constant, our results in Section 2 and 3 continue to hold true. That is, our estimators in previous section are robust to groupwise heteroskedasticity in the cross-sectional unit, a result supported by our Monte Carlo simulations in Table 4 in the next section.

The intuition of the validity of this estimator comes from the standard observation that

$$\frac{\widehat{\sigma}_2(\varrho_t)}{\sigma_2^2(\varrho_t)} \xrightarrow{P} 1,$$

so that

$$\widehat{u}_{pt} = \frac{\widehat{v}_{pt}}{\widehat{\sigma}_2(\varrho_t)} \simeq \frac{v_{pt}}{\widehat{\sigma}_2(\varrho_t)} = \frac{v_{pt}}{\sigma_2(\varrho_t)}(1+o_p(1)) =: \sigma_1(w_p)u_{pt}(1+o_p(1)),$$

and

$$\frac{\widehat{v}_{pt}}{\widehat{\sigma}_2(\varrho_t)\sigma_1(w_p)} =: u_{pt}(1+o_p(1))$$

from the above arguments. Of course the details can be lengthy, but have been considered in other contexts many times.

Our bootstrap algorithms also require some obvious and minimal change. The only adjustment to the wild bootstrap algorithm relates to the use of the robust estimator of Φ provided in (4.6). For the naïve bootstrap, a straightforward modification involves the following steps

***STEP 1″*:** Obtain the residuals

$$\widehat{v}_{pt} = \widetilde{y}_{pt} - \widetilde{\beta}'\tilde{\dot{x}}_{pt},\ \ p = 1, ..., n;\ \ t = 1, ..., T,$$

compute $\widehat{\sigma_1^2(w_p)\sigma_2^2(\varrho_t)} = T^{-1}\sum_{t=1}^{T}\widehat{v}_{pt}^2 \cdot n^{-1}\sum_{p=1}^{n}\widehat{v}_{pt}^2$, and obtain the standardized residuals

$$\widehat{u}_{pt} = \widehat{v}_{pt}/\widehat{\sigma_1(w_p)\sigma_2(\varrho_t)}$$

***STEP 2′′*:** Denoting $\hat{U}_t = \{\hat{u}_{pt}\}_{p=1}^{n}$, do standard random sampling from the empirical distribution of the residuals $\{\hat{U}_t\}_{t=1}^{T}$. That is, we assign probability T^{-1} to each $n \times 1$ vector $\hat{U}_t$. Denote the bootstrap sample by $\{U_t^*\}_{t=1}^{T}$, where $U_t^* = \left\{u_{pt}^*\right\}_{p=1}^{n}$. Let $V_t^* = \left\{\widehat{\sigma_1(w_p)\sigma_2(\varrho_t)}u_{pt}^*\right\}_{p=1}^{n}$ Compute the bootstrap analogue of (2.3) as

$$\mathcal{J}_{y*,p}(\lambda_j) = \widetilde{\beta}'\mathcal{J}_{\widetilde{x},p}(\lambda_j) + \left(\frac{1}{n}\sum_{q=1}^{n}\mathcal{I}_{\hat{u},q}(\lambda_j)\right)^{1/2}\mathcal{J}_{v^*,p}(\lambda_j)$$

for $p = 1, ..., n$ and $j = 1, ..., T-1$.

***STEP 3′′*:** Compute the corresponding bootstrap analogue of (2.7) as

$$\widetilde{\beta}^* = \left(\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{J}_{\widetilde{x},p}(\lambda_j)\mathcal{J}'_{\widetilde{x},p}(-\lambda_j)\right)^{-1}\left(\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{J}_{\widetilde{x},p}(\lambda_j)\mathcal{J}_{\widetilde{y}*,p}(-\lambda_j)\right),$$

with $\mathcal{J}_{\widetilde{y}*,p}(\lambda_j) = \mathcal{J}_{y*,p}(\lambda_j) - \frac{1}{n}\sum_{q=1}^{n}\mathcal{J}_{y*,q}(\lambda_j)$

Remark 9. *Step 2′′ assumes that the temporal dependence of the $n\times 1$ vector $\{v_{pt}/(\sigma_1(w_p)\sigma_2(\varrho_t))\}_{p=1}^{n}$ is homogeneous so we can use the average periodogram to impose proper dependence structure on u_{pt}^* (drawings from the empirical distribution of $\{\hat{v}_{pt}/(\widehat{\sigma_1(w_p)\sigma_2(\varrho_t)})\}_{p=1}^{n}$).*

We now discuss the scenario where the conditional moment of the error term depends on the regressors $\acute{x}_{pt} =: x_{pt}$ themselves, i.e.,

$$y_{pt} = \beta'\acute{x}_{pt} + \eta_p + \alpha_t + v_{pt}, \text{ with } v_{pt} =: \sigma(\acute{x}_{pt})u_{pt}$$

As mentioned in the introduction, this would require us to estimate the conditional expectation, $\sigma^2(\acute{x}_{pt})$, nonparametrically. Several methods are available such as the Kernel regression method or sieve estimation. As this approach would require the selection of a bandwidth parameter which we set out to avoid in this paper, we do not consider this in detail although we outline how to proceed. Regardless of the approach used, we anticipate that the estimator would be pretty accurate as the number of observations in large panel data will normally be huge. For instance in a typical data set, with $T = 20$ and $n = 1000$, we can use $20,000$ observations to estimate the nonparametric function. The estimator for Φ, see also (4.6), becomes

$$\breve{\Phi} = \frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}(\lambda_j)\mathcal{J}_{\hat{u},p}(-\lambda_j)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}'_{\acute{x},p}(-\lambda_j)\mathcal{J}_{\hat{u},p}(\lambda_j)\right)\right\},$$

where $\hat{u}_{pt} := \hat{v}_{pt}/\hat{\sigma}(\acute{x}_{pt})$ and $\widetilde{\tilde{x}}_{pt} = \hat{\sigma}(\acute{x}_{pt})\widetilde{x}_{pt}$. For the associated naïve bootstrap procedure, we can proceed as above where $\widehat{\sigma_1(w_p)\sigma_2(\varrho_t)}$ is replaced by $\hat{\sigma}(\acute{x}_{pt})$.[4]

[4]As this setting requires the use of a bandwidth, we may also consider an alternative bootstrap based on the MBB methodology. For its validity, though, the resampling would need to be done on the residuals and regressors jointly (a pairs MBB, see also Politis et al. (1997), that is we need to perform the bootstrap on $\{\mathcal{Z}_t\}_{t=1}^{T}$, where $\mathcal{Z}_t = \left\{(\acute{x}'_{pt}, \hat{u}_{pt})'\right\}_{p=1}^{n}$.

5. FINITE SAMPLE BEHAVIOUR

In this section, we discuss the finite sample performance of our cluster-based inference procedure in the presence of cross-sectional and temporal dependence of unknown form. We contrast this performance with the HAC-based inference procedure proposed by Driscoll and Kraay (1998), which unlike ours, requires the choice of smoothing parameters that may be arbitrary and erroneous. We also provide evidence of the potential finite sample improvements of our frequency domain bootstrap schemes and implement the MBB time domain bootstrap to the vector containing all individual observations for each t. Our frequency domain approaches have the benefit that they do not rely on the choice of any smoothing parameter or require an ordering of cross-sectional units, which, as we argued before, may be arbitrary and erroneous. Another benefit of our estimator we address in our simulations is the fact that our estimator permits heterogeneity in the temporal dependence. In our simulations, we also consider a multiplicative error structure that permits groupwise heteroskedasticity and we reveal the robustness of our estimator to this setting.

In our Monte Carlo experiments, we first consider the following data generating process

$$y_{pt} = \alpha_t + \eta_p + \beta x_{pt} + u_{pt} \text{ for } p = 1, ..., n \text{ and } t = 1, .., T.$$

The time fixed effects α_t and individual fixed effects η_p are drawn independently ($\alpha_t \sim IIDN(1,1)$ and $\eta_p \sim IIDN(1,1)$) and are held fixed across replications and without loss of generality β is set equal to zero. The independently drawn errors and regressors are postulated to exhibit a variety of scenarios for the temporal and cross-sectional dependence that are assumed to be the same for simplicity.

To evaluate the performance of our proposed cluster estimator, we analyze the empirical size and power for testing the significance of our parameter, $H_0 : \beta = 0$ against $H_A : \beta \neq 0$, at the nominal 5% level for various pairs of n and T using 5,000 simulations. In addition to presenting the rejection rates based on the asymptotic distribution of the Wald statistic $nT\hat{\beta}_{FE}'\hat{V}^{-1}\hat{\beta}_{FE}$, with $\widehat{V} =: \widetilde{\Sigma}_x^{-1}\breve{\Phi}\widetilde{\Sigma}_x^{-1}$ where $\breve{\Phi}$ is defined in (2.17) (or equivalently the asymptotic t-test as β is scalar), we present rejection rates based on the empirical distribution of the bootstrapped test statistic

$$nT\left(\hat{\beta}_{FE}^* - \hat{\beta}_{FE}\right)'\left[\hat{V}^*\right]^{-1}\left(\hat{\beta}_{FE}^* - \hat{\beta}_{FE}\right),$$

where $\hat{\beta}_{FE}^*$ and $\hat{V}^*$ are the bootstrapped estimators of β and V defined in Section 3. As inference based on the asymptotic distribution might not provide a good approximation to the finite sample one, this allows us to assess the finite sample improvements our bootstrap schemes may yield.

We compare the finite sample performance of our cluster-based inference procedure to the HAC based inference procedure and select the bandwidth parameter, denoted m_T, using the parametric AR(1) plug-in method suggested in Andrews (1991).[5] This lag window is designed to minimize (approximately) the mean square of the standard error.[6] For the HAC based inference we provide rejection rates of the Wald statistic $nT\hat{\beta}_{FE}'\hat{V}_{m_T}^{-1}\hat{\beta}_{FE}$ based on the asymptotic critical

[5] m_T is chosen to be upward rounded integers.

[6] Kiefer and Vogelsang (2002) discuss the use of HAC estimators with bandwidth equal to the sample size ($b = 1$). This bandwidth free approach does come at the cost of power relative to Andrews' popular data driven optimal bandwidth selection, see also Vogelsang (2012).

values (asy) and the critical values based on the fixed-b asymptotics (fixb) of Kiefer and Vogelsang (2005) as this is shown to lead to more reliable inference, see also Vogelsang (2012). With $\widehat{V}_{m_T} =: \widetilde{\Sigma}_x^{-1}\hat{\Phi}_{m_T}\widetilde{\Sigma}_x^{-1}$, $\hat{\Phi}_{m_T}$ is defined as

$$\hat{\Phi}_{m_T} = \frac{1}{nT}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{|t-s|}{m_T}\right)\hat{z}_{pt}\hat{z}'_{qs},$$

where $\hat{z}_{pt} = \widetilde{x}_{pt}\widehat{u}_{pt}$ and $K(h) = (1-|h|)\,\mathbf{1}\,(|h| \leq 1)$ is the Bartlett kernel. The fixed-b asymptotic distribution is non-standard and our critical values are obtained by simulation.[7]

We also provide critical values for HAC based inference that rely on the pairs moving block bootstrap proposed by Gonçalves (2011). She obtained bootstrapped samples $z_{it}^* = (y_{it}^*, x_{it}^{*\prime})'$ by arranging k resampled blocks of ℓ observations from the set of $T-\ell+1$ overlapping blocks $\{B_{1,\ell}, .., B_{T-\ell+1,\ell}\}$ with $B_{t,\ell} = \{z_{t,n}, z_{t+1,n}, .., z_{t+\ell-1,n}\}$ and $z_{t,n} = (z_{1t}, .., z_{nt})'$ in sequence (for notational simplicity $T = k\ell$). When $\ell = 1$ this corresponds to the standard iid bootstrap on $\{z_{t,n}\}_{t=1}^T$. The MMB based critical value are based on the standardized test statistic $T\left(\hat{\beta}_{FE}^* - \hat{\beta}_{FE}\right)'$ $\left[\hat{V}_\ell^*\right]^{-1}\left(\hat{\beta}_{FE}^* - \hat{\beta}_{FE}\right)$. Here $\hat{V}_\ell^* = \left(\widetilde{\Sigma}_x^*\right)^{-1}\breve{\Phi}_\ell^*\left(\widetilde{\Sigma}_x^*\right)^{-1}$, $\widetilde{\Sigma}_x^* = \frac{1}{nT}\sum_{p=1}^{n}\sum_{t=1}^{T}\widetilde{x}_{pt}^*\widetilde{x}_{pt}^*$ and

$$\breve{\Phi}_\ell^* = \frac{1}{k}\sum_{j=1}^{k}\left(\ell^{-1/2}\sum_{t=1}^{\ell} n^{-1}\hat{s}_{n,(j-1)\ell+t}^*\right)\left(\ell^{-1/2}\sum_{t=1}^{\ell} n^{-1}\hat{s}_{n,(j-1)\ell+t}^*\right)',$$

where $\hat{s}_{nt}^* = \sum_{p=1}^{n}\widetilde{x}_{pt}^*\left(\widetilde{y}_{pt}^* - \widetilde{x}_{pt}^{*\prime}\hat{\beta}_{FE}^*\right)$ with $\widetilde{y}_{pt}^* = y_{pt}^* - \overline{y}_{p\cdot}^* - \overline{y}_{\cdot t}^* + \overline{\overline{y}}_{\cdot\cdot}^*$ and $\widetilde{x}_{pt}^* = x_{pt}^* - \overline{x}_{p\cdot}^* - \overline{x}_{\cdot t}^* + \overline{\overline{x}}_{\cdot\cdot}^*$ (see also Götze and Künsch, 1996). The block size used is given by the integer part of the automatic bandwidth chosen by the Andrews (1991) as proposed by Gonçalves (2011).

5.1. **Simulations with Homogeneous Time Dependence.** In the first set of simulations, we assume that the time dependence is homogeneous among individuals $p = 1, .., n$. In particular, we assume that the error and regressors are mutually independent, homoskedastic, first order autoregressive random variables with $\rho = 0.7$ or $\rho = 0.9$. The error term, therefore, takes the form

$$u_{pt} = \rho u_{p,t-1} + \sqrt{1-\rho^2}\eta_{pt}, \text{ with } \rho = 0.7, 0.9$$

where η_{pt} characterizes the spatial dependence inherent in the error.[8] We consider both a weak and a strong cross-sectional dependence scenario for u_{pt} (η_{pt}). To describe the cross-sectional dependence, we follow Lee and Robinson (2013) and draw random locations for individual units along a line, denoted $s = (s_1, ...s_n)'$ with $s_p \sim IIDU[0, n]$ for $p = 1, .., n$. Using the linear time dependence representation, $\eta_{pt} = \sigma_p\left(\sum_{\ell=1}^{\infty} c_\ell(p)\, e_{\ell t}\right)$ with $e_{\ell t} \sim IIDN(0,1)$, we set $c_\ell(p) = (1+|s_\ell - s_p|_+)^{-10}$ to permit weak dependence; σ_p is such that $Var(\eta_{pt}) = 1$. For the strong spatial dependence setting, we use $c_\ell(p) = (1+|s_\ell - s_p|_+)^{-0.7}$ instead, see also Hidalgo and Schafgans (2017). The same discussion holds for the independently drawn, strictly exogenous regressor x_{pt},

[7]Let $W_q(r)$ denote a q dimensional vector of independent standard Wiener processes and define $\tilde{W}_q(r) = W_q(r) - rW_q(1)$. The limiting distribution of the t-test is $W_1(1)/\sqrt{C_1}$ with $C_q = \frac{2}{b}\int_0^1 \tilde{W}_q(r)\tilde{W}_q(r)'dr$ $-\frac{1}{b}\int_0^{1-b}\left[\tilde{W}_q(r+b)\tilde{W}_q(r)' + \tilde{W}_q(r)\tilde{W}_q(r+b)'\right]dr$ given the use of the Bartlet kernel, where $b \in (0,1]$ with $m_T = bT$ (see Theorem 4, Vogelsang, 2012); the limiting distribution of the Wald test is $W_q(1)'C_q^{-1}W_q(1)$. We obtain the critical values using 500,000 simulations.

[8]We generated the spatial data with $49+T$ periods and take the last T periods as our sample using 0 as the starting value.

where, to allow for some time heterogeneity, we may, without loss of generality add μ_t, which is independently drawn $(\mu_t \sim IIDN(1,1))$.

In Table 1, we report the empirical size for testing the significance of β at the 5% level of significance based on our cluster estimator of the variance of $\tilde{\beta}$ in the columns labelled HS (Cluster). In addition to presenting the rejection rates based on the asymptotic critical values (asy), we report the empirical size based on the naïve bootstrap (nb), and the wild bootstrap (nb). The empirical size based on the HAC based inference procedure proposed by Driscoll and Kraay (1998) are reported in the columns labelled DK (HAC). For the HAC based inference, we provide rejection rates based on the asymptotic critical values (asy), the critical values based on the fixed-b asymptotics (fixb) of Kiefer and Vogelsang (2005), and Gonçalves' (2011) MBB (mbb). We used the parametric AR(1) plug-in method suggested by Andrews (1991) to determine the window lag m_T and the block length ℓ.

Insert Table 1 around here

The results from Table 1 reveal that our cluster based inference performs remarkably well even in the presence of strong cross sectional dependence for moderately large panels. As before, the rejection rates based on the asymptotic critical values tend to be closer to the nominal rejection rates as n and T increase. The finite sample performance using these asymptotic critical values does suffer, in particular, from T being small, more so when the temporal dependence is stronger. This suggests that the cluster variance's finite sample performance, in particular, appears to require larger T, in order for us to be able to rely on the asymptotic critical value. Nevertheless, finite sample improvements in inference can be made using either frequency domain bootstrap schemes as rejection rates based on them are typically closer to the nominal rejection rates, with the differences typically smaller as sample sizes increase. Given that we assume the temporal dynamics to be the same for all individuals in this simulation, both bootstrap schemes are valid. The naïve bootstrap approach tends to perform better in the sense of providing a size closer to the nominal rejection rate.

Our cluster based inference, using the naïve bootstrap for small panels, suggests large improvements in size relative to HAC based inference. While the use of fixed-b asymptotic critical values for HAC based inference does indeed improve its performance, in accordance with Vogelsang, (2012), the gains in improvement in size achieved by our cluster based estimator remain significant and are larger when the temporal or spatial dependence is stronger. Our cluster based inference, however, does not necessarily perform superior to the HAC based inference that use the critical values based on Gonçalves' pairwise MBB. Her approach indeed performs very well in this setting where the temporal dependence is homogeneous across individuals. As we will see in Table 3 relaxing this assumption, which is more realistic, does reveal a marked improvement of our cluster based performance over HAC based inference using the MBB. But even in the homogeneus setting, it should be noted that the MBB approach is sensitive to the chosen block size, and its selection here was appropriate given the imposed AR(1) temporal dependence (which is unknown in practical applications). Contrary to the MBB we do not need to choose a block size.

In Table 2, we present the empirical power of our test for the significance of the slope when $\beta = 0.1$ for a selection of (n, T) pairs and compare the performance of our cluster-based inference procedure to the HAC based inference procedure proposed by Driscoll and Kraay as before.

Insert Table 2 around here

The results from Table 2 show that our cluster based inference has good power to reject $H_0 : \beta = 0$ when $\beta = 0.1$ in both time dependence scenarios, even for small panels, in particular when the spatial dependence is not strong. For the reported sample sizes, the cluster based inference using the naïve bootstrap only showed limited size distortions. As expected its power approaches one as the sample size, and therefore the precision of our estimator, increases. This improved power performance comes about faster when the cross sectional and/or temporal dependence is lower and improved power performance appears stronger with increases in T relative to n. The power for our cluster based inference, using the naïve bootstrap, compares well with that of power of HAC based inference. Where the size-distortions for HAC based inference are smallest, any apparent power loss of cluster based inference disappears. Both cluster based inference and HAC based inference have a comparable loss of power when both the spatial and the temporal dependence are large.

5.2. **Simulations with Heterogeneous Time Dependence.** In our second set of simulations, we allow individual heterogeneity in the time dependence of the error and the strictly exogenous regressor. The error term u_{pt} is generated using various heterogeneous ARMA processes

$$(1 - \rho_{1,p}L)(1 + \rho_2 L + \rho_3 L^2)u_{pt} = (1 + \theta_{1,p}L + \theta_2 L^2 + \theta_3 L^3)\eta_{pt},$$

with L denoting the lag operator, such that, e.g., $Lu_{pt} = u_{p,t-1}$, $\rho_{1,p}$ and $\theta_{1,p}$ are individual specific AR and MA coefficients, and (ρ_2, ρ_3) and (θ_2, θ_3) are additional non-varying higher order AR and MA coefficients. As before η_{pt} characterizes the spatial dependence. A similar description holds for the independently drawn, strictly exogenous regressor, x_{it}, which is assumed to have the same spatial temporal dependence as the error for simplicity. We allow the variance of u_{pt} to vary across individuals $p = 1, ..., n$.

We consider four heterogeneous specifications: Mixed AR(1), Mixed AR(1)/MA(1), Mixed AR(3), and Mixed AR(3)/MA(3). The individual specific parameters $\rho_{1,p}$ and $\theta_{1,p}$, where non-zero, reflect equidistant points on $[0.5, 0.9]$. The full details of these heterogeneous specifications are provided at the bottom of Table 3.

Insert Table 3 around here

In Table 3, we report the empirical size for testing the significance of β in the presence of heterogeneous time dependence for panels where $n = 100$ and $T = 64$, 128, and 256. As before, we consider both weak and strong spatial dependence scenarios. For HAC based inference we used the parametric AR(1) plug-in method suggested by Andrews (1991) again, to determine the window lag m_T and the block length ℓ. A common approach, which neither recognizes the temporal heterogeneity nor the higher order (autoregressive) nature of the temporal dependence under consideration.

The results in Table 3 show that our cluster estimator of the variance is robust to the presence of individual specific time dependence. The rejection rates based on the asymptotic critical values in the heterogeneous AR(1) time dependence setting, with $\{\rho_{1,p}\}_{p=1}^{n}$ in the range $[0.5, 0.9]$, are comparable to the rejection rates in the homogeneous AR(1) setting with $\rho = 0.7$.[9] As in the homogeneous time dependence setting, the rejection rates based on the asymptotic critical values approach the nominal rejection rate of 5% as the sample size increases. The rejection rates based on both frequency-based bootstrap schemes show that finite sample improvements in inference can be made. The improvements achieved when applying the wild bootstrap, proven to be valid in the heterogeneous time dependence scenario, are more modest than those suggested by the naïve bootstrap, which assumes homogeneous time dependence. Our cluster based inference reveals a similar pattern when we permit higher order heterogeneous autoregressive/moving average temporal dependence with the naïve bootstrap performing remarkably well again, suggesting that the naïve bootstrap may be robust to violations of the homogeneous time dependence such as those considered in these simulations. Whereas the wild bootstrap does perform less well then expected, in particular in the presence of strong spatial dependence, the discrepancy between the rejection rates based on the two bootstrap schemes does appear to be smaller than in the homogeneous time dependence scenario.

Importantly, our cluster based inference suggests large improvements over HAC based inference in these heterogeneous time dependence settings, whether we use the asymptotic, the fixed-b asymptotic critical values or base its rejection rates on the MBB. The inferior HAC based inference may be explained by the inappropriate use of a single smoothing parameter in these heterogeneous settings, as is common practice, in addition to the fact that the parametric AR(1) plug-in method does not account for other, and possible higher order (autoregressive) processes, than AR(1). Our cluster based inference benefits from not requiring the choice of any smoothing parameter, and is therefore not subject to this deterioration in size. Aside from the ease of implementation, the robustness of our approach to the presence of individual specific time dependence is a particularly attractive feature of our cluster robust inference.

5.3. **Simulations with (Conditional) Heteroskedasticity.** Finally, we consider simulations that make use of our "modified" cluster based inference that permits general forms of heteroskedasticity. Here the data generating process is given by

$$y_{pt} = \alpha_t + \eta_p + \beta \acute{x}_{pt} + \sigma_1(w_p)\sigma_2(\varrho_t)\, u_{pt} \text{ for } p = 1, ..., n \text{ and } t = 1, .., T.$$

We consider both an additive and multiplicative specification for $\acute{x}_{pt}$, in particular

$$\acute{x}_{pt} = x_{pt} + w_p + \varrho_t \text{ and } \acute{x}_{pt} = x_{pt}\,(w_p \varrho_t)^2 .$$

Here u_{pt} and x_{pt} are drawn independently with weak temporal and weak cross-sectional dependence; w_p and ϱ_t are additional regressors where w_p exhibits strong spatial dependence and ϱ_t follows an AR(1) with coefficient equal to 0.7. Without loss of generality $\beta = 0$ again. Due to the presence of the multiplicative error $\sigma_1(w_p)\,\sigma_2(\varrho_t)\, u_{pt} := v_t$, this setting does permit (conditional)

[9] Associated simulations considering the power to reject $H_0 : \beta = 0$ when $\beta = 0.1$ in the presence of heterogenous temporal dependence show comparable results as in the homogenous time dependence setting, see also Hidalgo and Schafgans (2018).

heteroskedasticity with $Var\left(v_{pt}|x_{pt}, w_p\right) = \sigma_1^2\left(w_p\right)\sigma_2^2\left(\varrho_t\right)$ after a normalization of the variance of u_{pt} to one, for simplicity. In particular, we consider

$$\sigma_1\left(w_p\right)\sigma_2\left(\varrho_t\right) = \sigma \cdot \left[\exp(\delta_1 w_p) + 1\right]\left[\exp(\delta_2 \varrho_t) + 1\right] \text{ with } \delta_1 = 0.5,\ 2.0 \text{ and } \delta_2 = 0, 0.2, 0.5.$$

The severity of heteroskedasticity, which we can measure using the coefficient of variation of $\sigma_1^2\left(w_p\right)\sigma_2^2\left(\varrho_t\right)$, increases with the values of δ_1 and δ_2. The coefficient of variation is defined as the ratio of the standard deviation of $\sigma_1^2\left(w_p\right)\sigma_2^2\left(\varrho_t\right)$ to its mean. The average coefficient of variation of $\sigma_1^2\left(w_p\right)\sigma_2^2\left(\varrho_t\right)$ over our simulations with $\delta_2 = 0.5$ ranges from 42% ($\delta_1 = 0.5$) to 260% ($\delta_1 = 2.0$). The constant σ in chosen in such a way that the expected variability of $\sigma_1^2\left(w_p\right)\sigma_2^2\left(\varrho_t\right)$, equals one for comparability across simulations.

In Table 4, we report the empirical size for testing the significance of β in the presence of (conditional) heteroskedasticity. The average coefficient of variation for each specification across the simulations is given in the first column. We provide two sets of simulations for our cluster based inference: first we apply the original cluster based inference, which is robust to the presence of heteroskedasticity that is only cross-sectional in nature, followed by the heteroskedasticity robust cluster based inference. In the top panel, we report the results based on the additive specification of the regressor. The multiplicative specification of the regressors is in the bottom panel. As before, we will compare the empirical size of our (robust) cluster based inference with the HAC based inference, in particular those using the MMB based critical values. As we impose an AR(1) temporal dependence, we have ensured that the use of the parametric AR(1) plug-in method suggested by Andrews (1991) to determine the window lag m_T and the block length ℓ required for this approach is suitable.

Insert Table 4 around here

The results in Table 4 show that under the additive formulation of the regressor, the performance of the cluster based inference and robust cluster based inference (which accounts for a non-constant $\sigma_2\left(\rho_t\right)$), is quite similar. The robust cluster based inference is required for our first three formulations, where $\sigma(w_p, \varrho_t) = \sigma_1(w_p)\sigma_2(\varrho_t)$, whereas the final formulation permits the original cluster based inference. Compared to the results in Table 1 (case with weak spatial and temporal dependence), the rejection rates in the presence of (conditional) heteroskedasticity are only slightly larger (in part explained by the need to use estimates for $\sigma_2\left(\rho_t\right)$). Since $\widetilde{\widetilde{x}}_{pt} = \widetilde{x}_{pt}$ under the additive formulation of the regressor, a comparison with the results in Table 1 is more straightforward here than under the multiplicative formulation. Rejection rates that rely on the naïve bootstrap compare favourably with those of the HAC based inference that uses the MBB and there does not appear a serious deterioration in the performance of the (robust) cluster based inference when the severity of heteroskedasticity increases, either via δ_1 or δ_2, in this setting.

Under the multiplicative formulation of the regressor, the performance of the robust cluster based inference is clearly superior in our first three specifications where robust cluster based inference is required. The cluster based inference that does not account for (conditional) heteroskedasticity that is not purely cross-sectional in nature (i.e., in the presence of non-constant $\sigma_2\left(\rho_t\right)$), deteriorates quite quickly with δ_2 (parameter reflecting the severity of temporal heteroskedasticity). The rejection rates that use the robust estimator of the long-run variance, 4.6, are much closer to the nominal 5% rejection rates, whether we use the asymptotic critical values or the bootstrap algorithms. In fact, rejection rates that rely on the naïve bootstrap compare again quite well with the HAC based inference that uses the MBB, which reveals the robustness of our estimator to this type of (conditional) heteroskedasticity.[10] This is a welcome result, given that our estimator is simple to apply and does not require the choice of any smoothing parameter.

6. CONCLUSIONS

In this paper we extend the literature on inference in panel data models in the presence of both temporal and cross-sectional dependence of unknown form. While a standard methodology, based on the HAC estimator, is often invoked and used in the context of time series regression models, in the presence of cross-sectional dependence its implementation has only recently been considered, see Kim and Sun (2013), Driscoll and Kraay (1998) or Vogelsang (2012). To deal with various potential caveats of the HAC estimator, we propose a cluster based estimator which is able to take into account both types of dependence and allows the temporal dependence to be heterogeneous across individuals, extending the work of Arellano (1987) and Driscoll and Kraay (1998) in a substantial way. We provide a new CLT that accounts for an unknown and general temporal spatial dependence structure that permits strong spatial dependence. We thereby provide primitive conditions that guarantee Kim and Sun's (2013) , Driscoll and Kraay's (1998) and Gonçalves' (2011) assumption of the existence of a suitable CLT.

Our approach is based on the insightful observation that the spectral representation of the fixed effect panel data model is such that the errors become approximately temporally uncorrelated and heteroskedastic allowing the use of a cluster estimator of the long run variance in the frequency domain. As the cluster estimator may not be reliable in small samples, and therefore may not provide a good approximation to make accurate inferences, we present and examine bootstrap schemes in the frequency domain that are also bandwidth parameter free.

Our simulation results reveal that our cluster estimator performs quite well even in the presence of strong spatial dependence. For large panels, inference based on our cluster estimator is properly sized even in the presence of heterogeneous time dependence unlike Driscoll and Kraay's HAC based inference of cross sectional averages that ignores such heterogeneity. Our bootstrap schemes provide small sample improvements, where inference that uses the naïve bootstrap, in particular, is well sized, and reveal large improvement in size relative to HAC based inference when fixed-b asymptotic critical values are used. Improvements over MBB based inference are more limited, except in the presence of heterogeneous time dependence. We have shown the robustness of our cluster based inference to the presence of "groupwise" heteroskedasticity. To enable us to adapt to the presence of "groupwise" heteroskedasticity that is not purely cross-sectional in nature, a simple robust cluster based inference procedure was proposed that also does not require the selection of any smoothing parameter.

[10]While more general forms of heteroskedasticity could have been considered, this would have required non-parametric estimates for $\sigma(\acute{x}_{pt})$ and bandwidth selection (see also footnote 4) and was not attempted here. For moderately large sized panels, there is no reason to expect the MBB to outperform our proposal.

Appendix A: PROOF OF MAIN RESULTS

We first introduce some notation. For a generic function h, we shall abbreviate $h(\lambda_j)$ by $h(j)$ and for generic sequences $\{\psi_{pt}\}_{t=1}^{T}$, $p=1,...,n$,

$$\mathcal{J}_{\overline{\psi},\cdot}(j)=\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(\frac{1}{n}\sum_{q=1}^{n}\psi_{qt}\right)e^{-it\lambda_j}.$$

Using expression (10.3.12) of Brockwell and Davis (1991), we also have the useful relation

$$\begin{aligned}\mathcal{J}_{u,p}(j) &= \mathcal{B}_{u,p}(-j)\,\mathcal{J}_{\xi,p}(j)+\Upsilon_{u,p}(j) \\ \mathcal{J}_{x,p}(j) &= \mathcal{B}_{x,p}(-j)\,\mathcal{J}_{\chi,p}(j)+\Upsilon_{x,p}(j), \qquad p=1,...,n,\end{aligned} \tag{A.1}$$

where $\mathcal{B}_{u,p}(j)=:\mathcal{B}_{u,p}\left(e^{i\lambda_j}\right)$, $\mathcal{B}_{x,p}(j)=:\mathcal{B}_{x,p}\left(e^{i\lambda_j}\right)$ and

$$\begin{aligned}\Upsilon_{u,p}(j) &= \sum_{\ell=0}^{\infty}d_{\ell}(p)\,e^{-i\ell\lambda_j}\left(\frac{1}{T^{1/2}}\left\{\sum_{t=1-\ell}^{T-\ell}-\sum_{t=1}^{T}\right\}\xi_{pt}e^{-it\lambda_j}\right) \\ \Upsilon_{x,p}(j) &= \sum_{\ell=0}^{\infty}c_{\ell}(p)\,e^{-i\ell\lambda_j}\left(\frac{1}{T^{1/2}}\left\{\sum_{t=1-\ell}^{T-\ell}-\sum_{t=1}^{T}\right\}\chi_{pt}e^{-it\lambda_j}\right).\end{aligned} \tag{A.2}$$

Finally, we shall make use of the well know result

$$\begin{aligned}E\mathcal{J}_{\chi,p}(j)\,\mathcal{J}_{\chi,q}(-k) &= \varphi_x(p,q)\,\mathbf{1}(j=k) \\ E\mathcal{J}_{\xi,p}(j)\,\mathcal{J}_{\xi,q}(-k) &= \varphi_u(p,q)\,\mathbf{1}(j=k).\end{aligned} \tag{A.3}$$

6.1. PROOF OF THEOREM 1.

For completeness, we provide the proof using the time domain estimator, $\hat{\beta}$, and the frequency domain estimator, $\tilde{\beta}$.

We begin with $\hat{\beta}$. Without loss of generality assume that x_{pt} is scalar. Using (2.2) and standard arguments, we obtain

$$\begin{aligned}\sum_{t=1}^{T}\sum_{p=1}^{n}\widetilde{x}_{pt}\widetilde{u}_{pt} &= \sum_{t=1}^{T}\sum_{p=1}^{n}x_{pt}u_{pt}-\sum_{t=1}^{T}\sum_{p=1}^{n}\left(\overline{x}_{\cdot t}+\overline{x}_{p\cdot}-\overline{x}_{\cdot\cdot}\right)u_{pt} \\ &-\sum_{t=1}^{T}\sum_{p=1}^{n}\left(\overline{u}_{\cdot t}+\overline{u}_{p\cdot}-\overline{u}_{\cdot\cdot}\right)x_{pt}+o_p\left((nT)^{1/2}\right).\end{aligned} \tag{A.4}$$

Because the second and third terms on the right of $(A.4)$ are handled similarly, we shall only look at the second. Now

$$\begin{aligned}E\left(\sum_{t=1}^{T}\sum_{p=1}^{n}\overline{x}_{\cdot t}u_{pt}\right)^2 &= \sum_{t,s=1}^{T}\sum_{p,q=1}^{n}E\left(\overline{x}_{\cdot t}\overline{x}_{\cdot s}\right)\gamma_{u,pq}(t-s)\,\varphi_u(p,q) \\ &= \frac{1}{n^2}\sum_{p_2,q_2,p_1,q_1=1}^{n}\varphi_x(p_2,q_2)\,\varphi_u(p_1,q_1)\sum_{t,s=1}^{T}\gamma_{x,p_2q_2}(t-s)\,\gamma_{u,p_1q_1}(t-s) \\ &\le C\frac{T}{n^2}\left(\sum_{p_2,q_2=1}^{n}\left|\varphi_x(p_2,q_2)\right|\right)\left(\sum_{p_1,q_1=1}^{n}\left|\varphi_u(p_1,q_1)\right|\right) \\ &= o(nT).\end{aligned}$$

The latter displayed expression holds true because Conditions $C1$ and $C2$ imply that

$$\sum_{t,s=1}^{T} \sup_{p,q} \left|\gamma_{x,pq}(t-s)\right| + \sup_{p,q} \left|\gamma_{u,pq}(t-s)\right| < C, \tag{A.5}$$

whereas Condition $C3$, see also Remark 2, implies that[11]

$$\sum_{q=1}^{n} \varphi_u(p,q) \sum_{q=1}^{n} \varphi_x(p,q) = o(n) \tag{A.6}$$

so that

$$\sum_{p_1,p_2=1}^{n} \varphi_u(p_1,p_2) \sum_{q_1,q_2=1}^{n} \varphi_x(q_1,q_2) = o\left(n^3\right). \tag{A.7}$$

Proceeding similarly with $\sum_{t=1}^{T}\sum_{p=1}^{n} \overline{x}_{p\cdot} u_{pt}$ and $\overline{x}_{\cdot\cdot} \sum_{t=1}^{T}\sum_{p=1}^{n} u_{pt}$, we can conclude using $(A.4)$ that

$$\frac{1}{(nT)^{1/2}} \sum_{t=1}^{T}\sum_{p=1}^{n} \widetilde{x}_{pt}\widetilde{u}_{pt} = \frac{1}{(nT)^{1/2}} \sum_{t=1}^{T}\sum_{p=1}^{n} x_{pt}u_{pt} + o_p(1) \overset{d}{\to} \mathcal{N}(0,\Phi)$$

by Lemma B.8. From here it is standard to conclude that $(nT)^{1/2}\left(\widehat{\beta}-\beta\right) \to_d \mathcal{N}\left(0,\Sigma^{-1}\Phi\Sigma^{-1}\right)$.

We now show that $(nT)^{1/2}\left(\widetilde{\beta}-\beta\right) \to_d \mathcal{N}\left(0,\Sigma^{-1}\Phi\Sigma^{-1}\right)$. Proceeding similarly as we did above, we shall examine

$$\frac{1}{(nT)^{1/2}} \sum_{p=1}^{n}\sum_{j=1}^{T-1} \mathcal{J}_{x,p}(j)\,\mathcal{J}_{u,p}(-j) - \frac{1}{(nT)^{1/2}} \sum_{p=1}^{n}\sum_{j=1}^{T-1} \mathcal{J}_{x,p}(j)\,\mathcal{J}_{\overline{u},\cdot}(-j) \tag{A.8}$$
$$-\frac{1}{(nT)^{1/2}} \sum_{p=1}^{n}\sum_{j=1}^{T-1} \mathcal{J}_{\overline{x},p}(j)\,\mathcal{J}_{u,\cdot}(-j).$$

The first term of $(A.8)$ converges in distribution to $\mathcal{N}(0,\Phi)$ by Lemma B.9. So, to complete the proof it suffices to show that the last two terms of $(A.8)$ are $o_p(1)$. We examine the second term only, with the third term being handled similarly. By standard algebra and $(A.1)$, this term is

$$\frac{1}{n^{3/2}} \sum_{p,q=1}^{n} \frac{1}{T^{1/2}} \sum_{j=1}^{T-1} \mathcal{B}_{x,p}(j)\,\mathcal{B}_{u,q}(j)\,\mathcal{J}_{\chi,p}(j)\,\mathcal{J}_{\xi,q}(-j)$$
$$+\frac{1}{n^{3/2}} \sum_{p,q=1}^{n} \frac{1}{T^{1/2}} \sum_{j=1}^{T-1} \mathcal{B}_{x,p}(j)\,\mathcal{J}_{\chi,p}(j)\left\{\mathcal{J}_{u,q}(-j) - \mathcal{B}_{u,q}(j)\,\mathcal{J}_{\xi,q}(-j)\right\}$$
$$+\frac{1}{n^{3/2}} \sum_{p,q=1}^{n} \frac{1}{T^{1/2}} \sum_{j=1}^{T-1} \mathcal{B}_{u,p}(j)\,\mathcal{J}_{\xi,q}(-j)\left\{\mathcal{J}_{x,q}(-j) - \mathcal{B}_{x,q}(j)\,\mathcal{J}_{\chi,p}(j)\right\} \tag{A.9}$$
$$+\frac{1}{n^{3/2}} \sum_{p,q=1}^{n} \frac{1}{T^{1/2}} \sum_{j=1}^{T-1} \left(\mathcal{J}_{x,q}(-j) - \mathcal{B}_{x,q}(j)\,\mathcal{J}_{\chi,p}(j)\right) \times \left(\mathcal{J}_{u,q}(-j) - \mathcal{B}_{u,q}(j)\,\mathcal{J}_{\xi,q}(-j)\right).$$

[11]For two nonnegative sequences $\{\alpha_p\}$ and $\{\beta_p\}$, $\sum \alpha_p\beta_p < C$ implies that $\sum \alpha_p \sum \beta_p = o(n)$ if $\sum(\alpha_p+\beta_p) = o(n)$.

We examine the second term of $(A.9)$ first. Using $(A.3)$, we have that its second moment is bounded by

$$\begin{aligned}
& \frac{1}{Tn^3}\sum_{p_1,p_2,q_1,q_2=1}^{n}\varphi_u\left(q_1,q_2\right)\varphi_x\left(p_1,p_2\right)\frac{1}{T}\sum_{j=1}^{T-1}\sup_{p_1,p_2}\left|f_{x,p_1p_2}\left(j\right)\right| \\
= & \frac{1}{Tn^3}\sum_{q_1,q_2=1}^{n}\varphi_u\left(q_1,q_2\right)\sum_{p_1,p_2=1}^{n}\varphi_x\left(p_1,p_2\right) \\
= & o\left(T^{-1}\right),
\end{aligned}$$

by Lemma B.1 and $(A.7)$. Likewise the third and fourth terms of $(A.9)$ are $o_p\left(T^{-1/2}\right)$. So to complete the proof we need to examine the first term of $(A.9)$, whose second moment is, by $(A.7)$ and using $\left(\sup_{p,q}\left|f_{x,pq}\left(j\right)\right|+\sup_{p,q}\left|f_{u,pq}\left(j\right)\right|\right)\leq C$, bounded by

$$\frac{1}{Tn^3}\sum_{j=1}^{T-1}\sup_{p,q}\left|f_{x,pq}\left(j\right)\right|\left|f_{u,pq}\left(j\right)\right|\sum_{p_1,p_2=1}^{n}\varphi_x\left(p_1,p_2\right)\sum_{q_1,q_2=1}^{n}\varphi_u\left(q_1,q_2\right)=o\left(1\right).$$

This concludes the proof of the theorem. □

6.2. PROOF OF PROPOSITION 1.

We begin with part (**a**). We need to show that, for any $k_1,k_2=1,...,k$,

$$\begin{aligned}
\breve{\Phi}_{k_1,k_2} = & \frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p,k_1}\left(j\right)\mathcal{J}_{\widehat{u},p}\left(-j\right)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p,k_2}\left(-j\right)\mathcal{J}_{\widehat{u},p}\left(j\right)\right)\right\} \\
& \overset{P}{\to}\Phi_{k_1,k_2}.
\end{aligned}$$

To simplify the notation we shall assume that $k=1$. Now, after observing that

$$\mathcal{J}_{\widehat{u},p}\left(j\right)=\mathcal{J}_{\widetilde{u},p}\left(j\right)-\left(\widetilde{\beta}-\beta\right)\mathcal{J}_{\widetilde{x},p}\left(j\right),$$

we have that $\breve{\Phi}=:\breve{\Phi}_{1,1}$ is

$$\begin{aligned}
& \frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}\left(j\right)\mathcal{J}_{u,p}\left(-j\right)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}\left(-j\right)\mathcal{J}_{u,p}\left(j\right)\right)\right\} \\
& +2\left(\widetilde{\beta}-\beta\right)\frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{I}_{\widetilde{x},p}\left(j\right)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}\left(-j\right)\mathcal{J}_{u,p}\left(j\right)\right)\right\} \\
& +\left(\widetilde{\beta}-\beta\right)^2\frac{1}{T}\sum_{j=1}^{T-1}\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{I}_{\widetilde{x},p}\left(j\right)\right)^2. \qquad (A.10)
\end{aligned}$$

The third term of $(A.10)$ is $O_p\left(T^{-1}\right)$ by Lemma B.7 and $\widetilde{\beta}-\beta=O_p\left((nT)^{-1/2}\right)$. The second term of $(A.10)$ is also $o_p\left(1\right)$ by Cauchy-Schwarz inequality if we show that the first term converges in probability to Φ. Since

$$\mathcal{J}_{\widetilde{x},p}\left(j\right)=\mathcal{J}_{x,p}\left(j\right)-\mathcal{J}_{\overline{x},\cdot}\left(j\right), \qquad (A.11)$$

this result holds true if we show that

$$\frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{x,p}\left(j\right)\mathcal{J}_{u,p}\left(-j\right)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{x,p}\left(-j\right)\mathcal{J}_{u,p}\left(j\right)\right)\right\}\overset{P}{\to}\Phi \qquad (A.12)$$

and

$$
\begin{aligned}
& \frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\overline{x},\cdot}\left(j\right)\mathcal{J}_{u,p}\left(-j\right)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{x,p}\left(-j\right)\mathcal{J}_{u,p}\left(j\right)\right)\right\} \\
& +\frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\overline{x},\cdot}\left(j\right)\mathcal{J}_{u,p}\left(-j\right)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\overline{x},\cdot}\left(-j\right)\mathcal{J}_{u,p}\left(j\right)\right)\right\} \\
= \;& o_p\left(1\right). \qquad \text{(A.13)}
\end{aligned}
$$

First we examine $(A.13)$. We begin with the first term on the left of $(A.13)$, whose first moment is

$$
\begin{aligned}
& \frac{1}{T}\sum_{j=1}^{T-1}\sum_{p=1}^{n}E\left(\mathcal{J}_{\overline{x},\cdot}\left(j\right)\mathcal{J}_{x,p}\left(-j\right)\right)E\left(\mathcal{J}_{\overline{u},p}\left(-j\right)\mathcal{J}_{u,p}\left(j\right)\right) \\
= \;& \frac{C}{Tn^2}\sum_{j=1}^{T-1}\sum_{p=1}^{n}\sum_{r=1}^{n}\varphi_x\left(p,r\right)\sum_{q=1}^{n}\varphi_u\left(p,q\right)\left\{1+\frac{C}{T}\right\}.
\end{aligned}
$$

using Lemma B.1, after we observe that the factor in brackets is $n^{1/2}\mathcal{J}_{\overline{x},\cdot}\left(j\right)\mathcal{J}_{\overline{u},\cdot}\left(-j\right)$. Using $(A.6)$, we conclude that the last displayed expression is $o\left(1\right)$. Next, we observe that Lemma B.5 implies, for instance, that

$$
\begin{aligned}
& E\left(\mathcal{J}_{\overline{u},\cdot}\left(-j\right)\mathcal{J}_{u,p}\left(j\right)\mathcal{J}_{\overline{u},\cdot}\left(-k\right)\mathcal{J}_{u,q}\left(k\right)\right)-E^2\left(\mathcal{J}_{\overline{u},\cdot}\left(-j\right)\mathcal{J}_{u,p}\left(j\right)\right) \\
= \;& \varphi_u\left(p,q\right)\frac{1}{n^2}\sum_{p_1,q_1=1}^{n}\varphi_u\left(p_1,q_1\right)\left\{\mathbf{1}\left(j=k\right)+\frac{C}{T}\right\}.
\end{aligned}
$$

The variance of the first term on the left of $(A.13)$, therefore, is bounded by

$$
\frac{1}{T^2}\sum_{j,k=1}^{T-1}\sum_{p,q=1}^{n}\varphi\left(p,q\right)\frac{1}{n^4}\sum_{p_1,q_1=1}^{n}\varphi_u\left(p_1,q_1\right)\sum_{p_2,q_2=1}^{n}\varphi_x\left(p_2,q_2\right)\left\{\mathbf{1}\left(j=k\right)+\frac{C}{T}\right\}=o\left(\frac{1}{T}\right)
$$

using Condition $C3$ and $(A.7)$. Hence the first term on the left of $(A.13)$ is $o_p\left(1\right)$. The same conclusion holds true for the second term of $(A.13)$.

To complete the proof of part (**a**), it remains to show $(A.12)$. Using $(A.1)$, we have that $(A.12)$ holds true if the following expressions $(A.14)-(A.16)$ are $o_p\left(1\right)$;

$$
\begin{aligned}
& \frac{1}{nT}\sum_{j=1}^{T-1}\left\{\left(\sum_{p=1}^{n}\mathcal{B}_{x,p}\left(-j\right)\mathcal{B}_{u,p}\left(j\right)\mathcal{J}_{\chi,p}\left(j\right)\mathcal{J}_{\xi,p}\left(-j\right)\right)\right. \\
& \qquad \left.\left(\sum_{p=1}^{n}\mathcal{B}_{x,p}\left(-j\right)\mathcal{B}_{u,p}\left(j\right)\mathcal{J}_{\chi,p}\left(j\right)\mathcal{J}_{\xi,p}\left(-j\right)\right)\right\}-\Phi, \qquad \text{(A.14)}
\end{aligned}
$$

$$
\frac{1}{nT}\sum_{j=1}^{T-1}\left(\sum_{p=1}^{n}\mathcal{B}_{x,p}\left(-j\right)\mathcal{J}_{\chi,p}\left(j\right)\Upsilon_{u,p}\left(-j\right)\right)\left(\sum_{p=1}^{n}\mathcal{B}_{u,p}\left(j\right)\mathcal{J}_{\xi,p}\left(-j\right)\Upsilon_{x,p}\left(j\right)\right), \qquad \text{(A.15)}
$$

$$
\frac{1}{nT}\sum_{j=1}^{T-1}\left(\sum_{p=1}^{n}\Upsilon_{x,p}\left(j\right)\Upsilon_{u,p}\left(-j\right)\right)\left(\sum_{p=1}^{n}\Upsilon_{u,p}\left(-j\right)\Upsilon_{x,p}\left(j\right)\right) \qquad \text{(A.16)}
$$

We begin by showing that $(A.14)$ is $o_p(1)$. First, the expectation of $(A.14)$ is

$$\frac{1}{n}\sum_{p,q=1}^{n}\varphi(p,q)\frac{1}{T}\sum_{j=1}^{T-1}\mathcal{B}_{x,p}(-j)\mathcal{B}_{x,q}(j)\mathcal{B}_{u,p}(j)\mathcal{B}_{u,q}(-j)-\Phi=O\left(T^{-1}\right)$$

because, by continuous differentiability of $f_{x,pq}(-\lambda)f_{u,pq}(\lambda)$, we have that

$$\frac{1}{T}\sum_{j=1}^{T-1}\mathcal{B}_{x,p}(-j)\mathcal{B}_{x,q}(j)\mathcal{B}_{u,p}(j)\mathcal{B}_{u,q}(-j)-\int_0^{2\pi}f_{x,pq}(-\lambda)f_{u,pq}(\lambda)d\lambda=O\left(T^{-1}\right).$$

Next, because $(A.3)$ implies that

$$\begin{aligned}
& E\{(\mathcal{J}_{\chi,p_1}(j)\mathcal{J}_{\xi,p_1}(-j)\mathcal{J}_{\chi,q_1}(-j)\mathcal{J}_{\xi,q_1}(j)-E(\cdot)) \\
&\qquad (\mathcal{J}_{\chi,p_2}(-k)\mathcal{J}_{\xi,p_2}(k)\mathcal{J}_{\chi,q_2}(k)\mathcal{J}_{\xi,q_2}(-k)-E(\cdot))\} \\
= \;& \varphi_x(p_1,p_2)\varphi_x(q_1,q_2)\varphi_u(q_1,p_2)\varphi_u(p_1,q_2)\mathbf{1}(j=k) \\
& +\varphi_x(p_1,p_2)\varphi_x(q_1,q_2)\varphi_u(p_1,p_2)\varphi_u(q_1,q_2)\mathbf{1}(j=k) \\
& +2\varphi_x(p_1,p_2)\varphi_x(q_1,q_2)\sum_{\ell=1}^{\infty}c_\ell(p_1)c_\ell(p_2)c_\ell(q_1)c_\ell(q_2)\mathbf{1}(j=k) \\
& +\sum_{\ell=1}^{\infty}c_\ell(p_1)c_\ell(p_2)c_\ell(q_1)c_\ell(q_2)\sum_{\ell=1}^{\infty}d_\ell(p_1)d_\ell(p_2)d_\ell(q_1)d_\ell(q_2)\left(\mathbf{1}(j=k)+\frac{\kappa_{4,\xi}\kappa_{4,\chi}}{T}\right),
\end{aligned}$$

standard algebra yields that the second moment of $(A.14)$ is $o(1)$, when recognizing

$$\begin{aligned}
\sum_{\ell=1}^{\infty}d_\ell(p_1)d_\ell(p_2)d_\ell(q_1)d_\ell(q_2) &\leq \sum_{\ell=1}^{\infty}d_\ell(p_1)d_\ell(p_2)\sum_{\ell=1}^{\infty}d_\ell(q_1)d_\ell(q_2) \\
&= \varphi_u(p_1,p_2)\varphi_u(q_1,q_2) \qquad (A.17)
\end{aligned}$$

$$\begin{aligned}
\sum_{\ell=1}^{\infty}c_\ell(p_1)c_\ell(p_2)c_\ell(q_1)c_\ell(q_2) &\leq \sum_{\ell=1}^{\infty}c_\ell(p_1)c_\ell(p_2)\sum_{\ell=1}^{\infty}c_\ell(q_1)c_\ell(q_2) \\
&= \varphi_x(p_1,p_2)\varphi_x(q_1,q_2) \qquad (A.18)
\end{aligned}$$

and

$$\begin{aligned}
\sum_{p_1=1}^{n}\varphi_x(p_1,p_2)\varphi_u(p_1,q_2) &\leq \left(\sum_{p_1=1}^{n}\varphi_x^{1/\alpha}(p_1,p_2)\right)^{\alpha}\left(\sum_{p_1=1}^{n}\varphi_u^{1/1-\alpha}(p_1,q_2)\right)^{1-\alpha} \\
&= O(1) \qquad (A.19)
\end{aligned}$$

since $\sum_{p_1=1}^{n}\varphi_x(p_1,p_2)\varphi_u(p_1,p_2)=O(1)$ implies $\varphi_x(p_1,p_2)=O\left(p_1^{-\alpha}\right)$ and $\varphi_u(p_1,p_2)=O\left(p_1^{-\beta}\right)$ with $\alpha+\beta>1$.

Next consider $(A.15)$. Because $\sup_p|\mathcal{B}_{x,p}(-j)\mathcal{B}_{u,p}(j)|<C$, the second moment of $(A.15)$ is bounded by

$$\begin{aligned}
&\frac{1}{(nT)^2}\sum_{j,k=1}^{T-1}\sum_{p_1,q_1,p_2,q_2=1}^{n}|E\{\mathcal{J}_{\chi,p_1}(j)\mathcal{J}_{\chi,q_1}(-k)\mathrm{Y}_{x,p_2}(j)\mathrm{Y}_{x,q_2}(-k)\} \\
&E\{\mathrm{Y}_{u,p_1}(-j)\mathrm{Y}_{u,q_1}(k)\mathcal{J}_{\xi,p_2}(-j)\mathcal{J}_{\xi,q_2}(k)\}|.
\end{aligned}$$

From here, proceeding as with $(A.14)$ but using Lemmas B.1 and B.2 as needed, we easily conclude that $(A.15) = o_p(1)$ by Markov's inequality, since for instance

$$\begin{aligned}
& E\left\{\mathcal{J}_{\chi,p_1}(j)\,\mathcal{J}_{\chi,q_1}(k)\,\Upsilon_{x,p_2}(-j)\,\Upsilon_{x,q_2}(-k)\right\} \\
= \; & E\left(\mathcal{J}_{\chi,p_1}(j)\,\mathcal{J}_{\chi,q_1}(k)\right) E\left(\Upsilon_{x,p_2}(-j)\,\Upsilon_{x,q_2}(-k)\right) \\
& + E\left(\mathcal{J}_{\chi,p_1}(j)\,\Upsilon_{x,p_2}(-j)\right) E\left(\mathcal{J}_{\chi,q_1}(k)\,\Upsilon_{x,q_2}(-k)\right) \\
& + E\left(\mathcal{J}_{\chi,p_1}(j)\,\Upsilon_{x,q_2}(-k)\right) E\left(\mathcal{J}_{\chi,q_1}(k)\,\Upsilon_{x,p_2}(-j)\right) \\
& + cum\left(\mathcal{J}_{\chi,p_1}(j)\,;\mathcal{J}_{\chi,q_1}(k)\,;\Upsilon_{x,p_2}(-j)\,;\Upsilon_{x,q_2}(-k)\right).
\end{aligned}$$

The proof of part (**a**) now concludes since $(A.16) = o_p(1)$ by standard algebra and Lemmas B.1 and B.2.

Part (**b**). Because the continuous differentiability of $f_{x,p}(\lambda)$, we have that $T^{-1}\sum_{j=1}^{T} f_{x,p}(j) \to \int_0^{2\pi} f_{x,p}(\lambda)\,d\lambda =: \Sigma_{x,p}$, see Brillinger (1981, p. 15), so we can conclude by Lemma B.6 and $(A.11)$, that to finish the proof, it suffices to show that

$$\frac{1}{nT}\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{I}_{\overline{x},\cdot}(j) \text{ and } \frac{2}{nT}\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{J}_{\overline{x},\cdot}(j)\,\mathcal{J}_{x,p}(j) = o_p(1).$$

are both $o_p(1)$. However this is the case proceeding similarly as with the proof of $(A.13)$, so it is omitted. □

6.3. PROOF OF THEOREM 2.

Because Lemma B.7 implies that $(nT)^{-1}\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{I}_{\widetilde{x},p}(j) \xrightarrow{P} \Sigma_x$ and abbreviating $\widehat{f}_u(j) = \frac{1}{n}\sum_{q=1}^{n}\mathcal{I}_{\widehat{u},q}(j)$, it suffices to show

$$(\mathbf{i}) \quad \frac{1}{T^{1/2}n^{1/2}}\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{J}_{\widetilde{x},p}(j)\left(\widehat{f}_u^{1/2}(j) - f_u^{1/2}(j)\right)\mathcal{J}_{u^*,p}(-j) = o_{p^*}(1) \tag{A.20}$$

$$(\mathbf{ii}) \quad \frac{1}{T^{1/2}n^{1/2}}\sum_{p=1}^{n}\sum_{j=1}^{T-1}\mathcal{J}_{\widetilde{x},p}(\lambda_j)\,f_u^{1/2}(j)\,\mathcal{J}_{u^*,p}(-j) \xrightarrow{d^*} \mathcal{N}(0,\Phi) \tag{A.21}$$

We begin with part (**ii**). The left hand side of $(A.21)$ is

$$\begin{aligned}
& \frac{1}{T^{1/2}n^{1/2}}\sum_{p=1}^{n}\sum_{j=1}^{T-1} f_u^{1/2}(j)\,\mathcal{B}_{x,p}(j)\,\mathcal{J}_{\chi,p}(j)\,\mathcal{J}_{u^*,p}(-j) \qquad (A.22)\\
& + \frac{1}{T^{1/2}n^{1/2}}\sum_{p=1}^{n}\sum_{j=1}^{T-1} f_u^{1/2}(j)\left(\mathcal{J}_{\widetilde{x},p}(j) - \mathcal{B}_{x,p}(j)\,\mathcal{J}_{\chi,p}(j)\right)\mathcal{J}_{u^*,p}(-j).
\end{aligned}$$

The second (bootstrap) moment of the second term of $(A.22)$ is

$$\frac{1}{nT}\sum_{p,q=1}^{n}\sum_{j=1}^{T-1} f_u(j)\,\widehat{\sigma}_{u,pq}\left(\mathcal{J}_{\widetilde{x},p}(j) - \mathcal{B}_{x,p}(j)\,\mathcal{J}_{\chi,p}(j)\right)\left(\mathcal{J}_{\widetilde{x},q}(-j) - \mathcal{B}_{x,q}(-j)\,\mathcal{J}_{\chi,q}(-j)\right) \tag{A.23}$$

using

$$E^*\left(\mathcal{J}_{u^*,p}(j)\,\mathcal{J}_{u^*,q}(-k)\right) = \widehat{\sigma}_{u,pq}\mathbf{1}(j=k)\,; \qquad \widehat{\sigma}_{u,pq} = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{pt}\widehat{u}_{qt}, \tag{A.24}$$

By Lemma B.1 and $(A.1)$,

$$E\left(\mathcal{J}_{\widetilde{x},p}(j) - \mathcal{B}_{x,p}(j)\mathcal{J}_{\chi,p}(j)\right)\left(\mathcal{J}_{\widetilde{x},q}(-j) - \mathcal{B}_{x,p}(-j)\mathcal{J}_{\chi,p}(-j)\right) = \frac{C}{T}\varphi_x(p,q);$$
$$\widehat{\sigma}_{u,pq} = \varphi_u(p,q)\left(1 + O_p\left(T^{-1/2}\right)\right).$$

Hence it easily follows that the expected value of equation $(A.23)$ is $o(1)$ and consequently the second term of $(A.22)$ is $o_{p^*}(1)$, after we observe that $(A.23)$ is a nonnegative expression.

Turning to the first term of $(A.22)$, let us denote

$$\Xi^*_{s,t}(n) = \frac{1}{n^{1/2}}\sum_{p=1}^{n}\chi_{ps}u^*_{pt}; \qquad \mathcal{G}(j) =: \mathcal{B}_{x,p}(j)\, f^{1/2}_{u,p}(j). \tag{A.25}$$

Standard algebra yields that the first term of $(A.22)$ is

$$\frac{1}{\widetilde{T}^{1/2}}\frac{1}{T}\sum_{t,s=1}^{T}\Xi^*_{s,t}(n)\sum_{j=1}^{\widetilde{T}}\mathcal{G}(j)\, e^{i(t-s)\lambda_j} = \frac{1}{T^{1/2}}\sum_{t,s=1}^{T}\phi(|t-s|)\,\Xi^*_{s,t}(n) + \frac{C}{T^{3/2}}\sum_{t,s=1}^{T}\Xi^*_{s,t}(n), \tag{A.26}$$

where to simplify the notation we assume that $\varphi_x(p,p) = \varphi_u(p,p) = 1$ for all $p = 1,...,n$ and $\phi(r)$ is the rth Fourier coefficient of $\mathcal{G}(j)$. Hence the right hand side of $(A.26)$ can now be written as

$$\frac{\phi(0)}{T^{1/2}}\sum_{t=1}^{T-\ell}\frac{1}{n^{1/2}}\sum_{p=1}^{n}\chi_{pt}u^*_{pt} + \sum_{\ell=1}^{T-1}\frac{\phi(\ell)}{T^{1/2}}\sum_{t=1}^{T-\ell}\frac{1}{n^{1/2}}\left\{\sum_{p=1}^{n}\chi_{pt}u^*_{p,t+\ell} + \sum_{p=1}^{n}\chi_{p,t+\ell}u^*_{pt}\right\}. \tag{A.27}$$

Because $\phi(r) = O\left(r^{-2}\right)$ by Conditions $C1$ and $C2$, given the independence of the sequences of random variables $n^{-1/2}\sum_{p=1}^{n}\chi_{pt}u^*_{p,t+\ell}$ and $n^{-1/2}\sum_{p=1}^{n}\chi_{p,t+\ell}u^*_{pt}$ in t, to complete the proof of part (**ii**), it suffices to show that

$$\Lambda^*_{t,n} =: \frac{1}{n^{1/2}}\sum_{p=1}^{n}\chi_{pt}u^*_{p,t+\ell} \overset{d^*}{\to} \mathcal{N}\left(0, \frac{T-\ell}{T}\lim_{n\to\infty}\frac{1}{n}\sum_{p,q=1}^{n}\varphi(p,q)\right).$$

The second bootstrap moment of $\Lambda^*_{t,n}$ is

$$\frac{1}{n}\sum_{p,q=1}^{n}\chi_{pt}\chi_{qt}\frac{1}{T}\sum_{r=1}^{T-\ell}\widehat{u}_{p,r+\ell}\widehat{u}_{q,r+\ell} = \frac{1}{n}\sum_{p,q=1}^{n}\chi_{pt}\chi_{qt}\frac{1}{T}\sum_{r=1}^{T-\ell}u_{p,r+\ell}u_{q,r+\ell}\left(1 + o_p(1)\right),$$

by standard algebra and Theorem 1. Now, Conditions $C1$ and $C2$ imply that

$$\frac{1}{n}\sum_{p,q=1}^{n}\left\{E\left(\chi_{pt}\chi_{qt}\right)\frac{1}{T}\sum_{r=1}^{T-\ell}E\left(u_{p,r+\ell}u_{q,r+\ell}\right)\right\} = \frac{T-\ell}{T}\frac{1}{n}\sum_{p,q=1}^{n}\varphi(p,q).$$

Moreover, because $E\left(u_{p_1,t+\ell}u_{q_1,t+\ell}u_{p_2,s+\ell}u_{q_2,s+\ell}\right)=E\left(u_{p_1t}u_{q_1t}u_{p_2s}u_{q_2s}\right)$

$$
\begin{aligned}
& E\left(\frac{1}{n}\sum_{p,q=1}^{n}\chi_{pt}\chi_{qt}\frac{1}{T}\sum_{t=1}^{T-\ell}u_{pt}u_{qt}\right)^2 \\
= & \frac{1}{n^2}\sum_{p_1,q_1,p_2,q_2=1}^{n}E\left(\chi_{p_1t}\chi_{q_1t}\chi_{p_2t}\chi_{q_2t}\right)\frac{1}{T^2}\sum_{t,s=1}^{T-\ell}E\left(u_{p_1t}u_{q_1t}u_{p_2s}u_{q_2s}\right) \\
= & \frac{1}{n^2}\sum_{p_1,q_1,p_2,q_2=1}^{n}\frac{1}{T^2}\sum_{t,s=1}^{T-\ell}\left\{E\left(\chi_{p_1t}\chi_{q_1t}\right)E\left(\chi_{p_2t}\chi_{q_2t}\right)+E\left(\chi_{p_1t}\chi_{q_2t}\right)E\left(\chi_{p_2t}\chi_{q_1t}\right)\right. \\
& \left.+E\left(\chi_{p_1t}\chi_{p_2t}\right)E\left(\chi_{q_1t}\chi_{q_2t}\right)+cum\left(\chi_{p_1t};\chi_{q_1t};\chi_{p_2t};\chi_{q_2t}\right)\right\} \\
& \times\left\{E\left(u_{p_1t}u_{q_1t}\right)E\left(u_{p_2s}u_{q_2s}\right)+E\left(u_{p_1t}u_{q_2s}\right)E\left(u_{p_2s}u_{q_1t}\right)\right. \\
& \left.+E\left(u_{p_1t}u_{p_2s}\right)E\left(u_{q_1t}u_{q_2s}\right)+cum\left(u_{p_1t};u_{q_1t};u_{p_2s};u_{q_2s}\right)\right\} \\
= & \frac{1}{n^2T^2}\sum_{p_1,q_1,p_2,q_2=1}^{n}\sum_{t,s=1}^{T-\ell}E\left(\chi_{p_1t}\chi_{q_1t}\right)E\left(\chi_{p_2t}\chi_{q_2t}\right)E\left(u_{p_1t}u_{q_1t}\right)E\left(u_{p_2s}u_{q_2s}\right)(1+o(1)) \\
= & \left(\frac{T-\ell}{T}\frac{1}{n}\sum_{p,q=1}^{n}\varphi(p,q)\right)^2(1+o(1))
\end{aligned}
$$

because $E\left(u_{ps}u_{qr}\right)=\varphi_u(p,q)\gamma_{u,pq}(r-s)$, $\sum_{r,s=1}^{T}\left|\gamma_{u,pq}(r-s)\right|=O(T)$ and $(A.19)$. This shows that the second moment converges to the square of the first moment, and hence $E^*\left|\Lambda_{t,n}^*\right|^2-\frac{T-\ell}{T}\frac{1}{n}\sum_{p,q=1}^{n}\varphi(p,q)=o_p(1)$.

Thus, it remains to show the Lindeberg's condition to complete the proof of part (**ii**). To that end, it suffices to show that

$$
\frac{1}{n^2}\sum_{p=1}^{n}E^*\left(\chi_{pt}u_{p,t+\ell}^*\right)^4=o_p(1)\,.
$$

The left hand side of the last displayed expression is

$$
\begin{aligned}
\frac{1}{n^2}\sum_{p=1}^{n}\left\|\chi_{pt}\right\|^4\frac{1}{T}\sum_{t=1}^{T-\ell}\widehat{u}_{p,t+\ell}^4 &= \frac{1}{n^2}\sum_{p=1}^{n}\left\|\chi_{pt}\right\|^4\frac{1}{T}\sum_{t=1}^{T-\ell}u_{p,t+\ell}^4(1+o_p(1)) \\
&= O_p\left(n^{-1}\right),
\end{aligned}
$$

which completes the proof of part (**ii**).

Next we prove part (**i**). The left side of $(A.20)$ is

$$
\begin{aligned}
& \frac{1}{T^{1/2}n^{1/2}}\sum_{p=1}^{n}\sum_{j=1}^{T-1}\left(\widehat{f}_u^{1/2}(j)-f_u^{1/2}(j)\right)\mathcal{B}_{x,p}(j)\,\mathcal{J}_{\chi,p}(j)\,\mathcal{J}_{u^*,p}(-j) \qquad (A.28)\\
& +\frac{1}{T^{1/2}n^{1/2}}\sum_{p=1}^{n}\sum_{j=1}^{T-1}\left(\widehat{f}_u^{1/2}(j)-f_u^{1/2}(j)\right)\left(\mathcal{J}_{\widetilde{x},p}(j)-\mathcal{B}_{x,p}(j)\,\mathcal{J}_{\chi,p}(j)\right)w_{u^*,p}(-j)\,.
\end{aligned}
$$

We shall only show explicitly that the first term of $(A.28)$ is $o_{p^*}(1)$, the second term following similarly if not easier proceeding as with the second term of $(A.22)$ and Lemma B.1. Now by

$(A.24)$, the first term of $(A.28)$ has second bootstrap moment given by

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{nT}\sum_{j=1}^{T-1}\left\{\widehat{f}_u^{1/2}(j)-f_u^{1/2}(j)\right\}^2 f_x(j)\sum_{p,q=1}^{n}\widehat{u}_{pt}\widehat{u}_{qt}\mathcal{J}_{\chi,p}(j)\mathcal{J}_{\chi,q}(-j).$$

Because the last displayed expression is a nonnegative expression, to show that it is $o_p(1)$, it suffices to show that its first moment converges to zero. To that end, we first observe that

$$\left\{\widehat{f}_u^{1/2}(j)-f_u^{1/2}(j)\right\}^2 \le \left|\frac{1}{n}\sum_{q=1}^{n}\mathcal{I}_{\widehat{u},q}(j)-f_u(j)\right| = o_p(1) \tag{A.29}$$

using standard arguments and Theorem 1 under Condition $C4$. On the other hand, proceeding similarly as in Proposition 1, we obtain easily that

$$\frac{1}{n}\sum_{p,q=1}^{n}\widehat{u}_{pt}\mathcal{J}_{\chi,p}(j)\widehat{u}_{qt}\mathcal{J}_{\chi,q}(-j)=\frac{1}{n}\sum_{p,q=1}^{n}u_{pt}\mathcal{J}_{\chi,p}(j)u_{qt}\mathcal{J}_{\chi,q}(-j)(1+o_p(1)),$$

and thus the proof of part (**i**), and thereby the theorem, is completed if

$$E\left(\sum_{p,q=1}^{n}u_{pt}u_{qt}\mathcal{J}_{\chi,p}(j)\mathcal{J}_{\chi,q}(-j)\right)=O(n).$$

But the left hand side of the last displayed expression is

$$\sum_{p,q=1}^{n}\varphi_u(p,q)\frac{1}{T}\sum_{t,s=1}^{T}E(x_{pt}x_{qs})e^{-i(t-s)\lambda_j}=C\sum_{p,q=1}^{n}\varphi_u(p,q)\varphi_x(p,q)=O(n)$$

by Condition $C3$, which completes the proof of the theorem. □

6.4. PROOF OF PROPOSITION 2.

As with the proof of Proposition 1, we shall assume that $k=1$. Now, after observing that

$$\mathcal{J}_{u^*p}(j)=\mathcal{J}_{\widetilde{u}^*,p}(j)-\left(\widetilde{\beta}^*-\widetilde{\beta}\right)\mathcal{J}_{\widetilde{x},p}(j),$$

we have that $\breve{\Phi}^*$ equals the sum of the following expressions $(A.30)-(A.32)$;

$$\frac{1}{T}\sum_{j=1}^{T-1}\widehat{f}_u(j)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}(j)\mathcal{J}_{u^*,p}(-j)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}(-j)\mathcal{J}_{u^*,p}(j)\right)-\breve{\Phi} \tag{A.30}$$

$$2\left(\widetilde{\beta}^*-\widetilde{\beta}\right)\frac{1}{T}\sum_{j=1}^{T-1}\widehat{f}_u^{1/2}(j)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{I}_{\widetilde{x},p}(j)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}(-j)\mathcal{J}_{u^*,p}(j)\right) \tag{A.31}$$

$$\left(\widetilde{\beta}^*-\widetilde{\beta}\right)^2\frac{1}{T}\sum_{j=1}^{T-1}\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{I}_{\widetilde{x},p}(j)\right)^2. \tag{A.32}$$

That $(A.32)$ is $o_{p^*}(1)$ follows straightforwardly by Theorem 2 and Lemma B.7 and $(A.31)$ is $o_{p^*}(1)$ by Cauchy-Schwarz inequality if we show that $(A.30)$ is $o_{p^*}(1)$. To that end, using $(A.11)$

and $(A.24)$, we have

$$E^{*}(A.30) = \frac{1}{nT}\sum_{j=1}^{T-1}\widehat{f}_u(j)\sum_{p,q=1}^{n}\mathcal{J}_{x,p}(j)\mathcal{J}_{x,q}(-j)\widehat{\sigma}_{u,pq}-\breve{\Phi} + \frac{1}{nT}\sum_{j=1}^{T-1}\widehat{f}_u(j)\sum_{p,q=1}^{n}\mathcal{J}_{\overline{x},\cdot}(j)\mathcal{J}_{\overline{x},\cdot}(-j)\widehat{\sigma}_{u,pq}.$$

Because $\widehat{\sigma}_{u,pq}=\varphi_u(p,q)(1+o_p(1))$ and $\breve{\Phi}-\Phi=o_p(1)$ by Proposition 1, proceeding as in the proof of Theorem 2 part (**i**), it suffices to examine the behaviour of

$$\frac{1}{nT}\sum_{j=1}^{T-1}f_u(j)\sum_{p,q=1}^{n}\{\varphi_u(p,q)\mathcal{J}_{x,p}(j)\mathcal{J}_{x,q}(-j)\}-\Phi \tag{A.33}$$

$$+\frac{1}{T}\sum_{j=1}^{T-1}f_u(j)\mathcal{J}_{\overline{x},\cdot}(j)\mathcal{J}_{\overline{x},\cdot}(-j)\frac{1}{n}\sum_{p,q=1}^{n}\varphi_u(p,q). \tag{A.34}$$

$(A.34)$ is $o_p(1)$ as we now show. As it is a nonnegative sequence, it suffices to show that its first mean converges to zero. Using $(A.1)$ and then Lemmas B.1 and B.2, we have that its first moment is proportional to

$$\frac{1}{n^2}\sum_{p,q=1}^{n}\varphi_x(p,q)\frac{1}{n}\sum_{p,q=1}^{n}\varphi_u(p,q)=o(1)$$

by $(A.6)$. Because the first moment of $(A.33)$ is $o(1)$, it then remains to show that the (bootstrap) variance of $(A.30)$, with $\mathcal{J}_{\widetilde{x},p}(j)$ replaced by $\mathcal{J}_{x,p}(j)$, converges to zero. Using $(A.24)$, the (bootstrap) variance is

$$\frac{1}{T^2}\sum_{j=1}^{T-1}\widehat{f}_u^2(j)\left(\frac{1}{n^2}\sum_{p_1,q_1p_2,q_2=1}^{n}\mathcal{J}_{x,p_1}(j)\mathcal{J}_{x,q_1}(-j)\mathcal{J}_{x,p_2}(-j)\mathcal{J}_{x,q_2}(j)\widehat{\sigma}_{u,p_1p_2}\widehat{\sigma}_{u,q_1q_2}\right)$$
$$+\frac{\kappa_{4,\xi}(1+o_p(1))}{T^3n^2}\sum_{j,k=1}^{T-1}\left\{\widehat{f}_u(j)\widehat{f}_u(k)\right.$$
$$\left.\times\sum_{p_1,q_1p_2,q_2=1}^{n}\varphi_u(p_1,q_1)\varphi_u(p_2,q_2)\mathcal{J}_{x,p_1}(j)\mathcal{J}_{x,q_1}(-j)\mathcal{J}_{x,p_2}(-k)\mathcal{J}_{x,q_2}(k)\right\},$$

with Lemma B.4 guaranteeing

$$cum^{*}\left(u^{*}_{p_1t},u^{*}_{q_1t},u^{*}_{p_2t},u^{*}_{q_2t}\right)=\kappa_{4,\xi}\varphi_u(p_1,q_1)\varphi_u(p_2,q_2)(1+o_p(1)).$$

From here we proceed as before after noticing that $\widehat{\sigma}_{u,p_1p_2}=\varphi_u(p_1,p_2)(1+o_p(1))$. This completes the proof of the proposition. □

6.5. PROOF OF PROPOSITION 3.

As with the proof of Theorem 2, it suffices to show that

$$\frac{1}{T^{1/2}n^{1/2}}\sum_{j=1}^{T-1}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}(j)\mathcal{J}_{\widehat{u},p}(-j)\eta_j\overset{d^*}{\to}\mathcal{N}(0,\Phi). \tag{A.35}$$

Because η_j are normally distributed it suffices to show

$$E^*\left(\frac{1}{T^{1/2}n^{1/2}}\sum_{j=1}^{T-1}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}(j)\,\mathcal{J}_{\widehat{u},p}(-j)\,\eta_j\right)^2 \xrightarrow{P} \Phi.$$

This is the case as we now show. The left hand side of the last displayed expression is

$$\begin{aligned}
&\frac{1}{nT}\sum_{j=1}^{T-1}\sum_{p,q=1}^{n}\mathcal{J}_{\widetilde{x},p}(j)\,\mathcal{J}_{\widetilde{x},q}(-j)\,\mathcal{J}_{\widehat{u},p}(-j)\,\mathcal{J}_{\widehat{u},q}(j)\\
=\;&\frac{1}{nT}\sum_{j=1}^{T-1}\sum_{p,q=1}^{n}\mathcal{J}_{\widetilde{x},p}(j)\,\mathcal{J}_{\widetilde{x},q}(-j)\,\mathcal{J}_{u,p}(-j)\,\mathcal{J}_{u,q}(j)+o_p(1)
\end{aligned}$$

as $\widehat{u}_{pt}-u_{pt}=\left(\widetilde{\beta}-\beta\right)x_{pt}$ and $\widetilde{\beta}-\beta=O_p\left(T^{-1/2}n^{-1/2}\right)$. Using $(A.11)$ and proceeding as in the proof of part (**a**) of Proposition 1, we now have that the right hand side is

$$\begin{aligned}
&\frac{1}{nT}\sum_{j=1}^{T-1}\sum_{p,q=1}^{n}\mathcal{J}_{x,p}(j)\,\mathcal{J}_{x,q}(-j)\,\mathcal{J}_{u,p}(-j)\,\mathcal{J}_{u,q}(-j)\\
&+\frac{2}{nT}\sum_{j=1}^{T-1}\sum_{p,q=1}^{n}\mathcal{J}_{x,p}(j)\,\mathcal{J}_{\overline{x},\cdot}(-j)\,\mathcal{J}_{u,p}(-j)\,\mathcal{J}_{u,q}(-j)\\
&+\frac{1}{nT}\sum_{j=1}^{T-1}\sum_{p,q=1}^{n}\mathcal{J}_{\overline{x},\cdot}(j)\,\mathcal{J}_{\overline{x},\cdot}(-j)\,\mathcal{J}_{u,p}(-j)\,\mathcal{J}_{u,q}(-j)+o_p(1)\,.
\end{aligned}$$

The first term converges in probability to Φ, whereas the second term follows by Cauchy-Schwarz inequality if the third term is also $o_p(1)$. But that term is $o_p(1)$ proceeding as in the proof of part (**a**) of Proposition 1 using Lemma B.5. Again observe that the expression is nonnegative. This concludes the proof. □

Appendix B: LEMMAS

First denoting $\Upsilon_{\ell,p}(j)=\left\{\sum_{t=1-\ell}^{T-\ell}-\sum_{t=1}^{T}\right\}\xi_{pt}e^{-it\lambda_j}$ and $\Psi_{\ell,p}(j)=\left\{\sum_{t=1-\ell}^{T-\ell}-\sum_{t=1}^{T}\right\}\chi_{pt}e^{-it\lambda_j}$, we have that $\mathrm{Y}_{u,p}(j)$ and $\mathrm{Y}_{x,p}(j)$ given in $(A.2)$ can be decomposed as

$$\begin{aligned}
\mathrm{Y}_{u,p}(j) &= \mathrm{Y}^{(1)}_{u,p}(j)+\mathrm{Y}^{(2)}_{u,p}(j) \qquad\qquad (B.1)\\
\mathrm{Y}_{x,p}(j) &= \mathrm{Y}^{(1)}_{x,p}(j)+\mathrm{Y}^{(2)}_{x,p}(j)\,,
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{Y}^{(1)}_{u,p}(j) &= \frac{1}{T^{1/2}}\sum_{\ell=0}^{T}d_\ell(p)\,e^{-i\ell\lambda_j}\Upsilon_{\ell,p}(j)\,;\quad \mathrm{Y}^{(2)}_{u,p}(j)=\frac{1}{T^{1/2}}\sum_{\ell=T+1}^{\infty}d_\ell(p)\,e^{-i\ell\lambda_j}\Upsilon_{\ell,p}(j)\\
\mathrm{Y}^{(1)}_{x,p}(j) &= \frac{1}{T^{1/2}}\sum_{\ell=0}^{T}c_\ell(p)\,e^{-i\ell\lambda_j}\Psi_{\ell,p}(j)\,;\quad \mathrm{Y}^{(2)}_{x,p}(j)=\frac{1}{T^{1/2}}\sum_{\ell=T+1}^{\infty}c_\ell(p)\,e^{-i\ell\lambda_j}\Psi_{\ell,p}(j)\,.
\end{aligned}$$

Lemma B.1. *Assuming $C1$ and $C2$, we have that for $p,q=1,..,n$ and some $\upsilon_u,\upsilon_x>0$ finite,*

$$E\left(\mathrm{Y}^{(1)}_{w,p}(j)\,\mathrm{Y}^{(1)}_{w,q}(-k)\right) = \frac{\upsilon_w\varphi_w(p,q)}{T};\qquad w=:u\text{ or }x \qquad (B.2)$$

$$E\left(\mathrm{Y}^{(2)}_{w,p}(j)\,\mathrm{Y}^{(2)}_{w,q}(-k)\right) = o\left(T^{-2}\right)\varphi_w(p,q)\,\mathbf{1}(j=k)\,;\qquad w=:u\text{ or }x. \qquad (B.3)$$

Proof. We examine only the case when $w =: u$, with the proof for $w =: x$ similarly handled. We begin with $(B.3)$. Because for $\ell \geq T$, $E\left(\Upsilon_{\ell,p}(j)\Upsilon_{\ell,q}(-k)\right) = 2T\varphi_u(p,q)\mathbf{1}(j=k)$, we obtain that the left hand side of $(B.3)$ is

$$2\sum_{\ell_1,\ell_2=T+1}^{\infty} d_{\ell_1}(p)\,d_{\ell_2}(q)\,\varphi_u(p,q)\,\mathbf{1}(j=k)\,.$$

The conclusion then follows because Condition $C1$ implies that $\sum_{\ell=T+1}^{\infty}\sup_p|d_\ell(p)| = o\left(T^{-1}\right)$.

Next we consider $(B.2)$. By definition, the left side is

$$\frac{1}{T}\sum_{\ell_1,\ell_2=0}^{T} d_{\ell_1}(p)\,d_{\ell_2}(q)\,E\left(\Upsilon_{\ell,p}(j)\Upsilon_{\ell,q}(-k)\right) = \varphi_u(p,q)\frac{v_u}{T}$$

since $\Upsilon_{\ell,p}(j) = \left\{\sum_{t=1-\ell}^{0} - \sum_{t=T-\ell+1}^{T}\right\}\xi_{pt}e^{it\lambda_j}$ when $\ell \leq T$, so that

$$E\left(\Upsilon_{\ell,p}(j)\Upsilon_{\ell,q}(-k)\right) = 2\varphi_u(p,q)\sum_{t=1}^{\ell} e^{it(\lambda_j-\lambda_k)}.$$

We now conclude because $\sum_{\ell=0}^{\infty}\ell\sup_p|d_\ell(p)| < \infty$ by Condition $C1$. □

Lemma B.2. *Assuming $C1$ and $C2$, we have that for $p,q = 1,..,n$,*

$$\begin{aligned}
(\mathbf{a})\quad E\left(\Upsilon^{(1)}_{u,p}(j)\,\mathcal{J}_{\xi,q}(-k)\right) &= \varphi_u(p,q)\frac{1}{T}\sum_{\ell=0}^{T} d_\ell(p)\,e^{-i\ell\lambda_j}\sum_{t=1}^{\ell} e^{it\lambda_{j-k}}\\
E\left(\Upsilon^{(2)}_{u,p}(j)\,\mathcal{J}_{\xi,q}(-k)\right) &= \varphi_u(p,q)\,\mathbf{1}(j=k)\,o\left(T^{-2}\right)\\
(\mathbf{b})\quad E\left(\Upsilon^{(1)}_{x,p}(j)\,\mathcal{J}_{\chi,q}(-k)\right) &= \varphi_x(p,q)\frac{1}{T}\sum_{\ell=0}^{T} c_\ell(p)\,e^{-i\ell\lambda_j}\sum_{t=1}^{\ell} e^{it\lambda_{j-k}}\\
E\left(\Upsilon^{(2)}_{x,p}(j)\,\mathcal{J}_{\chi,q}(-k)\right) &= \varphi_x(p,q)\,\mathbf{1}(j=k)\,o\left(T^{-2}\right).
\end{aligned}$$

Proof. As in the proof of Lemma B.1 we shall only show part (**a**). To that end, we first notice that Condition $C1$ implies that

$$E\left(\Upsilon_{\ell,p}(j)\,\mathcal{J}_{\xi,q}(-k)\right) = \frac{\varphi_u(p,q)}{T^{1/2}}\left(\mathbf{1}(j=k)\,\mathbf{1}(\ell\geq T) + \sum_{t=T-\ell+1}^{T} e^{it\lambda_{j-k}}\mathbf{1}(\ell<T)\right).$$

From here the proof concludes by standard algebra. □

Lemma B.3. *Assuming $C1$ and $C2$, we have that*

$$\begin{aligned}
\left|cum\left(\xi_{p_1t};\xi_{p_2t};\xi_{p_3t};\xi_{p_4t}\right)\right| &\leq |\kappa_{4,\xi}|\,\varphi_u(p_1,p_2)\,\varphi_u(p_3,p_4)\\
\left|cum\left(\chi_{p_1t};\chi_{p_2t};\chi_{p_3t};\chi_{p_4t}\right)\right| &\leq |\kappa_{4,\chi}|\,\varphi_x(p_1,p_2)\,\varphi_x(p_3,p_4)
\end{aligned}\qquad\text{(B.4)}$$

Proof. Using inequality $(A.17)$, the proof follows easily since by definition

$$cum\left(\xi_{p_1t};\xi_{p_2t};\xi_{p_3t};\xi_{p_4t}\right) = \kappa_{4,\xi}\sum_{\ell=1}^{\infty} a_\ell(p_1)\,a_\ell(p_2)\,a_\ell(p_3)\,a_\ell(p_4)\,.$$

The proof is similar for the second expression in $(B.4)$, where inequality $(A.18)$ is used instead of $(A.17)$. □

Lemma B.4. *Assuming $C1$ and $C2$, for some $\tau > 2$,*

$$\begin{aligned} |cum(u_{p_1t_1}; u_{p_2t_2}; u_{p_3t_3}; u_{p_4t_4})| &\leq C\frac{|\kappa_{4,\xi}|\,\varphi_u(p_1,p_2)\,\varphi_u(p_3,p_4)}{(t_2-t_1)^\tau (t_3-t_1)^\tau (t_4-t_1)^\tau} \\ |cum(x_{p_1t_1}; x_{p_2t_2}; x_{p_3t_3}; x_{p_4t_4})| &\leq C\frac{|\kappa_{4,\chi}|\,\varphi_x(p_1,p_2)\,\varphi_x(p_3,p_4)}{(t_2-t_1)^\tau (t_3-t_1)^\tau (t_4-t_1)^\tau}. \end{aligned}$$

Proof. As in the proof of Lemma B.3, we handle the first displayed inequality only. Without loss of generality we take $t_1 \leq t_2 \leq t_3 \leq t_4$. Condition $C1$ and the definition of the fourth cumulant then yield that

$$\begin{aligned} cum(u_{p_1t_1}; u_{p_2t_2}; u_{p_3t_3}; u_{p_4t_4}) &= \sum_{k=1}^{\infty} d_k(p_1)\, d_{k+t_2-t_1}(p_2)\, d_{k+t_3-t_1}(p_3)\, d_{k+t_4-t_1}(p_4) \\ &\quad \times cum(\xi_{p_1t}; \xi_{p_2t}; \xi_{p_3t}; \xi_{p_4t}). \end{aligned}$$

From here we conclude using Lemma B.3 and the fact that Condition $C1$ implies that $\sup_p |d_k(p)| = O(k^{-\tau})$ for some $\tau > 2$. □

Lemma B.5. *Assuming $C1$ and $C2$, we have that for $w =: u$ or x,*

$$E(\mathcal{J}_{w,p_1}(j)\,\mathcal{J}_{w,p_2}(-k)) = f_{w,p_1p_2}(j)\,\varphi_w(p_1,p_2)\left\{\mathbf{1}(j=k) + \frac{C}{T}\right\} \tag{B.5}$$

and

$$\begin{aligned} & E(\mathcal{J}_{w,p_1}(j)\,\mathcal{J}_{w,p_2}(-j)\,\mathcal{J}_{w,p_3}(k)\,\mathcal{J}_{w,p_4}(-k)) \\ &= \varphi_w(p_1,p_2)\,\varphi_w(p_3,p_4)\left\{1+\mathbf{1}(j=k) + \frac{C}{T}\right\}. \end{aligned} \tag{B.6}$$

Proof. Consider $w =: u$, say. By $(A.1)$, we have that the left hand side of $(B.5)$ is

$$E((\mathcal{B}_{u,p_1}(-j)\,\mathcal{J}_{\xi,p_1}(j) + \Upsilon_{u,p_1}(j))(\mathcal{B}_{u,p_2}(k)\,\mathcal{J}_{\xi,p_2}(-k) + \Upsilon_{u,p_2}(-k))),$$

which using $(A.3)$ equals the right hand side of $(B.5)$ by Lemmas B.1 and B.2.

Next, the left hand side of $(B.6)$ is

$$\begin{aligned} & E(\mathcal{J}_{u,p_1}(j)\,\mathcal{J}_{u,p_2}(-j))\,E(\mathcal{J}_{u,p_3}(k)\,\mathcal{J}_{u,p_4}(-k)) + E(\mathcal{J}_{u,p_1}(j)\,\mathcal{J}_{u,p_3}(k))\,E(\mathcal{J}_{u,p_2}(-j)\,\mathcal{J}_{u,p_4}(-k)) \\ & + E(\mathcal{J}_{u,p_1}(j)\,\mathcal{J}_{u,p_4}(-k))\,E(\mathcal{J}_{u,p_3}(k)\,\mathcal{J}_{u,p_2}(-j)) + cum(\mathcal{J}_{u,p_1}(j); \mathcal{J}_{u,p_2}(-j); \mathcal{J}_{u,p_3}(k); \mathcal{J}_{u,p_4}(-k)). \end{aligned}$$

Using $(B.5)$, the first three terms of the last displayed expression are proportional to

$$f_{u,p_1p_2}(j)\,f_{u,p_3p_4}(j)\,\varphi_u(p_1,p_2)\,\varphi_u(p_3,p_4)\,\mathbf{1}(j=k),$$

while the absolute value of the last term is bounded by

$$\begin{aligned} \frac{1}{T^2}\sum_{t_1,t_2,t_3,t_4=1}^{T} |cum(u_{p_1t_1}; u_{p_2t_2}; u_{p_3t_3}; u_{p_4t_4})| &\leq C\frac{|\kappa_{4,\xi}|}{T^2}\sum_{t_1,t_2,t_3,t_4=1}^{T} \frac{\varphi_u(p_1,p_2)\,\varphi_u(p_3,p_4)}{(t_2-t_1)^\tau (t_3-t_1)^\tau (t_4-t_1)^\tau} \\ &\leq \frac{C}{T}\varphi_u(p_1,p_2)\,\varphi_u(p_3,p_4) \end{aligned}$$

because $\tau > 2$ using Lemma B.4. From here the conclusion follows easily. □

Lemma B.6. *Assuming $C2$–$C3$, we have that as $n, T \to \infty$,*

$$E\left(\frac{1}{n}\sum_{p=1}^{n}\mathcal{I}_{x,p}(j)-f_{x,p}(j)\right)^{2}=o(1). \tag{B.7}$$

Proof. Standard algebra yields that the left hand side of $(B.7)$ is bounded by

$$E\left(\frac{1}{n}\sum_{p=1}^{n}\left\{\mathcal{J}_{x,p}(j)\mathcal{J}_{x,p}'(-j)-E\left(\mathcal{J}_{x,p}(j)\mathcal{J}_{x,p}'(-j)\right)\right\}\right)^{2}+\left(\frac{1}{n}\sum_{p=1}^{n}E\mathcal{I}_{x,p}(j)-f_{x,p}(j)\right)^{2}.$$

Now $n^{-1}\sum_{p=1}^{n}E\mathcal{I}_{x,p}(j)-f_{x,p}(j)=O\left(T^{-1}\right)$ is standard as $f_{x,p}(\lambda)$ is twice continuously differentiable, whereas Lemma B.5 implies that the first term of the last displayed expression is

$$\frac{C}{n^{2}}\sum_{p,q=1}^{n}\varphi_{x}^{2}(p,q)\left(1+\frac{C}{T}\right)=o(1)$$

by Condition $C3$, see also Remark 1. □

Lemma B.7. *Under $C1$–$C3$, we have that as $n, T \to \infty$,*

$$\frac{1}{T}\sum_{j=1}^{T-1}\left(\frac{1}{n}\sum_{p=1}^{n}\mathcal{I}_{\widetilde{x},p}(j)\right)^{2}-\left(\frac{1}{n}\sum_{p=1}^{n}\mathcal{I}_{x,p}(j)\right)^{2} = o_{p}(1) \tag{B.8}$$

$$\frac{1}{T}\sum_{j=1}^{T-1}\left(\frac{1}{n}\sum_{p=1}^{n}\mathcal{I}_{x,p}(j)\right)^{2}-\int_{-\pi}^{\pi}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{p=1}^{n}f_{x,p}(\lambda)\right)^{2}d\lambda = o_{p}(1). \tag{B.9}$$

Proof. Noticing that

$$\frac{1}{n}\sum_{p=1}^{n}\mathcal{I}_{\widetilde{x},p}(j)-\mathcal{I}_{x,p}(j)=-\mathcal{I}_{\overline{x},\cdot}(j),$$

we obtain that the left hand side of $(B.8)$ equals

$$\frac{1}{T}\sum_{j=1}^{T-1}\mathcal{I}_{\overline{x},\cdot}^{2}(j)-\frac{2}{T}\sum_{j=1}^{T-1}\mathcal{I}_{\overline{x},\cdot}(j)\frac{1}{n}\sum_{p=1}^{n}\mathcal{I}_{x,p}(j).$$

We shall examine the first term of the last displayed expression, with the second one being handled similarly, if not easier. Now, by definition

$$\mathcal{I}_{\overline{x},\cdot}(j)=\frac{1}{n^{2}}\sum_{p,q=1}^{n}\mathcal{J}_{x,p}(j)\mathcal{J}_{x,q}(-j),$$

so that Lemma B.5, in particular $(B.6)$, implies that

$$E\mathcal{I}_{\overline{x},\cdot}^{2}(j)=\frac{1}{n^{4}}\sum_{p_{1},\ldots,p_{4}=1}^{n}\varphi_{x}(p_{1},p_{2})\varphi_{x}(p_{3},p_{4})\left\{1+\mathbf{1}(j=k)+\frac{C}{T}\right\}=o(1)$$

because $n^{-2}\sum_{p_{1},p_{2}=1}^{n}\varphi_{x}(p_{1},p_{2})=o(1)$ by ergodicity. This completes the proof of $(B.8)$.

Regarding $(B.9)$, it suffices to show that

$$\frac{1}{T}\sum_{j=1}^{T-1}\left(\frac{1}{n}\sum_{p=1}^{n}\mathcal{I}_{x,p}(j)-E(\mathcal{I}_{x,p}(j))\right)^{2} = o_p(1) \tag{B.10}$$

$$\frac{1}{T}\sum_{j=1}^{T-1}\left(\frac{1}{n}\sum_{p=1}^{n}\mathcal{I}_{x,p}(j)-E(\mathcal{I}_{x,p}(j))\right)\frac{1}{n}\sum_{p=1}^{n}E(\mathcal{I}_{x,p}(j)) = o_p(1), \tag{B.11}$$

because the continuous differentiability of $f_{x,p}(\lambda)$ implies

$$\frac{1}{T}\sum_{j=1}^{T-1}\frac{1}{n}\sum_{p=1}^{n}E(\mathcal{I}_{x,p}(j))-\int_{-\pi}^{\pi}\lim_{n\to\infty}\frac{1}{n}\sum_{p=1}^{n}f_{x,p}(\lambda)=o(1)$$

by standard arguments. Now $(B.10)$ holds true by Lemma B.6 and $(B.11)$ follows by Cauchy-Schwarz inequality. □

The next lemma extends a Central Limit Theorem in Phillips and Moon (1999) when their independence condition fails.

Lemma B.8. *Let $\{u_{pt}\}_{t\in\mathbb{Z}}$ and $\{x_{pt}\}_{t\in\mathbb{Z}}$, $p\in\mathbb{N}^{+}$, satisfy Conditions $C1$–$C3$. Then as $n,T\to\infty$,*

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{1}{n^{1/2}}\sum_{p=1}^{n}x_{pt}u_{pt}\xrightarrow{d}\mathcal{N}(0,\Phi). \tag{B.12}$$

Proof. First, Hidalgo and Schafgans' (2017) Theorem 1 implies that

$$z_{n,t}=\frac{1}{n^{1/2}}\sum_{p=1}^{n}x_{pt}u_{pt}\xrightarrow{d}\mathcal{N}(0,\Omega_t),\ t=1,...,T, \tag{B.13}$$

and also for any $r,s\geq 0$,

$$\frac{1}{n^{1/2}}\sum_{p=1}^{n}\chi_{p,t+r}\xi_{p,t+s}\xrightarrow{d}\mathcal{N}(0,\Omega_{t,r,s}).$$

Now, Phillips and Moon's (1999) Theorem 2 cannot be employed as the latter result requires that the left hand side of $(B.13)$, that is $\{z_{n,t}\}_{t\geq 1}$, is a sequence of independent random variables.

Dropping the subscript "p" for notational convenience, we have that

$$u_t x_t=(D_u(L)\xi_t)(C_x(L)\chi_t), \tag{B.14}$$

where

$$D_u(L)=\sum_{\ell=0}^{\infty}d_\ell L^\ell;\qquad C_x(L)=\sum_{\ell=0}^{\infty}c_\ell L^\ell$$

by Conditions $C1$ and $C2$. We now employ a "second-order" BN decomposition similar to that in Phillips and Solo (1992, pp. 978-979). First, we notice that standard algebra yields that the right hand side of $(B.14)$ is

$$\begin{aligned}
& \sum_{\ell=0}^{\infty} d_\ell c_\ell \xi_{t-\ell}\chi_{t-\ell} + \left(\sum_{\ell=0}^{\infty}\sum_{k=\ell+1}^{\infty} + \sum_{k=0}^{\infty}\sum_{\ell=k+1}^{\infty}\right) d_\ell c_k \xi_{t-\ell}\chi_{t-k} \\
= & \sum_{\ell=0}^{\infty} d_\ell c_\ell \xi_{t-\ell}\chi_{t-\ell} + \sum_{k=1}^{\infty}\left(\sum_{\ell=0}^{\infty} d_\ell c_{\ell+k}\xi_{t-\ell}\chi_{t-k-\ell}\right) + \sum_{\ell=1}^{\infty}\left(\sum_{k=0}^{\infty} c_k d_{k+\ell}\chi_{t-k}\xi_{t-k-\ell}\right) \\
= & \sum_{\ell=0}^{\infty} d_\ell c_\ell \xi_{t-\ell}\chi_{t-\ell} + \sum_{k=1}^{\infty}\left(\sum_{\ell=0}^{\infty} d_\ell c_{\ell+k} L^\ell\right)\xi_t\chi_{t-k} + \sum_{\ell=1}^{\infty}\left(\sum_{k=0}^{\infty} c_k d_{k+\ell} L^k\right)\chi_t\xi_{t-\ell} \\
= & \varrho_0(L)\,\xi_t\chi_t + \sum_{k=1}^{\infty}\varrho_k(L)\,\xi_t\chi_{t-k} + \sum_{\ell=1}^{\infty} g_\ell(L)\,\chi_t\xi_{t-\ell},
\end{aligned}$$

where $\varrho_k(L) = \sum_{\ell=0}^{\infty} d_\ell c_{\ell+k}L^\ell$ and $g_\ell(L) = \sum_{k=0}^{\infty} c_k d_{k+\ell}L^k$. Observe that $\varrho_0(L) = g_0(L)$.

Next, because for a generic polynomial $h(L) = \sum_{\ell=0}^{\infty} h_\ell L^\ell$, we have the identity $h(L) = h(1) - (1-L)\,\widetilde{h}(L)$, where $\widetilde{h}(L) = \sum_{\ell=0}^{\infty}\widetilde{h}_\ell L^\ell$ with $\widetilde{h}_\ell = \sum_{p=\ell+1}^{\infty} h_p$, we can write the right hand side of the last displayed equality as

$$\begin{aligned}
& \varrho_0(1)\,\xi_t\chi_t + \xi_t\sum_{k=1}^{\infty}\varrho_k(1)\,\chi_{t-k} + \chi_t\sum_{\ell=1}^{\infty} g_\ell(1)\,\xi_{t-\ell} \\
& -(1-L)\sum_{k=1}^{\infty}\widetilde{dc}_k\xi_{t-k}\chi_{t-k} - (1-L)\sum_{k=1}^{\infty}\widetilde{\varrho}_k(L)\,\xi_t\chi_{t-k} - (1-L)\sum_{\ell=1}^{\infty}\widetilde{g}_\ell(L)\,\chi_t\xi_{t-\ell}.
\end{aligned} \tag{B.15}$$

Observe that

$$\begin{aligned}
\widetilde{dc}_k &= \widetilde{\varrho}_0(L)\,, \quad \widetilde{\varrho}_k(L) = \sum_{\ell=0}^{\infty}\widetilde{\upsilon}_{\ell,k}L^\ell \ \text{ with } \widetilde{\upsilon}_{\ell,k} = \sum_{p=\ell+1}^{\infty} d_p c_{p+k}, \\
\widetilde{g}_\ell(L) &= \sum_{k=0}^{\infty}\widetilde{\omega}_{k,\ell}L^\ell \ \text{ with } \widetilde{\omega}_{k,\ell} = \sum_{p=k+1}^{\infty} c_p d_{p+\ell},
\end{aligned}$$

and $\xi_t\sum_{k=1}^{\infty}\varrho_k(1)\,\chi_{t-k}$ and $\chi_t\sum_{\ell=1}^{\infty} g_\ell(1)\,\xi_{t-\ell}$ are mutually independent martingale differences.

Given $(B.15)$, we can write the left hand side of $(B.12)$ as the sum of six terms. The contribution due to the fourth term of $(B.15)$ is

$$\sum_{k=1}^{\infty}\widetilde{dc}_k\frac{1}{T^{1/2}}\frac{1}{n^{1/2}}\sum_{p=1}^{n}\xi_{p,t-k}\chi_{p,t-k} = O_p\left(T^{-1/2}\right)$$

because $E\left(n^{-1/2}\sum_{p=1}^{n}\xi_{p,t-k}\chi_{p,t-k}\right)^2 < C$ and by summability of the sequence $\left\{\widetilde{dc}_k\right\}_{k\in\mathbb{N}^+}$. Next, the contribution due to the fifth and sixth terms of $(B.15)$ follow similarly and hence they are $o_p(1)$.

So, we need to examine the contribution due to the first three terms of $(B.15)$ on the left side of $(B.12)$, that is

$$\frac{\varrho_0(1)}{(Tn)^{1/2}}\sum_{t=1}^{T}\sum_{p=1}^{n}\xi_{pt}\chi_{pt} + \frac{1}{(Tn)^{1/2}}\sum_{t=1}^{T}\sum_{p=1}^{n}\xi_{pt}\widetilde{\chi}_{pt} + \frac{1}{(Tn)^{1/2}}\sum_{t=1}^{T}\sum_{p=1}^{n}\widetilde{\xi}_{pt}\chi_{pt}, \tag{B.16}$$

where

$$\widetilde{\chi}_{pt} =: \sum_{k=1}^{\infty}\varrho_k(1)\,\chi_{p,t-k}; \quad \widetilde{\xi}_{pt} =: \sum_{\ell=1}^{\infty} g_\ell(1)\,\xi_{p,t-\ell}.$$

The result that the first term of $(B.16)$ converges to a normal random variable follows by (the proof of) Hidalgo and Schafgans' (2017) Theorem 1 and Phillips and Moon's (1999) Theorem 2 as $n^{-1/2}\sum_{p=1}^{n}\xi_{pt}\chi_{pt}$ are independent sequences in t. Because the second and third terms of $(B.16)$ are similar, we only handle the second one explicitly. Now, that term is

$$\sum_{k=1}^{K}\varrho_k(1)\frac{1}{(Tn)^{1/2}}\sum_{t=1}^{T}\sum_{p=1}^{n}\xi_{pt}\chi_{p,t-k}+\sum_{k=K+1}^{\infty}\varrho_k(1)\frac{1}{(Tn)^{1/2}}\sum_{t=1}^{T}\sum_{p=1}^{n}\xi_{pt}\chi_{p,t-k}. \tag{B.17}$$

By summability of $\varrho_k(1)$ and given that

$$E\left(\frac{1}{(Tn)^{1/2}}\sum_{t=1}^{T}\sum_{p=1}^{n}\xi_{pt}\chi_{p,t-k}\right)^2=\frac{1}{Tn}\sum_{t=1}^{T}\sum_{p,q}\varphi(p,q)\le C$$

by Condition $C3$, we obtain that by choosing K large enough the second term of $(B.17)$ is $o_p(1)$. The first term of $(B.17)$ on the other hand converges to a normal random variable proceeding as with the first term of $(B.16)$. The proof is then completed using Bernstein's lemma. □

Lemma B.9. *Under the same conditions of Lemma B.8, we have that*

$$\frac{1}{\widetilde{T}^{1/2}}\sum_{j=1}^{\widetilde{T}}\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{x,p}(j)\,\mathcal{J}_{u,p}(-j)\xrightarrow{d}\mathcal{N}(0,\Phi). \tag{B.18}$$

Proof. Using $(A.1)$ and $(B.5)$ of Lemma B.5, we have that the left side of $(B.18)$ is governed by

$$\frac{1}{\widetilde{T}^{1/2}}\sum_{j=1}^{\widetilde{T}}\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{B}_{x,p}(j)\,\mathcal{B}_{u,p}(-j)\,\mathcal{J}_{\chi,p}(j)\,\mathcal{J}_{\xi,p}(-j)$$

$$=\frac{1}{\widetilde{T}^{1/2}}\sum_{j=1}^{\widetilde{T}}\frac{1}{T}\sum_{t,s=1}^{T}\Xi_{s,t}(n;j)\,e^{i(t-s)\lambda_j}, \tag{B.19}$$

where

$$\Xi_{s,t}(n;j)=\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{G}_p(j)\,\chi_{ps}\xi_{pt};\qquad \mathcal{G}_p(j)=:\mathcal{B}_{x,p}(j)\,\mathcal{B}_{u,p}(-j). \tag{B.20}$$

Because $\{\chi_{pt}\}_{t\in\mathbb{Z}}$ and $\{\xi_{pt}\}_{t\in\mathbb{Z}}$, $p\in\mathbb{N}^+$, are mutually independent *iid* zero mean sequences, we have that $\Xi_{s,t}(n)$ is independent of $\Xi_{r,m}(n)$ if $s\neq r$ and $t\neq m$ and uncorrelated if $s\neq r$ and $t=m$ or $s=r$ and $t\neq m$. By Lemma B.8, it follows that $\Xi_{s,t}(n;j)\to_d\mathcal{N}\left(0,\widetilde{V}(j)\right)$, where

$$\widetilde{V}(j)=\lim_{n\to\infty}\frac{1}{n}\sum_{p,q=1}^{n}f_{x,pq}(j)\,f_{u,pq}(j)\,\varphi(p,q)$$

and $E\|\Xi_{s,t}(n)\|^4<C$.

Next, the right hand side of $(B.19)$ is

$$\frac{2^{1/2}}{T^{3/2}}\sum_{t,s=1}^{T}\frac{1}{n^{1/2}}\sum_{p=1}^{n}\chi_{ps}\xi_{pt}\left\{\sum_{j=1}^{\widetilde{T}}g_p(j)\,e^{i(t-s)\lambda_j}\right\}$$

$$=\frac{1}{T^{1/2}}\sum_{t,s=1}^{T}\frac{1}{n^{1/2}}\sum_{p=1}^{n}\phi_p(t-s)\,\chi_{ps}\xi_{pt}\left(1+\frac{C}{T}\right) \tag{B.21}$$

using Brillinger's (1981) Exercise 1.7.14(b), where $\phi_p(s)$ denotes the s-th Fourier coefficient of $g_p(\lambda_j)$ defined in (B.20). Note also that Parseval's equality, see Fuller's (1996) Theorem 3.1.6, implies that

$$\sum_{\ell=-\infty}^{\infty} \phi_p^2(\ell) = \frac{1}{2n} \int_{-\pi}^{\pi} g_p^2(\lambda)\, d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{x,p}(\lambda)\, f_{u,p}(\lambda)\, d\lambda.$$

Now, the right hand side of (B.21) can be written as

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T-\ell} \frac{1}{n^{1/2}} \sum_{p=1}^{n} \phi_p(0)\, \chi_{pt}\xi_{pt} + \frac{1}{T^{1/2}} \sum_{\ell=1}^{T-1} \sum_{t=1}^{T-\ell} \frac{1}{n^{1/2}} \left\{ \sum_{p=1}^{n} \phi_p(\ell) \left(\chi_{pt}\xi_{p,t+\ell} + \chi_{p,t+\ell}\xi_{pt} \right) \right\}.$$

From here, we conclude the proof proceeding as we did in Lemma B.8 since, say,

$$\frac{1}{n^{1/2}} \sum_{p=1}^{n} \phi_p(\ell)\, \chi_{pt}\xi_{p,t+\ell}$$

is a sequence of independent random variables in the t dimension which converges to a Gaussian random variable by arguments similar to those in the proof of Hidalgo and Schafgans' (2017) Theorem 1 and

$$\frac{1}{T^{1/2}} \sum_{\ell=b}^{T-1} \sum_{t=1}^{T-\ell} \frac{1}{n^{1/2}} \sum_{p=1}^{n} \phi_p(\ell)\, \chi_{pt}\xi_{p,t+\ell} = o_p(1)$$

by choosing b large enough since $\phi_p(\ell) = O\left(\ell^{-2}\right)$. □

Acknowledgements: We would like to thank Professor Serena Ng, Silvia Gonçalves, two anonymous referees, and participants at the Bristol ESG Meetings for their helpful comments.

References

[1] Andrews, D.W.K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, **59**, 817-858.

[2] Andrews, D.W.K. (2005): "Cross-Section Regression with Common Shocks," *Econometrica*, **73**, 1551-1585.

[3] Arellano, M. (1987): "Computing robust standard errors for within group estimators," *Oxford Bulletin of Economics and Statistics*, **49**, 431-434.

[4] Bester, C., Conley, J. and Hansen, C. (2011): "Inference with dependent data using cluster covariance estimators," *Journal of Econometrics*, **165**, 137-151.

[5] Brillinger, D. (1981): *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, New York.

[6] Brockwell, P.J. and Davis, R.A. (1991): *Time Series: Theory and Methods.* Spinger-Verlag, New York.

[7] Chan, C. and Ogden, R.D. (2009): "Bootstrapping sums of independent but not identically distributed continuous processes with applications to functional data," *Journal of Multivariate Analysis*, **100**, 1291-1303.

[8] Cliff, A. and Ord, J. (1973): *Spatial Autocorrelation.* Pion, London.

[9] Conley, T.G. (1999): "GMM estimation with cross sectional dependence," *Journal of Econometrics*, **92**, 1-44.

[10] Cressie, N. and Huang, H.-C. (1999): "Classes of nonseparable, spatio-temporal stationary covariance functions," *Journal of the American Statistical Association*, **94**, 1330-1340.

[11] Deaton, A. (1996): *The Analysis of Household. Surveys: A Microeconometric Approach to Development Policy.* Baltimore MD: Johns Hopkins University Press

[12] Driscoll, J.C. and Kraay, A.C. (1998) "Consistent covariance matrix estimation with spatially dependent panel data," *The Review of Economics and Statistics*, **80**, 549-560.

[13] Engle, R.F. (1974): "Specification of the Disturbance Term for Efficient Estimation," *Econometrica*, **42**, 135-146.

[14] Fuentes, M. (2006): "Testing for separability of spatial-temporal covariance functions," *Journal of Statistics Planning and Inference*, **136**, 447-466.

[15] Fuller, W.A. (1996): Introduction to statistical time series. Second Edition. John Wiley: New York.

[16] Gneiting, T. (2002): "Nonseparable, stationary covariance functions for space-time data," *Journal of the American Statistical Association*, **97**, 590-600.

[17] Gonçalves, S. (2011): "The moving blocks bootstrap for panel linear regression models with individual fixed effects," *Econometric Theory*, **27**, 1048-1082.

[18] Götze, F. and Künsch, H.R. (1996): "Second-order correctness of the blockwise bootsrap for stationary observations," *Annals of Statistics*, **24**, 1914-1933.

[19] Greene, W.H. (2018): *Econometric Analysis.* Pearson, New York, NY.

[20] Hahn, J. and Kuersteiner, G. (2002): "Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large, " *Econometrica*, **70**, 1639-1657.

[21] Hannan, E.J. (1963): "Regression for time series, " in Rosenblatt, M. (Ed.), *Time Series Analysis.* Wiley, New York, pp. 17-36.

[22] Hidalgo, J. (2003): "An alternative bootstrap to moving blocks for time series regression models," *Journal of Econometrics*, **117**, 369-399.

[23] Hidalgo, J. and Schafgans, M. (2017): "Inference and testing breaks in large dynamic panels with strong cross sectional dependence," *Journal of Econometrics,* **196**, 259-274.

[24] Hidalgo, J. and Schafgans, M. (2018): "Inference without smoothing for large panels with cross sectional and temporal dependence," *Sticerd Discussion Paper,* EM/2018/597, London School of Economics.

[25] Ibragimov, I.A. and Rozanov, Y.A. (1978): Gaussian random processes. Springer. New York.

[26] Jenish, N. and Prucha, I.R. (2009): "Central limit theorems and uniform laws of large numbers for arrays of random fields," *Journal of Econometrics*, **150**, 86-98.

[27] Jenish, N. and Prucha, I.R. (2012): "On spatial processes and asymptotic inference under near-epoch dependence," *Journal of Econometrics*, **170**, 178-190.

[28] Kelejian, H.H. and Prucha, I.R. (2007): "HAC estimation in a spatial framework," *Journal of Econometrics*, **140**, 131-154.

[29] Kiefer, N.M. and Vogelsang, T.J. (2002): "Heteroskedasticity-autocorrelation robust standard errors using the Bartlett Kernel without Truncation," *Econometica*, **70**, 2093-2095.

[30] Kiefer, N.M. and Vogelsang, T.J. (2005): "A new asymptotic theory for heteroskedasticity-autocorrelation robust tests," *Econometric Theory*, **21**, 1130-1164.

[31] 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.

[32] Lazarus, E., Lewis, D.J., Stock, J.H. and Watson, M.W. (2018): "HAR Inference: recommendations for practice," *Journal of Buisness and Economic Statistics*, **36**, 541–559.

[33] Lee, J. and Robinson, P.M. (2013): "Series estimation under cross-sectional dependence," *Sticerd Discussion Paper,* EM/2013/570, London School of Economics.

[34] Matsuda, Y. and Yajima, Y. (2004): "On testing for separable correlations of multivariate time series," *Journal of Time Series Analysis*, **25**, 501-528.

[35] Nicholls, D.F. and Pagan, A.R. (1977): "Specification of the Disturbance for Efficient Estimation–An Extended Analysis," *Econometrica*, 2**45**, 211-217.

[36] Pesaran, M.H. (2006): "Estimation and inference in large heterogeneous panels with a multifactor error structure," *Econometrica*, **74**, 967-1012.

[37] Pesaran, M.H. and Yamagata, T. (2008): "Testing slope heterogeneity in large panels," *Journal of Econometrics*, **142**, 50-93.

[38] Phillips, P.C.B. (2005), "HAC Estimation by Automated Regression," *Econometric Theory,* 21, 116–142.

[39] Phillips, P.C.B. and Moon, R. (1999): "Linear regression limit theory for nonstationary panel data," *Econometrica*, **67**, 1057-1111.

[40] Phillips, P.C.B. and Solo. (1992): "Asymptotics for linear processes," *Annals ofStatistics*, **20**, 971-1001.

[41] Politis, D.N., Romano, J.P. and Wolf, M. (1997): "Subsampling for heteroskedastic time series," *Journal of Econometrics*, **81,** 281–317.

[42] Robinson, P.M. (1998): "Inference without smoothing in the presence of autocorrelation," *Econometrica*, **66**, 1163-1182

[43] Robinson, P.M. (2011): "Asymptotic theory for nonparametric regression with spatial data," *Journal of Econometrics*, **165**, 5-19.

[44] Robinson, P.M. and Hidalgo, J. (1997): "Time series regression with long-range dependence," *Annals of Statistics*, **24**, 77-104.

[45] Sun, Y. (2013), "Heteroscedasticity and autocorrelation robust F test using orthonormal series variance estimator," *The Econometrics Journal*, 16, 1–26.

[46] Vogelsang, T.J. (2012): "Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects," *Journal of Econometrics*, **166**, 303-319.

[47] White, H. (1980): "A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity," *Econometrica*, **48**, 817-838.

[48] Whittle , P. (1954): "On stationary processes in the plane," *Biometrika*, **41**, 434-449.

[49] Zellner, A. (1962): "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," *Journal of the American Statistical Association*, **57**, 348–368.

Table 1. Monte Carlo Simulations with Homogeneous Time Dependence
Empirical size of test for significance of β

Spatial Dependence	Weak Spatial Dependence						Strong Spatial Dependence					
Estimator	HS (Cluster)			DK (HAC)			HS (Cluster)			DK (HAC)		
	asy	nb	wb	asy	fixb	mbb	asy	nb	wb	asy	fixb	mbb
(n, T)	Time Dependence: AR(1), $\rho = 0.7$											
$(50, 16)$	.180	.074	.134	.253	.163	.028	.177	.068	.133	.261	.176	.030
$(50, 32)$	.126	.067	.091	.192	.131	.042	.129	.056	.091	.210	.148	.043
$(50, 64)$	.080	.054	.068	.128	.092	.049	.091	.050	.076	.158	.119	.056
$(50, 128)$	.067	.049	.062	.108	.084	.056	.068	.046	.057	.116	.089	.060
$(50, 256)$	.055	.048	.055	.087	.073	.058	.060	.051	.050	.096	.083	.065
$(100, 16)$	.172	.070	.120	.249	.153	.033	.183	.073	.134	.261	.174	.031
$(100, 32)$	.122	.057	.094	.185	.126	.050	.121	.053	.088	.200	.143	.037
$(100, 64)$	.082	.056	.070	.132	.098	.064	.096	.055	.084	.153	.110	.054
$(100, 128)$	.065	.047	.056	.108	.082	.066	.072	.052	.060	.114	.091	.062
$(100, 256)$	.058	.050	.063	.088	.074	.065	.062	.054	.058	.089	.078	.059
(n, T)	Time Dependence: AR(1), $\rho = 0.9$											
$(50, 16)$	.320	.131	.276	.410	.258	.009	.312	.106	.257	.415	.279	.013
$(50, 32)$	.242	.097	.189	.327	.209	.013	.260	.093	.201	.368	.246	.022
$(50, 64)$	.168	.058	.107	.261	.169	.026	.174	.068	.124	.281	.195	.037
$(50, 128)$	.111	.057	.084	.192	.132	.046	.115	.059	.089	.199	.147	.050
$(50, 256)$	.081	.055	.067	.142	.107	.062	.085	.055	.069	.149	.114	.061
$(100, 16)$	.316	.125	.254	.414	.255	.007	.302	.130	.253	.400	.268	.011
$(100, 32)$	.252	.084	.204	.350	.224	.017	.242	.086	.173	.344	.229	.015
$(100, 64)$	.174	.067	.118	.249	.167	.026	.174	.069	.139	.269	.188	.033
$(100, 128)$	.112	.054	.091	.181	.131	.052	.118	.060	.083	.198	.140	.056
$(100, 256)$	.075	.049	.068	.132	.096	.057	.088	.047	.071	.146	.115	.061

TABLE 2. Monte Carlo Simulations with Homogeneous Time Dependence
Empirical Power of test for significance of β when $\beta = 0.1$

Spatial Dependence	Weak Spatial Dependence						Strong Spatial Dependence					
Estimator	HS (Cluster)			DC (HAC)			HS (Cluster)			DK (HAC)		
	asy	nb	wb	asy	fixb	mbb	asy	nb	wb	asy	fixb	mbb
(n, T)	Time Dependence: AR(1), $\rho = 0.7$											
$(50, 64)$	.852	.794	.830	.919	.882	.802	.243	.158	.212	.337	.277	.162
$(50, 128)$	.980	.971	.979	.992	.986	.978	.386	.322	.357	.465	.428	.356
$(50, 256)$	1.00	1.00	1.00	1.00	1.00	1.00	.549	.527	.525	.619	.588	.558
$(100, 64)$	.972	.959	.967	.991	.987	.971	.318	.239	.297	.418	.354	.241
$(100, 128)$	1.00	1.00	1.00	1.00	1.00	1.00	.451	.395	.426	.539	.495	.429
$(100, 256)$	1.00	1.00	1.00	1.00	1.00	1.00	.690	.662	.675	.747	.721	.679
(n, T)	Time Dependence: AR(1), $\rho = 0.9$											
$(50, 64)$	.605	.401	.513	.707	.605	.232	.248	.116	.195	.359	.268	.065
$(50, 128)$	.716	.575	.659	.812	.745	.539	.253	.164	.214	.361	.287	.133
$(50, 256)$	.884	.849	.867	.942	.917	.856	.290	.227	.253	.381	.333	.231
$(100, 64)$	.784	.602	.710	.872	.799	.408	.287	.142	.238	.402	.305	.082
$(100, 128)$	.926	.857	.905	.969	.949	.867	.288	.184	.230	.401	.321	.152
$(100, 256)$	.995	.989	.994	.999	.998	.995	.348	.256	.319	.453	.393	.272

TABLE 3. Monte Carlo Simulations with Heterogeneous Time Dependence
Empirical size of test for significance of β

Spatial Dependence	Weak Spatial Dependence						Strong Spatial Dependence					
Estimator	HS (Cluster)			DC (HAC)			HS (Cluster)			DK (HAC)		
	asy	nb	wb	asy	fixb	mbb	asy	nb	wb	asy	fixb	mbb
(n,T)	Time Dependence: Mixed AR(1)											
$(100,64)$	.101	.055	.079	.189	.132	.062	.114	.065	.094	.207	.148	.052
$(100,128)$	.082	.055	.071	.144	.114	.078	.079	.055	.067	.146	.116	.065
$(100,256)$	.064	.046	.052	.121	.098	.073	.069	.053	.066	.117	.098	.073
(n,T)	Time Dependence: Mixed AR(1)/MA(1)											
$(100,64)$	.080	.049	.066	.176	.132	.083	.092	.061	.087	.185	.136	.064
$(100,128)$	.068	.051	.059	.149	.120	.097	.069	.053	.063	.140	.110	.074
$(100,256)$	.058	.049	.052	.112	.095	.084	.067	.053	.063	.111	.095	.081
(n,T)	Time Dependence: Mixed AR(3)											
$(100,64)$	.064	.051	.062	.147	.135	.120	.074	.055	.062	.150	.134	.125
$(100,128)$	.057	.048	.056	.146	.137	.122	.062	.052	.056	.143	.134	.121
$(100,256)$	.053	.048	.050	.137	.132	.127	.063	.061	.063	.138	.133	.135
(n,T)	Time Dependence: Mixed AR(3)/MA(3)											
$(100,64)$	.068	.048	.062	.134	.113	.094	.072	.058	.068	.148	.129	.106
$(100,128)$	.057	.049	.049	.109	.094	.089	.063	.052	.057	.120	.107	.091
$(100,256)$	.053	.046	.050	.094	.087	.080	.061	.056	.061	.106	.098	.094

Note: With $(1-\rho_{1,p}L)(1+\rho_2 L+\rho_3 L^2)u_{pt} = (1+\theta_{1,p}L+\theta_2 L^2+\theta_3 L^3)\eta_{pt}$, the following parameterizations are used: Denoting $\rho_p = \left(\rho_{1,p},\rho_2,\rho_3\right)'$ and $\theta_p = (\theta_{1,p},\theta_2,\theta_3)'$

Mixed AR(1): $\left\{\rho_p = \left(0.5+0.4\frac{p-1}{n-1},0,0\right)', \theta_p = 0\right\}_{p=1}^{n}$

Mixed AR(1)/MA(1): $\left\{\rho_p = \left(0.5+0.4\frac{p-1}{n/2-1},0,0\right)', \theta_p = 0\right\}_{p=1}^{n/2}$
$\left\{\rho_p = 0, \theta_p = \left(0.5+0.4\frac{p-n/2-1}{n/2-1},0,0\right)\right\}_{p=n/2+1}^{n}$

Mixed AR(3): $\left\{\rho_p = \left(0.5+0.4\frac{p-1}{n-1},0.3,0.6\right)', \theta_p = 0\right\}_{p=1}^{n}$

Mixed AR(3)/MA(3): $\left\{\rho_p = \left(0.5+0.4\frac{p-1}{n/2-1},0.3,0.6\right)', \theta_p = 0\right\}_{p=1}^{n/2}$
$\left\{\rho_p = 0,\ \theta_p = \left(0.5+0.4\frac{p-n/2-1}{n/2-1},0.3,0.6\right)\right\}_{p=1}^{n/2}$

TABLE 4. Monte Carlo Simulations with Conditional Heteroskedasticity
Empirical size of test for significance of β

Spatial Dep.		Weak Spatial Dependence								
Estimator		HS (Cluster)$^{\text{orig}}$			HS (Cluster)^{robust}			DC (HAC)		
		asy	nb	wb	asy	nb	wb	asy	fixb	mbb
	CV	$\acute{x}$ - additive								
(n,T)		$\sigma(w_p,\varrho_t)=\sigma\cdot[\exp(0.5w_p)+1][\exp(0.2\varrho_t)+1]$								
$(100,64)$	.428	.085	.054	.068	.089	.054	.068	.138	.100	.052
$(100,128)$	.424	.069	.053	.058	.071	.052	.060	.110	.089	.065
$(100,256)$	.432	.060	.052	.059	.062	.052	.060	.090	.076	.061
(n,T)		$\sigma(w_p,\varrho_t)=\sigma\cdot[\exp(0.5w_p)+1][\exp(0.5\varrho_t)+1]$								
$(100,64)$	.733	.086	.054	.068	.089	.052	.069	.132	.100	.055
$(100,128)$	.760	.069	.053	.060	.073	.056	.061	.109	.087	.061
$(100,256)$	.788	.059	.050	.060	.060	.050	.060	.086	.074	.060
(n,T)		$\sigma(w_p,\varrho_t)=\sigma\cdot[\exp(2w_p)+1][\exp(0.5\varrho_t)+1]$								
$(100,64)$	2.560	.090	.057	.070	.099	.055	.073	.133	.098	.051
$(100,128)$	2.626	.068	.054	.058	.075	.053	.060	.106	.086	.063
$(100,256)$	2.588	.055	.047	.056	.062	.045	.059	.086	.074	.062
(n,T)		$\sigma(w_p,\varrho_t)=\sigma\cdot[\exp(2w_p)+1]$								
$(100,64)$	2.117	.088	.052	.070	.093	.050	.070	.134	.101	.056
$(100,128)$	2.135	.068	.052	.056	.074	.050	.057	.106	.085	.057
$(100,256)$	2.074	.059	.052	.059	.066	.054	.063	.088	.075	.064
	CV	$\acute{x}$ - multiplicative								
(n,T)		$\sigma(w_p,\varrho_t)=\sigma\cdot[\exp(0.5w_p)+1][\exp(0.2\varrho_t)+1]$								
$(100,64)$	.428	.077	.062	.077	.070	.057	.069	.160	.142	.066
$(100,128)$	.424	.076	.065	.081	.065	.054	.073	.143	.129	.079
$(100,256)$	.432	.067	.062	.069	.057	.057	.056	.105	.098	.068
(n,T)		$\sigma(w_p,\varrho_t)=\sigma\cdot[\exp(0.5w_p)+1][\exp(0.5\varrho_t)+1]$								
$(100,64)$	.733	.128	.113	.125	.076	.063	.071	.162	.140	.051
$(100,128)$	.760	.134	.122	.137	.063	.052	.064	.141	.131	.055
$(100,256)$	.788	.141	.137	.145	.057	.057	.057	.105	.098	.056
(n,T)		$\sigma(w_p,\varrho_t)=\sigma\cdot[\exp(2w_p)+1][\exp(0.5\varrho_t)+1]$								
$(100,64)$	2.560	.131	.115	.123	.077	.063	.069	.162	.140	.049
$(100,128)$	2.626	.135	.130	.133	.070	.061	.062	.138	.123	.049
$(100,256)$	2.588	.145	.139	.146	.062	.058	.055	.108	.102	.054
		$\sigma(w_p,\varrho_t)=\sigma\cdot[\exp(2w_p)+1]$								
$(100,64)$	2.117	.074	.049	.067	.086	.051	.068	.129	.099	.063
$(100,128)$	2.135	.064	.052	.060	.075	.053	.059	.111	.093	.077
$(100,256)$	2.074	.052	.049	.052	.061	.051	.053	.084	.076	.071

Note: The time dependence assumed is AR(1) with $\rho = 0.7$.
σ is chosen to ensure that the expected $\sigma^2\left(w_p\right)\sigma^2(\varrho_t)$ equals 1.