Javier Hidalgo and Marcia M. A. Schafgans

Inference without smoothing for large panels with cross-sectional and temporal dependence

Discussion paper

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 without relying 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 a valid bootstrap which accommodates 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 temporal uncorrelated. Our proposed bootstrap can be viewed as a wild bootstrap in the frequency domain. We present some Monte-Carlo simulations to shed some light on the small sample performance of our inferential procedure and illustrate our results using an empirical example.

JEL classification: C12, C13, C23

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

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 of the model in this scenario. 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, to specify an appropriate model in the presence of cross-sectional dependence is quite difficult as there are ample generic models that are able to justify such a dependence. Some examples are the Simultaneous Autoregressive (SAR) model of Cliff and Ord (1973), which has its origins in Whittle (1954), Andrews' (2005) proposal who captures common shocks across observations and Pesaran's (2006) factor structure model. In Conditions $C1$ and $C2$ below, we shall give a generalization of the SAR model. Second, in many settings it may be quite unrealistic to assume that the temporal dependence is the same for all individuals, so to find a correct specification may be infeasible as n increases with no limit. In addition it is worth recognizing that the inferential properties based on parameter estimators that use a specific (wrong) structure 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 inferences, 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 inferences in panel data not only when the error term exhibits (potentially) both temporal and cross-sectional dependence, but 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). While this approach is often invoked and used in the context of time series regression models, in the presence of cross-sectional dependence its implementation has recently been considered in Kim and Sun (2013) or Vogelsang (2015). Unfortunately, the implementation typically requires not only the selection of a bandwidth parameter but, more importantly, an associated measure of distance between the cross-sectional units. This route has two major drawbacks. First, it explicitly assumes that there is some type of ordering among the individuals or cross-sectional units which, as opposed to the time dimension, is not unambiguous. Even if one accepts the existence of such an ordering, there is no theoretical reason to restrict it to a single measure as various economics and/or geographical distance measures may be required. 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. Second, even the selection of a bandwidth parameter to account for the temporal dependence may become impossible as we recognize that it might not be the same for all individuals. Any cross-validation algorithm used to determine the bandwidth parameter for temporal dependence may then need to be performed for each individual. In Section 2.1 we shall describe these and other drawbacks in more detail.

To deal with the potential caveats of the HAC estimator, we shall propose a cluster based estimator which is able to take into account both types of dependence, extending the work of Arellano (1987) and Driscoll and Kraay (1998) in a substantial way. 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 temporal uncorrelated although heteroscedastic. As the asymptotic distribution of the LSE might provide a poor approximation to the finite sample distribution when we employ the cluster estimator, we present and examine a bootstrap algorithm which does not require the choice of any bandwidth parameter, contrary to the sieve or moving block bootstraps. In fact, our proposed bootstrap can be viewed as a wild bootstrap but in the frequency domain.

The remainder of the paper is organized as follows. In the next section, we discuss the regularity conditions for our model and describe the main results. Section 3 discusses two bootstrap algorithms and we demonstrate their validity. Section 4 presents a Monte Carlo simulation experiment to shed some light on the finite sample performance of our cluster estimator and we illustrate the finite sample benefits of the bootstrap algorithm. We also compare the relative performance of the cluster estimator with various HAC estimators and we provide an empirical implementation. Section 5 gives a summary. 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 consider 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 variance $E\left(u_{pt}^2\right) = \sigma_p^2$, $p \in \mathbb{N}^+$. We shall assume that the sequences $\{x_{pt}\}_{t \in \mathbb{Z}}$, $p \in \mathbb{N}^+$, are mutually independent of the error term $\{u_{pt}\}_{t \in \mathbb{Z}}$, $p \in \mathbb{N}^+$, although not necessarily independent from the fixed effects η_p or α_t. More specific conditions describing the temporal and cross-sectional dependence structures of the sequences $\{u_{pt}\}_{t \in \mathbb{Z}}$ and $\{x_{pt}\}_{t \in \mathbb{Z}}$, $p \in \mathbb{N}^+$, will be given in Conditions $C1$ and $C2$ below respectively.

Our first aim in the paper is to perform inferences 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 being that in their work the sequences of random variables, say $\left\{\psi_{pt}\right\}_{t \in \mathbb{Z}}$, $p \in \mathbb{N}^+$, are assumed to be independent, that is $\left\{\psi_{pt}\right\}_{t \in \mathbb{Z}}$ and $\left\{\psi_{qt}\right\}_{t \in \mathbb{Z}}$ are mutually independent for any $p \neq q$, which is ruled out in our context as we permit cross-sectional dependence. In addition, as we allow for "*strong-dependence*", 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 implemented in our framework either. We also extend the results in Hidalgo and Schafgans (2017) by allowing the errors u_{pt} to exhibit temporal dependence as well. A second aim of the paper is to extend the work of Arellano (1987) and Driscoll and Kraay (1998) by examining a cluster estimator for the asymptotic covariance of the slope parameters estimators that does not require the ordering of the observations (in the cross-sectional dimension) or the selection of a bandwidth parameter.

Our estimator is the usual fixed effect estimator and a reformulation thereof based on the frequency domain formulation of the model. The usual fixed effect estimator of β is given by the LSE after removing the fixed effects η_p and α_t from the model. Denoting for any generic sequence $\{\varsigma_{pt}\}_{t=1}^{T}$, $p = 1, ..., n$, the required transformation by

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

we rewrite (2.1) as

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

The fixed effect estimator, $\widehat{\beta}$, is then given by

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

In view of Conditions $C1$ and $C2$ below, 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}.

The frequency domain formulation of (2.4) employs the *Discrete Fourier Transform (DFT)* of our model (2.3). This formulation, as will become clear later, proves instrumental in describing both the cluster estimator of the asymptotic covariance matrix of $\widehat{\beta}$, or $\widetilde{\beta}$ given in (2.7) below, and the bootstrap algorithm described in Section 3. Denoting 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 since $\mathcal{J}_{\varsigma,p}(\lambda_j)=\mathcal{J}_{\varsigma,p}(-\lambda_{T-j})$, $j=1,...,\widetilde{T}$, we can reformulate (2.3) as

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

and β is then estimated by

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

Under suitable regularity conditions it is well known that, $\widetilde{\beta}$ is an approximation of $\widehat{\beta}$ in that $\widetilde{\beta}-\widehat{\beta}=o_p\left(T^{-1/2}\right)$ when $n=1$. It is worth recalling that the reason not to include the frequencies λ_j for $j=0$, or T, is related to the centering of the sequences $\{\vartheta_{pt}\}_{t=1}^{T}$, $p=1,...,n$, around their sample means $T^{-1}\sum_{t=1}^{T}\vartheta_{pt}e^{it\lambda_\ell}$ as $\sum_{t=1}^{T}e^{it\lambda_\ell}=0$ if $1\leq\ell\leq T-1$.

We introduce the following regularity conditions.

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

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

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

(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*

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

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

(**i**)
$$x_{pt}=\sum_{k=0}^{\infty}c_k(p)\,\chi_{p,t-k},\qquad \sum_{k=0}^{\infty}kc_k<\infty,\quad c_k=:\sup_p\|c_k(p)\|,$$
where $\|B\|$ denotes the norm of the matrix B and $E\left(\chi_{pt}\mid\Upsilon_{p,t-1}\right)=0$; $Cov\left(\chi_{pt}\mid\Upsilon_{p,t-1}\right)=\Sigma_{\chi,p}$ and $E\left\|\chi_{pt}\right\|^4<\infty$, with $\Upsilon_{p,t}$ denoting the $\sigma-$algebra generated by $\left\{\chi_{ps},\ s\leq t\right\}$.

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

$$\chi_{pt}=\sum_{\ell=1}^{\infty}b_\ell(p)\,\eta_{\ell t},\qquad \sup_{p\in\mathbb{N}^+}\sum_{\ell=1}^{\infty}|b_\ell(p)|^2<\infty,\quad \sup_{\ell\geq1}\sum_{p=1}^{n}|b_\ell(p)|^2<\infty,$$

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 that*

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

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

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

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

For generic sequences $\{z_{pt}\}_{t\in\mathbb{Z}}$, $p\in\mathbb{N}^+$, we denote

$$\varphi_z(p,q)=\mathrm{Cov}\left(z_{pt};z_{qt}\right),\ \text{for any } p,q\geq1.$$

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*

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

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

C4: *$T,n\to\infty$ such that $n^{-1}=o\left(T^{-\xi}\right)$ for any $\xi>0$.*

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 and permit the usual SAR (or more generally $SARMA$) model. Indeed, by definition of the SAR model, 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 becomes a particular model of that allowed in Conditions $C1$ or $C2$. However Condition $C1$ does permit the sequence $\sum_{p=1}^{n}|a_\ell(p)|$ to grow with n, which is not the case with the SAR model. Of course one can allow the weights $a_\ell(p)$ to depend also 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, appears to be weaker than those typically assumed when cross-sectional dependence is allowed. It is worth pointing out that our Conditions $C1$ and $C2$ can be relaxed to some extend 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 papers. Conditions $C1$ and $C2$ do permit, though, 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 is a consequence of our conditions because $E\left(\xi_{pt}^2 \mid \mathcal{V}_{p,t-1}\right) = \sum_{\ell=1}^{\infty} |a_\ell(p)|^2$ 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 $p, q \in \mathbb{N}^+$,

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

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), however these are more complicated to model and quite difficult to handle. Despite this, there is some work on testing for separability, see Fuentes (2006) or Matsuda and Yajima (2004). If there were no cross-sectional dependence, i.e. $E\left(\xi_{p1}\xi_{q1}\right) = \sigma_{\xi p}^2 \mathbf{1}(p=q)$, then $E(u_{pt}u_{qs}) = \sigma_{\xi p}^2 \gamma_{u;pp}(t-s)\,\mathbf{1}(p=q)$. Here, and in what follows, $\mathbf{1}(A)$ denotes the indicator function.

Observe that the spectral density function of $\{u_{pt}\}_{t\in\mathbb{Z}}$ is

$$f_{u,p}(\lambda) = \frac{\varphi_u(p,p)}{2\pi} \sum_{k=-\infty}^{\infty} \left(\sum_{\ell=0}^{\infty} d_{|k|+\ell}(p)\, d_\ell(p)\right) e^{-ik\lambda}; \qquad p \in \mathbb{N}^+,$$

which is continuously differentiable as $\sum_{k=0}^{\infty} k d_k < \infty$. When $d_k(p) = d_k$ for all p, the spectral density function is the same for all cross-sectional units up to a multiplicative constant. The arguments also hold for the sequences $\{x_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$, denoting its spectral density matrix by $f_{x,p}(\lambda)$.

We now comment on Condition $C3$. As we assume that the errors and regressors are uncorrelated, we have that 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. Recall that 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

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

With the convention that $\gamma_{u,pq}(0) = \gamma_{x,pq}(0) = 1$,

$$Cov(z_{pt}, z_{qt}) = \varphi(p,q) =: \varphi_u(p,q)\varphi_x(p,q).$$

We can relax Condition $C3$ to assume that the sequences $\{x_{pt}\}_{t\in\mathbb{Z}}$ and $\{u_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$, are conditional independent in their first and second moments. However, to simplify the arguments somewhat, we have preferred to keep the condition as it stands. The condition does rule out long memory 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. On the other hand, because 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 is 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 $> 1/2$, which appears to be a necessary condition for standard asymptotic results. Nevertheless, the combined cross-sectional dependence, that is the dependence of the sequence $\{z_{pt} = u_{pt}x_{pt}\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}^+$, is required to be "weakly-dependent" as we impose (2.9), see also Hidalgo and Schafgans (2017).

Here and in what follows, we have adopted the convention that $\gamma_{u,pq}(t-s) = E(u_{pt}u_{ps})/\varphi_u(p,p)$ without loss of generality.

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

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

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*

$$\frac{1}{n^2} \sum_{p,q=1}^{n} \left(\varphi_u(p,q) + \varphi_x(p,q)\right) < C,$$

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* $\varsigma_u + \varsigma_x = 1 + \delta > 0$.

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

Finally Condition $C4$ is very weak as $\xi > 0$ effectively means that n increases to infinity at least as $\log\log T$ say. This relaxes significantly the condition given in Pesaran and Yamagata (2008), who needed that $n^{1/2}/T \to 0$ or even $n^{1/4}/T \to 0$. It appears that most panel data satisfy the condition.

Before presenting our first main result, denote

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

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

and

$$V = \Sigma_x^{-1}\Phi\Sigma_x^{-1}, \tag{2.13}$$

where $\Sigma_x > 0$ was defined in Condition $C2$.

It is standard to see that

$$\begin{aligned}\Phi &=: \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}E\left\{\left(\sum_{p=1}^{n}\sum_{t=1}^{T}x_{pt}u_{pt}\right)\left(\sum_{p=1}^{n}\sum_{t=1}^{T}x_{pt}'u_{pt}\right)\right\}\\ &= \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}\sum_{p,q=1}^{n}\sum_{t,s=1}^{T}E\left(x_{pt}x_{qs}'\right)E\left(u_{pt}u_{qs}\right). \end{aligned}\tag{2.14}$$

or, using its spectral domain formulation,

$$\begin{aligned}\Phi &= \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}E\left\{\left(\sum_{j=1}^{T-1}\sum_{p=1}^{n}\mathcal{J}_{x,p}(\lambda_j)\mathcal{J}_{u,p}(-\lambda_j)\right)\left(\sum_{j=1}^{T-1}\sum_{p=1}^{n}\mathcal{J}_{x,p}'(-\lambda_j)\mathcal{J}_{u,p}(\lambda_j)\right)\right\}\\ &= \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{nT}\sum_{j=1}^{T-1}\sum_{p,q=1}^{n}E\left(\mathcal{J}_{x,p}(\lambda_j)\mathcal{J}_{x,q}'(-\lambda_j)\right)E\left(\mathcal{J}_{u,p}(-\lambda_j)\mathcal{J}_{u,q}(\lambda_j)\right). \end{aligned}\tag{2.15}$$

We now give our main result of this section.

Theorem 1. *Under Conditions $C1-C4$, we have that as $n,T\to\infty$,*

$$\text{(i)}\quad (Tn)^{1/2}\left(\widehat{\beta}-\beta\right)\xrightarrow{d}\mathcal{N}(0,V)$$

$$\text{(ii)}\quad (Tn)^{1/2}\left(\widetilde{\beta}-\beta\right)\xrightarrow{d}\mathcal{N}(0,V).$$

Proof. The proof of this result or any other will be given in Appendix A. □

Recalling our definition of V in (2.13), the results of Theorem 1 indicate that to make inferences on β, we need to provide a consistent estimator of Φ. A first glance at (2.14) or (2.15) 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 dependence, that is to employ a HAC type of estimator, has various and potential drawbacks in the presence of cross-sectional dependence. While choosing a bandwidth parameter associated with the cross-sectional dependence requires an ordering of individuals which is non-trivial, individual heterogeneous temporal dependence (as assumed in Conditions $C1$ and

$C2$) would render intractable any cross validation method to choose the temporal bandwidth parameter.

While Kim and Sun's (2013) approach is subject to both these criticisms, Driscoll and Kraay (1989) avoid the need to specify an ordering of individuals by introducing a HAC estimator of cross-sectional averages, so that one can consider their estimator as a hybrid between a HAC and a cluster one: they employ the HAC methodology to accommodate the temporal dependence whereas they employ a cluster type of estimator to account for the cross-sectional dependence. In our Monte-Carlo experiment we compare inferences when using Driscoll and Kraay's approach with either a fixed or automated temporal bandwidth choice as suggested in Andrews (1991) against our proposed methodology. The sensitivity of relying on Kim and Sun's approach to an inappropriate ordering is indubitable.

The approach we want to advocate does not require any ordering or the selection of bandwidth parameter and it permits a more general spatio-temporal dependence structure than that allowed by either Driscoll and Kraay (1989) or Kim and Sun (2013). It can be regarded as a natural extension of earlier work by Robinson (1998) in a time series regression model context. In his case, abstracting from cross-sectional dependence

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

Applying his estimator to our model then yields the estimator

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

where $\widehat{\gamma}_{x,p}(j)$ and $\widehat{\gamma}_{u,p}(j)$ are respectively the standard sample moment estimators of $\gamma_{x,p}(j)$ and $\gamma_{u,p}(j)$. When cross-sectional dependence is allowed, the latter arguments suggest that (2.16) is not a consistent (cluster) estimator for Φ. The reason for this (see also the proof of Proposition 1 below) is that

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

as expected since the first moment of (2.16) does not capture the cross-sectional dependence. The purpose of the next section therefore is to provide a consistent "cluster" estimator for Φ in 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 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) 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 $p \neq q$, $\mathcal{J}_{u,p}(\lambda_j)$ and $\mathcal{J}_{u,q}(\lambda_k)$ can be considered as being uncorrelated although heteroscedastic. The observation was employed in the landmark paper by Hannan (1963) on adaptive estimation in a time series regression models. So the fact that we can consider $\mathcal{J}_{\widetilde{x},p}(\lambda_j)\mathcal{J}_{\widehat{u},p}(-\lambda_j)$ as a sequence of uncorrelated and heteroscedastic random variables in j, although not in p, suggests that, in a spirit similar to White's (1980) estimator, we may estimate Φ by

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

Notice that when $n=1$, (2.17) becomes the estimator given in Robinson (1998) so that we might consider Φ as a natural extension of his estimator.

Denote the estimator of Σ_x by

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

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

$$(\mathbf{a}) \qquad \breve{\Phi} - \Phi = o_p(1)$$
$$(\mathbf{b}) \qquad \widetilde{\Sigma}_x - \Sigma_x = o_p(1).$$

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*

$$(\mathbf{i}) \quad (Tn)^{1/2}\hat{V}^{-1/2}\left(\widehat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}(0,I)$$
$$(\mathbf{ii}) \quad (Tn)^{1/2}\hat{V}^{-1/2}\left(\widetilde{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}(0,I).$$

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

We now describe the time domain analogue 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),

$$\widehat{\gamma}_{x,pq}(\ell) = \frac{1}{T}\sum_{t=1}^{T-|\ell|}\widetilde{x}_{pt}\widetilde{x}'_{q,t+\ell};$$

$$\widehat{\gamma}_{u,pq}(\ell) = \frac{1}{T}\sum_{t=1}^{T-|\ell|}\widehat{u}_{pt}\widehat{u}_{q,t+\ell}\mathbf{1}(\ell>0) + \frac{1}{T}\sum_{t=1}^{T-|\ell|}\widehat{u}_{qt}\widehat{u}_{p,t+\ell}\mathbf{1}(\ell<0),$$

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

3. THE BOOTSTRAP ALGORITHM

Our motivation to introduce a bootstrap algorithm is due to the findings in our Monte-Carlo experiment, which suggest that the asymptotic distribution of $(Tn)^{1/2}\hat{V}^{-1/2}\left(\widetilde{\beta}-\beta\right)$ 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 no clear how the sieve bootstrap can be implemented under cross-sectional dependence as there is no clear ordering of the data. Even the BB suffers to some extend from this as we expect some sensitivity of the block bootstrap to any particular ordering chosen by the practitioner, let alone its validity. A second potential drawback of the BB method is that the covariance structures of $(x_{1t},...,x_{pt})'$ and $(x_{2t},...,x_{p+1,t})'$ do not need to be the same, that is we have a lack of "weak" stationarity. A third drawback is the sensitivity of the outcome of the moving block algorithm to the choice of the block size. Although some cross-validation techniques are available, see Politis and White (2004), it may not be useful for testing purposes and its implementation calls for a time series type of dependence. These drawbacks are further compounded by the fact that in our context we even need to choose two block sizes, one to deal with the time dependence and a second one to deal with the cross-sectional one, which renders its use in empirical applications quite hard to implement and the outcome can be sensitive to the choice of the bandwidth parameter.

In light of these drawbacks, we propose here a valid bootstrap algorithm with the interesting features that it is 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 it does not require the choice of any bandwidth parameter for its implementation, thereby avoiding any level of arbitrariness. We describe two bootstrap algorithms. The first one assumes that the time dependence is homogeneous among the cross-sectional units, while the second one drops the latter assumption.

The first bootstrap is described in the following simple *3 STEPS*.

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

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

(or $\widehat{u}_{pt} = \widetilde{y}_{pt} - \widehat{\beta}'\widetilde{x}_{pt}$), and compute the periodogram of the centered and scaled residuals $\{\check{u}_{pt}\}_{t=1}^{T}$, $p = 1,...,n$,

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

where, denoting $\widetilde{\sigma}^2_{\widehat{u}}(p) = T^{-1}\sum_{t=1}^{T}\widehat{u}^2_{pt}$,

$$\check{u}_{pt} = \left(\widehat{u}_{pt} - \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{pt}\right)/\widetilde{\sigma}_{\widehat{u}}(p).$$

Remark 3. *It is worth noticing that the same outcome would have been achieved if we used* $\widetilde{\sigma}^{-1}_{\widehat{u}}(p)\widehat{u}_{pt}$. *This is the case as* $\sum_{t=1}^{T} e^{it\lambda_j} = 0$ *if* $j \neq 0, T$. *The motivation to scale the centered residuals* $\widehat{u}_{pt}$ *by* $\widetilde{\sigma}_{\widehat{u}}(p)$ *is due to the fact that the variance is not the same for all individuals.*

***STEP 2*:** Denoting $\check{U}_t = \{\check{u}_{pt}\}_{p=1}^{n}$, do standard random resampling from the empirical distribution of $\{\check{U}_t\}_{t=1}^{T}$, that is we assigned probability T^{-1} to each $n \times 1$ vector $\check{U}_t$. Denote the bootstrap sample by $\{U^*_t\}_{t=1}^{T}$, which is $\{u^*_{pt}\}_{t=1}^{T}$, $p = 1,...,n$. Compute the bootstrap analogue of (2.3) as

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

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

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

$$\widetilde{\beta}^* = \left(\sum_{p=1}^{n}\sum_{j=1}^{T-1} \mathcal{J}_{\widetilde{x},p}(\lambda_j)\, \mathcal{J}'_{\widetilde{x},p}(-\lambda_j)\right)^{-1} \left(\sum_{p=1}^{n}\sum_{j=1}^{T-1} \mathcal{J}_{\widetilde{x},p}(\lambda_j)\, \mathcal{J}_{\widetilde{y}*,p}(-\lambda_j)\right) \tag{3.1}$$

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

We now comment on the bootstrap. The most important feature is that there is no need to choose any bandwidth parameter for its implementation. Also uniformly in $j = 1, ..., T$, we have that

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

and

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

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

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

We now state our main result of this section.

Theorem 2. *Under Conditions $C1 - C4$, we have that*

$$(Tn)^{1/2}\left(\widetilde{\beta}^* - \widetilde{\beta}\right) \overset{d^*}{\to} \mathcal{N}(0, V), \qquad \text{(in probability).}$$

Remark 4. *The results of Theorem 2 still hold true if $\widetilde{\beta}$ were replaced by $\widehat{\beta}$, as we have already established that $(Tn)^{1/2}\left(\widetilde{\beta} - \widehat{\beta}\right) = o_p(1)$.*

The previous results can be extended to incorporate the more realistic situation where the temporal dynamics might be different for different individuals, as given in Conditions $C1$ and $C2$. For this we modify the above bootstrap by replacing *STEP 2*, with *STEP 2'* :

***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

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

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

Remark 5. *We refer to Hidalgo* (2003) *for a discussion regarding the requirement that* $\eta_j = \eta_{T-j}$ *for* $j = \widetilde{T} + 1, ..., T$.

The latter bootstrap approach merges ideas in Hidalgo (2003) and Chan and Ogden (2009) and can be regarded as a wild-type bootstrap approach with increasing dimensional vectors.

The (bootstrap) cluster estimator of the asymptotic covariance is given by

$$\breve{\Phi}^* = \frac{1}{T}\sum_{j=1}^{T-1}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}_{\widetilde{x},p}(\lambda_j)\,\mathcal{J}_{\widehat{u}^*,p}(-\lambda_j)\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{J}'_{\breve{x},p}(-\lambda_j)\,\mathcal{J}_{\widehat{u}^*,p}(\lambda_j)\right)\right\}. \tag{3.2}$$

Proposition 2. *Assuming* $C1 - C4$*, we have that*

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

Together, these results yield the following proposition.

Proposition 3. *Under the same conditions of Theorem 2, we have*

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

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

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

4. FINITE SAMPLE BEHAVIOUR AND EMPIRICAL EXAMPLE

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 HAC-based inference procedures, which unlike ours, require the choice of smoothing parameters that may be arbitrary and erroneous. We also provide evidence of the potential finite sample improvements of our bootstrap algorithmns and illustrate its implementation with real data.

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

$$y_{pt} = \alpha_t + \eta_p + \beta x_{pt} + u_{pt}$$

for $p = 1, ..., n$ 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 β is set equal to zero. We postulate a variety of scenarios for the temporal and cross-sectional dependence for both the strictly exogenous regressor x_{pt} and error term u_{pt}.

4.1. **Simulations with Homogeneous Time Dependence.** In the first set of simulations, we consider the time dependence to be the same (homogenous) for all individuals $p = 1, .., n$. In particular, we consider the settings of no temporal dependence, autoregressive and moving average time dependence, where for the error term

$$u_{pt} = \rho_u u_{p,t-1} + \sqrt{1-\rho_u^2}\eta_{pt}, \text{ with } \rho_u = 0 \text{ or } 0.7$$

and

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

with η_{pt} characterizing the spatial dependence inherent in the error. Several cross-sectional dependence scenarios are considered for u_{pt} (η_{pt}): no spatial dependence, weak spatial dependence and strong spatial dependence. In the absence of cross-sectional dependence, η_{pt} (and u_{pt}) is $IIDN(0,1)$ for $p = 1, .., n$. Two weak spatial dependence formulations are considered. First we follow Lee and Robinson (2013), where random locations for individual units are drawn along a line, denoted $s = (s_1, ...s_n)'$ with $s_p \sim IIDU[0,n]$ for $p = 1, .., n$. Keeping these locations fixed across replications, η_{pt} are generated independently as scalar normal variables with mean zero and covariances $cov(\eta_{pt}, \eta_{qt}) = (0.5)^{|s_p - s_q|}$ (see also Hidalgo and Schafgans, 2017). This ensures that u_{pt} exhibits an exponential cross sectional decay in dependence with distance across individuals in addition to the assumed time dependence. Second, we consider a polynomial decay of cross sectional dependence in u_{pt} with distance across individuals. In the latter case, using the linear time dependence representation, $\eta_{pt} = \sigma_p \left(\sum_{\ell=1}^{\infty} c_\ell(p)\, e_{\ell t}\right)$, we chose $c_\ell(p) = |s_\ell - s_p|_+^{-10}$ with s_p and s_ℓ the random locations as before and $e_{\ell t} \sim IIDN(0,1)$; σ_p is such that $Var(\eta_{pt}) = 1$. For the strong spatial dependence setting, we use $c_\ell(p) = |s_\ell - s_p|^{-0.7}$ instead, see also Hidalgo and Schafgans (2017).[1] The same discussion holds for the independently drawn, strictly exogenous regressor x_{pt}, where, to allow for some time heterogeneity, we add μ_t which is independently drawn ($\mu_t \sim IIDN(1,1)$), such that, say, under autoregressive time dependence

$$x_{pt} = \mu_t + \rho_x x_{p,t-1} + \sqrt{1-\rho_x^2}\vartheta_{pt},$$

where ϑ_{pt} characterizes the spatial dependence inherent in the regressor.

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 Table 1, the empirical size based on our cluster estimator of the variance of $\widehat{\beta}_{FE}$ is reported in the columns labelled $\widehat{V}$. In addition to presenting the rejection rates based on the asymptotic critical values, we report the empirical size based on the naive bootstrap algorithm in the column labelled $\widehat{V}^{nb}$, and the wild bootstrap algorithm in the column labelled $\widehat{V}^{wb}$. 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 these bootstrap algorithms may yield. For comparison, we report the empirical size obtained using a variety of other estimators of the variance; the column indicates the particular estimator of the variance used. Specifically, we consider the time-cluster estimator of the variance $\widehat{V}_{Ct}$ (Hidalgo and Schafgans, 2017), the individual-cluster estimator

[1] In the polynomial case, we use $\max(1, |s_\ell - s_p|)$ as our measure of distance; not imposing such a censoring would remove all dependence in settings where for some (ℓ, p) s_ℓ and s_p lie very close together.

of the variance $\widehat{V}_{Cp}$, and Driscoll and Kraay's proposal to use a (time) HAC on cross sectional averages $\widehat{V}_{Ht}^{m_T}$. With $\widehat{V} = \widetilde{\Sigma}_x^{-1}\breve{\Phi}\widetilde{\Sigma}_x^{-1}$ where $\breve{\Phi}$ is defined in (2.17), formulae for the other estimators of the variance of $\widehat{\beta}_{FE}$ under consideration are given by $\widehat{V}_{Ct} =: \widetilde{\Sigma}_x^{-1}\hat{\Phi}_{Ct}\widetilde{\Sigma}_x^{-1}$, $\widehat{V}_{Cp} =: \widetilde{\Sigma}_x^{-1}\hat{\Phi}_{Cp}\widetilde{\Sigma}_x^{-1}$ and $\widehat{V}_{Ht}^{m_T} =: \widetilde{\Sigma}_x^{-1}\hat{\Phi}_{Ht}^{m_T}\widetilde{\Sigma}_x^{-1}$ with

$$\begin{aligned}
\hat{\Phi}_{Ct} &= \frac{1}{T}\sum_{t=1}^{T}\left\{\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\hat{z}_{pt}\right)\left(\frac{1}{n^{1/2}}\sum_{p=1}^{n}\hat{z}_{pt}\right)\right\}, \\
\hat{\Phi}_{Cp} &= \frac{1}{n}\sum_{p=1}^{n}\left\{\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\hat{z}_{pt}\right)\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\hat{z}_{pt}\right)\right\}, \\
\hat{\Phi}_{Ht}^{m_T} &= \frac{1}{nT}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}K\left(\frac{|t-s|}{m_T+1}\right)\hat{z}_{pt}\hat{z}_{qs},
\end{aligned}$$

where $\hat{z}_{pt} = \widetilde{x}_{pt}\widehat{u}_{pt}$ and $K(h) = (1-|h|)\,\mathbf{1}\,(|h|\leq 1)$ is the Bartlett kernel. Unlike our estimator, $\widehat{V}_{Ct}$ ignores any time-dependence, $\hat{V}_{Cp}$ ignores any cross-sectional dependence, and $\widehat{V}_{Ht}^{m_T}$ restricts the time-dependence and requires the selection of the lag window, m_T, for which we implement the parametric AR(1) plug-in method suggested in Andrews (1991). We also use a fixed lag window, m_T^*, equalling 5, 7, and 9 (optimal when the time dependence is AR(1) (with $\rho_u = \rho_x = 0.7$) for $T = 64$, 128, and 256 respectively, see also Andrews, 1991). None of these estimators require an ordering of cross-sectional units, which, as we argued before, may be arbitrary and erroneous. Below we will consider some simulations to address this issue.

The results from Table 1 reveal that the use of a cluster estimator that ignores time dependence ($\widehat{V}_{Ct}$) clearly results in a deterioration in size (becoming oversized, reflective of standard errors being too small) as the time dependence increases, and similarly inference that use a cluster estimator that ignores cross sectional dependence ($\widehat{V}_{Cp}$) result in a deterioration in size as the cross sectional dependence increases. Our cluster estimator, which accounts for both types of dependence, does not suffer from these obvious defects and performs remarkably well even in the presence of strong cross sectional dependence. The rejection rates based on the asymptotic critical values do tend to be closer to the nominal rejection rates when n and T both increase. Finite sample improvements in inference can be made by using either of the bootstrap algorithms as rejection rates based on them are typically closer to the nominal rejection rates, with the differences typically smaller as sample sizes increase. As the temporal dynamics is the same for all individuals here, either algorithm is valid and there does not seem to be a clear preference of these two approaches in terms of their relative closeness to the nominal size.

The simulations point to an interesting result which indicates that there is little evidence to suggest that small sample inference based on $\widehat{V}_{Ct}$ is better than $\widehat{V}$ when there is no time-dependence. In the presence of strong spatial dependence, the empirical size associated with $\widehat{V}$ is closer to the nominal rate compared to $\widehat{V}_{Ct}$ for all n, T pairs considered; for $n = 50$ and $T = 64$ the size drops from 0.066 to 0.058. The improvement is even more pronounced when we contrast the size associated with $\widehat{V}_{Ct}$ with the bootstrap based rejection rate, say $\widehat{V}^{nb}$. While there does appear to be some evidence to suggest that small sample inference based on $\widehat{V}_{Cp}$ is better than $\widehat{V}$ when there is no cross-sectional dependence, certainly when the time dependence is stronger, we

TABLE 1. Monte Carlo Simulations with Homogeneous Time Dependence - Size

Time Dep.	AR(1) $\rho_u = \rho_x = 0.7$							MA(1) $\theta_u = \theta_x = 0.7$							None						
	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}_{Ct}$	$\widehat{V}_{Cp}$	$\widehat{V}_{Ht}^{m_T}$	$\widehat{V}_{Ht}^{m_T^*}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}_{Ct}$	$\widehat{V}_{Cp}$	$\widehat{V}_{Ht}^{m_T}$	$\widehat{V}_{Ht}^{m_T^*}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}_{Ct}$	$\widehat{V}_{Cp}$	$\widehat{V}_{Ht}^{m_T}$	$\widehat{V}_{Ht}^{m_T^*}$
(n,T)																					
	No Spatial dependence																				
$(50,64)$	.089	.056	.070	.266	.063	.145	.145	.062	.051	.052	.116	.062	.087	.095	.059	.052	.055	.063	.061	.062	.092
$(50,128)$	.067	.059	.060	.268	.062	.113	.113	.058	.050	.050	.112	.065	.079	.084	.049	.049	.042	.049	.059	.050	.067
$(50,256)$	.060	.057	.058	.246	.060	.086	.086	.045	.040	.041	.094	.052	.054	.057	.055	.051	.051	.054	.065	.055	.066
$(100,64)$	.089	.049	.082	.267	.055	.142	.142	.063	.050	.067	.115	.052	.086	.095	.053	.051	.054	.052	.049	.053	.079
$(100,128)$	.073	.051	.068	.263	.057	.110	.110	.055	.051	.053	.110	.055	.073	.079	.056	.055	.051	.057	.059	.057	.078
$(100,256)$	.059	.049	.057	.260	.053	.091	.091	.055	.047	.056	.110	.058	.067	.070	.050	.049	.048	.051	.053	.051	.064
$(200,64)$	.087	.054	.071	.255	.055	.140	.140	.062	.052	.058	.112	.054	.084	.091	.058	.048	.051	.061	.055	.061	.089
$(200,128)$	.067	.050	.061	.263	.054	.115	.114	.051	.046	.052	.106	.049	.072	.074	.050	.048	.050	.055	.051	.056	.070
$(200,256)$	.067	.058	.067	.257	.057	.094	.093	.057	.049	.059	.110	.056	.067	.069	.051	.052	.051	.051	.050	.051	.063
	Weak Spatial dependence (Exponential)																				
$(50,64)$	.085	.051	.069	.262	.235	.145	.145	.061	.046	.058	.111	.224	.085	.096	.055	.050	.055	.058	.240	.060	.085
$(50,128)$	.069	.058	.060	.251	.212	.102	.103	.053	.051	.050	.109	.216	.071	.076	.046	.044	.045	.050	.215	.050	.069
$(50,256)$	.055	.050	.055	.255	.237	.082	.081	.054	.048	.057	.102	.230	.064	.066	.049	.043	.056	.049	.233	.049	.063
$(100,64)$	.090	.051	.073	.275	.242	.149	.150	.062	.046	.060	.115	.237	.090	.097	.060	.052	.063	.057	.233	.058	.083
$(100,128)$	.070	.061	.067	.260	.207	.109	.108	.061	.051	.064	.111	.205	.077	.086	.046	.046	.042	.046	.198	.046	.067
$(100,256)$	.059	.054	.060	.254	.216	.089	.089	.050	.049	.052	.101	.221	.059	.061	.049	.046	.051	.049	.221	.050	.064
$(200,64)$	.087	.055	.079	.262	.206	.140	.140	.064	.058	.064	.117	.209	.089	.097	.052	.045	.048	.054	.202	.055	.078
$(200,128)$	.063	.044	.051	.254	.210	.098	.098	.058	.046	.055	.108	.215	.074	.080	.055	.055	.055	.055	.221	.057	.073
$(200,256)$	.064	.055	.063	.262	.227	.094	.095	.054	.047	.057	.107	.222	.071	.071	.050	.051	.050	.051	.217	.052	.061

TABLE 1. Monte Carlo Simulations with Homogeneous Time Dependence - Size (cont'd)

Time Dep.	AR(1) $\rho_u = \rho_x = 0.7$							MA(1) $\theta_u = \theta_x = 0.7$							None						
	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}_{Ct}$	$\widehat{V}_{Cp}$	$\widehat{V}_{Ht}^{m_T}$	$\widehat{V}_{Ht}^{m_T^*}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}_{Ct}$	$\widehat{V}_{Cp}$	$\widehat{V}_{Ht}^{m_T}$	$\widehat{V}_{Ht}^{m_T^*}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}_{Ct}$	$\widehat{V}_{Cp}$	$\widehat{V}_{Ht}^{m_T}$	$\widehat{V}_{Ht}^{m_T^*}$
(n,T)																					
	Weak Spatial dependence (Polynomial)																				
$(50,64)$	.080	.054	.068	.251	.091	.128	.128	.056	.048	.053	.108	.089	.082	.093	.062	.051	.062	.059	.101	.061	.086
$(50,128)$	.067	.049	.062	.257	.103	.107	.107	.058	.051	.059	.107	.100	.073	.077	.057	.046	.050	.055	.105	.056	.070
$(50,256)$	.055	.048	.055	.266	.106	.086	.084	.054	.053	.057	.112	.107	.068	.069	.053	.049	.055	.054	.104	.055	.065
$(100,64)$	.082	.056	.070	.263	.106	.131	.131	.064	.054	.061	.120	.109	.091	.101	.059	.049	.056	.061	.105	.061	.087
$(100,128)$	.065	.047	.056	.256	.087	.107	.108	.055	.052	.055	.107	.089	.070	.078	.053	.047	.055	.050	.087	.050	.066
$(100,256)$	.058	.050	.063	.252	.085	.087	.086	.058	.056	.059	.117	.096	.072	.074	.049	.051	.054	.049	.089	.049	.060
$(200,64)$	.084	.050	.074	.263	.099	.142	.142	.059	.048	.057	.115	.094	.090	.096	.055	.045	.054	.059	.091	.058	.086
$(200,128)$	.066	.048	.062	.262	.131	.112	.111	.054	.044	.054	.111	.128	.073	.080	.052	.050	.049	.053	.120	.052	.067
$(200,256)$	.061	.051	.055	.267	.109	.095	.094	.055	.052	.054	.109	.106	.071	.073	.048	.048	.041	.048	.096	.048	.057
	Strong Spatial dependence (Polynomial)																				
$(50,64)$	.091	.050	.076	.270	.628	.156	.156	.064	.045	.058	.121	.635	.090	.100	.058	.044	.053	.066	.635	.072	.093
$(50,128)$	.068	.046	.057	.268	.571	.114	.113	.059	.053	.055	.117	.576	.079	.087	.054	.052	.057	.058	.568	.059	.072
$(50,256)$	.060	.051	.050	.264	.598	.094	.094	.059	.057	.054	.117	.600	.071	.073	.050	.046	.042	.054	.590	.052	.063
$(100,64)$	.096	.055	.084	.265	.641	.152	.154	.062	.050	.060	.111	.648	.088	.094	.058	.050	.052	.059	.649	.061	.087
$(100,128)$	.072	.052	.060	.258	.641	.114	.114	.056	.050	.054	.111	.627	.078	.085	.053	.047	.053s	.059	.656	.061	.073
$(100,256)$	.062	.054	.058	.249	.647	.087	.086	.057	.051	.052	.108	.648	.069	.070	.051	.049	.048	.051	.659	.053	.063
$(200,64)$	.086	.047	.073	.267	.684	.143	.143	.065	.055	.061	.117	.678	.086	.094	.060	.055	.052	.062	.689	.063	.090
$(200,128)$	.070	.057	.069	.263	.672	.114	.115	.059	.054	.059	.112	.676	.073	.080	.052	.049	.047	.056	.681	.057	.075
$(200,256)$	.059	.038	.059	.255	.681	.086	.086	.058	.049	.052	.109	.698	.069	.070	.049	.048	.048	.052	.684	.051	.066

recognize that we have to be cautious as the empirical size based on the naive bootstrap algorithm for our cluster estimator does tend to be closer to the nominal rate than $\widehat{V}_{Cp}$.

As pointed out, in the absence of time-dependence in u_{pt}, the size based on our proposed estimator for the variance of $\widehat{\beta}_{FE}$ compares well with that based on $\widehat{V}_{Ct}$. Inference based on $\widehat{V}_{Ht}^{m_T}$, which limits the time-dependence and requires the selection of the lag-window m_T, also does not have better size properties than our bandwidth parameter free estimator. In the presence of strong spatial dependence, for $n = 100$ and $T = 128$ the size associated with $\widehat{V}_{Ht}^{m_T}$ equals 0.061 against 0.053 using $\widehat{V}$ and 0.059 using $\widehat{V}_{Ct}$. Unsurprisingly, in the absence of time-dependence in u_{pt}, inferences based on either $\widehat{V}_{Ct}$, $\widehat{V}_{Ht}^{m_T}$ and $\widehat{V}$ outperform the $\widehat{V}_{Cp}$ based inference when there is cross-sectional dependence, as the latter ignores this dependence.

In the presence of time-dependence in u_{it}, improvements in size are observed when using our proposed estimator for the variance of $\widehat{\beta}_{FE}$ vis-a-vis $\widehat{V}_{Ht}^{m_T}$, which signals that accounting for insufficient time dependence, through an inappropriate lag-window $m(T)$ required for $\widehat{V}_{Ht}^{m_T}$, negatively impacts the inference on our slope parameter. The size improvements observed when using $\widehat{V}$ instead of $\widehat{V}_{Ht}^{m_T}$ are larger in the setting where there is autoregressive time dependence compared to moving average dependence and are larger when the spatial dependence is stronger. Unsurprisingly, in the presence of AR(1) time dependence, the sizes associated with $\widehat{V}_{Ht}^{m_T}$ (automatic) and $\widehat{V}_{Ht}^{m_T^*}$ (fixed) are close as m_T^* is the optimal choice in this setting, see Andrews (1991). On the other hand, we do observe larger differences in size reflective of the sensitivity to the lag-window choice, when there is MA(1) or no time dependence at all.

Kim and Sun's (2013) recent proposal to deal with both temporal and cross sectional dependence does not only suffer from the sensitivity associated with the selection of the lag-window but the actual ordering of cross-sectional units as well. This is also the case for the estimator based on (individual) HAC on time averages. We demonstrate this next. Assuming that individuals are ordered on the basis of s_p, such that $s_1 \leq s_2 \leq \cdots \leq s_n$, we consider the following two (individual) HAC on time averages The first one, $\widehat{V}_{Hp,p}^{m_n} =: \widetilde{\Sigma}_x^{-1}\hat{\Phi}_{Hp,p}^{m_n}\widetilde{\Sigma}_x^{-1}$, uses the ranking of each individual, denoted by the subscript p, to measure distance, whereas the latter, $\widehat{V}_{Hp,p}^{m_n} =: \widetilde{\Sigma}_x^{-1}\hat{\Phi}_{Hp,p}^{m_n}\widetilde{\Sigma}_x^{-1}$, uses the actual distance measure used to generate the cross-sectional dependence, s_p, $p = 1,..,n$, where

$$\begin{aligned}
\hat{\Phi}_{Hp,p}^{m_n} &= \frac{1}{nT}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{|p-q|}{m_n+1}\right)\hat{z}_{pt}\hat{z}_{qs} \\
\hat{\Phi}_{Hp,s_p}^{m_n} &= \frac{1}{nT}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{|s_p-s_q|}{m_n+1}\right)\hat{z}_{pt}\hat{z}_{qs},
\end{aligned}$$

To see the sensitivity to an incorrect ordering of individuals, we randomly reorder the individuals and implement the erroneous formulae

$$\begin{aligned}
\hat{\Phi}_{Hp,p^*}^{m_n} &= \frac{1}{nT}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{|p^*-q^*|}{m_n+1}\right)\hat{z}_{pt}\hat{z}_{qs} \\
\hat{\Phi}_{Hp,s_p^*}^{m_n} &= \frac{1}{nT}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{|s_{p^*}-s_{q^*}|}{m_n+1}\right)\hat{z}_{pt}\hat{z}_{qs}
\end{aligned}$$

Table 2. HAC Simulations with Homogenous Time Dependence - Size

HAC	Time	Cross-Sectional				Time & Cross-Sectional	
		valid		invalid		valid	invalid
	$\hat{V}^{m_T}_{Ht}$	$\hat{V}^{m_n}_{Hp,p}$	$\hat{V}^{m_n}_{Hp,s_p}$	$\hat{V}^{m_n}_{Hp,p^*}$	$\hat{V}^{m_n}_{Hp,s^*_p}$	$\hat{V}^{m_T,m_n}_{HAC,p}$	$\hat{V}^{m_T,m_n}_{HAC,p^*}$
(n,T)							
	AR(1) & Weak Spatial dependence (Polynomial)						
$(100,64)$	0.131	0.090	0.110	0.131	0.160	0.160	0.217
$(100,128)$	0.108	0.080	0.086	0.114	0.121	0.141	0.186
$(100,256)$	0.086	0.083	0.080	0.111	0.106	0.136	0.175
	AR(1) & Strong Spatial dependence (Polynomial)						
$(100,64)$	0.154	0.323	0.324	0.655	0.557	0.370	0.585
$(100,128)$	0.114	0.315	0.291	0.662	0.535	0.360	0.577
$(100,256)$	0.086	0.326	0.278	0.662	0.524	0.349	0.567
	MA(1) & Weak Spatial dependence (Polynomial)						
$(100,64)$	0.101	0.091	0.073	0.129	0.113	0.087	0.135
$(100,128)$	0.078	0.081	0.060	0.113	0.089	0.073	0.116
$(100,256)$	0.074	0.088	0.068	0.124	0.096	0.088	0.120
	MA(1) & Strong Spatial dependence (Polynomial)						
$(100,64)$	0.094	0.321	0.270	0.658	0.509	0.280	0.512
$(100,128)$	0.085	0.317	0.259	0.644	0.491	0.278	0.498
$(100,256)$	0.070	0.323	0.255	0.664	0.493	0.273	0.498

where p^* and s_{p^*} denote the erroneously assumed location of individual $p = 1,..,n$. Finally, we apply HAC in both directions as suggested by Kim and Sun (2013), in particular we consider $\widehat{V}^{m_T,m_n}_{HAC,p} =: \widetilde{\Sigma}_x^{-1}\hat{\Phi}^{m_T,m_n}_{HAC,p}\widetilde{\Sigma}_x^{-1}$ with

$$\hat{\Phi}^{m_T,m_n}_{HAC,p} = \frac{1}{nT}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{|t-s|}{m_T+1}\right) K\left(\frac{|p-q|}{m_n+1}\right)\hat{z}_{pt}\hat{z}_{qs},$$

where both the true p and the erroneous locations p^* are considered. For m_n we select fixed lag windows m_n^*, equalling 5, 7 and 8 for $n = 50$, 100 and 200 respectively; for m_T we select the fixed lag window m_T^* as before. Clearly, the individual and time HAC estimators are special cases hereof with, e.g., $\hat{\Phi}^{m_T}_{Ht} := \hat{\Phi}^{m_T,\infty}_{HAC,p}$.

With our estimator typically outperforming these HAC estimators (and being robust to incorrect specification of the cross-sectional order/distance), we primarily focus here on the relative performance of the HAC estimators in the presence of both temporal and spatial dependence. In Table 2 we report the results of these simulations for $n = 100$ and a variety of choices for T.

The results in Table 2 indicate that the (individual) HAC estimator of time averages based on the true location, $\widehat{V}_{Hp}$ and $\widehat{V}_{Hps}$, perform comparably across different time dependencies, just as the (time) HAC estimator of cross sectional averages, $\widehat{V}_{Ht}$, performs comparably across different cross-sectional dependencies. As expected, in the presence of strong spatial dependence, inference based on $\widehat{V}_{Hp}$ is particularly bad as it ignores cross sectional dependence, similarly, to the aforementioned deterioration of the performance of inference based on $\widehat{V}_{Ht}$ with time dependence. There appears some evidence that using the locations s_p in place of the rank order enhances the size of our test. This evidence, in particular, appears when the accompanying time dependence is not too strong, or better yet, absent. Incorrect ordering of individuals clearly has a strong impact on the size with the impact increasing with the level of cross sectional dependence. When keeping the lag window m_n fixed while increasing the temporal dependence weakens the performance of the individual HAC on time averages $\widehat{V}_{Hp}$ and $\widehat{V}_{Hps}$, just as the performance of the (time) HAC on individual averages $\widehat{V}_{Ht}^{m_T}$ deteriorates with stronger time dependence when keeping m_T fixed. Clearly the lag window need to be chosen appropriately to reflect the cross-sectional and/or temporal dependence. No attempt was made to implement an automated choice for the lag length, as such an approach is not obvious in general. There appears little evidence that in finite samples, inference based on $\widehat{V}_{HAC,p}^{m_T,m_n}$ tends to outperform $\widehat{V}_{Ht}^{m_T}$ and $\widehat{V}_{Hp,p}^{m_n}$.

In Table 3, we present the empirical power for testing the significance of β using our proposed estimator of the variance at the nominal 5% level when $\beta = 0.1$. In addition to presenting the rejection rates based on the asymptotic critical values (column labelled $\widehat{V}$), we report the empirical power based on the naive bootstrap algorithm in the column labelled $\widehat{V}^{nb}$, and the wild bootstrap algorithm in $\widehat{V}^{wb}$.

The results indicate that even for small panels, when $n = 50$ and $T = 64$, we have high power to reject $H_0 : \beta = 0$ when $\beta = 0.1$ when using our bandwidth parameter free estimator for the variance in all cross-sectional and time dependence scenarios. The power appears to be negatively related to the level of time- and cross-sectional dependence. In the presence of weak (polynomial) spatial dependence, say, the power decreases from 0.999 in the absence of time dependence to 0.983 under MA(1) time dependence and 0.852 under AR(1) time dependence. In the presence of AR(1) time dependence (worst case scenario) the power decreases from 0.919 in the absence of spatial dependence to 0.852 under weak (polynomial) dependence and 0.243 under strong (polynomial) dependence. As the sample sizes increase, the power approaches one, faster when the cross sectional and/or temporal dependence is lower. The empirical power of the test based on using the asymptotic critical values is comparable to the empirical power based on either the naive or wild bootstrap algorithm.

4.2. **Simulations with Heterogeneous Time Dependence.** In our second set of simulations, we allow individual heterogeneity in the time dependence in both the error term and the strictly exogenous regressor. The error term u_{pt} is generated as

$$u_{pt} = \rho_{u,p} u_{p,t-1} + \eta_{pt} \text{ or } u_{pt} = \eta_{pt} + \theta_{u,p} \eta_{p,t-1}$$

with $\rho_{u,p}$ and $\theta_{u,p}$ individual specific AR and MA coefficients respectively and η_{pt}, as before, characterizing the spatial dependence. A similar description holds for the independently drawn, strictly exogenous regressor x_{it}, with $\rho_{x,p}$ and $\theta_{x,p}$ denoting the individual specific AR and MA coefficients respectively and ϑ_{pt} characterizing its spatial dependence. As before, we allow for some time heterogeneity in the exogenous regressor as well. Unlike in our first set of simulations, we allow the variances to exhibit heteroskedasticity as well.

In Table 4, we report the empirical size for testing the significance of our parameter in the presence of heterogenous time dependence when $n = 100$ and $T = 64$, 128, and 256. We consider here two heterogeneous specifications. For the first specifications we assume that the time dependence in u_{pt} and x_{pt} for all individuals is AR(1), with corresponding correlations

TABLE 3. Monte Carlo Simulations with Homogeneous Time Dependence - Power $\beta = 0.1$

Time Dep.	AR(1) $\rho_u = \rho_x = 0.7$			MA(1) $\theta_u = \theta_x = 0.7$			None				AR(1) $\rho_u = \rho_x = 0.7$			MA(1) $\theta_u = \theta_x = 0.7$			None		
	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$		$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$
(n,T)										(n,T)									
	No Spatial dependence										Weak Spatial dependence (Polynomial)								
$(50,64)$	.919	.878	.900	.995	.995	.995	1.00	1.00	1.00	$(50,64)$	.852	.794	.830	.983	.979	.982	.999	.998	.999
$(50,128)$	.995	.995	.995	1.00	1.00	1.00	1.00	1.00	1.00	$(50,128)$	.980	.971	.979	1.00	1.00	1.00	1.00	1.00	1.00
$(50,256)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$(50,256)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$(100,64)$	.994	.987	.993	1.00	1.00	1.00	1.00	1.00	1.00	$(100,64)$	.972	.959	.967	1.00	1.00	1.00	1.00	1.00	1.00
$(100,128)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$(100,128)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$(100,256)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$(100,256)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$(200,64)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$(200,64)$	1.00	.999	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$(200,128)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$(200,128)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$(200,256)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$(200,256)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Weak Spatial dependence (Exponential)										Strong Spatial dependence (Polynomial)								
$(50,64)$	.629	.540	.598	.853	.825	.848	.943	.938	.943	$(50,64)$	.243	.158	.212	.310	.260	.302	.396	.359	.384
$(50,128)$	.874	.858	.861	.992	.991	.991	.999	.999	.999	$(50,128)$	.386	.322	.357	.580	.556	.565	.719	.712	.731
$(50,256)$	.984	.982	.983	1.00	1.00	1.00	1.00	1.00	1.00	$(50,256)$	.549	.527	.525	.822	.818	.814	.933	.928	.926
$(100,64)$	.847	.781	.822	.982	.974	.981	.999	.999	.999	$(100,64)$	.318	.239	.297	.430	.390	.422	.536	.507	.511
$(100,128)$	.992	.990	.991	1.00	1.00	1.00	1.00	1.00	1.00	$(100,128)$	.451	.395	.426	.687	.663	.679	.834	.820	.834
$(100,256)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$(100,256)$	.690	.662	.675	.921	.912	.914	.980	.979	.978
$(200,64)$	.988	.981	.987	1.00	1.00	1.00	1.00	1.00	1.00	$(200,64)$	.414	.316	.382	.577	.550	.566	.711	.697	.690
$(200,128)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$(200,128)$	.603	.559	.597	.844	.836	.845	.944	.941	.937
$(200,256)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$(200,256)$	.834	.790	.834	.984	.982	.983	.999	.999	.999

TABLE 4. Monte Carlo Simulations with Heterogeneous Time Dependence - Size

Time Dependence	Mixed AR(1) $\rho_{x,p}=\rho_{u,p}=0.4+\frac{p-1}{2(n-1)}$ for $p=1,..,n$							Mixed AR(1)/MA(1) $\rho_{x,p}=\rho_{u,p}=0.4+\frac{p-1}{2(n/2-1)}$ for $p=1,..,n/2$; 0 else $\theta_{x,p}=\theta_{u,p}=0.4+\frac{p-1-n/2}{2(n/2-1)}$ for $p=n/2+1,..,n$; 0 else						
Estimator	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}_{Ct}$	$\widehat{V}_{Cp}$	$\widehat{V}_{Ht}^{m_T}$	$\widehat{V}_{Ht}^{m_T^*}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}_{Ct}$	$\widehat{V}_{Cp}$	$\widehat{V}_{Ht}^{m_T}$	$\widehat{V}_{Ht}^{m_T^*}$
(n,T)	No spatial dependence													
$(100,64)$	.097	.053	.074	.374	.056	.195	.198	.084	.060	.070	.309	.062	.179	.179
$(100,128)$	.078	.055	.067	.395	.058	.157	.161	.066	.051	.060	.312	.061	.143	.141
$(100,256)$	.069	.056	.063	.382	.059	.117	.124	.059	.048	.053	.330	.059	.122	.122
	Weak dependence (exponential)													
$(100,64)$	.076	.049	.069	.221	.155	.140	.140	.086	.061	.078	.243	.205	.145	.146
$(100,128)$	.074	.057	.065	.241	.127	.122	.121	.065	.049	.062	.236	.184	.119	.116
$(100,256)$	.058	.045	.056	.227	.120	.092	.091	.065	.053	.059	.266	.215	.116	.112
	Weak dependence (Polynomial)													
$(100,64)$	.099	.054	.074	.350	.087	.185	.189	.077	.050	.065	.291	.092	.167	.168
$(100,128)$	.078	.054	.069	.368	.074	.140	.144	.065	.047	.056	.315	.082	.144	.141
$(100,256)$	.063	.047	.053	.385	.075	.118	.122	.058	.047	.052	.312	.072	.111	.110
	Strong dependence (Polynomial)													
$(100,64)$	.103	.061	.084	.342	.475	.191	.192	.087	.062	.080	.274	.464	.168	.169
$(100,128)$	.077	.053	.064	.352	.470	.140	.144	.066	.052	.061	.276	.413	.130	.131
$(100,256)$	.068	.052	.063	.340	.455	.115	.119	.062	.056	.062	.272	.415	.104	.105

$\rho_{z,p}=\rho_{u,p}=0.4+\frac{p-1}{2(n-1)}$ for $p=1,..,n$ (equidistant on the interval $[0.4,0.9]$). For the second specification we assume that half the individuals have an AR(1) time dependence and half the individuals have an MA(1) time dependence, with the coefficients ranging from $[0.4,0.9]$ for both dependence processes.

The results in Table 4, suggest that our cluster estimator of the variance is robust to the presence of individual specific time dependence. Compared to the homogeneous AR(1) time dependence (see Table 1), there are only moderate increases in the size of our test associated with our cluster estimator in both the heterogeneous AR(1) and heterogeneous AR(1)/MA(1) setting. As in the homogeneous time dependence setting, the rejection rates based on the asymptotic critical values do tend to be closer to the nominal rejection rates when n and T both increase. While the rejection rates based on both bootstrap algorithms suggest that finite sample improvements in inference can be made using these algorithms, we should be more cautious here as only the wild bootstrap will be valid in the heterogenous setting. The improvements achieved when applying the wild bootstrap are more modest than those suggested by the naive bootstrap.

In the absence of spatial dependence, inference based on the cluster estimator that ignores such dependence, $\widehat{V}_{Cp}$, has better size properties than ours in these heterogenous settings as well. Our estimator, though, is robust to the presence of spatial dependence, while inference based on $\widehat{V}_{Cp}$ clearly is not. In the presence of strong spatial dependence, the size based on $\widehat{V}_{Cp}$ is 0.475 when $n=100$ and $T=64$, whereas our proposal (based on the rejection rates using the wild bootstrap algorithm) yields a size of 0.084. Unsurprisingly, inference based on the cluster estimator that ignores temporal dependence, $\widehat{V}_{Ct}$, is oversized in the presence of heterogenous time dependence as well. While improvements can be made by using a (time) HAC estimator of group averages $\widehat{V}_{Ht}^{m_T}$, the gains are less in these heterogeneous settings, and our cluster estimator, which does not require the selection of a lag window m_T and accounts for the heterogeneity, has size closer to the nominal rate in both the heterogeneous AR(1) and heterogenous AR(1)/MA(1) setting. In the presence of strong spatial dependence, when $n = 100$ and $T = 128$ the rejection rates for our cluster estimator (0.077 using the asymptotic critical values and 0.064 based on the wild bootstrap algorithm) compare well to a size equalling 0.115 based on $\widehat{V}_{Ht}^{m_T}$.

In our final set of simulations, we generalize the individual heterogeneity in the time dependence for both the strictly exogenous regressor and the error term to permit a higher order autoregressive/moving average process. Specifically, we consider the following heterogeneous AR(3) and MA(3) processes for the error term (appropriately adjusted when describing the dependence for the strictly exogenous regressor)

$$(1-\rho_{u1,p}L)(1+\rho_{u2}L+\rho_{u3}L^2)u_{pt}=\eta_{pt} \text{ or}$$
$$u_{pt}=(1+\theta_{u1,p}L+\theta_{u2}L^2+\theta_{u3}L^3)\eta_{pt},$$

with $\rho_{u1,p}$ and $\theta_{u1,p}$ individual specific (equidistant on the interval $[0.4, 0.9]$) and η_{pt} characterizing the spatial dependence inherent in the error term. The two heterogeneous specification we consider here are: one where the time dependence for all individuals is AR(3), and one where we assume that half the individuals have an AR(3) time dependence and half the individuals have an MA(3) time dependence. The empirical size of the test for significance of our parameter for these two heterogeneous specifications are given in Table 5 for $n = 100$ and $T = 64$, 128, and 256.

Table 5 shows that our cluster estimator of the variance also performs well when we permit higher order heterogeneous autoregressive/moving average temporal dependence. While inference based on the cluster estimator that ignores temporal dependence, $\widehat{V}_{Ct}$, is less oversized in these heterogeneous settings compared to the heterogeneous AR(1) or heterogeneous AR(1)/MA(1) (indicative that the temporal dependence is weaker here), there is now a much larger difference in the performance of the (time) HAC estimator of group averages based on $\widehat{V}_{Ht}^{m_T}$ (automatic) and $\widehat{V}_{Ht}^{m_T^*}$ (fixed). This is not surprising, since our chosen lag windows m_T and m_T^* still rely on a, now incorrect, AR(1) dependence assumption. In the presence of strong spatial dependence, the rejection rates for our cluster estimator in the heterogeneous AR(3) setting (0.064 using the asymptotic critical values and 0.059 based on the wild bootstrap algorithm when $n = 100$ and $T = 128$) compare again well to the rejection rate based on $\widehat{V}_{Ht}^{m_T}$ which equals 0.131. While in the homogeneous AR setting finite sample improvements can be achieved by implementing either the naive of wild bootstrap algorithm, here we require the use of the wild bootstrap algorithm.

In Table 6, we report the empirical power of the test for significance of β at the nominal 5% level when $\beta = 0.1$ for the first two heterogeneous time dependence considered: heterogeneous AR(1) and heterogeneous AR(1)/MA(1).

The results in Table 6 indicate that also in the presence of heterogeneous temporal dependence, the power to reject $H_0 : \beta = 0$ when $\beta = 0.1$ is large, even in small panels. As before, the power appears to be negatively related to the level of time-dependence and cross-sectional dependence.

TABLE 5. Monte Carlo Simulations with Heterogeneous Time Dependence - Size

	Mixed AR(3): $(1-\rho_{1,p}L)(1+\rho_2 L+\rho_3 L^2)$; $\rho_{x,p}=\rho_{u,p}=\rho_p$ for $p=1,...,n$; $\rho_{1,p}=0.4+\frac{p-1}{2(n-1)}$; $\rho_2=0.3;\rho_3=0.6$							Mixed AR(3)/MA(3): $(1-\rho_{1,p}L)(1+\rho_2 L+\rho_3 L^2)$; $\rho_{x,p}=\rho_{u,p}=\rho_p$ for $p=1,...n/2$; $=0$ else; $(1+\theta_{1,p}L+\theta_2 L^2+\theta_3 L^3)$; $\theta_{x,p}=\theta_{u,p}=\theta_p$ for $p=n/2+1,...n$; $=0$ else; $\rho_{1,p}=\theta_{1,p}=0.4+\frac{p-1}{2(n/2-1)}$; $\rho_2=\theta_2=0.3;\rho_3=\theta_3=0.6$						
Estimator	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}_{Ct}$	$\widehat{V}_{Ci}$	$\widehat{V}_{Ht}^{m_T}$	$\widehat{V}_{Ht}^{m_T^*}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}_{Ct}$	$\widehat{V}_{Ci}$	$\widehat{V}_{Ht}^{m_T}$	$\widehat{V}_{Ht}^{m_T^*}$
(n,T)	No spatial dependence													
$(100,64)$	.063	.051	.061	.156	.053	.156	.139	.067	.056	.058	.177	.055	.130	.125
$(100,128)$	.061	.054	.054	.172	.058	.150	.121	.058	.051	.055	.181	.057	.114	.105
$(100,256)$	.057	.052	.058	.164	.056	.135	.093	.053	.047	.054	.174	.053	.093	.081
	Weak dependence (exponential)													
$(100,64)$	.066	.051	.064	.142	.230	.143	.128	.064	.054	.063	.139	.212	.127	.127
$(100,128)$	.068	.064	.063	.150	.207	.149	.113	.059	.056	.053	.139	.190	.122	.105
$(100,256)$	.060	.059	.057	.148	.225	.147	.093	.060	.049	.059	.146	.206	.110	.087
	Weak dependence (Polynomial)													
$(100,64)$	.066	.052	.060	.153	.093	.149	.131	.067	.048	.061	.174	.092	.128	.122
$(100,128)$	.057	.048	.053	.159	.083	.141	.109	.057	.049	.047	.165	.074	.106	.098
$(100,256)$	.051	.048	.047	.156	.078	.132	.091	.054	.047	.049	.173	.076	.092	.080
	Strong dependence (Polynomial)													
$(100,64)$	.069	.051	.069	.148	.531	.144	.133	.072	.058	.066	.183	.510	.142	.132
$(100,128)$	.064	.051	.059	.157	.525	.141	.114	.062	.052	.058	.180	.498	.118	.106
$(100,256)$	.059	.055	.060	.162	.517	.131	.100	.056	.051	.056	.175	.510	.099	.086

The most noteworthy finding here is that in the presence of strong (polynomial) spatial dependence the power is larger in the heterogeneous AR(1) setting than in the homogeneous AR(1) setting, see also Table 3. When $n = 100$ and $T = 64$, in the heterogeneous AR(1) setting the size equals 0.386 versus 0.318 in the homogenous AR(1) setting.

4.3. **Empirical Example: Bid-Ask-Spread of Stocks.** The empirical example is taken from Hoechle (2007), who introduced the Stata programme *xtscc* that implements the $\hat{V}_{Ht}^{m_T}$ robust standard errors for panel regressions with cross-sectional dependence. Following Hoechle (2007), we consider the following linear panel regression model

$$BA_{pt} = \alpha + \beta_1 aVol_{pt} + \beta_2 Size_{pt} + \beta_3 TRMS_{pt}^2 + \beta_4 TRMS_{pt} + \varepsilon_{pt}$$

to investigate whether information differentials can partially explain the cross-sectional differences in quoted bid-ask spreads, as suggested by Glosten (1987). The dependent variable is the relative bid-ask spread, BA, and the independent variables are the stock's abnormal trading volume, $aVol$, the stock's size decile, *Size,* the monthly return of the MSCI Europe total return index in USD

Table 6. Monte Carlo Simulations with heterogeneous time dependence - Power $\beta = 0.1$

Time Dependence	Mixed AR(1) $\rho_{x,p} = \rho_{u,p} = 0.4 + \frac{p-1}{2(n-1)}$ for $p = 1,..,n$			Mixed AR(1)/MA(1) $\rho_{x,p} = \rho_{u,p} = 0.4 + \frac{p-1}{2(n/2-1)}$ for $p = 1,..,n/2$; 0 else $\theta_{x,p} = \theta_{u,p} = 0.4 + \frac{p-1-n/2}{2(n/2-1)}$ for $p = n/2+1,..,n$; 0 else		
Estimator	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$	$\widehat{V}$	$\widehat{V}^{nb}$	$\widehat{V}^{wb}$
(n,T)	No spatial dependence					
$(100,64)$	.939	.898	.923	.969	.957	.963
$(100,128)$	.993	.991	.992	.998	.997	.998
$(100,256)$	1.00	1.00	1.00	1.00	1.00	1.00
	Weak dependence (exponential)					
$(100,64)$	.931	.896	.923	.879	.845	.873
$(100,128)$	.997	.995	.996	.992	.989	.992
$(100,256)$	1.00	1.00	1.00	1.00	1.00	1.00
	Weak dependence (Polynomial)					
$(100,64)$	.906	.843	.876	.945	.925	.937
$(100,128)$	.990	.984	.989	.996	.994	.995
$(100,256)$	1.00	1.00	1.00	1.00	1.00	1.00
	Strong dependence (Polynomial)					
$(100,64)$	.386	.300	.353	.441	.387	.424
$(100,128)$	.498	.429	.464	.646	.608	.633
$(100,256)$	.753	.723	.745	.862	.851	.861

(in %), *TRMS,* and its square, *TRMS*^{2}, as a simply proxy for the stock market risk.[2] Using monthly data from December 2000 to December 2005 on 195 stocks selected from the MSCI Europe constituents, Hoechle (2007) revealed that the impact of accounting for cross-sectional dependence on the standard errors of the parameter estimators was large whether individual stock fixed effects were included or not. No time fixed effects were considered, as *TRMS* is constant across stocks. In order to implement our frequency domain based cluster estimator of the covariance matrix, we modify the sample slightly.[3] To permit the computational advantage of the Discrete Fast Fourier Transform based on a prime factor algorithm we ignored the last observation, so that $T = 60$.

In Table 7 we provide various estimates of the standard errors for the fixed effect estimator together with their associated p-values for their individual significance. Aside from using Driscoll and Kraay's *HAC* estimator, $\widehat{V}_{Ht}^{m_T}$, we provide standard errors based on our cluster estimator,

[2]The bid-ask spread, *BA,* is defined as $100.\frac{Ask_{pt}-Bid_{pt}}{0.5\left(Ask_{pt}+Bid_{pt}\right)}$, and the stock's abnormal trading volume, *aVol*, is defined $100\left(\ln\left(Vol_{pt}\right) - \frac{1}{T}\sum_s \ln\left(Vol_{ps}\right)\right)$, where Vol_{pt} denotes the number (in thousands) of stocks traded on the last trading day of month t.

[3]While the xtscc procedure allows for unbalanced panel datasets, our focus on the frequency domain specification makes it more natural to focus on balanced panels. To reduce the impact, we have replaced missing observations for the bid-ask spread by their predicted values based on the unbalanced FE estimates. We have dropped one stock due to missing data on abnormal trading volume and 23 stocks that that do not cover the whole period.

TABLE 7. Bid-Ask-Spread of Stocks

	aVol	*Size*	*TRMS²*	*TRMS*
Parameter estimates, $\widehat{\beta}_j$	$-.0012$	$-.1389$	.0030	$-.0047$
Standard error				
$\widehat{V}_{Ht}^{m_T^*}$	.0007	.0298	.0009	.0050
$\widehat{V}_{Ht}^{m_T}$	.0009	.0365	.0008	.0081
$\widehat{V}$	.0005	.0237	.0011	.0084
$\widehat{V}_{Ct}$	.0009	.0333	.0008	.0078
$\widehat{V}_{Cp}$	.0008	.0358	.0005	.0042
p-value (significance)				
$\widehat{V}_{Ht}^{m_T^*}$	.068	.000	.000	.349
$\widehat{V}_{Ht}^{m_T}$	.168	.000	.000	.558
$\widehat{V}$ asymptotic	.022	.000	.006	.576
naive bootstrap	.056	.000	.012	.583
wild bootstrap	.003	.000	.008	.624
$\widehat{V}_{Ct}$	.198	.000	.000	.543
$\widehat{V}_{Cp}$	.148	.000	.000	.262

$n = 195$, $T = 60$

$\widehat{V}$, together with cluster estimators that either ignore time dependence, $\widehat{V}_{C_t}$, or ignore spatial dependence, $\widehat{V}_{Cp}$. For the lag window required for the HAC estimator, we consider two choices: m_T^* (fixed) equal to the 8 months lag chosen by Hoechle and m_T (automatic) which implements the parametric AR(1) plug-in method. As our simulations suggest finite sample improvements can be made using our bootstrap algorithms, we report p-values associated with our proposal based on the asymptotic distribution and as obtained using the naive and wild bootstrap algorithms.

As in Hoechle, the HAC standard errors (with $m_T^* = 8$) of all parameter estimates, with the exception of the variable *Size*, are considerably larger than the usual LSE standard errors. As the associated p-values for the significance of *aVol* exceed the 5% level of significance, consequently, there is only weak evidence that information differentials help explain differences in quoted bid-ask spread ceteris paribus once spatial dependence is accounted for using the HAC standard errors. This result, however, is sensitive to the particular choice of m_T^*, with the p-values and standard errors for *aVol* decreasing when larger window lags are considered. Our cluster estimator, which does not require the selection of a window lag and thereby does not restrict the time dependence, has p-values (and standard errors) for *aVol* that are smaller than those indicated by HAC with window lag m_T^* suggesting that the window lag indeed may have been chosen too small. Computing the p-values using the wild bootstrap algorithm (indicated here in view of observed heterogeneity in the sample) lends further support of this. The automated window lag, based on implementing the parametric AR(1) plug-in estimator, on the other hand suggests that much less time dependence should have been taken into account as $m_T = 1$; this suggests that the parametric plug-in method does not work particularly well in this case. While our cluster based inference therefore finds evidence that information differentials help explain differences in quoted bid-ask spread ceteris paribus, cluster based inference that ignore either cross sectional or temporal dependence, on the other hand, do not permit the support of Glosten's hypothesis.

The results in Table 7 also lend support for return differentials between small and large stocks (*Size)* (e.g., see also Fama and French, 1993) and a correlation between stock market risk ($TRMS^2$) and the bid-ask spread. While the p-values associated with their significance are small regardless of which estimator for the covariance of the fixed effect estimator is used, it is clear that the standard errors based on using the HAC estimator for *Size* and $TRMS^2$ exhibit a high sensitivity to the particular choice of the window lag, m_T^* versus m_T, which renders the limitation of bandwidth based inference obvious. When removing the restriction on the time dependence imposed by Driscoll and Kraay's HAC estimator, we obtain standard errors on $TRMS$ and $TRMS^2$ that are larger than the presented HAC standard errors in accordance with results obtained in our simulations.

5. **CONCLUSIONS**

In this paper we expand the literature on inference in panel data models in the presence of both temporal and cross-sectional dependence without relying on any parametric functional form of such dependences. 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 (2015). To deal with various serious caveats of the HAC estimator, we propose a cluster based estimator which is able to take into account both types of dependence, extending the work of Arellano (1987) and Driscoll and Kraay (1998) in a substantial way. Our approach is based on the realization that the spectral representation of the fixed effect panel data model is such that the errors become approximately temporally uncorrelated and heteroscedastic. As the cluster estimator may not be reliable in small samples, and therefore it may not provide a good approximation to make accurate inferences, we present and examine a bootstrap algorithm in the frequency domain. Simulation results reveal that our estimator performs quite well, even in the presence of strong spatial dependence, and our bootstrap algorithms provide small sample improvements. In light of the sensitivity of the HAC estimator to the choice of the window lag and, more importantly, the associated measure of distance between the cross-sectional units, we feel that our approach offers a welcome contribution in this literature.

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

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

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

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

Finally, we shall make use of the well know result

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

A.1. **PROOF OF THEOREM 1.**

We begin with part (**i**). Without loss of generality assume that x_{pt} is scalar. Using (2.2) and standard arguments, we obtain

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

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

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

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 1, implies that[4]

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

[4]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 $\sum(\alpha_p + \beta_p) = o(n)$.

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 that the left hand side of $(A.4)$ is

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

by Lemma B.8. This concludes the proof of part (**i**) of the theorem.

We now show part (**ii**). Proceeding similarly as in part (**i**), 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

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

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

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

by Lemma B.1 and $(A.6)$. 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 bounded by

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

by $(A.7)$ and using $\left(\sup_{p,q}|f_{x,pq}(j)|+\sup_{p,q}|f_{u,pq}(j)|\right)\le C$. This concludes the proof of the theorem. □

A.2. **PROOF OF PROPOSITION 1.**

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

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

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

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

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

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

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 $o_p(1)$ by Cauchy-Schwarz's inequality if we show that the first term converges in probability to Φ. Since

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

this result holds true if we show that

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

and

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

First we examine $(A.13)$. The first term on the left of $(A.13)$, which can be rewritten as

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

has its first moment given by

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

using Lemma B.1. Using $(A.6)$, we conclude that the last displayed expression is $o(1)$. Next, we observe that Lemma B.5 implies, for instance, that

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

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

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

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

To complete part (**a**), we examine $(A.12)$. Using $(A.1)$, we have that $(A.12)$ holds true if the following expressions $(A.14)-(A.16)$ are $o_p(1)$:

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

$$\frac{1}{Tn}\sum_{j=1}^{T-1}\left(\sum_{p=1}^{n}\mathcal{B}_{x,p}(-j)\,\mathcal{J}_{\chi,p}(j)\,\mathrm{Y}_{u,p}(-j)\right)\left(\sum_{p=1}^{n}\mathcal{B}_{u,p}(j)\,\mathcal{J}_{\xi,p}(-j)\,\mathrm{Y}_{x,p}(j)\right),\tag{A.15}$$

$$\frac{1}{Tn}\sum_{j=1}^{T-1}\left(\sum_{p=1}^{n}\mathrm{Y}_{x,p}(j)\,\mathrm{Y}_{u,p}(-j)\right)\left(\sum_{p=1}^{n}\mathrm{Y}_{u,p}(-j)\,\mathrm{Y}_{x,p}(j)\right)\tag{A.16}$$

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

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

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

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

we have, by standard algebra, that the second moments of $(A.14)$ are $o(1)$, when recognizing

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

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

and

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

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

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

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

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

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**). Using Lemma B.6 and $(A.11)$, it suffices to show that

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

This holds true proceeding as with the proof of part (**a**) and by recognizing that, by the continuous differentiability of $f_{x,p}(\lambda)$, $T^{-1}\sum_{j=1}^{T} f_{x,p}(j) \to \int_0^{2\pi} f_{x,p}(\lambda)\,d\lambda =: \Sigma_{x,p}$. □

A.3. **PROOF OF THEOREM 2.**

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

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

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

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

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

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

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

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 $o_{p^*}(1)$. We observe that $(A.23)$ is a nonnegative expression.

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

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

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

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

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 $\mathcal{G}(j)$. The right hand side of $(A.26)$ now can be written as

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

As $\phi(r)=O\left(r^{-2}\right)$ by Conditions $C1$ and $C2$, given the independence of the sequences of random variables $\frac{1}{n^{1/2}}\sum_{p=1}^{n}\chi_{pt}u_{p,t+\ell}^{*}$ and $\frac{1}{n^{1/2}}\sum_{p=1}^{n}\chi_{p,t+\ell}u_{pt}^{*}$ in t for completion of the proof it suffices to show that.

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

Observe that $E^{*}\left|\Lambda_{t,n}^{*}\right|^{4}=O_p(1)$.

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

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

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

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

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

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

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

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

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 moments 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. Moreover, as for instance

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

because the left side is a nonnegative expression that has an expectation which is $o(n)$, we have by arguing as in the proof of Proposition 1 that

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

The proof of part (**i**), and thereby the theorem, therefore, 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

$$\begin{aligned}
&\sum_{p,q=1}^{n} \varphi_u(p,q) E\left(\mathcal{J}_{\chi,p}(j) \mathcal{J}_{\chi,q}(-j)\right) \\
= &\sum_{p,q=1}^{n} \varphi_u(p,q) \frac{1}{T} \sum_{t,s=1}^{T} E\left(x_{pt} x_{qs}\right) e^{-i(t-s)\lambda_j} \\
= &\; C \sum_{p,q=1}^{n} \varphi_u(p,q) \varphi_x(p,q) = O(n)
\end{aligned}$$

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

A.4. **PROOF OF PROPOSITION 2.**

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

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

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

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

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

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

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's inequality if we show that $(A.30)$ is $o_{p^*}(1)$. To that end, using $(A.11)$ and $(A.24)$, we have

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

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

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

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

$(A.34)$ 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

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

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. □

A.5. **PROOF OF PROPOSITION 3.**

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

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

Because η_j are normally distributed it suffices to show

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

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

$$\frac{1}{Tn}\sum_{j=1}^{T-1}\sum_{p,q=1}^{n}\mathcal{J}_{\widetilde{x},p}(j)\mathcal{J}_{\widetilde{x},q}(-j)\mathcal{J}_{\widehat{u},p}(-j)\mathcal{J}_{\widehat{u},q}(j)$$
$$=\frac{1}{Tn}\sum_{j=1}^{T-1}\sum_{p,q=1}^{n}\mathcal{J}_{\widetilde{x},p}(j)\mathcal{J}_{\widetilde{x},q}(-j)\mathcal{J}_{u,p}(-j)\mathcal{J}_{u,q}(j)+o_p(1)$$

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

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

The first term converges in probability to Φ, whereas the second term follows by Cauchy-Schwarz's 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

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

where

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

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

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

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

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

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 $p, q = 1, .., n$,*

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

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*

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

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 $\tau > 2$,*

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

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

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

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,*

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

and

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

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

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

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

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

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

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

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

Lemma B.6. *Assuming $C2-C4$, we have that for some $\eta > 0$,*

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

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 $\frac{1}{n}\sum_{p=1}^{n}E\mathcal{I}_{x,p}(j)-f_{x,p}(j) = O(n^{-\eta})$ is standard as $f_{x,p}(\lambda)$ is twice continuously differentiable and Condition $C4$ holds. Using Lemma B.5 ensures that the first term of the last displayed expression is

$$\frac{C}{n^2}\sum_{p,q=1}^{n}\varphi_x^2(p,q)\left(1+\frac{C}{T}\right) = O\left(n^{-\eta}\right)$$

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

Lemma B.7. *Under $C1 - C3$, we have that*

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

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

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

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

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

$$E\mathcal{I}^2_{\overline{x},\cdot}(j)=\frac{1}{n^4}\sum_{p_1,\ldots,p_4=1}^{n}\varphi_x(p_1,p_2)\varphi_x(p_3,p_4)\left\{1+\mathbf{1}(j=k)+\frac{C}{T}\right\}=o(1)$$

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 because Lemma B.5, in particular expression $(B.6)$, yields that

$$
\begin{aligned}
E\left(\frac{1}{n}\sum_{p=1}^{n}\mathcal{I}_{x,p}(j)-E\left(\mathcal{I}_{x,p}(j)\right)\right)^{2} &= \frac{1}{n^{2}}\sum_{p,q=1}^{n}E\left\{\mathcal{J}_{x,p}(j)\,\mathcal{J}_{x,p}'(-j)-E\left(\mathcal{J}_{x,p}(j)\,\mathcal{J}_{x,p}'(-j)\right)\right.\\
&\qquad\left.\times\,\mathcal{J}_{x,q}(j)\,\mathcal{J}_{x,q}'(-j)-E\left(\mathcal{J}_{x,q}(j)\,\mathcal{J}_{x,q}'(-j)\right)\right\}\\
&= \frac{1}{n^{2}}\sum_{p,q=1}^{n}\varphi_{x}^{2}(p,q)\left\{2+\frac{C}{T}\right\}\\
&= o(1)
\end{aligned}
$$

by Condition $C3$. Next $(B.11)$ follows by Cauchy-Schwarz's 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* $n,T\to\infty$,

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

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

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

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

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

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

$$
u_{t}x_{t}=\left(D_{u}(L)\,\xi_{t}\right)\left(C_{x}(L)\,\chi_{t}\right), \tag{B.14}
$$

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, p. 978-979). First, we notice that standard algebra yields that the right hand side of $(B.14)$ is

$$\begin{aligned}
&\sum_{\ell=0}^{\infty} d_\ell c_\ell \xi_{t-\ell}\chi_{t-\ell} + \left(\sum_{\ell=0}^{\infty}\sum_{k=\ell+1}^{\infty} + \sum_{k=0}^{\infty}\sum_{\ell=k+1}^{\infty}\right) d_\ell c_k \xi_{t-\ell}\chi_{t-k} \\
= &\sum_{\ell=0}^{\infty} d_\ell c_\ell \xi_{t-\ell}\chi_{t-\ell} + \sum_{k=1}^{\infty}\left(\sum_{\ell=0}^{\infty} d_\ell c_{\ell+k}\xi_{t-\ell}\chi_{t-k-\ell}\right) + \sum_{\ell=1}^{\infty}\left(\sum_{k=0}^{\infty} c_k d_{k+\ell}\chi_{t-k}\xi_{t-k-\ell}\right) \\
= &\sum_{\ell=0}^{\infty} d_\ell c_\ell \xi_{t-\ell}\chi_{t-\ell} + \sum_{k=1}^{\infty}\left(\sum_{\ell=0}^{\infty} d_\ell c_{\ell+k}L^\ell\right)\xi_t\chi_{t-k} + \sum_{\ell=1}^{\infty}\left(\sum_{k=0}^{\infty} c_k d_{k+\ell}L^k\right)\chi_t\xi_{t-\ell} \\
= &f_0(L)\,\xi_t\chi_t + \sum_{k=1}^{\infty} f_k(L)\,\xi_t\chi_{t-k} + \sum_{\ell=1}^{\infty} g_\ell(L)\,\chi_t\xi_{t-\ell},
\end{aligned}$$

where $f_k(L) = \sum_{\ell=0}^{\infty} d_\ell c_{\ell+k}L^\ell$ and $g_\ell(L) = \sum_{k=0}^{\infty} c_k d_{k+\ell}L^k$. Observe that $f_0(L) = g_0(L)$.

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

$$\begin{aligned}
&f_0(1)\,\xi_t\chi_t + \xi_t\sum_{k=1}^{\infty} f_k(1)\,\chi_{t-k} + \chi_t\sum_{\ell=1}^{\infty} g_\ell(1)\,\xi_{t-\ell} \\
&-(1-L)\sum_{k=1}^{\infty}\widetilde{dc}_k\xi_{t-k}\chi_{t-k} - (1-L)\sum_{k=1}^{\infty}\widetilde{f}_k(L)\,\xi_t\chi_{t-k} - (1-L)\sum_{\ell=1}^{\infty}\widetilde{g}_\ell(L)\,\chi_t\xi_{t-\ell}.
\end{aligned} \tag{B.15}$$

Now, we observe that

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

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

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 $(B.15)$,

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

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

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

$$\begin{aligned}
&f_0(1)\frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{1}{n^{1/2}}\sum_{p=1}^{n}\xi_{pt}\chi_{pt} + \frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{1}{n^{1/2}}\sum_{p=1}^{n}\xi_{pt}\widetilde{\chi}_{pt} \\
&+\frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{1}{n^{1/2}}\sum_{p=1}^{n}\widetilde{\xi}_{pt}\chi_{pt},
\end{aligned} \tag{B.16}$$

where

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

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 (2002) 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

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

By summability of $f_k(1)$ and given that

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

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 hand side of $(B.18)$ is governed by

$$\begin{aligned} & \frac{1}{\widetilde{T}^{1/2}}\sum_{j=1}^{\widetilde{T}}\frac{1}{n^{1/2}}\sum_{p=1}^{n}\mathcal{B}_{x,p}(j)\,\mathcal{B}_{u,p}(-j)\,\mathcal{J}_{\chi,p}(j)\,\mathcal{J}_{\xi,p}(-j) \\ = & \frac{1}{\widetilde{T}^{1/2}}\sum_{j=1}^{\widetilde{T}}\frac{1}{T}\sum_{t,s=1}^{T}\Xi_{s,t}(n;j)\,e^{i(t-s)\lambda_j}, \end{aligned} \tag{B.19}$$

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 $s\ne r$ and $t\ne m$ and uncorrelated if $s\ne r$ and $t=m$ or $s=r$ and $t\ne m$. 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 $E\left\|\Xi_{s,t}(n)\right\|^4 < C$.

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

$$\begin{aligned} & \frac{2^{1/2}}{T^{3/2}} \sum_{t,s=1}^{T} \frac{1}{n^{1/2}} \sum_{p=1}^{n} \chi_{ps} \xi_{pt} \left\{ \sum_{j=1}^{\widetilde{T}} g_p(j)\, e^{i(t-s)\lambda_j} \right\} \\ = \; & \frac{1}{T^{1/2}} \sum_{t,s=1}^{T} \frac{1}{n^{1/2}} \sum_{p=1}^{n} \phi_p(t-s)\, \chi_{ps} \xi_{pt} \left(1 + \frac{C}{T}\right) \end{aligned} \tag{B.21}$$

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 $\phi_p(\ell) = O\left(\ell^{-2}\right)$. □

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

[10] 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.

[11] Efron, B. (1979): "Bootstrap methods: another look at the jackknife," *Annals of Statistics*, **7**, 1-26.

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

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

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

[15] Glosten L.R. (1987): "Components of the bid-ask spread and the statistical properties of transaction prices," *Journal of Finance*, **42**, 1293-1307.

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

[17] 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.

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

[19] 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.

[20] Hoechle D, (2007): "Robust standard errors for panel regressions with cross-sectional dependence," *Stata Journal*, **7**, 281-312.

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

[22] 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.

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

[24] Kim, M.S. and Sun, Y. (2013): "Heteroskedasticity and spatiotemporal dependence X_I inference for linear panel models with fixed effects," *Journal of Econometrics*, **177**, 85-108.

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

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

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

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

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

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

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

[32] Politis, D.N. and White, H. (2004): "Automatic Block-Length Selection for the Dependent Bootstrap," *Econometric Reviews*, **23**, 53–70.

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

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

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

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

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

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

[39] 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.

ECONOMICS DEPARTMENT, LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON WC2A 2AE, UK
E-mail address: f.j.hidalgo@lse.ac.uk

ECONOMICS DEPARTMENT, LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON WC2A 2AE, UK
E-mail address: m.schafgans@lse.ac.uk