Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogeneity

Guido M. Kuersteiner[1] and Ingmar R. Prucha[2]

October 1, 2010

[1]Department of Economics, Georgetown University, Washington, DC 20057, Tel.: 202-687-0956, e-mail: gk232@georgetown.edu

[2]Department of Economics, University of Maryland, College Park, MD 20742, Tel.: 301-405-3499, e-mail: prucha@econ.umd.edu

Abstract

The paper derives a general Central Limit Theorem and asymptotic distributions for moment conditions related to panel data and spatial models with large n. The results allow for regressors to be only sequentially rather than strictly exogenous, while at the same time allowing the data to be cross sectionally dependent. The setup is sufficiently general to accommodate situations where cross sectional dependence stems from the presence of common factors, which leads to the need for random norming. The limit theorem for sample moments is derived by showing that the moment conditions can be recast such that a martingale difference array central limit theorem can be applied. We prove such a central limit theorem by first extending results for stable convergence in Hall and Hedye (1980) to non-nested martingale arrays relevant for our applications. We illustrate our result by establishing a generalized estimation theory for GMM estimators of a fixed effect panel model without imposing i.i.d. or strict exogeneity conditions.

Keywords: Cross-sectional dependence, spatial martingale difference sequence, Central Limit Theorem, spatial, panel, GMM.

1 Introduction[1]

In this paper we develop a central limit theory for data sets with cross-sectional dependence. Importantly, the theory is sufficiently general to cover panel data sets, allowing for regressors that are only sequentially (rather than strictly) exogenous, while at the same time allowing the data to be cross sectionally dependent. The paper considers cases where the time series dimension T is fixed. Our results also cover purely cross-sectional data-sets.

At the center of our results lies a cross-sectional conditional moment restriction that avoids any assumption of cross-sectional independence. The paper proves a central limit theorem for the corresponding sample moment vector by extending results of Hall and Heyde (1980) for stable convergence of martingale difference arrays to a situation of non-nested information sets. We then show that by judiciously constructing information sets in a way that preserves a martingale structure for the moment vector in the cross-section our martingale array central limit theorem is applicable to cross-sectionally dependent panel and spatial models.

The classical literature on dynamic panel data has generally assumed that the observations, including observations on the exogenous variables, which were predominantly treated as sequentially exogenous, are cross sectionally independent. The assumption of cross sectional independence will be satisfied in many settings where the cross sectional units correspond to individuals, firms, etc., and decisions are not interdependent or when dependent units are sampled at random as discussed in Andrews (2005). However in many other settings the assumption of cross-sectional independence may be violated. Examples where it seems appropriate to allow for cross sectional dependence in the exogenous variables may be situations where regressors constitute weighted averages of data that include neighboring units (as is common in spatial analysis), situations where the cross sectional units refer to counties, states, countries or industries, and situations where random sampling from the population is not feasible.

A popular approach to model cross sectional dependence is through common factors; see, e.g., Phillips and Sul (2007, 2003), Bai and Ng (2006a,b), Pesaran (2006), and Andrews (2005) for recent contributions. This represents an important class of models, however they are not geared towards modeling cross sectional interactions.[2] Our approach allows for factor structures in addition to general, unmodelled cross-sectional

---

[2]Bai and Ng (2006a,b) allow for cross sectional correlation in the idiosyncractic disturbances, but assume that the disturbance process is independent of the factors and loadings. The setups considered in the other papers imply that the observations are independent in the cross sectional dimension conditional on the common factors.

dependence. Using the GMM estimator for a linear panel model as an example, we illustrate that conventional inference methods remain valid under the conditions of our central limit theory when samples are not iid in the cross-section. These results extend findings in Andrews (2005) to situations where samples are not iid even after conditioning on a common factor. Given that our assumptions allow for factor structures, our limit theory involves and accommodates random norming. Technically this is achieved by establishing stable convergence in distribution and not just convergence in distribution for the underlying vector of sample moments. To this end we prove a martingale central limit theorem for stable convergence by extending results of Hall and Heyde (1980) to allow for non-nested σ-fields that naturally arise in our setting.

Another popular approach to model cross sectional dependence is to allow for spatial interactions in terms of spatial lags as is done in Cliff-Ord type models. Dynamic panel data models with spatial interactions have recently been considered by Mutl (2006), and Yu, de Jong and Lee (2007, 2008). All of those papers assume that the exogenous variables are fixed constants and thus maintain strict exogeneity. The methodology developed in this paper should be helpful in developing estimation theory for Cliff-Ord type spatial dynamic panel data models with sequentially exogenous regressors.

While some of the classical literature on dynamic panel data models allowed for cross sectional correlation in the exogenous variables, this was, to the best of our knowledge, always combined with the assumption that the exogenous variables are strictly exogenous. This may stem from the fact that strict exogeneity conveniently allows the use of limit theorems conditional on all of the exogenous variables. There are many important cases where the strict exogeneity assumption does not hold, and regressors, apart from time-lagged endogenous variables, or other potential instruments are only sequentially exogenous. Examples given by Keane and Runkle (1992) include rational expectations models or models with predetermined choice variables as regressors. Other example are the effects of children on the labor force participation of women considered by Arellano and Honore (2001, p. 3237) or the relationship between patents and R&D expenditure studied by Hausman, Hall and Griliches (1984); see, e.g., Wooldridge (1997) for further commentary on strict vs. sequential exogeneity.

Motivated by the above, the main aim of our paper is to develop a general central limit theory for sample moments of a panel data set, where we allow for cross sectional dependence in the explanatory variables and disturbances (and thus in the dependent variable), while allowing for some of the explanatory variables to be sequentially exogenous. The setup will be sufficiently general to accommodate cross sectional dependence due to common factors and/or spatial interactions.

The paper is organized as follows. In Section 2 we formulate the moment conditions, and give our basic result concerning the limiting distribution of the normalized sample moments. The analysis establishes not only convergence but stable convergence in distribution. In Section 3 we illustrate how the central

limit theory can be applied to efficient GMM estimators for linear panel models. We derive its limiting distribution, and give a consistent estimator for the limiting variance covariance matrix. Concluding remarks are given in Section 4. Basic results regarding stable convergence in distribution as well as all proofs are relegated to the appendices.

2 Central Limit Theory

2.1 Moment Conditions

In the following we develop a central limit theory (CLT) for a vector of sample moments for panel data, allowing for regressors that are only sequentially (rather than strictly) exogenous, while at the same time allowing the data to be cross sectionally dependent. Towards motivating our setup consider the following exemplary panel data model $(i = 1, \ldots, n; t = 1, \ldots, T)$:

$$y_{it} = x_{it}\beta_0 + z_{it}\gamma_0 + \mu_i + u_{it} \tag{1}$$

where y_{it}, x_{it} and z_{it} denote, respectively, the dependent variable, the sequentially exogenous and strictly exogenous explanatory variables (conditional on the unobserved components), $\theta_0 = (\beta_0', \gamma_0')'$ is the vector of unknown parameters, μ_i is an individual specific effect not observed by the econometrician and u_{it} is an unobserved error term. For the purpose of this section we assume that y_{it} and u_{it} are scalar valued, and x_{it} and z_{it} respectively are $1 \times k_x$ and $1 \times k_z$ vectors, and the vectors of parameters are defined conformably. Since x_{it} may include lagged endogenous variables the specification covers dynamic models.

For the CLT developed in this section we assume that sample averages are taken over n, with n tending to infinity and T fixed. We allow for purely cross-sectional data-sets by allowing $T = 1$ in the CLT. However, this condition may need to be strengthened to $T > T_0$ for some $T_0 > 1$ for specific models and data transformations.

The central limit theorem is stated for averages over the cross-section of random variables $x_i = (x_{i1} \ldots, x_{iT})$, $z_i = (z_{i1}, \ldots, z_{iT})$, and $u_i = (u_{i1}, \ldots, u_{iT})$ for $i = 1, \ldots, n$. We assume that the random variables (z_i, x_i, u_i, μ_i) for $i = 1, 2, \ldots$ are defined on a probability space $(\Omega, \mathcal{F}, P)$. Although not explicitly denoted, the random variables are allowed to depend on the sample size n, i.e., to form triangular arrays. Our setup accommodates fairly general forms of cross-sectional dependence. In particular, analogous to Andrews (2005), who considers static models, we allow in each period t for the possibility of regressors and disturbances (and thus for the dependent variable) to be affected by common shocks that are captured by a sigma field $\mathcal{C}_t \subset \mathcal{F}$. In particular, if λ_t denotes a vector of common shocks, then $\mathcal{C}_t = \sigma(\lambda_t)$. Alternatively or concurrently we allow for cross sectional dependence due to "spatial lags" in the sense that

3

some of the variables may be weighted cross-sectional averages of some basic variables.[3] In the following let $x_{it}^o = (x_{i1}, \ldots, x_{it})$, $u_{it}^o = (u_{i1}, \ldots, u_{it})$ and $\mathcal{C}_t^o = \mathcal{C}_1 \vee \ldots \vee \mathcal{C}_t$ where $\vee$ denotes the sigma field generated by the union of two sigma fields. For simplicity we will also write $\mathcal{C}$ for $\mathcal{C}_T^o$ in the following.

Our central limit theory is focused on providing general results concerning the limiting distribution of a vector of sample moments of the form

$$
\begin{align}
m_{(n)} &= n^{-1/2} \sum_{i=1}^{n} m_i, \tag{2} \\
m_i &= \left(h_{i1} u_{i1}^+, \ldots, h_{T^+ i} u_{T^+ i}^+ \right)',
\end{align}
$$

with $T^+ \leq T$, where $h_{it} = (x_{it}^o, z_i)$ denotes a vector of instruments corresponding to t, and $u_i^+ = \left(u_{i1}^+, \ldots, u_{iT^+}^+ \right)$ denotes a vector of transformed disturbances with $u_{it}^+ = \sum_{s=t}^{T} \pi_{ts} u_{is}$, where for the moment, for ease of presentation, we treat the π_{ts} as nonstochastic constants. The class of transformations considered includes first differences, $u_{it}^+ = u_{i,t+1} - u_{it}$, as well as the Helmert transformation, $u_{it}^+ = \alpha_t \left[u_{it} - (u_{i,t+1} + \ldots + u_{iT}) / (T - t) \right]$, $\alpha_t^2 = (T - t)/(T - t + 1)$, for $t = 1, \ldots, T$. As a special case we also have $u_{it}^+ = u_{it}$, for $t = 1, \ldots, T$. Given the fairly general formulation of the sample moments our CLT should be helpful in establishing the limiting distribution of a broad class of estimators. Further generalizations of the setup covering more general formulations of the instruments and transformations are discussed in a remark at the end of this section.

For the subsequent discussion it proves convenient to express the transformed disturbances more compactly as $u_i^{+\prime} = \Pi u_i'$ where Π is a $T^+ \times T$ matrix with t, s-th element π_{ts}. Observe that the lower diagonal elements of Π are zero. Furthermore, let $H_i = diag_{t=1}^{T^+}(h_{it})$, then we can express the moment vectors as

$$
m_i = H_i' u_i^{+\prime} = H_i' \Pi u_i' = \sum_{t=1}^{T} H_i' \pi_t u_{it}, \tag{3}
$$

where π_t denotes the t-th column of Π. Clearly, given the adopted setup, we have $H_i' \pi_t = [\pi_{1t} h_{i1}, \ldots, \pi_{tt} h_{it}, 0]'$.

We next state a set of mild regularity conditions. The conditions accommodate sequentially exogenous and strictly exogenous regressors, and allow those regressors to be cross-sectionally dependent. As one potential source of cross-sectional dependence the conditions account for the possibility of common shocks.

Definition 1 *Define the following sub-σ-fields of $\mathcal{F}$:*

$$
\mathcal{B}_{n,i,t} = \sigma \left\{ \left(x_{tj}^o, z_j, u_{t-1j}^o, \mu_j \right)_{j=1}^{n}, u_{-i,t} \right\},
$$

---

[3]We note that spatial lags will generally depend on the sample size, which motivates why the variables are allowed to form triangular arrays.

with $u_{-i,t} = (u_{1t}, \ldots, u_{i-1,t}, u_{i+1,t}, \ldots u_{nt})$ and $(i = 1, \ldots, n)$

$$
\begin{aligned}
\mathcal{F}_{n,0} &= \mathcal{C}, \\
\mathcal{F}_{n,i} &= \sigma \left\{ \left(x_{j1}^o, z_j, \mu_j \right)_{j=1}^n, (u_{j1})_{j=1}^{i-1} \right\} \vee \mathcal{C}, \\
\mathcal{F}_{n,n+i} &= \sigma \left\{ \left(x_{j2}^o, z_j, u_{j1}^o, \mu_j \right)_{j=1}^n, (u_{j2})_{j=1}^{i-1} \right\} \vee \mathcal{C}, \\
&\vdots \\
\mathcal{F}_{n,(T-1)n+i} &= \sigma \left\{ \left(x_{jT}^o, z_j, u_{j,T-1}^o, \mu_j \right)_{j=1}^n, (u_{jT})_{j=1}^{i-1} \right\} \vee \mathcal{C}.
\end{aligned}
\tag{4}
$$

We maintain the following assumption.

Assumption 1 *For some $\delta > 0$ and some finite constant K (which is taken, w.l.o.g., to be greater then one) the following conditions hold for all $t = 1, \ldots, T$, $i = 1, \ldots, n$, $n \geq 1$:*
(a) The $2 + \delta$ absolute moments of the random variables $x_{it}, z_{it}, u_{it},$ and μ_i exist, and the moments are uniformly bounded by K. In addition,

$$
E \left[|u_{it}|^{2+\delta} \mid \mathcal{B}_{n,i,t} \vee \mathcal{C} \right] \leq K.
\tag{5}
$$

(b) The following conditional moment restrictions hold:

$$
E \left[u_{it} \mid \mathcal{B}_{n,i,t} \vee \mathcal{C}_t^o \right] = 0.
\tag{6}
$$

(c) Let $\tilde{V}_{(n)} = \sum_{t=1}^T \tilde{V}_{t,n}$ with $\tilde{V}_{t,n} = n^{-1} \sum_{i=1}^n E \left[u_{it}^2 \mid \mathcal{F}_{n,(t-1)n+i} \right] H_i' \pi_t \pi_t' H_i$. There exists a matrix $V = \sum_{t=1}^T V_t$, where V has finite elements and is positive definite a.s., V_t is $\mathcal{C}$ measurable, and $\tilde{V}_{t,n} - V_t \xrightarrow{p} 0$ as $n \to \infty$.

Assumption 1(a) ensures the existence of various expectations considered subsequently. The condition in Assumption 1(b) that the conditional mean of the disturbances is zero implies, of course, that $E[u_{it}] = 0$ and that $cov(u_{it}, u_{j,s}) = 0$ for $i \neq j$ and/or $t \neq s$. Most importantly, observing that $\pi_{\tau t} h_{i\tau}$ for $\tau \leq t$ and thus $H_i' \pi_t$ are $\mathcal{B}_{n,i,t} \vee \mathcal{C}_t^o$ measurable, this cross-sectional conditional moment condition implies that

$$
E[m_i] = 0.
\tag{7}
$$

Furthermore, in light of this condition we have $E \left[m_i m_j' \right] = 0$ for $i \neq j$, and

$$
E \left[m_i m_i' \right] = \sum_{t=1}^T \sum_{s=1}^T E \left[u_{it} u_{is} H_i' \pi_t \pi_s' H_i \right] = \sum_{t=1}^T E \left[u_{it}^2 H_i' \pi_t \pi_t' H_i \right],
\tag{8}
$$

5

since $E u_{it} u_{is} H_i' \pi_t \pi_s' H_i = E \left\{ E \left[u_{it} \mid \mathcal{F}_{n,(t-1)n+i} \right] u_{is} H_i' \pi_t \pi_s' H_i \right\} = 0$ for $s < t$. The requirement in Assumption 1(a) that the conditional $2 + \delta$ moments are bounded could be replaced by an assumption that bounds higher order moments of the unconditional distribution.

Towards interpreting Assumption 1(c) consider the matrix of second order sample moments

$$V_{(n)} = n^{-1} \sum_{i=1}^{n} m_i m_i'. \tag{9}$$

Then in light of (8) and since $E \left[u_{it}^2 H_i' \pi_t \pi_t' H_i \right] = E \left\{ E \left[u_{it}^2 \mid \mathcal{F}_{n,(t-1)n+i} \right] H_i' \pi_t \pi_t' H_i \right\}$ we have $E \left[V_{(n)} \right] = E \left[\tilde{V}_{(n)} \right]$. Assumption 1(c) holds under a variety of low level conditions with

$$V_t = \operatorname{plim} n^{-1} \sum_{i=1}^{n} E \left[u_{it}^2 H_i' \pi_t \pi_t' H_i \mid \mathcal{C} \right],$$

given that $E \left\{ E \left[u_{it}^2 \mid \mathcal{F}_{n,(t-1)n+i} \right] H_i' \pi_t \pi_t' H_i \mid \mathcal{C} \right\} = E \left[u_{it}^2 H_i' \pi_t \pi_t' H_i \mid \mathcal{C} \right]$. An example of such a low level condition is the conditional i.i.d. assumption of Andrews (2005), in which case $V_t = E \left[u_{it}^2 H_i' \pi_t \pi_t' H_i \mid \mathcal{C} \right]$. Alternatively, one could postulate explicit factor structures for the elements of h_{it} and allow for cross-sectional mixing. A third possibility is to postulate cross-sectional stationarity and appeal to the ergodic theorem. If in addition one assumes that $E \left[u_{it}^2 \mid \mathcal{F}_{n,(t-1)n+i} \right] = \sigma^2$, where σ^2 is constant, we have $\tilde{V}_{t,n} = \sigma^2 n^{-1} \sum_{i=1}^{n} H_i' \pi_t \pi_t' H_i$ and $\tilde{V}_{t,n} - V_t \xrightarrow{p} 0$ can then be implied solely from convergence assumptions on the second order sample moments of the instruments h_{it}. We allow for the possibility that there are no common factors by allowing for $\mathcal{C}_t = \{\emptyset, \Omega\}$. In that case, V is a matrix of fixed coefficients.

The moment condition (6) in Assumption 1(b) is formulated for a situation where the common factors are only sequentially exogenous. The next condition strengthens (6) by requiring that the common factors are orthogonal to all innovations.

Assumption 2 *The following conditional moment restrictions hold:*

$$E \left[u_{it} \mid \mathcal{B}_{n,i,t} \vee \mathcal{C} \right] = 0. \tag{10}$$

Remark 1 *Condition (10) implies (6) because $\mathcal{B}_{n,i,t} \vee \mathcal{C}_t^o \subset \mathcal{B}_{n,i,t} \vee \mathcal{C}$. The moment condition (10) is satisfied in models where the common factors are strictly exogenous. As remarked, our analysis includes the important case where no common factors are present by allowing $\mathcal{C}_t = \{\emptyset, \Omega\}$, as is typical in the spatial literature.[4] In this case conditions (6) and (10) are identical, and Assumption 2 is automatically implied by Assumption 1. An xample of a model for u_{it}, where u_{it} depends on common shocks λ_t and satisfies (10) is $u_{it} = \varepsilon_{it} \lambda_t$ where (ε_{jt}) and (λ_t) are independent and $E \left[\varepsilon_{it} \mid \mathcal{B}_{n,i,t} \right] = 0$.*

---

[4] See, e.g., Baltagi at al. (2003, 2009), Kapoor et al. (2007), Lee and Yu (2010a), and Yu et al. (2007, 2008). For recent reviews, see, e.g., Anselin (2010) and Lee and Yu (2010b).

Remark 2 *To further illustrate the relevance of the key conditional moment restrictions postulated above Kuersteiner and Prucha (2009) consider a class of private information games recently analyzed by Rust (1994), Aguirregabiria and Mira (2007) and Bajari, Benkard and Levin (2007). Estimators discussed by these authors rely on independent samples from repeated play of the same game. Kuersteiner and Prucha (2009) show how to formulate moment conditions that are suitable for our central limit theory and avoid the need for independent sampling.*

When (6) holds but not (10) several cases leading to different limiting distributions for the central limit theorem below can be distinguished. To establish general results for a limit theory for the sample moment vector we consider the following basic conditions.

Assumption 3 *Let $\tilde{m}_i = \left(E\left[h_{i1} u_{i1}^+ \mid \mathcal{F}_{n,i} \right], ..., E\left[h_{T+i} u_{T+i}^+ \mid \mathcal{F}_{n,(T+-1)n+i} \right] \right)'$ and $b_n = n^{-1} \sum_{i=1}^{n} \tilde{m}_i$. One of the following statements holds:*
(a) $b_n \xrightarrow{p} b$ where b is finite a.s. and $\mathcal{C}$ measurable.
(b) $\sqrt{n} b_n \xrightarrow{p} b$ where b is finite a.s. and $\mathcal{C}$ measurable.
(c) $\sqrt{n} b_n \xrightarrow{p} 0$.

Remark 3 *Assumption 2 implies that $b_n = 0$, and thus Assumption 2 automatically implies Assumption 3(c). If no common shocks are present, Assumption 3(c) is also automatically implied by Assumption 1.*

Remark 4 *The specification of the instruments as $h_{it} = (x_{it}^o, z_i)$ was chosen for expositional simplicity. Clearly the above discussion also applies if the vectors of instruments h_{it} are more generally allowed to be $\sigma\left[\{x_{it}^o, z_i\}_{i=1}^n \right]$ measurable functions of the regressors. (The dimension of the vectors may depend on t, but not on n.) The above discussion also applies if the π_{st}, $s \leq t$, depend on θ_0 and are measurable w.r.t. $\sigma\left[\{x_{it}^o, z_i\}_{i=1}^n \right] \vee \mathcal{C}_t^o$. The crucial property, pertaining to the instruments and the data transformation, that is used in establishing the moment condition (7) is that $H_i' \pi_t$ is $\mathcal{B}_{n,i,t} \vee \mathcal{C}_t^o$ measurable, which clearly is the case even under the above generalized specifications. We note further that the sample moment vector (2) is a function of the true disturbances u_{it}, and thus the specific functional form of the model (1) does not enter in the derivation of the limiting distribution of the sample moment vector. (Of course, it affects the limiting distribution of a corresponding GMM estimator.) Thus the central limit theory developed below should be a useful basic module for establishing the limiting distribution of GMM estimators for a fairly general class of possibly nonlinear dynamic models with cross sectional dependencies, and a fairly general class of data transformations, including forward transformations that allow for time-varying unit specific effects.*[5]

---

[5]Examples include forward looking variants of transformations considered by Ahn, Lee and Schmidt (2001,2006).

2.2 Central Limit Theory

In this Section we establish the limiting distribution of the moment vector $m_{(n)} = n^{-1/2} \sum_{i=1}^{n} m_i$ and then give a discussion of the strategy by which the result is derived. In fact, we not only establish convergence in distribution of $m_{(n)}$, but we establish $\mathcal{C}$-stable convergence in distribution of $m_{(n)}$, which allows us to establish the joint limiting distribution for $(m_{(n)}, A)$ for any matrix valued random variable A that is $\mathcal{C}$ measurable. Establishing joint limits is a requirement for the continuous mapping theorem to apply.[6] Applying the continuous mapping theorem allows us then to establish the limiting distribution of transformations of $m_{(n)}$, e.g., of $Am_{(n)}$ or $m'_{(n)}Am_{(n)}$, which is important for establishing the limiting distribution of typical estimators and test statistics. In particular, the developed limit theory allows us to accommodate random norming, where the need for random norming arises from the presence of the common factors represented by $\mathcal{C}$.

To prove stable convergence in distribution of $m_{(n)}$ we first establish a general central limit theorem for zero mean, square integrable martingale arrays $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ with differences X_{ni}, which we expect to be useful in many other contexts. We next present a formal definition of stable convergence in distribution, cp. Daley and Vere-Jones (1988, p. 644).

Definition 2 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{B}(\mathbb{R}^p)$ denote the Borel σ-field on $\mathbb{R}^p$. If $\{Z_n : n = 1, 2, \ldots\}$ and Z are $\mathbb{R}^p$-valued random vectors on $(\Omega, \mathcal{F}, P)$, and $\mathcal{F}_0$ is a σ-field such that $\mathcal{F}_0 \subset \mathcal{F}$, then*

$$Z_n \xrightarrow{d} Z \ (\mathcal{F}_0\text{-stably})$$

if for all $U \in \mathcal{F}_0$ and all $A \in \mathcal{B}(\mathbb{R}^p)$ with $P(Z \in \partial A) = 0$,

$$P(\{Z_n \in A\} \cap U) \to P(\{Z \in A\} \cap U)$$

as $n \to \infty$.

Remark 5 *In the following we will apply Definition 2 to establish stable convergence for $Z_n = S_{nk_n}$. The definition generalizes the definition of Hall and Heyde (1980, p. 56) to allow for stable convergence on a sub σ-field $\mathcal{F}_0$ rather than on $\mathcal{F}$. Restricting stable convergence to $\mathcal{F}_0$ is important in our setting because σ-fields $\mathcal{F}_{ni}$ do not satisfy condition (3.21) maintained by the central limit theorem of Hall and Heyde (1980, p. 58). We note that when $\mathcal{F}_0 = \{0, \Omega\}$, $\mathcal{F}_0$-stable convergence in distribution is convergence in distribution. Thus the former always implies the latter.*

We now present the central limit theorem for martingale arrays. The theorem extends results in Hall and Heyde (1980) by establishing stable convergence without requiring that the σ-fields $\mathcal{F}_{ni}$ are nested in

---

[6]See, e.g., Pötscher and Prucha (2001), pp. 207, for discussions and examples of this requirement.

the sense of Hall and Heyde's condition (3.21). This can be achieved by restricting stable convergence to a 'small' sub-σ-field $\mathcal{F}_0 \subset \mathcal{F}$. In our application, $\mathcal{F}_0$ can be taken to equal $\mathcal{C}$. The central limit theorem then provides exactly the type of stable and joint convergence we need.

Theorem 1 *Let $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ be a zero mean, square integrable martingale array with differences X_{ni}. Let $\mathcal{F}_0 = \cap_{n=1}^{\infty} \mathcal{F}_{n0}$ with $\mathcal{F}_{n0} \subseteq \mathcal{F}_{n1}$ for each n and $E[X_{n1}|\mathcal{F}_{n0}] = 0$ and let η^2 be an a.s. finite random variable measurable w.r.t. $\mathcal{F}_0$. If*

$$\max_i |X_{ni}| \xrightarrow{p} 0, \tag{11}$$

$$\sum_{i=1}^{k_n} X_{ni}^2 \xrightarrow{p} \eta^2 \tag{12}$$

and

$$E\left(\max_i X_{ni}^2\right) \text{ is bounded in } n \tag{13}$$

then

$$S_{nk_n} = \sum_{i=1}^{k_n} X_{ni} \xrightarrow{d} Z \quad (\mathcal{F}_0\text{-stably})$$

where the random variable Z has characteristic function $E\left[\exp\left(-\frac{1}{2}\eta^2 t^2\right)\right]$. In particular, $S_{nk_n} \xrightarrow{d} \eta\xi$ ($\mathcal{F}_0$-stably) where $\xi \sim N(0,1)$ is independent of η (possibly after redefining all variables on an extended probability space).

In the following let $V_{(n)} = n^{-1} \sum_{i=1}^{n} m_i m_i'$ as defined above, and let $p = T^+ \left[(T^+ + 1)k_x/2 + Tk_z\right]$ denote the dimension of $m_{(n)}$. We now state the main result of the paper, a central limit theorem for the moment vector $m_{(n)}$.

Theorem 2 *(a) Suppose Assumptions 1 and 2 hold. Then*

$$m_{(n)} \xrightarrow{d} V^{1/2}\xi \quad (\mathcal{C}\text{-stably}), \tag{14}$$

where $\xi \sim N(0, I_p)$, and ξ and $\mathcal{C}$ (and thus ξ and V) are independent.
(b) Let A be some $p_* \times p$ matrix that is $\mathcal{C}$ measurable with finite elements and rank p_* a.s.. Suppose Assumption 1 and either Assumption 2 or 3(c) hold, then

$$Am_{(n)} \xrightarrow{d} (AVA')^{1/2}\xi_*, \tag{15}$$

where $\xi_* \sim N(0, I_{p_*})$, and ξ_* and $\mathcal{C}$ (and thus ξ_* and AVA') are independent. If Assumptions 1 and 3(a) hold, then

$$A\left(m_{(n)} - \sqrt{n}b_n\right) \xrightarrow{d} (AVA')^{1/2}\xi_* \tag{16}$$

9

and $Am_{(n)}$ diverges. If Assumptions 1 and 3(b) hold then

$$Am_{(n)} \xrightarrow{d} (AVA')^{1/2}\xi_* + Ab \tag{17}$$

(c) Suppose Assumption 1 and either Assumption 2 or 3(c) hold. Suppose furthermore that

$$n^{-1}\sum_{i=1}^{n} u_{it}u_{is}H_i'\pi_t\pi_s'H_i \xrightarrow{p} 0 \quad for \quad t \neq s, \tag{18}$$

then $V_{(n)} - V \xrightarrow{p} 0$ and $V_{(n)}^{-1/2}m_{(n)} \xrightarrow{d} \xi \sim N(0, I_p)$.

The proof of Theorem 2 employs Theorem 1 for martingale difference arrays and employes Propositions A.1 and A.2 in the Appendix in conjunction with the Cramer-Wold device. We illustrate the proof strategy here and assume for the remainder of this section that Assumption 2 holds to simplify the argument. A detailed proof is given in the appendix. Let $\lambda = (\lambda_1', \lambda_2', \ldots, \lambda_{T^+}')'$ be some nonstochastic vector, where λ_t is of dimension $(tk_x + k_z) \times 1$ and where $\lambda'\lambda = 1$. Then

$$\lambda'm_{(n)} = n^{-1/2}\sum_{i=1}^{n} c_i'u_i \tag{19}$$

where

$$c_i' = (c_{i1}, \ldots, c_{iT}) = \lambda'H_i'\Pi.$$

Given that the lower diagonal elements of the $T^+ \times T$ matrix Π are zero it follows that $c_{it} = \lambda'H_i'\pi_t = \sum_{s=1}^{\min(t,T^+)} \pi_{st}\lambda_s'h_{is}'$, and thus c_{it} is a function only of x_{it}^o, z_i, and elements of λ and Π.[7] Next, let $X_{n,1} = 0$, and for $i = 1, \ldots, n$ define

$$X_{n,i+1} = n^{-1/2}c_{i1}u_{i1},$$
$$X_{n,n+i+1} = n^{-1/2}c_{i2}u_{i2},$$
$$\vdots \tag{20}$$
$$X_{n,(T-1)n+i+1} = n^{-1/2}c_{iT}u_{iT},$$

such that we can express $\lambda'm_{(n)}$ as

$$\lambda'm_{(n)} = \sum_{v=1}^{Tn+1} X_{n,v}. \tag{21}$$

Towards establishing the limiting distribution of $\sum_{v=1}^{Tn+1} X_{n,v}$ through the martingale difference array CLT we need to construct appropriate information sets. This is accomplished in Definition 1 where $\mathcal{F}_{n,v}$ is defined.

---

[7] More generally the proof also holds if we allow for h_{it} and the π_{st} to be as discussed in the remark at the end of the last sub-section.

Clearly the construction of the information sets is such that $\mathcal{F}_{n,v-1} \subseteq \mathcal{F}_{n,v}$, $X_{n,v}$ is $\mathcal{F}_{n,v}$-measurable, and $E[X_{n,v} \mid \mathcal{F}_{n,v-1}] = 0$ in light of Assumption 2 and observing that $\mathcal{F}_{n,(t-1)n+i} \subseteq \mathcal{B}_{n,i,t} \vee \mathcal{C}$. The proof of the first part of Theorem 2 in the appendix proceeds by verifying that under the maintained assumptions the martingale difference array $\{X_{n,v}, \mathcal{F}_{n,v}, 1 \leq v \leq Tn+1, n \geq 1\}$ satisfies all remaining conditions postulated by Theorem 1. Given that this CLT delivers stable convergence in distribution (and not just convergence in distribution) the claims in (14) and (15) then follow from Propositions A.1 and A.2.

The claim that $V_{(n)} - V \overset{p}{\to} 0$ in the second part of the theorem follows in essence as a by-product of the proof of the first part, observing that $V_{(n)} = \sum_{t=1}^{T} n^{-1} \sum_{i=1}^{n} u_{it}^2 H_i' \pi_t \pi_t' H_i + o_p(1)$ in light of (18). As remarked above, under Assumption 1(b) we have $E[u_{it} u_{is} H_i' \pi_t \pi_s' H_i] = 0$ for $t \neq s$, condition (18) will hold again under a variety of low level conditions.

Remark 6 *The above construction of the information sets $\mathcal{F}_{n,(t-1)n+i}$ is crucial. At first glance it may seem unusual to include $(u_{jt,n})_{j=1}^{i-1}$ in the information set $\mathcal{F}_{n,(t-1)n+i}$, and one may be tempted to use the information sets $\mathcal{B}_{n,t} \vee \mathcal{C}$ where $\mathcal{B}_{n,t} = \sigma \left\{ \left(x_{jt}^o, z_j, u_{j,t-1}^o, \mu_j \right)_{j=1}^{n} \right\}$ instead. However, we emphasize that it is precisely because of the inclusion of $(u_{jt,n})_{j=1}^{i-1}$ that $X_{n,v}$ is indeed $\mathcal{F}_{n,v}$-measurable for all v, as required by the CLT.[8] Using $\mathcal{B}_{n,t} \vee \mathcal{C}$ for the sequence of information sets would have lead to a violation of this measurability condition. Alternatively, one may have been tempted to use $\mathcal{B}_{n,i,t} \vee \mathcal{C}$ for the sequence of information sets, i.e., to include $u_{-i,t}$ in place of $(u_{jt,n})_{j=1}^{i-1}$. However this would have lead to a violation of the assumption that the information sets are non-decreasing.*

3 GMM Estimators

In this section we illustrate the use of the CLT for the GMM estimator of the commonly used panel data model (1) with fixed effects. For the subsequent discussion it proves convenient to rewrite the model more compactly as

$$y_{it} = w_{it}\theta_0 + \mu_i + u_{it}, \tag{22}$$

where $w_{it} = (x_{it}, z_{it})$ and $\theta_0 = (\beta_0', \gamma_0')'$.

As discussed in the introduction, much of the dynamic panel data literature maintains that the data are distributed i.i.d. in the cross sectional dimension. That is, let $y_i = (y_{i1}, \ldots, y_{iT})$, $x_i = (x_{i1}, \ldots, x_{iT})$, $z_i =$

---

[8]Within the context of establishing the limiting distribution of linear quadratic forms composed of independent disturbances Kelejian and Prucha (2001) employed somewhat related ideas; cp. also Yu et al. (2007, 2008). However their setups differ substantially from ours, and these papers do not consider sequentially exogenous covariates, nor common factors and corresponding stable convergence.

$(z_{i1}, \ldots, z_{iT})$, and $u_i = (u_{i1}, \ldots, u_{iT})$, then in this setting (x_i, z_i, μ_i, u_i) or equivalently (y_i, x_i, z_i, μ_i) would be distributed independently and identically across i. As discussed, this assumption is appropriate for many micro-econometric applications but problematic in many other situations, e.g., where i corresponds to countries, states, regions, industries, etc. Also in many spatial settings it would not make sense to assume that x_i and/or z_i are independent over i because elements of x_i and /or z_i may be weighted averages of characteristics of neighboring units, i.e., be spatial lags in the sense of Cliff and Ord (1973, 1981).[9]

It is of interest to compare Assumption 1 with those typically maintained under the assumption that (x_i, z_i, μ_i, u_i) is i.i.d.. For this discussion we also assume the absence of common factors for simplicity. Clearly, under cross sectional independence the conditions in Assumption 1(b) can be stated equivalently by replacing the conditioning sets by $x_{it}^o, z_i, \mu_i, u_{it-1}^o$. In particular, Assumption 1(b) simplifies to

$$E\left[u_{it} \mid x_{it}^o, z_i, \mu_i, u_{it-1}^o\right] = 0. \tag{23}$$

This is in contrast to the assumption that

$$E\left[u_{it} \mid x_{it}^o, z_i, \mu_i\right] = 0, \tag{24}$$

which is typically maintained in the literature under cross sectional independence. Clearly condition (23) rules out autocorrelation in the disturbances, even if x_{it} does not contain a lagged endogenous variable, while condition (24) does not.[10] Of course, if the model is dynamic and linear also condition (24) rules out autocorrelation in the disturbances. In this case conditions (23) and (24) are equivalent, since then x_{it}^o already incorporates the information contained in u_{it-1}^o through the lagged values of the dependent variable. We note that the need to include u_{it-1}^o in the conditioning information set stems from the use of a martingale difference CLT, while the i.i.d. case can simply be handled by a CLT for i.i.d. random vectors.

We consider moment estimators that are based on first differences of (22) such that

$$E\left[h_{t-1i}^{\prime} \Delta u_{it}\right] = 0 \text{ for } t = 2, \ldots, T, \tag{25}$$

where Δ is the difference operator. Let $H_i = \operatorname{diag}(h_{i1}, \ldots, h_{i,T-1})$ be the $(T-1) \times (T-1)\left[Tk_x/2 + k_z\right]$ matrix containing the $T-1$ instrument vectors, and let D be the $(T-1) \times T$ matrix such that $Du_i^{\prime} =$

---

[9]As noted, to allow for situations where, e.g., a regressor variable corresponding to unit i is a cross sectional weighted average of some basic variable, we allow all variables to be triangular arrays. For example, if the dependent variable is wages and i corresponds to regions, then a regressor may be a weighted average of the unemployment rate of region i and the unemployment rates of surrounding regions, with weights declining with distance. To save on notational cost, we do not explicitly account for the dependence of the observations on n.

[10]Specific forms of autocorrelated disturbances such as AR(1) disturbances could be accommodated by reformulating the moment conditions w.r.t. to the basic innovations entering the disturbance process.

$\Delta u_i := (\Delta u_{i2}, ..., \Delta u_{iT})'$. The sample moment vector $m_{(n)}$ corresponding to the above moment conditions is then given by (2), where m_i is as in (3) with $\Pi = D$ and $T^+ = T - 1$. For the following discussion it proves helpful to define $\Delta y_i := (\Delta y_{i2}, ..., \Delta y_{iT})'$ and $\Delta w_i := (\Delta w'_{i2}, ..., \Delta w'_{iT})'$.

The GMM estimator corresponding to the moment conditions (25) is defined as

$$\tilde{\theta}_n = \left(G'_n \tilde{\Xi}_n G_n\right)^{-1} G_n \tilde{\Xi}_n g_n$$

where $G_n = n^{-1} \sum_{i=1}^{n} H'_i \Delta w_i$, $g_n = n^{-1} \sum_{i=1}^{n} H'_i \Delta y_i$, and $\tilde{\Xi}_n$ is some weight matrix. The above expression for the GMM estimator is consistent with expressions given in the dynamic panel data literature under the assumption of cross sectional independence of the observations; compare, e.g., Arellano and Bond (1991).

The asymptotic distribution of the GMM estimator $\tilde{\theta}_n$ is well established in case the observations are i.i.d. In case all explanatory variables (outside of time lags of the dependent variable) are strictly exogenous, cross sectional dependence between the explanatory variables across units can also be handled readily by performing the analysis conditional on all strictly exogenous variables, i.e., by conditioning on $z_1, \ldots, z_n$. This is essentially the approach taken in the early literature on static panel data models. This approach was also taken by Mutl (2006), and Yu, de Jong and Lee (2007, 2008) in analyzing Cliff-Ord type spatial dynamic panel data models. However, as discussed, strict exogeneity rules out many important cases where u_{it} affects future values of the regressor.

In the following we utilize the theory developed in Section 2 to derive the asymptotic distribution of $\tilde{\theta}_n$ for situations where some or all regressors are allowed to be only sequentially rather than strictly exogenous, while at the same time allowing the data to be cross sectionally dependent. Correspondingly our analysis will postulate Assumption 1, which also accommodates cross sectional dependence due to sequentially exogenous common factors.

For completeness we discuss the structure of the matrices in Assumption 1(c) as implied by the moment conditions (25). The matrices $\tilde{V}_{t,n}$ are readily seen to be block diagonal of the form $\tilde{V}_{t,n} = n^{-1} \sum_{i=1}^{n} diag(0, \ldots, 0, S_{it}, 0, \ldots, 0)$ with $S_{i1} = \sigma_{i1}^2 h'_{i1} h_{i1}$,

$$S_{it} = \sigma_{it}^2 \begin{bmatrix} h'_{i,t-1} h_{i,t-1} & -h'_{i,t-1} h_{it} \\ -h'_{it} h_{i,t-1} & h'_{it} h_{it} \end{bmatrix}, t = 2, \ldots, T - 1,$$

and $S_{iT} = \sigma_{iT}^2 h'_{i,T-1} h_{i,T-1}$, where $\sigma_{it}^2 = E\left[u_{it}^2 \mid \mathcal{F}_{n,(t-1)n+i}\right]$. (Note that there is partial overlap between the non-zero blocks of $\tilde{V}_{t,n}$ and $\tilde{V}_{t-1,n}$.) If we assume additionally Assumption 2, the limiting matrices corresponding to $\tilde{V}_{t,n}$ will typically be of the form $V_t = \text{plim}_{n \to \infty} n^{-1} \sum_{i=1}^{n} diag(0, \ldots, 0, E[S_{it} \mid \mathcal{C}], 0, \ldots, 0)$, where by a iterated expectations argument it is seen that $E[S_{i1} \mid \mathcal{C}] = E\left[u_{i1}^2 h'_{i1} h_{i1} \mid \mathcal{C}\right]$,

$$E[S_{it} \mid \mathcal{C}] = E\left\{u_{it}^2 \begin{bmatrix} h'_{i,t-1} h_{i,t-1} & -h'_{i,t-1} h_{it} \\ -h'_{it} h_{i,t-1} & h'_{it} h_{it} \end{bmatrix} \mid \mathcal{C}\right\}, t = 2, \ldots, T - 1,$$

13

and $E\left[S_{iT} \mid \mathcal{C}\right] = E\left[u_{iT}^2 h_{i,T-1}' h_{i,T-1} \mid \mathcal{C}\right]$.

The next theorem establishes the basic asymptotic properties of the GMM estimator $\tilde{\theta}_n$ when common factors are either strictly exogenous or have an asymptotically negligible effect on the estimator bias. Under the same conditions we also give a result in Theorem 5 that can be utilized to establish the limiting distribution of test statistics, allowing for random norming corresponding to the common factors captured by $\mathcal{C}$.

Theorem 3 *Suppose Assumption 1, and either Assumption 2 or 3(c) hold, and that $G_n \xrightarrow{p} G$, $\tilde{\Xi}_n \xrightarrow{p} \Xi$, where G and Ξ are $\mathcal{C}$-measurable, G and Ξ have finite elements and G has full column rank and Ξ is positive definite a.s.*
(a) Then

$$n^{1/2}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} \Psi^{1/2}\xi, \quad \text{as } n \to \infty,$$

where ξ is independent of $\mathcal{C}$ (and hence of Ψ), $\xi \sim N(0, I_{k_x+k_z})$, and

$$\Psi = (G'\Xi G)^{-1} G'\Xi V \Xi G (G'\Xi G)^{-1}.$$

If in addition, $E\left[u_{it}^2 \mid \mathcal{F}_{n,(t-1)n+i}\right] = \sigma^2$ for a constant σ^2 holds, then $V = \sigma^2 \operatorname{plim}_{n\to\infty}\left(n^{-1}\sum_{i=1}^n H_i' D D' H_i\right)$.

(b) Suppose B is some $q \times k_x + k_z$ matrix that is $\mathcal{C}$ measurable with finite elements and rank q a.s., then

$$Bn^{1/2}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} (B\Psi B')^{1/2}\xi_*,$$

where $\xi_* \sim N(0, I_q)$, and ξ_* and $\mathcal{C}$ (and thus ξ_* and $B\Psi B'$) are independent.

The next result considers cases where the common factors are only sequentially exogenous, i.e., only (6) but not necessarily (10) holds, and where the resulting effect on the bias of the estimator is asymptotically non-negligible. The first part of the theorem considers a case where the estimator is inconsistent and converges to a random limit while the second part of the theorem covers a case where the estimator is root-n consistent but not asymptotically mixed normal.

Theorem 4 *Suppose Assumption 1 holds, and that $G_n \xrightarrow{p} G$, $\tilde{\Xi}_n \xrightarrow{p} \Xi$, where G and Ξ are $\mathcal{C}$-measurable, G and Ξ have finite elements and G has full column rank and Ξ is positive definite a.s.*
(a) If in addition Assumption 3(a) holds then

$$n^{1/2}(\tilde{\theta}_n - \theta_0 - (G'\Xi G)^{-1} G'\Xi b_n) \xrightarrow{d} \Psi^{1/2}\xi$$

14

and $\tilde{\theta}_n - \theta_0 \xrightarrow{p} (G'\Xi G)^{-1}G'\Xi b.$

(b) If in addition Assumption 3(b) holds then

$$n^{1/2}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} \Psi^{1/2}\xi + (G'\Xi G)^{-1}G'\Xi b.$$

For efficiency (conditional on $\mathcal{C}$) we can select $\Xi = V^{-1}$, in which case $\Psi = (G'V^{-1}G)^{-1}$. To construct a feasible efficient GMM estimator consider the following estimator for V

$$\widetilde{V}_{\Delta(n)} = n^{-1}\sum_{i=1}^{n} H_i'\widetilde{\Delta u_i}\widetilde{\Delta u_i}'H_i$$

where $\widetilde{\Delta u_i} = (\widetilde{\Delta u_{i2}}, \ldots, \widetilde{\Delta u_{iT}})$ with $\widetilde{\Delta u_{it}} = \Delta y_{it} - \Delta w_{it}\tilde{\theta}_n$, and $\tilde{\theta}_n$ is the initial GMM estimator with weight matrix $\tilde{\Xi}_n = I$, or some other consistent estimator for θ_0. The GMM estimator with weight matrix $\tilde{\Xi}_n = \widetilde{V}_{\Delta(n)}^{-1}$ is then given by,

$$\hat{\theta}_n = \left(G_n'\widetilde{V}_{\Delta(n)}^{-1}G_n\right)^{-1}G_n\widetilde{V}_{\Delta(n)}^{-1}g_n.$$

The above expression for the GMM estimator $\hat{\theta}_n$ is again consistent with expressions given in the dynamic panel data literature under the assumption of cross sectional independence of the observations.

By Theorem 3 the limiting variance covariance matrix of $\hat{\theta}_n$ is then given by $\Psi = (G'V^{-1}G)^{-1}$, which can be estimated consistently by $\hat{\Psi}_n = \left(G_n'\widetilde{V}_{\Delta(n)}^{-1}G_n\right)^{-1}$, provided it is shown that $\widetilde{V}_{\Delta(n)}$ is indeed a consistent estimator for V. Next let R be a $q \times (k_x + k_z)$ full row rank matrix and r a $q \times 1$ vector, and consider the Wald statistic

$$T_n = \left\|\left(R\hat{\Psi}R'\right)^{-1/2}\sqrt{n}(R\hat{\theta}_n - r)\right\|^2$$

to test the null hypothesis $H_0 : R\theta_0 = r$ against the alternative $H_1 : R\theta_0 \neq r$. The next theorem establishes the consistency of $\widetilde{V}_{\Delta(n)}$, and shows that T_n is distributed asymptotically chi-square, even if Ψ is allowed to be random due to the presence of common factors represented by $\mathcal{C}$.

Theorem 5 *Suppose the assumptions of Theorem 3 hold, that $V_{(n)} = n^{-1}\sum_{i=1}^{n} H_i'\Delta u_i \Delta u_i'H_i - V \xrightarrow{p} 0$, that $\tilde{\theta}_n \xrightarrow{p} \theta_0$, and that the fourth moments of u_{it}, x_{it} and z_{it} are uniformly bounded by a finite constant. Then $\widetilde{V}_{\Delta(n)} - V \xrightarrow{p} 0$, and*

$$\hat{\Psi}_n^{-1/2}\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \xi \sim N(0, I_{k_x+k_z}).$$

Furthermore

$$P\left(T_n > \chi_{q,1-\alpha}^2\right) \to \alpha$$

where $\chi_{q,1-\alpha}^2$ is the $1 - \alpha$ quantile of the chi-square distribution with q degrees of freedom.

15

A sufficient condition for $V_{(n)} - V \xrightarrow{p} 0$ is given in Theorem 2. Theorem 5 extends results of Andrews (2005) to the case of generally dependent cross-sectional samples. It establishes that conventional statistical tests remain valid under the postulated assumptions.

4   Conclusion

Most of the literature on dynamic panel data models either assumed independence in the cross sectional dimension, or treats regressors as strictly exogenous when allowing for cross sectional correlation. While the assumption that observations are independently distributed in the cross sectional dimension is appropriate for many applications, there are many applications where this assumption will likely be violated. Also, as discussed in the introduction, there are many important cases where the strict exogeneity assumption does not hold, and regressors, apart from time-lagged endogenous variables, or other potential instruments are only sequentially exogenous.

Against this background the paper develops a new CLT for martingale difference sequences, and applies it to develop a general central limit theory for the sample moment vector (and transformation thereof) of a dynamic panel data model, where the regressors may be cross sectionally correlated as well as sequentially exogenous (but not necessarily strictly exogenous). The paper shows how the new CLT can be utilized in establishing the limiting distribution of GMM estimators in the generalized setting.

The specification of cross sectional dependence is kept general. In particular, the methodology developed in this paper will have natural application within the context of spatial/cross sectional interaction models. A widely used class of spatial models originates from Cliff and Ord (1973, 1981). In those models, which are often referred to as Cliff-Ord type models, spatial/cross sectional interaction and dependencies are modeled in terms of spatial lags, which represent weighted averages of observations from neighboring units. The weights are typically modeled as inversely related to some distance. Since space does not have to be geographic space, those models are fairly generally applicable and have been used in a wide range of empirical research; for a collection of recent contributions including references to applied work see, e.g., Baltagi, Kelejian and Prucha (2007). The methodology developed in this paper also allows for common factor as a potential source of cross sectional dependence.

A   Appendix A: Proofs for Section 2

A.1   Stable Convergence in Distribution

The following proposition is proven in Daley and Vere-Jones (1988), p. 645-646.

Proposition A.1   *Let $\{Z_n : n = 1, 2, \ldots\}$, Z and $\mathcal{F}_0$ be as in Definition 2. Then the following conditions are equivalent:*

(i) $Z_n \xrightarrow{d} Z$ ($\mathcal{F}_0$-stably).

(ii) For all $\mathcal{F}_0$-measurable P-essentially bounded random variables ζ and all bounded continuous functions $h : \mathbb{R}^p \to \mathbb{R}$,

$$E\left[\zeta h(Z_n)\right] \to E\left[\zeta h(Z)\right] \qquad \text{as } n \to \infty.$$

(iii) For all real valued $\mathcal{F}_0$-measurable random variables ϑ, the pair (Z_n, ϑ) converges jointly in distribution to the pair (Z, ϑ).

(iv) For all bounded continuous functions $g : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}$, and all real valued $\mathcal{F}_0$-measurable random variables ϑ,

$$g(Z_n, \vartheta) \xrightarrow{d} (Z, \vartheta) \ (\mathcal{F}_0\text{-stably}).$$

(v) For all real vectors $t \in \mathbb{R}^p$ and all $\mathcal{F}_0$-measurable P-essentially bounded random variables ζ

$$E\left[\zeta \exp(it'Z_n)\right] \to E\left[\zeta \exp(it'Z)\right] \qquad \text{as } n \to \infty.$$

The following proposition is helpful in establishing the limiting distribution of random vectors under random norming.

Proposition A.2   *Let $\{Z_n : n = 1, 2, \ldots\}$, and $\mathcal{F}_0$ be as in Definition 2, and let V be a $\mathcal{F}_0$-measurable, a.s. finite and positive definite $p \times p$ matrix. Suppose for $\lambda \in \mathbb{R}^p$ with $\lambda'\lambda = 1$ we have*

$$\lambda'Z_n \xrightarrow{d} v_\lambda^{1/2} \xi_\lambda \ (\mathcal{F}_0\text{-stably}),$$

with $v_\lambda = \lambda'V\lambda$, where ξ_λ is independent of $\mathcal{F}_0$ (and thus of V) and $\xi_\lambda \sim N(0,1)$, and consequently the characteristic function of $v_\lambda^{1/2} \xi_\lambda$ is given by $\phi_\lambda(s) = E\left[\exp\{-\frac{1}{2}(\lambda'V\lambda)s^2\}\right]$, $s \in \mathbb{R}$.

(a) Then

$$Z_n \xrightarrow{d} V^{1/2} \xi \ (\mathcal{F}_0\text{-stably}),$$

17

where ξ is independent of $\mathcal{F}_0$ (and thus of V) and where $\xi \sim N(0, I_p)$. The characteristic function of $V^{1/2}\xi$ is given by $\phi(t) = E\left[\exp\{-\frac{1}{2}(t'Vt)\}\right]$.

(b) Let A be some $p_* \times p$ matrix that is $\mathcal{F}_0$-measurable, a.s. finite and has full row rank. Then

$$AZ_n \xrightarrow{d} AV^{1/2}\xi$$

where ξ is as defined in part (a), and hence also

$$AZ_n \xrightarrow{d} (AVA')^{1/2}\xi_*$$

where ξ_* is independent of $\mathcal{F}_0$ (and thus of AVA') and where $\xi_* \sim N(0, I_{p_*})$. The characteristic function of $AV^{1/2}\xi$ and $(AVA')^{1/2}\xi_*$ is given by $\phi_*(s) = E\left[\exp\{-\frac{1}{2}(s'AVA's)\}\right]$, $s \in \mathbb{R}^{p_*}$.

Proof of Proposition A.2. (a) In light of Proposition A.1(v), for all real vectors $s \in \mathbb{R}$ and all $\mathcal{F}_0$-measurable P-essentially bounded random variables ζ we have $E\left[\zeta\exp(is\lambda'Z_n)\right] \to E\left[\zeta\exp(isv_\lambda^{1/2}\xi_\lambda)\right]$ as $n \to \infty$. Since ζ and V are $\mathcal{F}_0$-measurable, and $\xi_\lambda \sim N(0, 1)$ we have

$$E\left[\zeta\exp(isv_\lambda^{1/2}\xi_\lambda)\right] = E\left[\zeta E\left[\exp(isv_\lambda^{1/2}\xi_\lambda) \mid \mathcal{F}_0\right]\right] = E\left[\zeta\exp\{-\frac{1}{2}(s^2\lambda'V\lambda)\}\right],$$

and thus

$$E\left[\zeta\exp(is\lambda'Z_n)\right] \to E\left[\zeta\exp\{-\frac{1}{2}(s^2\lambda'V\lambda)\}\right]. \tag{A.1}$$

Now consider some $t \in \mathbb{R}^p$, then by analogous argumentation as above, $E\left[\zeta\exp(it'V^{1/2}\xi)\right] = E\left[\zeta\exp\{-\frac{1}{2}(t'Vt)\}\right]$. In light of Proposition A.1(v) it thus suffices to show that

$$E\left[\zeta\exp(it'Z_n)\right] \to E\left[\zeta\exp\{-\frac{1}{2}(t'Vt)\}\right]. \tag{A.2}$$

Choosing λ and s be such that $t = s\lambda$, this is seen to hold in light of (A.1).

(b) Since A is $\mathcal{F}_0$-measurable it follows from Proposition A.1(iii) that (Z_n, A) converge jointly in distribution to $(V^{1/2}\xi, A)$. Hence by the continuous mapping theorem $AZ_n \xrightarrow{d} AV^{1/2}\xi$. The characteristic function of $AV^{1/2}\xi$ is given by

$$\begin{aligned}
\phi_*(s) &= E\left[\exp\left(is'AV^{1/2}\xi\right)\right] = E\left\{E\left[\exp\left(is'AV^{1/2}\xi\right) \mid \mathcal{F}_0\right]\right\} \\
&= E\left[\exp\left\{-\frac{1}{2}s'AVA's\right\}\right],
\end{aligned}$$

observing that $AV^{1/2}\xi$ conditional on $\mathcal{F}_0$ is distributed multivariate normal $(0, AVA')$. Recognizing that $\phi_*(s)$ is also the characteristic function of $(AVA')^{1/2}\xi_*$ completes the proof. $\blacksquare$

A.2 Proof of Martingale Central Limit Theorem

Proof of Theorem 1. The proof follows, with appropriate modifications, the strategy used by Hall and Heyde (1980, pp. 57-58 and pp. 60) in proving their Lemma 3.1 and Theorem 3.2. First suppose that η^2 is a.s. bounded such that for some $C > 1$,

$$P\left(\eta^2 < C\right) = 1. \tag{A.3}$$

Define $X'_{ni} = X_{ni}\mathbf{1}\left\{\sum_{j=1}^{i-1} X_{nj}^2 \leq 2C\right\}$ with $X'_{n1} = X_{n1}$, and $S'_{ni} = \sum_{j=1}^{i} X'_{nj}$ for $1 \leq i \leq k_n$.

By assumption $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ is a zero mean, square integrable martingale array with differences X_{ni}, i.e., (i) S_{ni} is measurable w.r.t. $\mathcal{F}_{ni}$, (ii) $E[S_{ni}] = 0$ and $E[S_{ni}^2] < \infty$, (iii) $E[S_{ni} \mid \mathcal{F}_{nj}] = S_{nj}$ a.s. for all $1 \leq j < i$. The differences are defined as $X_{n1} = S_{n1}$, and $X_{ni} = S_{ni} - S_{ni-1}$ for $2 \leq i \leq k_n$. Clearly for any $j \leq i$ the random variable X_{nj} is measurable w.r.t. to $\mathcal{F}_{ni}$, since $\mathcal{F}_{nj} \subseteq \mathcal{F}_{ni}$. Furthermore $E[X_{ni} \mid \mathcal{F}_{nj}] = 0$ for $0 \leq j < i$ and $1 \leq i \leq k_n$, since $E[X_{n1} \mid \mathcal{F}_{n0}] = 0$ by assumption, and for $2 \leq j < i$

$$
\begin{aligned}
E[X_{ni} \mid \mathcal{F}_{nj}] &= E[S_{ni} - S_{ni-1} \mid \mathcal{F}_{nj}] = E[E[S_{ni} - S_{ni-1} \mid \mathcal{F}_{ni-1}] \mid \mathcal{F}_{nj}] \\
&= E[(S_{ni-1} - S_{ni-1}) \mid \mathcal{F}_{nj}] = 0.
\end{aligned}
$$

We now establish that $\{S'_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ is also a zero mean, square integrable martingale array with, by construction, differences X'_{ni}. Since the random variables $X_{n1}, \ldots, X_{ni}$ are measurable w.r.t. $\mathcal{F}_{ni}$, clearly $S'_{ni} = \sum_{j=1}^{i} X'_{nj} = \sum_{j=1}^{i} X_{nj}\mathbf{1}\left\{\sum_{l=1}^{j-1} X_{nl}^2 \leq 2C\right\}$ is measurable w.r.t. $\mathcal{F}_{ni}$. Also, since $|S'_{ni}| \leq |S_{ni}|$ clearly $E[S'^2_{ni}] \leq E[S_{ni}^2] < \infty$. Next observe that $E[X'_{n1} \mid \mathcal{F}_{n0}] = E[X_{n1} \mid \mathcal{F}_{n0}] = 0$ by assumption, and for $2 \leq j < i$

$$E\left[X'_{ni} \mid \mathcal{F}_{nj}\right] = E\left[E(X'_{ni} \mid \mathcal{F}_{ni-1}) \mid \mathcal{F}_{nj}\right] = E\left[[E(X_{ni} \mid \mathcal{F}_{ni-1})]\mathbf{1}\left\{\sum_{j=1}^{i-1} X_{nj}^2 \leq 2C\right\} \mid \mathcal{F}_{nj}\right] = 0.$$

By iterated expectations $E[X'_{ni}] = 0$ and thus $E[S'_{ni}] = 0$. Furthermore for $1 \leq j < i$

$$E\left[S'_{ni} \mid \mathcal{F}_{nj}\right] = \sum_{l=1}^{i} E\left[X'_{nl} \mid \mathcal{F}_{nj}\right] = \sum_{l=1}^{j} X'_{nl} = S'_{nj}.$$

This verifies that $\{S'_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ is indeed a zero mean, square integrable martingale array.

Next let $U^2_{nk_n} = \sum_{i=1}^{k_n} X_{ni}^2$, then clearly $P(U^2_{nk_n} > 2C) \to 0$ in light of (12). Consequently

$$P(X'_{ni} \neq X_{ni} \text{ for some } i \leq k_n) \leq P(U^2_{nk_n} > 2C) \to 0, \tag{A.4}$$

19

which in turn implies $P(S'_{nk_n} \neq S_{nk_n}) \to 0$, and furthermore

$$E\left[\left|\zeta \exp(itS'_{nk_n}) - \zeta \exp(itS_{nk_n})\right|\right] \to 0$$

for any P-essentially bounded and $\mathcal{F}_0$-measurable random variable ζ. Consequently by Proposition A.1(v), $S_{nk_n} \xrightarrow{d} Z$ ($\mathcal{F}_0$-stably) iff $S'_{nk_n} \xrightarrow{d} Z$ ($\mathcal{F}_0$-stably). Observe furthermore that in view of (A.4) the martingale differences $\{X'_{ni}\}$ satisfy that $\max_i |X'_{ni}| \xrightarrow{p} 0$ and $\sum_{i=1}^{k_n} X'^2_{ni} \xrightarrow{p} \eta^2$. Since $|X'_{ni}| \leq |X_{ni}|$ condition (13) implies furthermore that $E\left[\max_i X'^2_{ni}\right]$ is bounded in n.

We now show that $S'_{nk_n} \xrightarrow{d} Z$ ($\mathcal{F}_0$-stably). Let $U^2_{ni} = \sum_{j=1}^{i} X^2_{nj}$ and $T'_n(t) = \prod_{j=1}^{k_n}\left(1 + itX'_{nj}\right)$ with

$$J_n = \begin{cases} \min\left\{i \leq k_n | U^2_{ni} > 2C\right\} & \text{if } U^2_{nk_n} > 2C \\ k_n & \text{otherwise} \end{cases}.$$

Observing that $X'_{nj} = 0$ for $j > J_n$, and that for any real number a we have $(1 + ia)^2 = (1 + a^2)$ and $\exp(1 + a^2) > 1 + a^2$, it follows that

$$\begin{aligned} E\left[\left|T'_n(t)\right|^2\right] &= E\left[\prod_{j=1}^{k_n}\left(1 + t^2 X'^2_{nj}\right)\right] \leq E\left[\left\{\exp\left(t^2 \sum_{j=1}^{J_n-1} X'^2_{nj}\right)\left(1 + t^2 X'^2_{nJ_n}\right)\right\}\right] \\ &\leq \left\{\exp(2Ct^2)\right\}\left(1 + t^2 E\left[X'^2_{nJ_n}\right]\right). \end{aligned}$$

Since $E\left[X'^2_{nJ_n}\right] \leq E\left[X^2_{nJ_n}\right]$ is uniformly bounded it follows from the above inequality that $E\left[\left|T'_n(t)\right|^2\right]$ is uniformly bounded in n.

Now define $I_n = \exp\left(itS'_{nk_n}\right)$ and $W_n = \exp\left(-\frac{1}{2}t^2 \sum_{i=1}^{k_n} X'^2_{ni} + \sum_{i=1}^{k_n} r\left(tX'_{ni}\right)\right)$ where $r(.)$ is as defined in Hall and Heyde (1980), p. 57. Then

$$I_n = T'_n(t)\exp\left(-\eta^2 t^2/2\right) + T'_n(t)(W_n - \exp\left(-\eta^2 t^2/2\right)). \tag{A.5}$$

By Proposition A.1(v) for $S'_{nk_n} \xrightarrow{d} Z$ ($\mathcal{F}_0$ stably) it is enough to show that

$$E(I_n\zeta) \to E\left[\exp\left(-\eta^2 t^2/2\right)\zeta\right] \tag{A.6}$$

for any any P-essentially bounded $\mathcal{F}_0$-measurable random variable ζ. Because $\mathcal{F}_0 \subset \mathcal{F}_{ni}$ it follows that $\exp\left(-\eta^2 t^2/2\right)\zeta$ is $\mathcal{F}_{ni}$-measurable for all n and $i \leq k_n$. Hence,

$$\begin{aligned} E\left[T'_n(t)\exp\left(-\eta^2 t^2/2\right)\zeta\right] &= E\left[\exp\left(-\eta^2 t^2/2\right)\zeta \prod_j^{k_n}\left(1 + itX'_{nj}\right)\right] \\ &= E\left\{E\left[\exp\left(-\eta^2 t^2/2\right)\zeta \prod_j^{k_n}\left(1 + itX'_{nj}\right)|\mathcal{F}_{nk_n-1}\right]\right\} \\ &= E\left\{\exp\left(-\eta^2 t^2/2\right)\zeta \prod_j^{k_n-1}\left(1 + itX'_{nj}\right) E\left[\left(1 + itX'_{nk_n}\right)|\mathcal{F}_{nk_n-1}\right]\right\} \\ &= E\left\{\exp\left(-\eta^2 t^2/2\right)\zeta \prod_j^{k_n-1}\left(1 + itX'_{nj}\right)\right\} \\ &= \ldots \\ &= E\left\{\exp\left(-\eta^2 t^2/2\right)\zeta E\left[\left(1 + itX'_{n1}\right)|\mathcal{F}_{n0}\right]\right\} = E\left[\exp\left(-\eta^2 t^2/2\right)\zeta\right]. \end{aligned}$$

Thus, in light of (A.5), for (A.6) to hold it suffices to show that

$$E\left[T'_n(t)\left(W_n - \exp\left(-\eta^2 t^2/2\right)\right)\zeta\right] \to 0. \tag{A.7}$$

Let K be some constant such that $P(|\zeta| \le K) = 1$, then $E\left[\left|T'_n(t)\exp\left(-\eta^2 t^2/2\right)\zeta\right|^2\right] \le K^2 E\left[\left|T'_n(t)\right|^2\right]$ is uniformly bounded in n, since $E\left[\left|T'_n(t)\right|^2\right]$ is uniformly bounded as shown above. Observing that $|I_n| = 1$ we also have $E\left[|I_n\zeta|^2\right] \le K^2$. In light of (A.5) it follows furthermore that

$$E\left[\left|T'_n(W_n - \exp\left(-\eta^2 t^2/2\right))\zeta\right|^2\right] \le 2E\left[|I_n\zeta|^2\right] + 2E\left[\left|T'_n(t)\exp\left(-\eta^2 t^2/2\right)\zeta\right|^2\right]$$

is uniformly bounded in n, it follows that $T'_n(t)\left(W_n - \exp\left(-\eta^2 t^2/2\right)\right)\zeta$ is uniformly integrable. Having established uniform integrability, Condition (A.7) now follows since as shown by Hall and Heyde (1980, p. 58), $W_n - \exp\left(-\eta^2 t^2/2\right) \xrightarrow{p} 0$ and thus $T'_n\left(W_n - \exp\left(-\eta^2 t^2/2\right)\right)\zeta \xrightarrow{p} 0$. This completes the proof that $S'_{nk_n} \xrightarrow{d} Z$ ($\mathcal{F}_0$-stably) when η^2 is a.s. bounded.

The case where η^2 is not a.s. bounded can be handled in the same way as in Hall and Heyde (1980, p.62) after replacing their $I(E)$ with ζ.

Let $\xi \sim N(0,1)$ be some random variable independent of $\mathcal{F}_0$, and hence independent of η (possibly after redefining all variables on an extended probability space), then for any P-essentially bounded $\mathcal{F}_0$-measurable random variable ζ we have $E\left[\zeta \exp(it\eta\xi)\right] = E\left[\zeta \exp(-\frac{1}{2}\eta^2 t^2)\right]$ by iterated expectations, and thus $S_{nk_n} \xrightarrow{d} \eta\xi$ ($\mathcal{F}_0$-stably) in light of Proposition A.1(v). $\blacksquare$

A.3   Proof of Central Limit Theorem for Sample Moments

Proof of Theorem 2.   To prove **Part (a)** of the Theorem we use Proposition A.2 and follow the approach outlined after the theorem in the text to derive the limiting distribution of $\lambda' m_{(n)}$. In particular, we consider the representation $\lambda' m_{(n)} = \sum_{v=1}^{k_n} X_{n,v}$ with $k_n = Tn + 1$, defined by (20)-(21), and the corresponding information sets defined in (4). We recall the definitions ($t = 1, \ldots, T, i = 1, \ldots, n$)

$$X_{n,(t-1)n+i+1} = n^{-1/2} c_{it} u_{it},$$
$$\mathcal{F}_{n,(t-1)n+i} = \sigma\left\{\left(x^o_{tj}, z_j, u^o_{t-1j}, \mu_j\right)_{j=1}^n, (u_{jt})_{j=1}^{i-1}\right\} \vee \mathcal{C}$$

with

$$c_{it} = \lambda' H'_i \pi_t = \sum_{s=1}^{\min(t,T^+)} \pi_{st} \lambda'_s h'_{is}. \tag{A.8}$$

and $X_{n,1} = 0$. In the following, let $v = (t-1)n + i + 1$. We also use $\mathcal{F}_{n,0} = \mathcal{C}$, then clearly $\mathcal{F}_{n,0} \subset \mathcal{F}_{n,1}$ and $\mathcal{F}_0 = \cap_{n=1}^\infty \mathcal{F}_{n,0} = \mathcal{C}$. To prove part (a) of the theorem we verify that $\{X_{n,v}, \mathcal{F}_{n,v}, 1 \le v \le Tn + 1, n \ge 1\}$ is a square integrable martingale difference array that satisfies the assumptions maintained by Theorem

1 with $\eta^2 = \lambda'V\lambda$, observing that η^2 be an a.s. finite random variable measurable w.r.t. $\mathcal{F}_0$ in light of Assumption 1.

Observing that c_{it} defined in (A.8) is a function of x_{it}^o, z_i, λ and Π it is readily seen that $X_{n,v}$ is measurable w.r.t. to $\mathcal{F}_{n,v}$. Observing further that $\mathcal{F}_{n,(t-1)n+i} \subseteq \mathcal{B}_{n,i,t} \vee \mathcal{C}$ it follows from (10) that

$$
\begin{aligned}
E\left[X_{n,v} \mid \mathcal{F}_{n,v-1}\right] &= E\left[X_{n,(t-1)n+i+1} \mid \mathcal{F}_{n,(t-1)n+i}\right] \\
&= n^{-1/2}c_{it}E\left[u_{it} \mid \mathcal{F}_{n,(t-1)n+i}\right] = 0.
\end{aligned}
\tag{A.9}
$$

Next consider some γ with $0 \leq \gamma \leq \delta$, then

$$
E\left[|X_{n,v}|^{2+\gamma} \mid \mathcal{F}_{n,v-1}\right] = E\left[|X_{n,(t-1)n+i+1}|^{2+\gamma} \mid \mathcal{F}_{n,(t-1)n+i}\right]
$$
$$
= \frac{1}{n^{1+\gamma/2}}E\left[|c_{it}u_{it}|^{2+\gamma} \mid \mathcal{F}_{n,(t-1)n+i}\right] = \frac{1}{n^{1+\gamma/2}}|c_{it}|^{2+\gamma}E\left[|u_{it}|^{2+\gamma} \mid \mathcal{F}_{n,(t-1)n+i}\right].
$$

For $\gamma = 0$ this implies that

$$
E\left[|X_{n,v}|^2 \mid \mathcal{F}_{n,v-1}\right] = \frac{1}{n}|c_{it}|^2 E\left[|u_{it}|^2 \mid \mathcal{F}_{n,(t-1)n+i}\right],
\tag{A.10}
$$

and for $\gamma = \delta$ we have

$$
E\left[|X_{n,v}|^{2+\delta} \mid \mathcal{F}_{n,v-1}\right]
\tag{A.11}
$$
$$
= \frac{1}{n^{1+\delta/2}}|c_{it}|^{2+\delta}E\left[|u_{it}|^{2+\delta} \mid \mathcal{F}_{n,(t-1)n+i}\right] \leq \frac{K}{n^{1+\delta/2}}|c_{it}|^{2+\delta}.
$$

The last inequality follows from (5), observing again that $\mathcal{F}_{n,(t-1)n+i} \subseteq \mathcal{B}_{n,i,t} \vee \mathcal{C}$.

Let $V_{nk_n}^2 = \sum_{v=1}^{k_n} E\left[X_{n,v}^2 \mid \mathcal{F}_{n,v-1}\right]$ and $U_{nk_n}^2 = \sum_{v=1}^{k_n} X_{n,v}^2$, and consider the following conditions:

$$
\sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta}\right] \to 0,
\tag{A.12}
$$

$$
V_{nk_n}^2 = \sum_{v=1}^{k_n} E\left[X_{n,v}^2 \mid \mathcal{F}_{n,v-1}\right] \xrightarrow{p} \eta^2,
\tag{A.13}
$$

$$
\sup_n E\left[V_{nk_n}^{2+\delta}\right] = \sup_n E\left[\sum_{v=1}^{k_n} E\left[X_{n,v}^2 \mid \mathcal{F}_{n,v-1}\right]\right]^{1+\delta/2} < \infty.
\tag{A.14}
$$

We next show that these conditions are sufficient for Assumptions (11),(12), and (13) of Theorem 1. As discussed by Hall and Heyde (1980, p. 53) Condition (11) is equivalent to:

$$
\text{for any } \varepsilon > 0, \qquad \sum_{v=1}^{k_n} X_{n,v}^2 \mathbf{1}\left(|X_{n,v}| > \varepsilon\right) \xrightarrow{p} 0.
$$

Condition (11) is now seen to hold since

$$\sum_{v=1}^{k_n} E\left[X_{n,v}^2 \mathbf{1}\left(|X_{n,v}| > \varepsilon\right)\right] = \sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta} \mathbf{1}\left(|X_{n,v}| > \varepsilon\right) / |X_{n,v}|^\delta\right] \leq \varepsilon^{-\delta} \sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta}\right] \to 0$$

in light of condition (A.12). By an analogous argument we also have:

$$\text{for any } \varepsilon > 0, \qquad \sum_{v=1}^{k_n} E\left[X_{n,v}^2 \mathbf{1}\left(|X_{n,v}| > \varepsilon\right) \mid \mathcal{F}_{n,v-1}\right] \xrightarrow{p} 0.$$

Observing that Condition (A.14) implies that V_{nk_n} is uniformly integrable it now follows from Hall and Heyde (1980, Theorem 2.23) that

$$E\left[\left|V_{nk_n}^2 - U_{nk_n}^2\right|\right] \to 0. \tag{A.15}$$

Condition (12) is now seen to hold since for any $\varepsilon > 0$

$$P\left(\left|\sum_{v=1}^{k_n} X_{n,v}^2 - \eta^2\right| \geq \varepsilon\right) \leq P\left(\left|V_{nk_n}^2 - \eta^2\right| \geq \varepsilon/2\right) + P\left(\left|V_{nk_n}^2 - U_{nk_n}^2\right| \geq \varepsilon/2\right) \to 0$$

in light of (A.13) and (A.15). Condition (13) is seen to hold since $E\left[U_{nk_n}^2\right] = E\left[V_{nk_n}^2\right]$ is uniformly bounded in light of Condition (A.14), using Lyapunov's inequality.

We next verify Conditions (A.12), (A.13) and (A.14). Utilizing (A.10) we have

$$
\begin{aligned}
\sum_{v=1}^{k_n} E\left[X_{n,v}^2 | \mathcal{F}_{n,v-1}\right] &= \sum_{t=1}^{T} n^{-1} \sum_{i=1}^{n} c_{it}^2 E\left[u_{it}^2 \mid \mathcal{F}_{n,(t-1)n+i}\right] \\
&= \sum_{t=1}^{T} n^{-1} \sum_{i=1}^{n} E\left[u_{it}^2 \mid \mathcal{F}_{n,(t-1)n+i}\right] \lambda' H_i' \pi_t \pi_t' H_i \lambda \\
&= \lambda' \tilde{V}_n \lambda \xrightarrow{p} \lambda' V \lambda = \eta^2,
\end{aligned} \tag{A.16}
$$

by Assumption 1(c). This verifies (A.13).

Next observe that for all t and i

$$|c_{it}|^{2+\delta} \leq (T^+)^{1+\delta} \sum_{s=1}^{T^+} \left|\pi_{st} \lambda_s' h_{is}'\right|^{2+\delta} \leq (T^+)^{1+\delta} K_\pi^{2+\delta} \sum_{s=1}^{T^+} \left[\lambda_s' h_{is}' h_{is} \lambda_s\right]^{1+\delta/2}, \tag{A.17}$$

using inequality (1.4.3) in Bierens (1994), and where K_π is a bound for the absolute elements of Π.

Now let $a = (a_k)$ be some non-stochastic $p \times 1$ vector with $|a_k| \leq K_a$, and let $b = (b_k)$ be some $1 \times p$ random vector with $E\left[|b_k|^{2+\delta}\right] \leq K_b$, then

$$
\begin{aligned}
E\left[\left|a' b' b a\right|^{1+\delta/2}\right] &= p^\delta \sum_{l=1}^{p} \sum_{k=1}^{p} |a_k|^{1+\delta/2} |a_l|^{1+\delta/2} E\left[|b_k|^{1+\delta/2} |b_l|^{1+\delta/2}\right] \\
&\leq p^{2+\delta} K_a^{2+\delta} K_b
\end{aligned} \tag{A.18}
$$

23

using inequality (1.4.3) in Bierens (1994), and observing that

$$E\left[|b_k|^{1+\delta/2}|b_l|^{1+\delta/2}\right] \le \left[E\,|b_k|^{2+\delta}\,E\,|b_l|^{2+\delta}\right]^{1/2} \le K_b$$

by the Schwartz inequality.

Since $\lambda'\lambda = 1$ all elements of λ are bounded by one. Furthermore, by Assumption 1(a) the $2+\delta$ absolute moments of the element of h_{is} are uniformly bounded by some finite constant K. Observing further that the dimensions of λ_s and h_{is} are bounded by $Tk_x + k_z$ it follows from applying inequality (A.18) that $E\left[\lambda'_s h'_{is} h_{is} \lambda_s\right]^{1+\delta/2} \le [Tk_x + k_z]^{2+\delta} K$, and thus in light of (A.17):

$$E\left[|c_{it}|^{2+\delta}\right] \le (T^+)^{2+\delta} K_\pi^{2+\delta} [Tk_x + k_z]^{2+\delta} K. \tag{A.19}$$

Utilizing (A.11) it follows that

$$\sum_{v=1}^{k_n} E\,|X_{n,v}|^{2+\delta} = \sum_{v=1}^{k_n} E\left\{E\left[|X_{n,v}|^{2+\delta} \mid \mathcal{F}_{n,v-1}\right]\right\} \tag{A.20}$$

$$\le \frac{(T^+)^{2+\delta} K_\pi^{2+\delta} [Tk_x + k_z]^{2+\delta} K^2}{n^{1+\delta/2}}(Tn+1) \to 0$$

as $n \to \infty$. This establishes condition (A.12).

Next observe that in light of (A.10) and Assumption 1(a)

$$E\left[V_{nk_n}^{2+\delta}\right] = E\left[\sum_{v=1}^{k_n} E\left[X_{n,v}^2 \mid \mathcal{F}_{n,v-1}\right]\right]^{1+\delta/2} = E\left[\sum_{v=1}^{k_n} \frac{1}{n}|c_{it}|^2 E\left[|u_{it}|^2 \mid \mathcal{F}_{n,(t-1)n+i}\right]\right]^{1+\delta/2}$$

$$\le KE\left[\sum_{v=1}^{k_n} \frac{1}{n}|c_{it}|^2\right]^{1+\delta/2} \le K\frac{k_n^{\delta/2}}{n^{1+\delta/2}}\sum_{v=1}^{k_n} E\left[|c_{it}|^{2+\delta}\right]$$

using again inequality (1.4.3) in Bierens (1994). In light of (A.19) it follows further that

$$E\left[V_{nk_n}^{2+\delta}\right] \le \frac{(T^+)^{2+\delta} K_\pi^{2+\delta} [Tk_x + k_z]^{2+\delta} K^2}{n^{1+\delta/2}}(Tn+1)^{1+\delta/2}$$

$$\le (T+1)^{1+\delta/2}(T^+)^{2+\delta} K_\pi^{2+\delta} [Tk_x + k_z]^{2+\delta} K^2,$$

which establishes (A.14). Of course, in light of (A.10) and (A.19) the above discussion also establishes that $X_{n,v}$ is square integrable.

Having verified all conditions of Theorem 1 it follows from that theorem that $\lambda' m_{(n)} \xrightarrow{d} v_\lambda^{1/2} Z_\lambda$ ($\mathcal{C}$-stably), where $v_\lambda = \lambda' V \lambda$, Z_λ is independent of $\mathcal{C}$ (and thus of V) and $Z_\lambda \sim N(0,1)$, possibly after redefining all variables on an extended probability space.

24

To prove **Part (b)** of the theorem we note that for $\tilde{u}_{it} = u_{it} - E\left[u_{it} \mid \mathcal{F}_{n,(t-1)n+i}\right]$ it follows by construction that $\tilde{u}_{it}$ is $\mathcal{F}_{n,(t-1)n+i+1}$ measurable, $E\left[\tilde{u}_{it} \mid \mathcal{F}_{n,(t-1)n+i}\right] = 0$ and

$$
\begin{aligned}
E\left[|\tilde{u}_{it}|^{2+\delta} \mid \mathcal{F}_{n,v-1}\right] &\leq E\left[\left(|u_{it}| + \left|E\left[u_{it} \mid \mathcal{F}_{n,(t-1)n+i}\right]\right|\right)^{2+\delta} \mid \mathcal{F}_{n,(t-1)n+i}\right] \\
&\leq 2^{1+\delta} E\left[|u_{it}|^{2+\delta} + \left|E\left[u_{it} \mid \mathcal{F}_{n,(t-1)n+i}\right]\right|^{2+\delta} \mid \mathcal{F}_{n,(t-1)n+i}\right] \\
&\leq 2^{2+\delta} E\left[|u_{it}|^{2+\delta} \mid \mathcal{F}_{n,(t-1)n+i}\right] \leq 2^{2+\delta} K
\end{aligned}
$$

such that the proof in Part (a) can be applied to $\tilde{X}_{n,(t-1)n+i+1} = X_{n,(t-1)n+i+1} - E\left[X_{n,(t-1)n+i+1} \mid \mathcal{F}_{n,(t-1)n+i}\right]$ to show that, given Assumption 1 holds,

$$
\lambda'\left(m_{(n)} - \sqrt{n}b_n\right) \xrightarrow{d} v_\lambda^{1/2} Z_\lambda \ (\mathcal{C}\text{-stably}). \tag{A.21}
$$

Note that when additionally Assumption 2 holds, $E\left[X_{n,(t-1)n+i+1} \mid \mathcal{F}_{n,(t-1)n+i}\right] = 0$ and thus $b_n = 0$, the result follows trivially from Part (a). Of course, (A.21) implies further that $v_\lambda^{1/2} Z_\lambda$ has the characteristic function $\phi_\lambda(s) = E\exp\{-\frac{1}{2}(\lambda'V\lambda)s^2\}$, $s \in \mathbb{R}$. By Proposition Proposition A.2 it follows from (A.21) that

$$
\left(m_{(n)} - \sqrt{n}b_n\right) \xrightarrow{d} V^{1/2}\xi \ (\mathcal{C}\text{-stably}), \tag{A.22}
$$

where ξ is independent of C (and thus of V) and where $\xi \sim N(0, I_p)$, *and that*

$$
A\left(m_{(n)} - \sqrt{n}b_n\right) \xrightarrow{d} (AVA')^{1/2}\xi_*. \tag{A.23}
$$

The claim in (15) holds observing that under Assumption 2 $\sqrt{n}b_n = 0$ and under Assumption 3(c) we have $Am_{(n)} = A\left(m_{(n)} - \sqrt{n}b_n\right) + o_p(1)$. Obviously (A.23) also established the claim in (16) under Assumption 3(a). The claim that under Assumption 3(a) $Am_{(n)}$ diverges is obvious since under this assumption $b_n \xrightarrow{p} b$ and thus $Am_{(n)} = A\left(m_{(n)} - \sqrt{n}b_n\right) + \sqrt{n}Ab_n = O_p(n^{1/2})$, observing that the first term on the r.h.s. as well as Ab_n are $O_p(1)$. To verify the claim in (17) under Assumption 3(b) observe that by Proposition A.1(iii)

$$
\left(m_{(n)} - \sqrt{n}b_n, b, A\right) \xrightarrow{d} \left(V^{1/2}\xi, b, A\right).
$$

Since $\sqrt{n}b_n - b \xrightarrow{p} 0$ it follows furthermore that

$$
\left(m_{(n)} - \sqrt{n}b_n, \sqrt{n}b_n - b, b, A\right) \xrightarrow{d} \left(V^{1/2}\xi, 0, b, A\right).
$$

Having established joint convergence in distribution the claim in (17) now follows in light of the continuous mapping theorem.

To prove **Part (c)** we next show that $V_{(n)} \xrightarrow{p} V$. Since $V_{(n)}$ and V are symmetric it suffices to show that $\lambda'V_{(n)}\lambda \xrightarrow{p} \lambda'V\lambda$, where λ is defined as above. Observe that

$$
\lambda'V_{(n)}\lambda = n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} u_{it}u_{is}\lambda'H_i'\pi_t\pi_s'H_i\lambda = \sum_{v=1}^{k_n} X_{n,v}^2 + o_p(1),
$$

25

since $n^{-1} \sum_{i=1}^{n} u_{it} u_{is} H_i' \pi_t \pi_s' H_i \xrightarrow{p} 0$ for $t \neq s$ by assumption. As shown after (A.15), $\sum_{v=1}^{k_n} X_{n,v}^2 \xrightarrow{p}$ $\eta^2 = \lambda' V \lambda$. Consequently, $\lambda' V_{(n)} \lambda \xrightarrow{p} \lambda' V \lambda$, and thus $V_{(n)} \xrightarrow{p} V$. In light of Assumption1(c) we have $V_{(n)}^{-1/2} \xrightarrow{p} V^{-1/2}$. Since by part (a) $m_{(n)}$ converges $\mathcal{C}$-stably in distribution to a random vector, it follows that $(V_{(n)}^{-1/2} - V^{-1/2}) m_{(n)} = o_p(1)$ and hence $V_{(n)}^{-1/2} m_{(n)} = V^{-1/2} m_{(n)} + o_p(1) \xrightarrow{d} (V^{-1/2} V V^{-1/2}) \xi = \xi$. ∎

B   Appendix B: Proofs for Section 3

Proof of Theorem 3.   From the definition of the GMM estimator and the model given in (22) we have

$$
\begin{aligned}
n^{1/2}(\tilde{\theta}_n - \theta_0) &= \left(G_n' \tilde{\Xi}_n G_n \right)^{-1} G_n' \tilde{\Xi}_n m_{(n)}, \\
B n^{1/2}(\tilde{\theta}_n - \theta_0) &= B \left(G_n' \tilde{\Xi}_n G_n \right)^{-1} G_n' \tilde{\Xi}_n m_{(n)}, \\
m_{(n)} &= n^{1/2} g_n = n^{-1/2} \sum_{i=1}^{n} H_i' \Delta u_i = n^{-1/2} \sum_{i=1}^{n} H_i' \Pi u_i',
\end{aligned}
$$

with $\Pi = D$. It now follows directly from part (a) of Theorem 2 that $m_{(n)}$ converges $\mathcal{C}$-stably in distribution to a random vector. Thus

$$
\begin{aligned}
n^{1/2}(\tilde{\theta}_n - \theta_0) &= (G' \Xi G)^{-1} G' \Xi m_{(n)} + o_p(1), \\
B n^{1/2}(\tilde{\theta}_n - \theta_0) &= B(G' \Xi G)^{-1} G' \Xi m_{(n)} + o_p(1).
\end{aligned}
\tag{B.1}
$$

First assume that Assumption 2 holds. Observing further that by assumption Ξ, G, B and V are $\mathcal{C}$-measurable and that under the maintained assumptions $\tilde{\Xi}_n \xrightarrow{p} \Xi$ and $G_n \xrightarrow{p} G$ it follows from Part (a) of Theorem 2 and Proposition A.1(iii) that, jointly,

$$
\left(\tilde{\Xi}_n - \Xi, G_n - G, B, \Xi, G, m_{(n)} \right) \xrightarrow{d} \left(0, 0, B, \Xi, G, V^{1/2} \xi \right)
\tag{B.2}
$$

where $\xi \sim N(0, I_{k_x + k_z})$. Because $G' \Xi G$ is positive definite $a.s.$ and $\Psi = (G' \Xi G)^{-1} G' \Xi V \Xi G (G' \Xi G)^{-1}$ is positive definite $a.s.$, it follows furthermore from (B.2) and the continuous mapping theorem that

$$
\begin{aligned}
n^{1/2}(\tilde{\theta}_n - \theta_0) &\xrightarrow{d} \Psi^{1/2} \xi, \\
B n^{1/2}(\tilde{\theta}_n - \theta_0) &\xrightarrow{d} (B \Psi B)^{1/2} \xi_*,
\end{aligned}
$$

where ξ and ξ_* ares independent of $\mathcal{C}$, and $\xi \sim N(0, I_{k_x + k_z})$ and $\xi_* \sim N(0, I_q)$. If $E\left[u_{it}^2 \mid \mathcal{F}_{n,(t-1)n+i} \right] = \sigma^2$, then $V = p\lim_{n \to \infty} \tilde{V}_{(n)} = \sigma^2 p\lim_{n \to \infty} n^{-1/2} \sum_{i=1}^{n} H_i' DD' H_i$ as claimed.

Now assume that Assumption 3(c) hold. Then it follows from (A.22) and Proposition A.1(iii) that

$$
\left(\tilde{\Xi}_n - \Xi, G_n - G, B, \Xi, G, m_{(n)} - \sqrt{n} b_n, \sqrt{n} b_n \right) \xrightarrow{d} \left(0, 0, B, \Xi, G, V^{1/2} \xi, 0 \right).
$$

The remainder of the argument follows in the same way as in the first part of the proof.   ∎

26

Proof of Theorem 4. As in the proof of Theorem 3 we still have (B.1). For Part (a) we observe that in light of (A.22) and Proposition A.1(iii)

$$\left(\widetilde{\Xi}_n - \Xi, G_n - G, B, \Xi, G, \left(m_{(n)} - \sqrt{n}b_n\right)\right) \xrightarrow{d} \left(0, 0, B, \Xi, G, V^{1/2}\xi\right).$$

The result then follows immediately from (B.1) and the continuous mapping theorem. For Part (b) we observe that in light of (A.22) and Proposition A.1(iii)

$$\left(\widetilde{\Xi}_n - \Xi, G_n - G, \Xi, G, \left(m_{(n)} - \sqrt{n}b_n\right), b\right) \xrightarrow{d} \left(0, 0, B, \Xi, G, V^{1/2}\xi, b\right)$$

and the result again follows from (B.1) and the continuous mapping theorem. ∎

Proof of Theorem 5. We first show that $\widetilde{V}_{\Delta(n)} \xrightarrow{p} V$. Since $\widetilde{\Delta u}_{it} = \Delta u_{it} - \Delta w_{it}(\tilde{\theta}_n - \theta_0)$ we have

$$
\begin{aligned}
\widetilde{V}_{\Delta(n)} &= n^{-1}\sum_{i=1}^{n} H_i'\widetilde{\Delta u}_i\widetilde{\Delta u}_i'H_i = V_{(n)} - n^{-1}\sum_{i=1}^{n} H_i'\Delta w_i(\tilde{\theta}_n - \theta_0)\Delta u_i'H_i \\
&\quad - n^{-1}\sum_{i=1}^{n} H_i'\Delta u_i(\tilde{\theta}_n - \theta_0)'\Delta w_i'H_i + n^{-1}\sum_{i=1}^{n} H_i'\Delta w_i(\tilde{\theta}_n - \theta_0)(\tilde{\theta}_n - \theta_0)'\Delta w_i'H_i.
\end{aligned}
$$

By assumption $V_{(n)} \xrightarrow{p} V$. For the $(T-1) \times (T-1)$ matrix $n^{-1}\sum_{i=1}^{n} H_i'\Delta w_i(\tilde{\theta}_n - \theta_0)\Delta u_i'H_i$ consider the typical t, s-block given by

$$n^{-1}\sum_{i=1}^{n} h_{it}'h_{is}\Delta u_{is}\Delta w_{it}(\tilde{\theta}_n - \theta_0) \tag{B.3}$$

$$
\begin{aligned}
&= n^{-1}\sum_{i=1}^{n} h_{it}'h_{is}u_{is}w_{it}(\tilde{\theta}_n - \theta_0) - n^{-1}\sum_{i=1}^{n} h_{it}'h_{is}u_{is-1}w_{it}(\tilde{\theta}_n - \theta_0) \\
&\quad - n^{-1}\sum_{i=1}^{n} h_{it}'h_{is}u_{is}w_{it-1}(\tilde{\theta}_n - \theta_0) + n^{-1}\sum_{i=1}^{n} h_{it}'h_{is}u_{is-1}w_{it-1}(\tilde{\theta}_n - \theta_0)
\end{aligned}
$$

where

$$\left\| n^{-1}\sum_{i=1}^{n} h_{it}'h_{is}u_{is}w_{it}(\tilde{\theta}_n - \theta_0) \right\| \leq \left\| \tilde{\theta}_n - \theta_0 \right\| n^{-1}\sum_{i=1}^{n} \left\| h_{it}'h_{is} \right\| |u_{is}| \|w_{it}\|$$

and

$$
\begin{aligned}
E\left[\left\| h_{it}'h_{is} \right\| |u_{is}| \|w_{it}\|\right] &\leq E\left[\left\| h_{it}'h_{is} \right\|^2\right]^{1/2} E\left[|u_{is}|^2 \|w_{it}\|^2\right]^{1/2} \\
&\leq E\left[\|h_{it}\|^4\right]^{1/4} E\left[\|h_{is}\|^4\right]^{1/4} E\left[|u_{si}|^4\right]^{1/4} E\left[\|w_{ti}\|^4\right]^{1/4}
\end{aligned}
$$

by repeated application of the Cauchy-Schwarz inequality. By the boundedness of fourth moments all expectations are bounded and thus $n^{-1}\sum_{i=1}^{n} \|h_{it}'h_{is}\| |u_{is}| \|w_{it}\| = O_p(1)$. Since by assumption $\left\| \tilde{\theta}_n - \theta_0 \right\| = o_p(1)$ it follows that $n^{-1}\sum_{i=1}^{n}\sum_{i=1}^{n} h_{it}'h_{is}u_{is}w_{it}(\tilde{\theta}_n - \theta_0) = o_p(1)$. The other terms appearing in B.3 can

27

be treated in the same way. Therefore $\widetilde{V}_{\Delta(n)} \overset{p}{\rightarrow} V$ as claimed, and furthermore $\hat{\Psi}_n = \left(G_n' \widetilde{V}_{\Delta(n)}^{-1} G_n\right)^{-1} \overset{p}{\rightarrow}$ $\Psi = (G'V^{-1}G)^{-1}$.

By part (a) of Theorem 3 it now follows that

$$n^{1/2}(\hat{\theta}_n - \theta_0) \overset{d}{\rightarrow} \Psi^{1/2}\xi, \tag{B.4}$$

where ξ is independent of $\mathcal{C}$ (and hence of Ψ), $\xi \sim N(0, I_{k_x+k_z})$. In light of (B.4), the consistency of $\hat{\Psi}_n$, and given that R has full row rank q it follows furthermore that under H_0

$$\begin{aligned}
\left(R\hat{\Psi}R'\right)^{-1/2} n^{1/2}(R\hat{\theta}_n - r) &= \left(R\hat{\Psi}R'\right)^{-1/2} R\left[n^{1/2}(\hat{\theta}_n - \theta_0)\right] \\
&= (R\Psi R')^{-1/2} R\left[n^{1/2}(\hat{\theta}_n - \theta_0)\right] + o_p(1).
\end{aligned}$$

Since $B = (R\Psi R')^{-1/2} R$ is $\mathcal{C}$-measurable and $B\Psi B = I$ it then follows from part (b) of Theorem 3 that

$$\left(R\hat{\Psi}R'\right)^{-1/2} n^{1/2}(R\hat{\theta}_n - r) \overset{d}{\rightarrow} \xi_* \tag{B.5}$$

where $\xi_* \sim N(0, I_q)$. Hence, in light of the continuous mapping theorem, T_n converges in distribution to a chi-square random variable with q degrees of freedom. The claim that $\hat{\Psi}_n^{-1/2}\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\rightarrow} \xi$ is seen to hold as a special case of (B.5) with $R = I$ and $r = \theta_0$. ∎

References

[1] Aguirregabiria, V. and P. Mira, 2007, Sequential Estimation of Dynamic Discrete Games, *Econometrica 75*, 1-53.

[2] Ahn, S.C., Y.H. Lee and P. Schmidt, 2001, GMM estimation of linear panel data models with time-varying individual effects, *Journal of Econometrics* 101, 219-255.

[3] Ahn, S.C., Y.H. Lee and P. Schmidt, 2006, Panel Data Models with Multiple Time-Varying Individual Effects, Working Paper.

[4] Andrews, D.W.K., 2005, Cross-Section regression with common shocks, *Econometrica* 73, 1551-1585.

[5] Anselin, L., 2010, Thirty Years of Spatial Econometrics, *Papers in Regional Science* 89, 3-25.

[6] Arellano, M., and B. Honoré, 2001, Panel data Models: Some Recent Developments, in : J.J. Heckman and E.E. Leamer (ed.), *Handbook in Econometrics* 5, 3229-3296.

[7] Bai, J., and S. Ng, 2006a, Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions, *Econometrica* 74, 1133-1150.

[8] Bai, J., and S. Ng, 2006b, Evaluating latent and observed factors in macroeconomics and finance, *Journal of Econometrics* 131, 507-537.

[9] Bajari, P., C.L. Benkard and J. Levin, 2007, Estimating Dynamic Models of Imperfect Competition, *Econometrica 75*, 1331-1370.

[10] Baltagi, B. H., S.H. Song and W. Koh, 2003, Testing Panel Data Regression Models with Spatial Error Correlation, *Journal of Econometrics* 117, 123-150.

[11] Baltagi, B. H., S.H. Song and J. Kwon, 2009, Testing for Heteroskedasticity and Spatial Correlation in a Random Effects Panel Data Model, *Computational Statistics & Data Analysis* 53, 2897-2922.

[12] Baltagi, B.H., Kelejian, H.H., and I.R. Prucha, eds, 2007, Analysis of Spatially Dependent Data, Annals Issue, *Journal of Econometrics* 140.

[13] Bierens, H.J., 1994, *Topics in Advanced Econometrics*, Cambridge University Press, Cambridge.

[14] Cliff, A., and J. Ord, 1973, *Spatial Autocorrelation*, Pion, London.

[15] Cliff, A., and J. Ord, 1981, *Spatial Processes, Models and Applications*, Pion, London.

[16] Daley, D.J., and D. Vere-Jones, 1988, *An Introduction to the Theory of Point Processes*, Springer Verlag, New York.

[17] Hall, P., and C. Heyde, 1980, *Martingale Limit Theory and its Applications*, Academic Press, New York.

[18] Hausman, J.A., Hall, B.H., and Z. Griliches, 1984, Econometric models for count data with applications to the patents-R&D relationship, *Econometrica* 72, 909-938.

[19] Hotz, J. and R.A. Miller, 1993, Conditional Choice Probabilities and the Estimation of Dynamic Models, *Review of Economic Studies 60,* 497-529.

[20] Kapoor, M., H.H. Kelejian and I.R. Prucha, 2007, Panel DataModels with Spatially Correlated Error Components, *Journal of Econometrics* 140, 97-130.

[21] Keane, M.P., and D.E. Runkle, 1992, On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous, *Journal of Business and Economic Statistics* 10, 1992, 1-9.

[22] Kelejian, H.H., and I.R. Prucha, 2001, On the asymptotic distribution of the Moran I test statistic with applications, *Journal of Econometrics* 104, 219-257.

[23] Kuersteiner, G.M, and I.R. Prucha, 2009, Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogeneity, Working Paper (Version June 15, 2009).

[24] Lee, L.-F., and J. Yu, 2010a, Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects, *Journal of Econometrics* 154, 165-185.

[25] Lee, L.-F., and J. Yu, 2010b, Some Recent Developments in Spatial Panel Data Models, *Regional Science and Urban Economics*, 2010, forthcoming.

[26] Mutl, J., 2006, Dynamic panel data models with spatially correlated innovations, Ph.D. thesis, University of Maryland, College Park.

[27] Pesaran, M.H., 2006, Estimation and inference in large heterogeneous panels with multifactor error structure, *Econometrica* 74, 967-1012.

[28] Phillips, P.C.B., and D. Sul, 2003, Dynamic panel estimation and homogeneity testing under cross sectional dependence, *Econometrics Journal* 6, 217-259.

[29] Phillips, P.C.B., and D. Sul, 2007, Transition modeling and econometric convergence tests, *Econometrica* 75, 1771-1855.

[30] Pötscher, B.M. and I.R. Prucha, 1997, *Dynamic Nonlinear Econometric Models, Asymptotic Theory*, Springer Verlag, New York.

[31] Pötscher, B.M. and I.R. Prucha, 2001, Basic elements of asymptotic theory, in *A Companion to Theoretical Econometrics*, ed., B.H. Baltagi, Blackwell Publishers.

[32] Rust, J., 1994, Estimation of Dynamic Structural Models, Problems and Prospects: Discrete Decision Processes, in *Advances in Econometrics - Sixth World Congress, Vol II,* ed. C.A. Sims, Cambridge University Press.

[33] Wooldridge, J.M., 1997, Multiplicative panel data models without the strict exogeneity assumption, *Econometric Theory* 13, 667-678.

[34] Yu, J., R. de Jong and L.-F. Lee, 2007, Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large: A nonstationary case, Ohio State University Working Paper.

[35] Yu, J., R. de Jong and L.-F. Lee, 2008, Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large, *Journal of Econometrics* 146, 118-134.