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

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

July 25, 2012

[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 (CLT) and asymptotic distributions for sample moments related to panel data models with large n. The results allow for the data to be cross sectionally dependent, while at the same time allowing the regressors to be only sequentially rather than strictly exogenous. The setup is sufficiently general to accommodate situations where cross sectional dependence stems from spatial interactions and/or from the presence of common factors. The latter 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. We also discuss a class of Maximum Likelihood (ML) estimators that can be analyzed using our CLT.

Keywords: Cross-sectional dependence, spatial martingale difference sequence, Central Limit Theorem, spatial, panel, GMM, MLE, multinomial choice, social interaction.

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 the data to be cross sectionally dependent, while at the same time allowing for regressors that are only sequentially (rather than strictly) exogenous. 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 the 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 arising in cross-sections and panel datasets. 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 or in social interaction models), 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 (through

[1] Our thanks for very helpful comments are owed to David Drukker, Nazgul Jenish, Harry Kelejian, Benedikt Pötscher, Peter Robinson and two anonymous referees. We also thank participants of conferences at the Advanced Institute in Vienna, at Cemmap in London, at the Cowles Foundation, the Swiss Economists Abroad Conference and at seminars at Boston University, ETH Zurich, NYU Stern School of Business and Rutgers University for their helpful comments. Ingmar Prucha gratefully acknowledges financial support from the National Institute of Health through the SBIR grants R43 AG027622 and R44 AG027622.

[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

covariates) cross-sectional dependence of the observed sample. 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 i.i.d. in the cross-section. These results extend findings in Andrews (2005) to situations where samples are not i.i.d. 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 and not just convergence in distribution for the underlying vector of sample moments. 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 and Ord (1973, 1981) type models. Dynamic panel data models with spatial interactions have recently been considered by, e.g., Mutl (2006), and Yu, de Jong and Lee (2008, 2012). 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 examples 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, both of which can affect the covariates. Our results are different from central limit theorems for spatial process such as Bolthausen (1982) and Jenish and Prucha (2009, 2012) because we do not impose a spatial structure on the cross-sectional dimension of

are independent in the cross sectional dimension conditional on the common factors.

the panel. As a result the high level conditions that need to be checked to apply our CLT are relatively simple compared to the spatial CLT's. On the other hand, the conditional moment restrictions we impose are often synonymous with correct specification of an underlying model which may not be required by CLT's for mixing processes as in Bolthausen (1982).

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 in distribution but stable convergence. In Section 3 we illustrate how the central limit theory can be applied to efficient GMM estimators for linear panel models. We derive their limiting distribution, and give a consistent estimator for the limiting variance covariance matrix. In Section 4 we present regularity conditions for a class of maximum likelihood estimators (MLE) and show how our CLT can be applied. We give examples of specific multinomial choice models that fit our framework. Concluding remarks are given in Section 5. Basic results regarding stable convergence 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 where n and T denote the cross section and time dimension, respectively. 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 for $T = 1$ 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.

Our basic central limit theorem is stated for averages

$$\psi_{(n)} = n^{-1/2} \sum_{i=1}^{n} \psi_i, \tag{1}$$

over the cross-section of $p \times 1$ random vectors $\psi_i = (\psi'_{i1}, \ldots, \psi'_{iT})'$. The dimension of the sub-vectors ψ_{it} is $p_t \times 1$ and thus allowed to depend on t. The index i is an identifier for a particular unit, where units could be individuals, firms, industries, counties, etc. While units may refer to geographic entities, no spatial structure is explicitly imposed on ψ_i. On the other hand, the index t is given the conventional notion of sequential time.

In introducing our basic CLT the aim is to provide a convenient modul that can be readily used to establish, in particular, a CLT for the sample moment vector associated with GMM estimators and the score of the log-likelihood function of ML estimators. For GMM estimators ψ_{it} will typically refer to the, say, p_t sample moments between a vector of instruments and some basic disturbances for unit i in period

t. For ML estimation ψ_{it} will typically refer to the score of the log likelihood function corresponding to unit i and period t, with $p_t = d$, the dimension of the parameter vector of interest. In the following we set $p = \sum_{t=1}^{T} p_t$.

We next give some basic notational definitions used throughout the paper. All variables are assumed to be defined on a probability space $(\Omega, \mathcal{F}, P)$. With y_{it}, x_{it}, z_{it}, μ_i and u_{it} we denote, respectively, the dependent variable, the sequentially exogenous covariates, the strictly exogenous covariates, unit specific unobserved effects and idiosyncratic disturbances. The particular meaning of sequential and strict exogeneity will be made explicit below. Furthermore, it proves helpful to introduce the following notation: $y_i = (y_{i1}, \ldots, y_{iT})$, $x_i = (x_{i1} \ldots, x_{iT})$, $z_i = (z_{i1}, \ldots, z_{iT})$, $u_i = (u_{i1}, \ldots, u_{iT})$, $y_{it}^o = (y_{i1}, \ldots, y_{it})$, $x_{it}^o = (x_{i1}, \ldots, x_{it})$, $u_{it}^o = (u_{i1}, \ldots, u_{it})$, and $u_{-i,t} = (u_{1t}, \ldots, u_{i-1,t}, u_{i+1,t}, ...u_{nt})$. Although not explicitly denoted, these random variables as well as the ψ_{it} are allowed to depend on the sample size n, i.e., to form triangular arrays.

Our setup is aimed at accommodating fairly general forms of cross-sectional dependence in the data. 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 $\mathcal{C}_t \subset \mathcal{F}$. A special case arises when f_t denotes a vector of common shocks such that $\mathcal{C}_t = \sigma(f_t)$. Alternatively or concurrently we allow for cross sectional dependence due to "spatial lags" in the sense that some of the variables may be weighted cross-sectional averages of some basic variables.[3] In the following let $\mathcal{C}_t^o = \mathcal{C}_1 \vee \ldots \vee \mathcal{C}_t$ where $\vee$ denotes the sigma field generated by the union of two sigma fields. For simplicity we will also write $\mathcal{C} = \mathcal{C}_T^o$ in the following. In the important special case where common shocks are not present we have $\mathcal{C}_t = \mathcal{C} = \{\emptyset, \Omega\}$.

As remarked, the CLT developed in this section will provide a basic module towards deriving a generalized limit theory for GMM and ML estimators. The estimator specific details regarding $\psi_{(n)}$ are not directly relevant and will be suppressed for the discussion of the CLT. However, to further motivate the results of this section we note that if $m_{(n)}$ denotes either the sample moment vector associated with GMM estimators or the score of the log-likelihood function of ML estimators considered below, then $m_{(n)}$ can be expressed as $m_{(n)} = M\psi_{(n)}$, where M is a non-stochastic matrix of dimension $q \times p$ where $q \leq p$. Furthermore, towards establishing the limit properties of specific estimators interest will typically focus on expressions of the form $Bm_{(n)} = BM\psi_{(n)}$, where B is a $p_* \times q$ matrix, which may depend on $\mathcal{C}$ in light of the potential presence of common factors. This motivates that we will also provide limit theorems for $A\psi_{(n)}$, where A is some general (potentially) stochastic $p_* \times p$ matrix, which is measurable w.r.t. to $\mathcal{C}$.

We next state a set of mild regularity conditions. To formulate our main moment conditions we define

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

the following information sets.

Definition 1 *Let* $\{\mathcal{G}_{n,v}, v = 0, \ldots, Tn+1\}$ *with* $\mathcal{G}_{n,0} = \{\emptyset, \Omega\}$ *be a non-decreasing sequence of sub-*σ*-fields of* $\mathcal{F}$ *and let* $\{\mathcal{B}_{n,i,t}, , i = 1, \ldots, n, t = 1, \ldots, T\}$ *be sub-*σ*-fields of* $\mathcal{F}$*, with both* $\mathcal{G}_{n,v}$ *and* $\mathcal{B}_{n,i,t}$ *generated by subsets of the random vectors* $(y_i, x_i, z_i, \mu_i, u_i)_{i=1}^n$*. Let* s *be any permutation of* $(1, ..., n) \to (s(1), ..., s(n))$*. Then the sub-fields* $\mathcal{G}_{n,v}$ *and* $\mathcal{B}_{n,i,t}$ *satisfy*

$$\mathcal{G}_{n,(t-1)n+i} \subseteq \mathcal{B}_{n,i,t} \tag{2}$$

and $\mathcal{B}_{n,i,t}$ *is invariant to any reordering of the random variables* $(y_i, x_i, z_i, \mu_i, u_i) \to \left(y_{s(i)}, x_{s(i)}, z_{s(i)}, \mu_{s(i)}, u_{s(i)}\right)$.

We formulate our basic conditional moment restriction with respect to $\mathcal{B}_{n,i,t}$ which implies that the moment restriction holds for all reorderings of the sample. On the other hand, $\mathcal{G}_{n,(t-1)n+i}$ is *constructed* for a fixed, yet arbitrary, ordering of the sample with the purpose of forming a martingale to which our martinagale CLT can be applied.

For the results in this section the specifics of the information sets are not of interest. However to motivate the above definitions we note that the information sets considered in connection with our GMM estimator will be of the form

$$\mathcal{B}_{n,i,t} = \sigma\left\{\left(x_{jt}^o, z_j, u_{j,t-1}^o, \mu_j\right)_{j=1}^n, u_{-i,t}\right\} \qquad \text{and} \qquad \mathcal{G}_{n,(t-1)n+i} = \sigma\left\{\left(x_{jt}^o, z_j, u_{j,t-1}^o, \mu_j\right)_{j=1}^n, (u_{jt})_{j=1}^{i-1}\right\}. \tag{3}$$

The basic moment condition for our GMM estimator will be $E\left[u_{it}|\mathcal{B}_{n,i,t} \vee \mathcal{C}_t^o\right] = 0$. Now let h_{it} denote a $1 \times p_t$ vector of available instruments, with the property that they are measurable w.r.t. $\sigma\left\{\left(x_{jt}^o, z_j\right)_{j=1}^n\right\}$, and let $\psi_{it} = h_{it}' u_{it}$ denote the vector of sample moments. Then within this setting, the moment condition critical to deriving the CLT, $E\left[\psi_{it}|\mathcal{G}_{n,(t-1)n+i} \vee \mathcal{C}_t^o\right] = 0$, clearly holds. The specification of the above information sets will be discussed in more detail in the section on GMM estimation. At this point we only note (i) that the definition of $\mathcal{B}_{n,i,t}$ is such that the basic moment condition is independent of the ordering of the data and (ii) measurability of the ψ_{it} w.r.t. $\mathcal{G}_{n,(t-1)n+i+1}$ is satisfied for any particular ordering of the data because $\mathcal{G}_{n,(t-1)n+i+1}$ is *constructed* for that particular ordering. Also, the definition of $\mathcal{B}_{n,i,t}$ formalizes the distinction between sequentially exogenous variables x_{jt} and strictly exogenous variables z_{jt}, in that for the former the basic moment condition is only assumed to hold for x_{jt}^o but not for x_j. The sequential nature of the moment restriction allows for dynamic models where x_{it} could contain lagged dependent variables. The fact that the conditioning set contains variables not only for unit i, but also for $j \neq i$ reflects that we allow for cross sectional dependence (even if common shocks are not present).

The ML estimators we consider in Section 4 satisfy regularity conditions that imply that the score of the log-likelihood can be written as $m_{(n)} = \sum_{i=1}^n \sum_{t=1}^T \psi_{it}$ where the ψ_{it} are functions of the data and unknown parameters. Under the regularity conditions in Section 4 it follows that $E\left[\psi_{it}|\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = 0$

such that our martingale CLT can be applied. More specifically, we consider models where the ψ_{it} are such that $\mathcal{B}_{n,i,t}$ and $\mathcal{G}_{n,(t-1)n+i}$ can be specified as

$$\mathcal{B}_{n,i,t} = \sigma\left\{\left(x_{jt}^o, y_{j,t-1}^o\right)_{j=1}^n, y_{-i,t}\right\} \qquad \text{and} \qquad \mathcal{G}_{n,(t-1)n+i} = \sigma\left\{\left(x_{jt}^o, y_{j,t-1}^o\right)_{j=1}^n, (y_{jt})_{j=1}^{i-1}\right\}. \tag{4}$$

We impose the following moment restrictions on ψ_{it}.

Assumption 1 *The following conditions hold for all $t = 1, \ldots, T$, $i = 1, \ldots, n$, $n \geq 1$:*
(a) Let $\|.\|$ denote the Euclidean norm, then for some $\delta > 0$,

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

where B_{it} is uniformly integrable.
(b) The ψ_{it} are measurable with respect to $\mathcal{G}_{n,(t-1)n+i+1} \vee \mathcal{C}$ and the following conditional moment restriction holds:

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

(c) Let $\tilde{V}_{(n)} = \operatorname{diag}\left(\tilde{V}_{1,n}, \ldots, \tilde{V}_{T,n}\right)$ with $\tilde{V}_{t,n} = n^{-1}\sum_{i=1}^n \psi_{it}\psi_{it}'$, then there exist a matrix $V = \operatorname{diag}(V_1, \ldots, V_T)$, where for each t, V_t has finite elements and is positive definite a.s., V_t is $\mathcal{C}$ measurable, and as $n \to \infty$,

$$\tilde{V}_{t,n} - V_t \xrightarrow{p} 0.$$

Assumption 1(a) ensures the existence of various expectations considered subsequently. The condition in Assumption 1(b) is the key moment condition we impose on ψ_{it}. If ψ_{it} is the moment vector or the score of an estimator, as it will be in our leading examples, then Assumption 1(b) usually implies correct specification of a conditional moment restriction or of the likelihood used as the basis for estimation. Assumption 1(b) implies, of course, that

$$E\left[\psi_i\right] = 0. \tag{7}$$

More importantly, because $\mathcal{G}_{n,(t-1)n+i} \vee \mathcal{C}_t^o \subseteq \mathcal{B}_{n,i,t} \vee \mathcal{C}_t^o$, it follows that $E\left[\psi_{it}|\mathcal{G}_{n,(t-1)n+i} \vee \mathcal{C}_t^o\right] = 0$ which, together with the fact that $\mathcal{G}_{n,(t-1)n+i} \vee \mathcal{C}_t^o$ is non-decreasing, leads to the construction of the martingale difference sequence central to our CLT. Assumption 1(b) is thus slightly stronger than needed for the CLT. Nevertheless, it is a natural starting point because it is invariant to the ordering of the sample and is often implied by specific models.

As an important example consider again moment vectors of the form $\psi_{it} = h_{it}' u_{it}$ as discussed after Definition 1 where the u_{it} are i.i.d. conditional on $\mathcal{B}_{n,t} = \sigma\left\{\left(x_{jt}^o, z_j, u_{j,t-1}^o, \mu_j\right)_{j=1}^n\right\}$. Then (6) clearly holds. Of course, u_{it} being i.i.d. does not imply that ψ_{it} is independent in the cross-section because h_{it} could be cross-sectionally dependent. In addition, if the elements of h_{it} are only sequentially exogenous instead of strictly exogenous, one cannot condition on all of the exogenous variables when analyzing the limit distribution of $\psi_{(n)}$. Thus, in this setting a CLT for i.i.d. sequences cannot be applied to $\psi_{(n)}$, even when u_{it} is i.i.d.

The moment condition (6) in Assumption 1(b) is formulated for a situation where the common factors are only sequentially exogenous. While this assumption may be natural from a modelling perspective, it turns out that Assumption 1(b) is not quite strong enough to establish a central limit theorem. The next condition strengthens (6) by requiring that the common factors are orthogonal to all elements of the sequence ψ_{it}.

Assumption 2 *The following conditional moment restrictions hold:*

$$E\left[\psi_{it}|\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = 0. \tag{8}$$

Remark 1 *Condition (8) implies (6) because $\mathcal{B}_{n,i,t} \vee \mathcal{C}_t^o \subset \mathcal{B}_{n,i,t} \vee \mathcal{C}$. An example where moment condition (8) is satisfied are models where the common factors are strictly exogenous. This is a typical assumption in the panel data literature with common factors. An example of a model for ψ_{it}, where ψ_{it} depends on common shocks f_t and satisfies (8) is $\psi_{it} = u_{it} f_t$ where (u_{jt}) and (f_t) are independent and $E\left[u_{it}|\mathcal{B}_{n,i,t}\right] = 0$. As remarked, our analysis includes the important case where no common factors are present by allowing $\mathcal{C}_t = \{\Omega, \emptyset\}$, as is typical in the spatial literature.[4] In this case conditions (6) and (8) are identical, and Assumption 2 is automatically implied by Assumption 1.*

Additional implications of Assumption 1(b) are that $E\left[\psi_{it}\psi_{js}'\right] = 0$ for $i \neq j$ or $t \neq s$, and thus

$$E\left[\psi_i\psi_i'\right] = \operatorname{diag}\left(E\left[\psi_{i1}\psi_{i1}'\right], ..., E\left[\psi_{iT}\psi_{iT}'\right]\right). \tag{9}$$

To interpret Assumption 1(c) consider the matrix of second order sample moments

$$V_{(n)} = n^{-1}\sum_{i=1}^{n} \psi_i\psi_i'. \tag{10}$$

Then in light of (9) 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

$$V_t = \operatorname{plim} n^{-1}\sum_{i=1}^{n} E\left[\psi_{it}\psi_{it}'|\mathcal{C}\right],$$

An example of such a low level condition is the conditional i.i.d. assumption of Andrews (2005), in which case $V_t = E\left[\psi_{it}\psi_{it}'|\mathcal{C}\right]$. Another example would be a condition that imposes some form of cross-sectional mixing. A third example would be to postulate cross-sectional stationarity and appeal to the ergodic theorem. If ψ_{it} is a moment vector of the form $h_{it}'u_{it}$ where h_{it} is a vector of instruments, and u_{it} is conditionally

[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).

homoskedastic with $E\left[u_{it}^2|\mathcal{B}_{n,i,t}\vee\mathcal{C}\right]=\sigma^2$, then $\tilde{V}_{t,n}-V_t\xrightarrow{p}0$ with $V_t=\sigma^2\operatorname{plim}n^{-1}\sum_{i=1}^n E\left[h_{it}'h_{it}|\mathcal{C}\right]$ can be implied solely from convergence assumptions on the second order sample moments of the instruments. If $\mathcal{C}_t=\{\emptyset,\Omega\}$, i.e., no common factors are present, then V is a matrix of fixed constants.

The following lemma provides a sufficient condition for Assumption 1(c).

Lemma 1 *Suppose Assumptions 1(a),(b) and 2 hold, and*

$$E\left[\psi_{it}\psi_{it}'|\mathcal{B}_{n,i,t}\vee\mathcal{C}\right]=E\left[\psi_{it}\psi_{it}'|\mathcal{F}_{n,(t-1)n+i}\right] \tag{11}$$

for any cross sectional ordering of the data and where $\mathcal{F}_{n,(t-1)n+i}$ is defined in (12) below. Let $\bar{V}_{(n)}=\operatorname{diag}\left(\bar{V}_{1,n},...,\bar{V}_{T,n}\right)$ with

$$\bar{V}_{t,n}=n^{-1}\sum_{i=1}^n E\left[\psi_{it}\psi_{it}'|\mathcal{B}_{n,i,t}\vee\mathcal{C}\right],$$

and assume that there exist a matrix $V=\operatorname{diag}(V_1,...,V_T)$, where for each t, V_t has finite elements and is positive definite a.s., V_t is $\mathcal{C}$ measurable, and $\bar{V}_{t,n}-V_t\xrightarrow{p}0$ as $n\to\infty$. Then Assumption 1(c) holds.

One potential application of the lemma is in the context of GMM estimation with moment vectors $\psi_{it}=h_{it}'u_{it}$. For the exemplary setting discussed after Definition 1 we have $E\left[\psi_{it}\psi_{it}'|\mathcal{B}_{n,i,t}\vee\mathcal{C}\right]=E\left[u_{it}^2|\mathcal{B}_{n,i,t}\vee\mathcal{C}\right]h_{it}'h_{it}$. The condition (11) would then hold if, e.g., the u_{it} are homoskedastic conditionally on $\mathcal{B}_{n,i,t}\vee\mathcal{C}$ or independent conditionally on $\mathcal{B}_{n,t}=\sigma\left\{\left(x_{jt}^o,z_j,y_{jt-1}^o,\mu_j\right)_{j=1}^n\right\}$.

When (6) holds but not (8) several cases leading to different limiting distributions for the central limit theorem below can be distinguished. For the proof of the CLT and the statement of the next assumption we need to introduce additional σ-fields $(t=1,\dots,T,i=1,\dots,n)$:

$$\mathcal{F}_{n,(t-1)n+i}=\mathcal{G}_{n,(t-1)n+i}\vee\mathcal{C} \tag{12}$$

and $\mathcal{F}_{n,0}=\mathcal{C}$. The σ-fields $\mathcal{F}_{n,(t-1)n+i}$ differ in two important regards from $\mathcal{B}_{n,i,t}\vee\mathcal{C}_t^o$. First, they are arranged sequentially, to accommodate the application of the CLT. Second, they contain the entire σ-fields $\mathcal{C}$ rather than $\mathcal{C}_t^o$. The later gives rise to the following assumption.

Assumption 3 *Let $\check{\psi}_i=(\check{\psi}_{i1}',\dots,\check{\psi}_{iT}')'=\psi_i-\psi_i^c$ where $\psi_i^c=\left(E\left[\psi_{i1}'|\mathcal{F}_{n,i}\right],...,E\left[\psi_{Ti}'|\mathcal{F}_{n,(T-1)n+i}\right]\right)'$, and let $b_n=n^{-1}\sum_{i=1}^n\psi_i^c$. Furthermore, let $\check{V}_{(n)}=\operatorname{diag}\left(\check{V}_{1,n},...,\check{V}_{T,n}\right)$ with $\check{V}_{t,n}=n^{-1}\sum_{i=1}^n\check{\psi}_{it}\check{\psi}_{it}'$, then there exist a matrix $\check{V}=\operatorname{diag}(\check{V}_1,...,\check{V}_T)$, where for each t, $\check{V}_t$ has finite elements and is positive definite a.s., $\check{V}_t$ is $\mathcal{C}$ measurable, and*

$$\check{V}_{t,n}-\check{V}_t\xrightarrow{p}0 \tag{13}$$

as $n\to\infty$. In addition, 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 2 *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.*

Finally, note that our setting does not imply that the random variables ψ_{it} are exchangeable. For example, let $\psi_{it} = u_{it}\delta_i f$ where u_{it} is i.i.d. $N(0,1)$ and independent of f with $\mathcal{C}_t = \sigma(f)$, $|f| \leq K < \infty$ a.s. Furthermore, let δ_i be nonrandom factor loadings with $|\delta_i| \leq K < \infty$ such that $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\delta_i^2$ exists. It can be checked that ψ_{it} satisfies our conditions with $\mathcal{B}_{n,i,t} = \sigma(u_{-i,t})$, in particular, $V_t = f^2 \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\delta_i^2$ which is $\mathcal{C}$-measurable. On the other hand, ψ_{it} does not satisfy the conditions in Andrews (2005) because, conditional on $\mathcal{C}_t$, ψ_{it} is not identically distributed and ψ_{it} is not exchangeable. To see the latter, note that for example $E\left[\psi_{it}^2\right] = \delta_i^2 E\left[f_t^2\right]$ which depends on i and contradicts exchangeability (see Kingman, 1978, p.185). The reason why ψ_{it} is not necessarily exchangeable comes from the fact that we only restrict the conditional mean of the distribution while the concept of exchangeability imposes restrictions on the entire distribution.

2.2 Limit Theorems

In this Section we establish the limiting distribution of the moment vector $\psi_{(n)} = n^{-1/2}\sum_{i=1}^{n}\psi_i$ and then give a discussion of the strategy by which the result is derived. In fact, we not only establish convergence in distribution of $\psi_{(n)}$, but we establish $\mathcal{C}$-stable convergence of $\psi_{(n)}$, which allows us to establish the joint limiting distribution for $\left(\psi_{(n)}, A\right)$ 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.[5] Applying the continuous mapping theorem allows us to establish the limiting distribution of transformations of $\psi_{(n)}$, e.g., of $A\psi_{(n)}$ or $\psi_{(n)}' A\psi_{(n)}$, 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 of $\psi_{(n)}$ we first establish a general central limit theorem for zero mean, square integrable martingale arrays $\{S_{nv}, \mathcal{F}_{nv}, 1 \leq v \leq k_n, n \geq 1\}$ with differences X_{nv}, which we expect to be useful in many other contexts. We next present a formal definition of stable convergence, cf. Daley and Vere-Jones (1988, p. 644). Various equivalent conditions are summarized in Proposition A.1 in the appendix.

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

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 $n \to \infty$ and where ∂A denotes the boundary of A.

The next theorem extends results in Hall and Heyde (1980) by establishing stable convergence without requiring that the σ-fields $\mathcal{F}_{nv}$ are nested in the sense of Hall and Heyde's condition (3.21), in other words satisfy $\mathcal{F}_{nv} \subseteq \mathcal{F}_{n+1,v}$. This is achieved at the cost of restricting stable convergence to $\mathcal{F}_0$ rather than establishing it on all of $\mathcal{F}$. The nesting condition $\mathcal{F}_{nv} \subseteq \mathcal{F}_{n+1,v}$ may be quite natural in a time series setting but does not hold for panel data with increasing cross-sectional sample size. It is therefore necessary to prove a modified version of Hall and Heyde's CLT adapted to our panel data structure. We note that when $\mathcal{F}_0 = \{0, \Omega\}$, $\mathcal{F}_0$-stable convergence is convergence in distribution. Thus the former always implies the latter.

Theorem 1 *Let $\{S_{nv}, \mathcal{F}_{nv}, 1 \le v \le k_n, n \ge 1\}$ 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*

$$\max_i |X_{nv}| \xrightarrow{p} 0, \tag{14}$$

$$\sum_{v=1}^{k_n} X_{nv}^2 \xrightarrow{p} \eta^2 \tag{15}$$

and

$$E\left(\max_v X_{nv}^2\right) \text{ is bounded in } n \tag{16}$$

then

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

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).

The conditions imposed in Theorem 1 are identical to the conditions of Hall and Heyde (1980, Theorem 3.2, p.58) except for their condition (3.21) postulating $\mathcal{F}_{nv} \subseteq \mathcal{F}_{n+1,v}$, which we do not require. On the other hand, our conclusion is weaker because we only establish $\mathcal{F}_0$-stable rather than $\mathcal{F}$-stable convergence. Eagleson (1975) obtains a result similar to ours by also assuming measurability of η w.r.t $\mathcal{F}_0$ without requiring Hall and Heyde's condition (3.21). The corollary after Theorem 3 in Eagleson (1975) maintains an identical catalogue of assumptions as Hall and Heyde (1980, Corollary 3.1, pp.58) except for their condition (3.21). However, in contrast to Hall and Heyde's corollary, Eagleson's corollary only establishes convergence in distribution and not stable convergence. The corollary after Theorem 2 in Eagleson (1975) establishes stable convergence, however he derives this result only at the expense of assuming almost sure convergence and not just convergence in probability in (15). Our result then shows that Hall and Heyde's condition (3.21) is sufficient but not necessary to establish stable convergence, on a restricted σ-field, under (15). This is in some contrast to what is implied by Hall and Heyde's comment on Eagleson's result (see Hall and Heyde, 1980, p 59). The difference between Eagleson's result and ours lies in the proof strategy. While Eagleson establishes convergence of the conditional distributions we are modifying Hall and Heyde's proof which is based on showing joint convergence. Both convergence concepts imply stable convergence by Réyni (1963) and Aldous and Eagleson (1978) who's results are summarized in Proposition A.1 in the appendix. Dedecker and Merlevede (2002) establish a result similar to Eagleson's (1975) under weaker conditions but by imposing stationarity which we do not require.

Our next result applies Theorem 1 to the panel structure specific to ψ_{it}. This is possible because of the particular way we construct the filtrations $\mathcal{F}_{n,(t-1)n+i}$ defined in (12) to generate a martingale structure.

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

$$\psi_{(n)} \xrightarrow{d} V^{1/2}\xi \quad (\mathcal{C}\text{-stably}), \tag{17}$$

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 Assumptions 1 and 2 hold, then

$$A\psi_{(n)} \xrightarrow{d} (AVA')^{1/2}\xi_* \ (\mathcal{C}\text{-stably}), \tag{18}$$

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

$$A\left(\psi_{(n)} - \sqrt{n}b_n\right) \xrightarrow{d} (A\check{V}A')^{1/2}\xi_* \ (\mathcal{C}\text{-stably}), \tag{19}$$

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

$$A\psi_{(n)} \xrightarrow{d} (A\check{V}A')^{1/2}\xi_* + Ab \ (\mathcal{C}\text{-stably}). \tag{20}$$

If Assumptions 1 and 3(c) hold, then (20) holds with $b = 0$.

The proof of Theorem 2 employs Theorem 1 for martingale difference arrays and uses 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 $p_t \times 1$ and where $\lambda'\lambda = 1$. Then

$$\lambda'\psi_{(n)} = n^{-1/2} \sum_{i=1}^{n} \lambda'\psi_i \tag{21}$$

Next, let $X_{n,1} = 0$, and for $i = 1, \ldots, n$ define

$$\begin{aligned} &X_{n,i+1} = n^{-1/2}\lambda_1'\psi_{i1}, \\ &X_{n,n+i+1} = n^{-1/2}\lambda_2'\psi_{i2}, \\ &\vdots \\ &X_{n,(T-1)n+i+1} = n^{-1/2}\lambda_T'\psi_{iT}, \end{aligned} \tag{22}$$

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

$$\lambda'\psi_{(n)} = \sum_{v=1}^{Tn+1} X_{n,v}. \tag{23}$$

To establish the limiting distribution of $\sum_{v=1}^{Tn+1} X_{n,v}$ through the martingale difference array CLT we utilize the information sets in (12). Clearly the construction of these 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 $E\left[X_{n,v}|\mathcal{F}_{n,v-1}\right] = 0$ 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 (and not just convergence in distribution) the claims in (17) and (18) then follow from Propositions A.1 and A.2.

The construction of the information sets $\mathcal{F}_{n,(t-1)n+i} = \mathcal{G}_{n,(t-1)n+i} \vee \mathcal{C}$ as in (12) is crucial. To provide some additional insights into the ideas underlying the construction of the information sets it may be helpful to consider again the information sets (3) which will be used in connection with our GMM estimator. For our GMM estimator we will have $X_{n,(t-1)n+i+1} = n^{-1/2}\lambda_t' c_{it} u_{it}$, where c_{it} will be be a function of $\left(x_{jt}^o, z_j\right)_{j=1}^n$. At first glance it may seem unusual to include$(u_{jt})_{j=1}^{i-1}$ in the information set $\mathcal{G}_{n,(t-1)n+i}$, and one may be tempted to use the information sets $\mathcal{B}_{n,t} \vee \mathcal{C}$ where $\mathcal{B}_{n,t} = \sigma\left\{\left(x_{jt}^o, z_j, y_{jt-1}^o, \mu_j\right)_{j=1}^n\right\}$ instead. However, we emphasize that it is precisely because of the inclusion of $(u_{jt})_{j=1}^{i-1}$ that $X_{n,v}$ is indeed

$\mathcal{F}_{n,v}$-measurable for all v, as required by the CLT. [6] 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 $(u_{jt})_{j=1}^{i-1}$. However this would have lead to a violation of the assumption that the information sets are non-decreasing.

2.3 Examples and Special Cases

In this section we further discuss Theorem 2 by considering some examples and special cases. Obviously an important special case is $T = 1$, i.e., when we have a single cross section. We consider a further specialization by assuming exemplarily that $\psi_{i1} = x_{i1}'u_{i1}$, which can be interpreted as the moment vector of a cross sectional regression without fixed effects.[7] Dropping subscripts t for notational convenience we have $\psi_i = x_i'u_i$, and the information sets corresponding to (3) are given by

$$\begin{aligned} \mathcal{B}_{n,i} &= \sigma\left\{(x_j)_{j=1}^n, (u_j)_{j=1,j\neq i}^n\right\}, \\ \mathcal{G}_{n,i} &= \sigma\left\{(x_j)_{j=1}^n, (u_j)_{j=1}^{i-1}\right\}. \end{aligned}$$

Since $E\left[\|\psi_i\|^{2+\delta}|\mathcal{B}_{n,i} \vee \mathcal{C}\right] = E\left[|u_i|^{2+\delta}|\mathcal{B}_{n,i} \vee \mathcal{C}\right][x_i'x_i]^{1+\delta/2}$ and $E[\psi_i|\mathcal{B}_{n,i} \vee \mathcal{C}] = x_i'E[u_i|\mathcal{B}_{n,i} \vee \mathcal{C}]$, and assuming that the $2+\delta$ moments of the x_i are uniformly bounded, it is readily seen that the following conditions imply the conditions of Assumption 1:

(a) $E\left[|u_i|^{2+\delta}|\mathcal{B}_{n,i} \vee \mathcal{C}\right] \leq B_i$ with $\sup_i E[B_i^{\rho}] < \infty$ for some $\rho > 1$. (The measurability of ψ_i w.r.t. $\mathcal{G}_{n,i+1}$ holds trivially.)

(b) $E[u_i|\mathcal{B}_{n,i} \vee \mathcal{C}] = 0$

(c) $n^{-1}\sum_{i=1}^n u_i^2 x_i x_i' \rightarrow_p V$ where V is $\mathcal{C}$-measurable, has finite elements and is positive definite $a.s.$

Of course for $T = 1$ Assumption 1 also implies Assumption 2. By Theorem 2 we then have $\psi_{(n)} \xrightarrow{d} V^{1/2}\xi$ ($\mathcal{C}$-stably),where $\xi \sim N(0, I_p)$, and ξ and V are independent. When $T = 1$ the result that $\psi_{(n)} \xrightarrow{d} V^{1/2}\xi$ ($\mathcal{C}$-stably) also follows directly from Theorem 3.2 of Hall and Heyde (1980) with $\mathcal{F}_{n,i} = \mathcal{G}_{n,i} \vee \mathcal{C}$, because in this case the nesting condition $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$ is satisfied.

We next demonstrate that given the assumptions maintained above, the CLT is invariant to the ordering of the sample. Let s be any permutation of $(1, ..., n) \rightarrow (s(1), ..., s(n))$. We now fix the original ordering

[6] 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, 2012). However their setups differ substantially from ours, and these papers do not consider sequentially exogenous covariates, nor common factors and corresponding stable convergence.

[7] We thank one of the referees for suggesting this special case.

of the observations by identifying it with the sequence of indices $(1, ..., n)$. Define

$$\begin{aligned}
\mathcal{B}_{n,s(i)}^{s} &= \sigma\left\{(x_{s(j)})_{j=1}^{n}, (u_{s(j)})_{j=1,j\neq i}^{n}\right\} = \sigma\left\{(x_{j})_{j=1}^{n}, (u_{s(j)})_{j=1,j\neq i}^{n}\right\}, \\
\mathcal{G}_{n,s(i)}^{s} &= \sigma\left\{(x_{s(j)})_{j=1}^{n}, (u_{s(j)})_{j=1}^{i-1}\right\} = \sigma\left\{(x_{j})_{j=1}^{n}, (u_{s(j)})_{j=1}^{i-1}\right\},
\end{aligned}$$

then clearly $\mathcal{G}_{n,s(i)}^{s} \subseteq \mathcal{G}_{n,s(i+1)}^{s}$, $\mathcal{G}_{n,s(i)}^{s} \subseteq \mathcal{B}_{n,s(i)}^{s}$, and $\psi_{s(i)} = x_{s(i)}' u_{s(i)}$ is measurable w.r.t $\mathcal{G}_{n,s(i+1)}^{s}$. Observing further that $\mathcal{B}_{n,s(i)}^{s} = \mathcal{B}_{n,s(i)}$ and that the above conditions are assumed to hold for all i it is readily seen that conditions (a) and (b) above also hold with i and $\mathcal{B}_{n,i}$ replaced with $s(i)$ and $\mathcal{B}_{n,s(i)}^{s}$. Since condition (c) is clearly invariant to the ordering of the data it follows that the CLT is indeed invariant to the ordering of the data. The fact that the ordering is irrelevant also applies when $T > 1$ because $\mathcal{G}_{n,(t-1)+s(i)}^{s} \subseteq \mathcal{B}_{n,s(i),t}$ for all perturbations s and $t = 1, ..., T$.

The next special case concerns marginal convergence of components within $\lambda'\psi_{(n)}$. Noting that

$$\lambda'\psi_{(n)} = \sum_{v=1}^{Tn+1} X_{n,v} = \sum_{t=1}^{T}\left(n^{-1/2}\sum_{i=1}^{n}\lambda_t'\psi_{it}\right)$$

where $X_{n,(t-1)n+i+1} = n^{-1/2}\lambda_t'\psi_{it}$ one can consider the convergence of $n^{-1/2}\sum_{i=1}^{n}\lambda_t'\psi_{it}$ for a fixed t. As for $T = 1$, in this case we can apply Theorem 3.2 of Hall and Heyde (1980) with $\mathcal{F}_{n,i} = \mathcal{G}_{n,(t-1)n+i} \vee \mathcal{C}$ and $\mathcal{F}_{n,0} = \mathcal{C}$, because in this case $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$ is again satisfied. Given Assumption 1 and 2 we have $n^{-1/2}\sum_{i=1}^{n}\psi_{it} \to^{d} V_t^{1/2}\xi_t$ ($\mathcal{C}$-stably) where $\xi_t \sim N(0, I_{p_t})$, and ξ_t and V_t are independent. In other words, marginal convergence of $n^{-1/2}\sum_{i=1}^{n}\lambda_t'\psi_{it}$ for each component within $\lambda'\psi_{(n)}$ can be established using the existing limit theory of Hall and Heyde (1980). Unfortunately, marginal convergence of all the components in $\lambda'\psi_{(n)}$ has, as is well known, no bearing on the joint limit. The following elementary example, inspired by Hall and Heyde (1980, p. 65, Example 1) illustrates this point.

Example 1 *Let $W_1(s), W_2(s)$, for $s \in [0,1]$, be two independent standard Brownian Motions defined on $(\Omega, \mathcal{F}, P)$. Consider the partition $0 = s_{0,n} < s_{1,n} < s_{2,n} ... < s_{n-1,n} < s_{n,n} = 1$ of the interval $[0,1]$ with $s_{i,n} = i/n$. Define the triangular arrays $u_{n,i1} = \sqrt{n}(W_1(s_{i,n}) - W_1(s_{i-1,n}))$ and $u_{n,i2} = \sqrt{n}(W_2(s_{i,n}) - W_2(s_{i-1,n}))$ where in the following we set $u_{n,i1} = u_{i1}$ and $u_{n,i2} = u_{i2}$ to simplify notation. By the properties of standard Browning Motion it follows that u_{i1} and u_{i2} are mutually independent Gaussian random variables with distribution $N(0,1)$. Because by definition $W_t(0) = 0$ it follows that $n^{-1/2}\sum_{i=1}^{n}u_{it} = W_t(1)$ by construction for all n and $t = 1, 2$. It then follows trivially, letting $u_i = (u_{i1}, u_{i2})$ and $Y = (Y_1, Y_2) \sim N(0, I_2)$, that*

$$n^{-1/2}\sum_{i=1}^{n}u_i \xrightarrow{d} Y.$$

Now let $\psi_i = (u_{i1}, u_{i2}f)'$ with $f = W_1(1)$, and let $\lambda = (\lambda_1, \lambda_2)'$ be a vector of constants such that

$\lambda_1^2 + \lambda_2^2 = 1$. Define

$$X_{n,(t-1)n+i+1} = \begin{cases} n^{-1/2} u_{i1} \lambda_1 & \text{for } t = 1 \\ n^{-1/2} u_{i2} f \lambda_2 & \text{for } t = 2 \end{cases}$$

with $X_{n,1} = 0$. It follows immediately and by construction that

$$\begin{aligned} \lambda' \psi_{(n)} &= n^{-1/2} \sum_{i=1}^{n} \lambda' \psi_i = \textstyle\sum_{v=1}^{2n+1} X_{n,v} = n^{-1/2} \displaystyle\sum_{i=1}^{n} u_{i1} \lambda_1 + f \left(n^{-1/2} \sum_{i=1}^{n} u_{i2} \right) \lambda_2 \\ &= \lambda_1 W_1(1) + \lambda_2 W_1(1) W_2(1) \end{aligned}$$

which has the same distribution as $\lambda_1 Y_1 + \lambda_2 Y_1 Y_2$. This implies that trivially,

$$\textstyle\sum_{v=1}^{2n+1} X_{n,v} \to^d \lambda_1 Y_1 + \lambda_2 Y_1 Y_2,$$

and thus by the Cramer-Wold device

$$\psi_{(n)} = n^{-1/2} \sum_{i=1}^{n} \psi_i \xrightarrow{d} \begin{bmatrix} Y_1 \\ Y_1 Y_2 \end{bmatrix}. \tag{24}$$

This shows that the two components of ψ_i, while asymptotically uncorrelated, are not independent. Define the σ-fields

$$\mathcal{G}_{n,i} = \sigma \left\{ (u_{j1})_{j=1}^{i-1} \right\},$$
$$\mathcal{G}_{n,n+i} = \sigma \left\{ (u_{j1})_{j=1}^{n}, (u_{j2})_{j=1}^{i-1} \right\},$$

It follows for $v = 1, ..., 2n+1$ that $\mathcal{G}_{n,v} \subset \mathcal{G}_{n,v+1}$ such that $\mathcal{G}_{n,v}$ is an increasing sequence of σ-fields. We show in the appendix that, $\left\{ \sum_{v=1}^{k} X_{n,v}, \mathcal{G}_{n,k}, 1 \le k \le 2n+1 \right\}$ is a zero mean square integrable martingale which satisfies all conditions of Hall and Heyde (1980, Theorem 3.2) except for Condition (3.21) which requires that $\mathcal{G}_{n,v} \subseteq \mathcal{G}_{n+1,v}$. To see that this condition is violated, choose exemplarily $n = 3$, $v = 5$. Then $\mathcal{G}_{3,5} = \sigma \left\{ (u_{j1})_{j=1}^{3}, u_{12} \right\}$ but $\mathcal{G}_{4,5} = \sigma \left\{ (u_{j1})_{j=1}^{4} \right\}$, which does not include $\mathcal{G}_{3,5}$ as a subset because u_{12} is no longer included and because in moving from n to $n+1$ the definition of u_{i1} changed. This example makes clear that the nesting Condition (3.21), while quite natural in a pure time series context, cannot be imposed in a panel or spatial setting where the cross-section size increases and random variables may depend on the sample size n.

Theorem 3.2 of Hall and Heyde (1980), were it applied to $\sum_{v=1}^{k} X_{n,v}$ would imply that $\lambda' \psi_{(n)} = \sum_{v=1}^{2n+1} X_{n,v} \xrightarrow{d} \left(\lambda_1^2 + f^2 \lambda_2^2 \right)^{1/2} Z$ where $Z \sim N(0,1)$ and Z is independent of f. Of course, the theorem cannot be applied here because their Condition (3.21) is violated. We next show that indeed the limit of $\sum_{v=1}^{2n+1} X_{n,v}$ is not characterized by $\left(\lambda_1^2 + f^2 \lambda_2^2 \right)^{1/2} Z$. To see this note that the Cramer-Wold theorem applied to $\left(\lambda_1^2 + \lambda_2^2 f^2 \right)^{1/2} Z$ indicates that

$$\psi_{(n)} = n^{-1/2} \sum_{i=1}^{n} \psi_i \xrightarrow{d} \begin{bmatrix} Y_1 \\ Y_3 Y_2 \end{bmatrix} \tag{25}$$

where Y_1, Y_2, Y_3 are independent $N(0,1)$.[8] *This implies that the two components are asymptotically independent, contradicting (24). This example shows that the conclusion of Theorem 3.2 of Hall and Heyde (1980) may not hold without their Condition (3.21). As discussed earlier, marginal convergence can be established using Hall and Heyde and the limiting marginal distributions are identical to the marginals of (25). However, the marginal distributions contain no information about the dependence between the first and second component.*

As Example 1 shows, joint limit results can sometimes be derived from first principles when more information about the data distribution is available. An approach that works more generally is to establish the limiting distribution of $\sum_{v=1}^{2n+1} X_{n,v}$ and then appeal to the Cramer-Wold Theorem. We now show how Theorem 1 can be used to that end. We introduce the sigma field $\mathcal{C} = \sigma(W_1(1))$. To apply Theorem 1 we need to enlarge the filtration $\mathcal{G}_{n,(t-1)n+i}$ to $\mathcal{F}_{n,(t-1)n+i} = \mathcal{G}_{n,(t-1)n+i} \vee \mathcal{C}$. This change only affects the conditional means of $X_{n,v}$ for $v \leq n+1$. Given that $E[X_{n,v}|\mathcal{F}_{n,v-1}] \neq 0$ for $v \leq n+1$ the $\{X_{n,v}, \mathcal{F}_{n,v}\}$ do not form a martingale difference sequence. However, as for the construction of $\check{\psi}_{it}$ in Assumption 3, one can subtract the conditional mean $E[X_{n,v}|\mathcal{F}_{n,v-1}]$ from $X_{n,v}$ yielding

$$\check{X}_{n,v} = \begin{cases} X_{n,v} - E[X_{n,v}|\mathcal{F}_{n,v-1}] & v \leq n+1 \\ X_{n,v} & v > n+1 \end{cases}.$$

Then, by construction $\check{X}_{n,v}$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_{n,v}$, since $E\left[\check{X}_{n,v}|\mathcal{F}_{n,v-1}\right] = 0$ for all $v \leq 2n+1$ and $\mathcal{F}_{n,v} \subseteq \mathcal{F}_{n,v+1}$. The fields $\mathcal{F}_{n,v}$ do not satisfy Hall and Heyde's Condition 3.21, but we show in the appendix that $\sum_{v=1}^{2n+1} \check{X}_{n,v} \xrightarrow{d} fZ\lambda_2$ ($\mathcal{C}$-stably) by Theorem 1 and $\sum_{v=1}^{n+1} E[X_{n,v}|\mathcal{F}_{n,v-1}] = f\lambda_1$ for all n. Because stable convergence implies joint convergence with all $\mathcal{C}$ measurable random variables we can apply the continuous mapping theorem and conclude that

$$\lambda'\psi_{(n)} = \sum_{v=1}^{2n+1} X_{n,v} = \sum_{v=1}^{2n+1} \check{X}_{n,v} + \sum_{v=1}^{n+1} E[X_{n,v}|\mathcal{F}_{n,v}] \xrightarrow{d} f\lambda_1 + fZ\lambda_2.$$

By the Cramer Wold Theorem we obtain again the correct results (24) for the limiting distribution of $\psi_{(n)}$.

Finally consider two examples that illustrate the role of common factors.

Example 2 *Consider estimating the mean of y_i in $y_i = \varepsilon_i$ where $\varepsilon_i = u_i f$ with u_i i.i.d. standard Gaussian mutually independent of f, and $f^2 = v/\chi^2$, where χ^2 is distributed chi-square with $v > 2$ degrees of freedom. Then $(y_1, \ldots, y_n)$ is distributed multivariate Student t with $E[y_i] = \mu$ where $\mu = 0$. The y_i have variance $v/(v-2)$ and are uncorrelated (but not independent). The MLE for μ is $\bar{y}$; see Zellner (1976) and $E[y_i|f] = 0$. Theorem 2 can be applied directly to $\bar{y}$ and the limiting distribution is $\sqrt{n}\bar{y} \to_d f\xi$ ($\mathcal{C}$-stably)*

[8] This is readily confirmed by verifying that the characteristic functions of $\left(\lambda_1^2 + \lambda_2^2 f^2\right)^{1/2} Z$ and $\lambda_1 Y_1 + \lambda_2 Y_3 Y_2$ are identical.

where ξ is standard normal and independent of f. Thus, the limit is mixed normal. More specifically, the limiting distribution is Student t with v degrees of freedom. Of course, given the simplicity of the example the result can also be derived directly by observing that $\sqrt{n}\bar{y}$ is distributed Student t for each n; see Kelejian and Prucha (1985).

The next example is a further illustration of the role Assumption 3 plays in handling non-standard situations where Assumption 2 fails.

Example 3 *Now change the model and assume that $y_i = \varepsilon_i$ with $\varepsilon_i = f + u_i$ with u_i, f standard Gaussian and mutually independent. In this case $\mu = E[y_i] = 0$ and the MLE is again $\bar{y}$. However, now $E[\varepsilon_i|f] = f$ such that Theorem 2 needs to be applied to $\check{\psi}_i = \varepsilon_i - E[\varepsilon_i|f] = u_i$ where u_i satisfies all conditions of Theorem 2. Of course, $1/\sqrt{n}\sum_{i=1}^n \check{\psi}_i \rightarrow_d \xi$ ($\mathcal{C}$-stably) where ξ is is standard normal and independent of f. It follows that $\sqrt{n}(\bar{y} - b_n) = 1/\sqrt{n}\sum_{i=1}^n \check{\psi}_i \rightarrow_d \xi$ ($\mathcal{C}$-stably) and thus $\bar{y} \rightarrow_d b = f$ by Theorem 2 and (19).*

Examples 2 and 3 illustrate that common factors affecting the error term can have very different implications for the limiting distribution of an estimator. In the first case, the effects are quite benign while in the second case the bias remains dominant asymptotically. The examples illustrate that Theorem 2 is quite flexible and can even be used to analyze estimators that do not produce usable inference asymptotically. Example 3 highlights that the bias terms of the limiting distribution may not always be correctable.

The case where no common factors are present remains an important area application for the above limit theorems. The reason for focusing our examples on cases where common factors are present is to highlight some of the additional subtleties arising from such situations.

3 GMM Estimators

In this section we consider the following linear panel data model ($i = 1, \ldots, n; t = 1, \ldots, T$):

$$y_{it} = x_{it}\beta_0 + z_{it}\gamma_0 + \mu_i + u_{it} \tag{26}$$

where, consistent with the notation above, 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), β_0 and γ_0 are vectors 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 subsequent discussion it proves convenient to rewrite the model more compactly as

$$y_{it} = w_{it}\theta_0 + \mu_i + u_{it}, \tag{27}$$

where $w_{it} = (x_{it}, z_{it})$ and $\theta_0 = (\beta_0', \gamma_0')' \in \mathbb{R}^d$. We maintain the general setup of Section 2 and assume that sample averages are taken over n, with n tending to infinity and T fixed.

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 $y_i = (y_{i1}, \ldots, y_{iT})$, $w_i = (w_{i1}', \ldots, w_{iT}')$ and $u_i = (u_{i1}, \ldots, u_{iT})$, then in this setting (w_i, μ_i, u_i) or equivalently (y_i, w_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. This includes situations where i corresponds to countries, states, regions, industries, etc., and situations where common factors are present. 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). Common factors and spatial interactions may both be present at the same time.

We consider moment conditions that are based on linear transformations of (27) such that

$$E\left[h_{it}' u_{it}^+\right] = 0 \text{ for } t = 1, \ldots, T^+, \tag{28}$$

with $T^+ \leq T$, where $h_{it} = (x_{it}^o, z_i)$, denotes a $1 \times (tk_x + Tk_z)$ vector of instruments corresponding to t, and $u_i^+ = \left(u_{i1}^+, ..., u_{iT^+}^+\right)$ denotes a vector of transformed disturbances with $u_{it}^+ = \sum_{s=t}^T \pi_{ts} u_{is}$, where the π_{ts} are known, 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 $t = 1, \ldots, T$. The class of transformations is thus fairly general. As a special case we also have $u_{it}^+ = u_{it}$, for $t = 1, \ldots, T$.

The sample moment vector corresponding to the moment conditions (28) is given by

$$m_{(n)} = n^{-1/2} \sum_{i=1}^n m_i, \qquad m_i = \left(h_{i1} u_{i1}^+, ..., h_{iT^+} u_{iT^+}^+\right)'. \tag{29}$$

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 $m_i = H_i' u_i^{+\prime} = H_i' \Pi u_i' = \sum_{t=1}^T H_i' \pi_t u_{it}$, where π_t denotes the t-th column of Π. Let

$$\underset{p_t \times 1}{c_{it}} = \begin{cases} [\pi_{1t} h_{i1}, \ldots, \pi_{tt} h_{it}]' & \text{for} \quad t \leq T^+ \\ [\pi_{1t} h_{i1}, \ldots, \pi_{T^+t} h_{iT^+}]' & \text{for} \quad t > T^+ \end{cases} \tag{30}$$

where by construction $p_t = t(t+1)k_x/2 + tTk_z$ for $t \leq T^+$ and $p_t = p_{T^+}$ for $t > T^+$, then $H_i' \pi_t = \left[c_{it}', 0_{1 \times (p_{T^+} - p_t)}\right]'$. We next rewrite the sample moment vector such that it fits into the setup of Section 2. To that effect define the matrices $M_t = \left[I_{p_t}, 0_{p_t \times (p_{T^+} - p_t)}\right]'$ and let $\psi_i = \left[\psi_{i1}', ..., \psi_{iT}'\right]'$ with

$$\psi_{it} = [\pi_{1t} h_{i1}, \ldots, \pi_{tt} h_{it}]' u_{it} = c_{it} u_{it} \tag{31}$$

then $m_{(n)} = n^{-1/2} \sum_{i=1}^{n} m_i = M\psi_{(n)}$ with $M = [M_1, \ldots, M_T]$.

The GMM estimator corresponding to the moment conditions (28) is defined as

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

where $G_n = n^{-1} \sum_{i=1}^{n} H_i' \Pi w_i'$, $g_n = n^{-1} \sum_{i=1}^{n} H_i' \Pi y_i'$, 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 when the observations are i.i.d. When 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. It is also the approach taken by Mutl (2006), and Yu, de Jong and Lee (2008, 2012) 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 illustrate how the theory developed in Section 2 can be utilized 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. We maintain the following assumption, which will be shown to imply Assumption 1 with ψ_{it} as defined in (31).

Assumption 4 *For some $\delta > 0$, $\rho>1$ and $\kappa>1$ with $1/\rho + 1/\kappa < 1$, and some finite constant K (which is taken, w.l.o.g., to be greater than one) the following conditions hold with $\mathcal{B}_{n,i,t}$ and $\mathcal{G}_{n,(t-1)n+i}$ defined by (3) and $t = 1, \ldots, T$, $i = 1, \ldots, n$, $n \geq 1$:*
(a) For some random variables ϕ_{it},

$$E\left[|u_{it}|^{2+\delta} | \mathcal{B}_{n,i,t} \vee \mathcal{C}\right] \leq \phi_{it}, \tag{32}$$

where the L_ρ norms of the random variables ϕ_{it}, i.e., $\|\phi_{it}\|_\rho = [E|\phi_{it}|^\rho]^{1/\rho}$, are uniformly bounded by K. Furthermore, the $L_{(2+\delta)\kappa}$ norms of the random variables x_{it}, z_{it}, u_{it}, and μ_i are uniformly bounded by K.
(b) The following conditional moment restrictions hold:

$$E\left[u_{it} | \mathcal{B}_{n,i,t} \vee \mathcal{C}_t^o\right] = 0. \tag{33}$$

(c) Let $\tilde{V}_{(n)} = \operatorname{diag}\left(\tilde{V}_{1,n}, ..., \tilde{V}_{T,n}\right)$ with $\tilde{V}_{t,n} = n^{-1} \sum_{i=1}^{n} \psi_{it}\psi_{it}' = n^{-1} \sum_{i=1}^{n} u_{it}^2 c_{it} c_{it}'$, then there exist a matrix $V = \operatorname{diag}(V_1, ..., V_T)$, where for each t, V_t 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 $n \to \infty$. If $E\left[u_{it}^2 | \mathcal{B}_{n,i,t} \vee \mathcal{C}_t^o\right] = \sigma^2$ then $V_t = \sigma^2 \operatorname{plim} n^{-1} \sum_{i=1}^{n} c_{it} c_{it}'$.

A special case for which condition (32) of Assumption 4(a) holds is the case where the $2+\delta$-th conditional moments of u_{it} are bounded by some constant, in which case we can take $\rho = \infty$. A sufficient condition for this special case is for the u_{it} to be i.i.d. conditional on $\mathcal{B}_{n,t} = \sigma\left\{\left(x_{jt}^o, z_j, u_{j,t-1}^o, \mu_j\right)_{j=1}^n\right\}$ with finite $2+\delta$-th moment. Of course, ψ_{it}, which is what matters for the CLT, is not necessarily independent even under these stronger assumptions, and thus as discussed after Assumption 1, even in this case a CLT for i.i.d. sequences cannot be applied.

It is of interest to compare Assumption 4(b) with the moment conditions typically maintained under the assumption that (w_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 4(b) can be stated equivalently by replacing the conditioning sets by $w_{it}^o, \mu_i, u_{it-1}^o$. In particular, Assumption 4(b) simplifies to

$$E\left[u_{it}|w_{it}^o, \mu_i, u_{it-1}^o\right] = 0. \tag{34}$$

This is in contrast to the assumption that

$$E\left[u_{it}|w_{it}^o, \mu_i\right] = 0, \tag{35}$$

which is typically maintained in the literature under cross sectional independence. Clearly condition (34) rules out autocorrelation in the disturbances, even if x_{it} does not contain a lagged endogenous variable, while condition (35) does not.[9] If the model is dynamic and linear condition (35) also rules out autocorrelation in the disturbances. In this case conditions (34) and (35) are equivalent, since then w_{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.

The following assumptions play the same role as Assumptions 2 and 3 in Section 2.

Assumption 5 *The following conditional moment restrictions hold:*

$$E\left[u_{it}|\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = 0. \tag{36}$$

Assumption 6 *Let* $\check{\psi}_i = (\check{\psi}_{i1}', \dots, \check{\psi}_{iT}')' = \psi_i - \psi_i^c$ *where* $\psi_i^c = \left(c_{i1}' E\left[u_{i1}|\mathcal{F}_{n,i}\right], ..., c_{iT}' E\left[u_{Ti}|\mathcal{F}_{n,(T-1)n+i}\right]\right)'$, *and let* $b_n = n^{-1}\sum_{i=1}^n \psi_i^c$. *Furthermore, let* $\check{V}_{(n)} = \operatorname{diag}\left(\check{V}_{1,n}, ..., \check{V}_{T,n}\right)$ *with* $\check{V}_{t,n} = n^{-1}\sum_{i=1}^n \check{\psi}_{it}\check{\psi}_{it}' = n^{-1}\sum_{i=1}^n \left(u_{it} - E\left[u_{ti}|\mathcal{F}_{n,(t-1)n+i}\right]\right)^2 c_{it}c_{it}'$, *then there exist a matrix* $\check{V} = \operatorname{diag}(\check{V}_1, ..., \check{V}_T)$, *where for each* t, $\check{V}_t$ *has finite elements and is positive definite a.s.,* $\check{V}_t$ *is* $\mathcal{C}$ *measurable, and* $\check{V}_{t,n} - \check{V}_t \xrightarrow{p} 0$ *as* $n \to \infty$. *In addition, one of the following statements holds:*

[9] 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.

(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) $\sqrt{n}b_n \xrightarrow{p} 0$.

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 Assumptions 4 and 5 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 as $n \to \infty$,

$$n^{1/2}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} \Psi^{1/2}\xi_* \text{ ($\mathcal{C}$-stably),}$$

with

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

and where $\Phi = MVM' = \operatorname{plim}_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it}^2 H_i'\pi_t\pi_t' H_i$ *is positive definite a.s.,* ξ_* *is independent of* $\mathcal{C}$ *(and hence of* Ψ*) and* $\xi_* \sim N(0, I_d)$*. If in addition,* $E\left[u_{it}^2|\mathcal{F}_{n,(t-1)n+i}\right] = \sigma^2$ *for a constant* σ^2 *holds, then* $\Phi = \sigma^2 \operatorname{plim}_{n\to\infty}\left(n^{-1}\sum_{i=1}^{n} H_i'\Pi\Pi' H_i\right)$.
(b) Suppose B *is some* $p^* \times d$ *matrix,* $p^* \leq d$*, that is* $\mathcal{C}$ *measurable with finite elements and rank* p^* *a.s., then*

$$Bn^{1/2}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} (B\Psi B')^{1/2}\xi^* \text{ ($\mathcal{C}$-stably)},$$

where $\xi^* \sim N(0, I_{p^*})$*, and* ξ^* *and* $\mathcal{C}$ *(and thus* ξ^* *and* $B\Psi B'$*) are independent.*
(c) Let $\Phi_{(n)} = n^{-1}\sum_{i=1}^{n} H_i' u_i^+ u_i^{+\prime} H_i$ *and suppose that*

$$n^{-1}\sum_{i=1}^{n} u_{it}u_{is}H_i'\pi_t\pi_s' H_i \xrightarrow{p} 0 \quad \text{for} \quad t \neq s, \tag{37}$$

then $\Phi_{(n)} - \Phi \xrightarrow{p} 0$ *and* $\Phi_{(n)}^{-1/2} m_{(n)} \xrightarrow{d} \xi^+$ *($\mathcal{C}$-stably) with* $\xi^+ \sim N\left(0, I_{p_{T^+}}\right)$.

The next result considers cases where the common factors are only sequentially exogenous, i.e., only (33) but not necessarily (36) holds, and where the resulting effect on the bias of the estimator may be asymptotically non-negligible. Part(a) of the theorem considers a case where the estimator is inconsistent and converges to a random limit while the Part(b) of the theorem covers a case where the estimator is root-n consistent but not asymptotically mixed normal. Part(c) considers the case where the estimator remains asymptotically mixed normal, but allows for V and $\check{V}$ to differ.

Theorem 4 *Suppose Assumption 4 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 6(a) holds then

$$n^{1/2}(\tilde{\theta}_n - \theta_0 - (G'\Xi G)^{-1}G'\Xi M b_n) \xrightarrow{d} \check{\Psi}^{1/2}\xi_* \ (\mathcal{C}\text{-stably})$$

where $\check{\Psi} = (G'\Xi G)^{-1}G'\Xi M\check{V}M'\Xi G(G'\Xi G)^{-1}$, and where ξ_* is independent of $\mathcal{C}$ (and hence of $\check{\Psi}$) and $\xi_* \sim N(0, I_d)$. Furthermore $\tilde{\theta}_n - \theta_0 \xrightarrow{p} (G'\Xi G)^{-1}G'\Xi Mb$.
(b) If in addition Assumption 6(b) holds then $n^{1/2}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} \check{\Psi}^{1/2}\xi_* + (G'\Xi G)^{-1}G'\Xi Mb$ ($\mathcal{C}$-stably).
(c) If in addition Assumption 6(c) holds then $n^{1/2}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} \check{\Psi}^{1/2}\xi_*$ ($\mathcal{C}$-stably).

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

$$\widetilde{\Phi}_{(n)} = n^{-1}\sum_{i=1}^{n} H_i'\widetilde{u}_i^{+\prime}\widetilde{u}_i^{+}H_i$$

where $\widetilde{u}_i = (\widetilde{u}_{i1}, \ldots, \widetilde{u}_{iT})$ with $\widetilde{u}_{it} = y_{it} - w_{it}\tilde{\theta}_n$, and $\tilde{\theta}_n$ denotes 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 $\tilde{\Xi}_n = \widetilde{\Phi}_{(n)}^{-1}$ is then given by,

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

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 $\Psi = (G'\Phi^{-1}G)^{-1}$, which can be estimated consistently by $\hat{\Psi}_n = \left(G_n'\widetilde{\Phi}_{(n)}^{-1}G_n\right)^{-1}$, provided it is shown that $\widetilde{\Phi}_{(n)}$ is indeed a consistent estimator for Φ. Next, let R be a $p^* \times d$ full row rank matrix and r a $p^* \times 1$ vector, and consider the Wald statistic

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

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 $\widetilde{\Phi}_{(n)}$, 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 $\Phi_{(n)} = n^{-1}\sum_{i=1}^{n} H_i'u_i^{+\prime}u_i^{+}H_i - \Phi \xrightarrow{p} 0$, 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 $\widetilde{\Phi}_{(n)} - \Phi \xrightarrow{p} 0$, and*

$$\hat{\Psi}_n^{-1/2}\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \xi_* \sim N(0, I_d).$$

Furthermore

$$P\left(T_n > \chi^2_{p^*,1-\alpha}\right) \to \alpha$$

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

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.

Remark 3 *The specification of the instruments as $h_{it} = (x^o_{it}, z_i)$ was chosen for expositional simplicity. Clearly the above discussion also applies if the vectors of instruments h_{it} are more generally allowed to be $\sigma\left[\{x^o_{it}, z_i\}_{i=1}^n\right]$ measurable functions of the regressors where 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. $\sigma\left[\{x^o_{it}, z_i\}_{i=1}^n\right] \vee \mathcal{C}^o_t$. The crucial property, pertaining to the instruments and the data transformation, that is used in establishing the moment conditions (33) and (36) is that $H_i'\pi_t$ is $\mathcal{B}_{n,t} \vee \mathcal{C}^o_t$ measurable, which clearly is the case even under the above generalized specifications. We note further that the sample moment vector (29) is a function of the true disturbances u_{it}, and thus the specific functional form of the model (26) 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 above 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.*[10]

4 Maximum Likelihood Estimation

We consider a class of likelihood estimators that satisfy our assumptions in Section 2. For concreteness we assume that the econometrician has a model allowing him to specify the partial likelihood function of an outcome variable $y_i = (y_{i1}, ..., y_{iT})$ conditional on sequentially exogenous covariates $x_i = (x_{i1}, \ldots, x_{iT})$. For ease of notation we omit exogenous covariates z_i which could easily be added to the analysis.

The outcome variables y_{it} may depend on individual specific effects μ_i and common factors f_t, which are treated in this section as random effects, and are assume to be captured by the partial likelihood function. This setting is quite different from the recent panel literature on interactive effects such as Pesaran (2006) or Bai (2009) which essentially conditions on the factors and loadings and estimates them as nuisance parameters. In our setting this is not feasible in general because T is assumed fixed. In addition, an important special case covered are models without common factors.

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

Assume that the measure $\mu(A) = P\left((y_i, x_i)_{i=1}^n \in A\right)$ has a parametric density $p\left((y_i, x_i)_{i=1}^n; \theta, \overline{\theta}\right)$ with respect to a σ-finite product measure ν. Because we distinguish sequentially exogenous and endogenous variables the joint density is assumed to admit a representation

$$\begin{aligned} p\left((y_i, x_i)_{i=1}^n; \theta, \overline{\theta}\right) &= \prod_{t=1}^T p\left((y_{it}, x_{it})_{i=1}^n, \left(y_{it-1}^o, x_{it-1}^o\right)_{i=1}^n; \theta, \overline{\theta}\right) \\ &= \prod_{t=1}^T p\left((y_{it})_{i=1}^n \mid \left(y_{it-1}^o, x_{it}^o\right)_{i=1}^n; \theta\right) p\left((x_{it})_{i=1}^n \mid \left(y_{it-1}^o, x_{it-1}^o\right)_{i=1}^n; \overline{\theta}\right) \end{aligned} \quad (38)$$

where $\theta \in \Theta \subset \mathbb{R}^d$ is the finite dimensional parameter of interest and $\overline{\theta}$ is an additional parameter governing the covariates. The covariates x_{it} are then weakly exogenous in the terminology of Engle, Hendry and Richard (1983). In light of (38) the maximum likelihood estimator is the maximizer of the partial likelihood $L_n(\theta) = \prod_{t=1}^T p\left((y_{it})_{i=1}^n \mid \left(y_{it-1}^o, x_{it}^o\right)_{i=1}^n; \theta\right)$.[11] Assuming that $L_n(\theta)$ is twice continuously differentiable in θ and that the score of the partial log-likelihood $\partial \log L_n(\theta)/\partial\theta$ determines the maximum likelihood estimator, $\hat{\theta}_n$ can be written as the solution to

$$\partial \log L_n\left(\hat{\theta}_n\right)/\partial\theta = 0.$$

We assume that the score has a representation

$$\partial \log L_n(\theta)/\partial\theta = \sum_{i=1}^n \sum_{t=1}^T \psi_{it}(\theta) \quad (39)$$

for some function $\psi_{it}(\theta)$ that satisfies the conditions in Assumption 7 and that depends on $(y_{it}^o, x_{it}^o)_{i=1}^n$ and θ. We give examples indicating that these restrictions hold in many cases of interest. In addition, it should be stressed that ψ_{it} only needs to be specified to verify our regularity conditions for a particular model, but is not needed to compute the likelihood estimator. We use the short hand notation $\psi_{it} = \psi_{it}(\theta_0)$. The following assumptions ensure that Theorem 2 can be applied to ψ_{it}. We focus on correctly specified likelihood models in the sense of satisfying the restriction $E\left[\psi_{it} \middle| \mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = 0$.

Assumption 7 *For $t = 1, \ldots, T, i = 1, \ldots, n, n \geq 1$, the ψ_{it} satisfy Assumptions 1 and 2, and the $\psi_{it}(\theta)$ are continuously differentiable for all $\theta \in \operatorname{int}(\Theta)$.*

To establish a limiting distribution for $\hat{\theta}_n$ we postulate the following additional regularity conditions commonly used in the literature. The existence of respective conditional expectations and probability limits is implicitly assumed.

Assumption 8 *(a) The true parameter $\theta_0 \in \operatorname{int}(\Theta)$.*
(b) $\hat{\theta}_n \xrightarrow{p} \theta_0$.

[11] In calling $L_n(\theta)$ the partial log-likelihood function we adopt the terminology of Cox (1975); see, e.g., also Pötscher and Prucha (1997), pp. 157.

(c) For all θ in some open neighborhood of θ_0 and $t = 1, \ldots, T$, let

$$\begin{aligned} V_t(\theta) &= \operatorname*{plim}_{n\to\infty} n^{-1} \textstyle\sum_{i=1}^n E\left[\psi_{it}(\theta)\,\psi'_{it}(\theta)\,|\mathcal{C}\right], \\ B_t(\theta) &= \operatorname*{plim}_{n\to\infty} n^{-1} \sum_{i=1}^n E\left[\partial\psi_{it}(\theta)/\partial\theta'|\mathcal{C}\right]. \end{aligned}$$

Then for some $\delta > 0$,

$$\begin{aligned} \sup_{\|\theta-\theta_0\|\le\delta} \left\| n^{-1} \textstyle\sum_{i=1}^n \psi_{it}(\theta)\,\psi'_{it}(\theta) - V_t(\theta)\right\| &= o_p(1). \\ \sup_{\|\theta-\theta_0\|\le\delta} \left\| n^{-1} \textstyle\sum_{i=1}^n \partial\psi_{it}(\theta)/\partial\theta' - B_t(\theta)\right\| &= o_p(1). \end{aligned}$$

Furthermore $V_t(\theta)$ and $B_t(\theta)$ are continuous at θ_0 a.s., and $V_t = V_t(\theta_0)$ and $B_t = B_t(\theta_0)$ are nonsingular a.s.

Theorem 6 *Suppose Assumptions 7 and 8 hold, and let $\Omega = \sum_{t=1}^T V_t$ and $B = \sum_{t=1}^T B_t$.*
(a) Then

$$n^{1/2}\left(\hat\theta_n - \theta_0\right) \xrightarrow{d} \Psi^{1/2}\xi_*, \qquad \text{and} \qquad \Psi^{-1/2}n^{1/2}\left(\hat\theta_n - \theta_0\right) \xrightarrow{d} \xi_*$$

with $\Psi = B^{-1}\Omega B'^{-1}$, and where $\xi_* \sim N(0, I_d)$ and independent of $\mathcal{C}$, and thus of Ω, B and Ψ.
(b) Let $\hat\Omega_n = \sum_{t=1}^T \hat V_{t,n}$ and $\hat B_n = \sum_{t=1}^T \hat B_{t,n}$ with

$$\hat V_{t,n} = n^{-1} \textstyle\sum_{i=1}^n \psi_{it}\left(\hat\theta_n\right)\psi'_{it}\left(\hat\theta_n\right) \qquad \text{and} \qquad \hat B_{t,n} = n^{-1} \textstyle\sum_{i=1}^n \partial\psi_{it}\left(\hat\theta_n\right)/\partial\theta',$$

then $\hat\Omega_n \xrightarrow{p} \Omega$ and $\hat B_n \xrightarrow{p} B$. Furthermore, let $\hat\Psi_n = \hat B_n^{-1}\hat\Omega_n \hat B_n'^{-1}$, then $\hat\Psi_n \xrightarrow{p} \Psi$.

The leading example for the result in Theorem 6 are models without common factors. In that case, Ω and B are constants and the MLE has a standard limiting distribution. The importance of the result lies in the fact that it allows for fairly general dependence structures in the underlying data distributions. Under these circumstances the score ψ_{it},while uncorrelated, is generally not independent even conditional on covariates or factors. A martingale CLT then is an alternative way to establish an asymptotic limiting distribution. When common factors are present the limiting distribution of $\hat\theta$ in general is mixed asymptotic normal. These points are illustrated with the following example.

Example 4 *(Dynamic Game) We now discuss a class of models where the form of ψ_{it} and Assumption 2 easily follow from a conditional independence assumption. We consider a dynamic multinomial choice model that arises out of individual utility maximization or from a dynamic game. Individual i chooses alternative a_{it} from a set $\{a_1, ..., a_J\}$ based on observable state variables x_{it} which are common knowledge and private signals ε_{it}. Let $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{nt})$, $\mathbf{a}_t = (a_{1t}, ..., a_{nt})$ and $\mathbf{x}_t = (x_{1t}, ..., x_{nt})$ where $\mathbf{x}_t$ is observed* *by all individuals and the econometrician and ε_{it} is only observed by individual i but not by either the econometrician nor the other individuals. Let $y_{it,j} = \mathbf{1}\{a_{it} = a_j\}$ be the indicator variable that individual i chooses alternative a_j and set $y_{it} = (y_{it,1}, ..., y_{it,J})'$. There are no common shocks in this model. We denote by $\sigma_i(\mathbf{x}_t, \varepsilon_{it}) = a_{it}$ the optimal choice of individual i at time t as a function of common and private information. We also assume that conditional on $\mathbf{x}_t$, the private signals ε_{it} are independent across i. Given this assumption the choices $\sigma_1, ..., \sigma_n$ are independent of each other conditional on $\mathbf{x}_t$.*

Assuming a parametric model with a finite dimensional parameter θ the choice probabilities are given by

$$p_{ij}(\mathbf{x}_t, \theta) = P(y_{it,j} = 1|\mathbf{x}_t, \theta) = P(\sigma_i(\mathbf{x}_t, \varepsilon_{it}, \theta) = a_j|\mathbf{x}_t, \theta)$$

and $p_i(y_{it}|\mathbf{x}_t, \theta) = \prod_{j=1}^J p_{ij}(\mathbf{x}_t, \theta)^{y_{it,j}}$. Then, using conditional independence and assuming that $(\mathbf{x}_t, \mathbf{y}_t)$ is a Markov process and $\mathbf{x}_t$ is weakly exogenous, the transition density of $(\mathbf{x}_t, \mathbf{y}_t)$ can be written as

$$p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}|\mathbf{x}_t, \mathbf{y}_t, \theta) = \prod_{i=1}^n p_i(y_{it+1}|\mathbf{x}_{t+1}, \theta)\, p_x\left(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}_t, \overline{\theta}\right). \tag{40}$$

This is of the same form as the terms in the last product on the r.h.s. of (38) with the added simplification that $p_i(y_{it+1}|\mathbf{x}_{t+1}, \theta)$ does not depend on y_{jt+1}.

We next show that the score of the partial log likelihood based on (40) satisfies Assumption 2, if sufficient conditions on the differentiability of $p_i(y_{it+1}|\mathbf{x}_{t+1}, \theta)$ with respect to θ are imposed. The partial log likelihood of the sample $\{\mathbf{y}_t, \mathbf{x}_t\}_{t=1}^T$ is given by

$$\log L_n(\theta) = \log \prod_{t=1}^T \prod_{i=1}^n p_i(y_{it}|\mathbf{x}_t, \theta) = \sum_{t=1}^T \sum_{i=1}^n \sum_{j=1}^J y_{it,j} \log p_{ij}(\mathbf{x}_t, \theta).$$

From this we see that the score is of the form (39) with

$$\psi_{it}(\theta) \equiv \sum_{j=1}^J \frac{y_{it,j}}{p_{ij}(\mathbf{x}_t, \theta)} \frac{\partial p_{ij}(\mathbf{x}_t, \theta)}{\partial \theta}.$$

Observing that y_{it} is independent of y_{jt} conditional on $\mathbf{x}_t$ it follows that $E(y_{it,j} \mid \mathcal{B}_{n,i,t} \vee \mathcal{C}) = p_{ij}(\mathbf{x}_t, \theta)$ with $\mathcal{B}_{n,i,t}$ as in (4) and $\mathcal{C} = \{0, \Omega\}$. Since $\sum_{j=1}^J \partial p_{ij}(\mathbf{x}_t, \theta)/\partial\theta = 0$ we have $E(\psi_{it} \mid \mathcal{B}_{n,i,t} \vee \mathcal{C}) = 0$ and thus the score $\partial \log L_n(\theta)/\partial\theta$ satisfies Assumption 2.

The multinomial model accommodates random utility models of individuals independently maximizing utility. However, it also accommodates more general dynamic games where actors play stationary Markov strategies taking into account decisions of other actors. Examples of such models can be found in Rust (1994), Aguirregabiria and Mira (2007) and Bajari, Benkard and Levin (2007) who all impose the conditional independence assumption.[12] In these models σ_i is the equilibrium strategy function taking into

[12] In an earlier version of this paper, in Kuersteiner and Prucha (2009), we give a more detailed discussion of how these models fit into our framework.

account the strategic behavior of the other agents through $\mathbf{x}_t$*.*

Despite the fact that ε_{it} *conditional on* $\mathbf{x}_t$ *are independent across* i*, the data* $\{y_{it}, x_{it}\}$ *are not independent in the cross section. Consequently* ψ_{it} *is not independent and conditioning on* $\{\mathbf{x}_t\}_{t=1}^T$ *generally is not feasible unless* $\mathbf{x}_t$ *is strictly exogenous. Thus, a conventional CLT based on independence could not be applied to derive the limiting distribution of the ML estimator* $\hat{\theta} = \arg\max \log L_n(\theta)$*. The main result of Theorem 6 is to provide an asymptotic theory for* $\hat{\theta}$ *that does not rely on independent sampling assumptions and thus extends the framework of Rust (1994), Aguirregabiria and Mira (2007) and Bajari, Benkard and Levin (2007) who rely on independently sampled observations.*

The conditions of Theorem 6 impose additional implicit restrictions on $p_{ij}(\mathbf{x}_t, \theta)$ *which, in the case of the complex models analyzed by Aguirregabiria or Mira (2007) and Bajari, Benkard and Levin (2007), would require an analysis of the Markov equilibrium that determines the form of* $p_{ij}(\mathbf{x}_t, \theta)$ *as well as the transition density* $\pi(.,.)$*. Such an analysis is clearly beyond the scope of this paper. In simpler cases, such as the conditional logit model,* $p_{ij}(\mathbf{x}_t, \theta)$ *reduces to* $p_j(x_{it}, \theta)$ *where the functional form of* p_j *is known and the conditions of Theorem 6 can be guaranteed by imposing conditions on* x_{it} *directly.*

Example 5 *(Social Interactions) Another special case of the choice model considered here is the discrete choice model with social interactions of Brock and Durlauf (2001). Individuals* i *choose a binary action from the set* $\{-1, 1\}$ *and are subject to social interaction in their neighborhood* $n(i)$*. More specifically, individual* i*'s utility depends on the average subjective expected value of the other individuals' choices, denoted by* $m^e_{n(i)}$*. We observe a vector of individual characteristics* x_i *and neighborhood characteristics* z_i*. All individuals in the same neighborhood have the same realization of* z_i*. The probability of choosing* $y_i = 1$ *from the set* $\{-1, 1\}$ *is*

$$\Pr\left(y_i = 1|x_i, z_i, m^e_{n(i)}\right) = \frac{\exp\left(\beta h_i + \beta J m^e_{n(i)}\right)}{\exp\left(-\beta h_i - \beta J m^e_{n(i)}\right) + \exp\left(\beta h_i + \beta J m^e_{n(i)}\right)}$$

where $h_i = (k + \gamma' x_i + d' z_i)$ *,* (β, k, γ, d, J) *are parameters and* $m^e_{n(i)}$ *solves*

$$m^e_{n(i)} = E\left[\tanh\left(\beta\left(k + \gamma' x_i + d' z_i\right) + J m^e_{n(i)}\right)|z_i\right]$$

and thus is a function of z_i*. The conditional log likelihood of a sample* $\{y_i, x_i, z_i\}$ *then is*

$$\sum_{i=1}^{n} \frac{(y_i + 1)}{2} \log\left(\Pr\left(y_i = 1|x_i, z_i\right)\right) + \frac{(1 - y_i)}{2} \log\left(1 - \Pr\left(y_i = -1|x_i, z_i\right)\right).$$

where conditional independence of y_i *given* x_i *and* z_i *is due to the random utility specification of Brock and Durlauf (2001, Eq 1). Then,* ψ_{it} *is readily computed by taking derivatives and satisfies our regularity conditions. In particular, for* $\mathcal{B}_{n,i} = \sigma\left\{(x_j, z_j)_{j=1}^n, y_{-i}\right\}$*, it follows that our basic moment condition* $E[\psi_{it}|\mathcal{B}_{n,i}] = 0$ *is satisfied. This is despite the fact that a sample of randomly selected individuals*

$\{y_i, x_i, z_i\}_{i=i}^n$ across a set of neighborhoods is not independent due to the fact that z_i is the same for all individuals in the same neighborhood.

5 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 numerous 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 vectors (and transformations thereof) of panel data. We consider examples of GMM and ML estimators for models 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 and ML 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 factors 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,

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

(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 any $\lambda \in \mathbb{R}^p$ with $\lambda'\lambda = 1$ we have*

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

with $v_\lambda = \lambda' V \lambda$, where ξ_λ is independent of $\mathcal{F}_0$ (and thus of V) and $\xi_\lambda \sim N(0, 1)$. Then 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) The above statement holds iff

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

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 then given by $\phi(t) = E\left[\exp\{-\frac{1}{2}(t'Vt)\}\right]$, $t \in \mathbb{R}^p$.

(b) Let a be some $\mathcal{F}_0$-measurable vector, then (A.1) implies

$$(Z_n', a') \xrightarrow{d} (Z', a') \ (\mathcal{F}_0\text{-stably}). \tag{A.3}$$

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

$$AZ_n \xrightarrow{d} AV^{1/2}\xi \ (\mathcal{F}_0\text{-stably}) \tag{A.4}$$

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

$$AZ_n \xrightarrow{d} (AVA')^{1/2}\xi_* \ (\mathcal{F}_0\text{-stably}) \tag{A.5}$$

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 $\phi_*(t_*) = E\left[\exp\{-\frac{1}{2}(t_*'AVA't_*)\}\right]$, $t_* \in \mathbb{R}^{p_*}$.

Proof of Proposition A.2. (a) Suppose (A.1) holds. Then in light of Proposition A.1(v), for all $s \in \mathbb{R}$ and all $\mathcal{F}_0$-measurable P-essentially bounded random variables ζ we have

$$E\left[\zeta \exp(is\lambda' Z_n)\right] \to E\left[\zeta \exp(isv_\lambda^{1/2}\xi_\lambda)\right] \tag{A.6}$$

as $n \to \infty$. Since ζ and V are $\mathcal{F}_0$-measurable, and $\xi_\lambda \sim N(0,1)$ we have

$$E\left[\zeta \exp(isv_\lambda^{1/2}\xi_\lambda)\right] = E\left[\zeta E\left[\exp(isv_\lambda^{1/2}\xi_\lambda)|\mathcal{F}_0\right]\right] = E\left[\zeta \exp\{-\frac{1}{2}(s^2\lambda' V\lambda)\}\right], \tag{A.7}$$

and thus

$$E\left[\zeta \exp(is\lambda' Z_n)\right] \to E\left[\zeta \exp\{-\frac{1}{2}(s^2\lambda' V\lambda)\}\right]. \tag{A.8}$$

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

$$E\left[\zeta \exp(it'Z_n)\right] \to E\left[\zeta \exp\{-\frac{1}{2}(t'Vt)\}\right]. \tag{A.9}$$

Choosing λ and s to be such that $t = s\lambda$, this is seen to hold in light of (A.8). Next suppose that (A.2) and thus (A.9) holds. Then (A.6) is seen to hold in light of (A.7) and taking $t = s\lambda$.

(b) Let $\alpha = (a', \vartheta)$ where ϑ is some $\mathcal{F}_0$-measurable random variables, and let $Z = V^{1/2}\xi$. By Proposition A.1(iii) it follows that for any fixed $\gamma \in \mathbb{R}^{\dim(\alpha)}$ and $\lambda \in \mathbb{R}^p$ with $\gamma'\gamma + \lambda'\lambda = 1$ we have $(\gamma'\alpha, Z_n) \to^d (\gamma'\alpha, Z)$ jointly, because $\gamma'\alpha$ is $\mathcal{F}_0$-*measurable.* By the Continuous Mapping Theorem it follows that $\left(\gamma'\alpha, \lambda' Z_n\right) \to^d \left(\gamma'\alpha, \lambda' Z\right)$, and thus by the Cramer-Wold device

$$\left(a', Z_n', \vartheta\right) \to^d \left(a', Z', \vartheta\right). \tag{A.10}$$

Consequently we have $(a', Z_n',) \to^d (a', Z')$ ($\mathcal{F}_0$-stably) by Proposition A.1(iii), which establishes the first claim.

Now take $a = \text{vec}(A)$, then it follows from (A.10) and the Continuous Mapping Theorem that $(g(\text{vec}(A), Z_n), \vartheta) \to^d (g(\text{vec}(A), Z), \vartheta)$ for any continuous $g(.)$. Now take $g(\text{vec}(A), Z) = AZ$. By Proposition A.1(iii) it follows that $AZ_n \xrightarrow{d} AZ$ ($\mathcal{F}_0$-stably), which establishes the second claim.

To prove the third claim observe that by Proposition A.1(v) $AZ_n \xrightarrow{d} AZ$ ($\mathcal{F}_0$-stably) iff for all real vectors $t_* \in \mathbb{R}^{p_*}$ and all $\mathcal{F}_0$-measurable P-essentially bounded random variables ζ we have

$$E\left[\zeta \exp(it_*' AZ_n)\right] \to E\left[\zeta \exp(it_*' AV^{1/2}\xi)\right] \tag{A.11}$$

Since ζ, A and V are $\mathcal{F}_0$-measurable, and $\xi \sim N(0, I_p)$ we have

$$E\left[\zeta \exp(it_*' AV^{1/2}\xi)\right] = E\left[\zeta E\left[\exp(it_*' AV^{1/2}\xi)|\mathcal{F}_0\right]\right] = E\left[\zeta \exp\{-\frac{1}{2}(t_*' AVA' t_*)\}\right]. \tag{A.12}$$

By an analogous argument we also see that

$$E\left[\zeta \exp(it_*'(AVA')^{1/2}\xi_*)\right] = E\left[\zeta \exp\{-\frac{1}{2}(t_*' AVA' t_*)\}\right]. \tag{A.13}$$

Thus in light of (A.11)-(A.13)

$$E\left[\zeta \exp(it_*' AZ_n)\right] \to E\left[\zeta \exp(it_*'(AVA')^{1/2}\xi_*)\right],$$

which establishes the third claim in light of Proposition A.1(v). The claim concerning the characteristic functions is seen to hold as a special case of (A.12) and (A.13) with $\zeta = 1$. ■

A.2 Proofs for Section 2.1

Proof for Lemma 1. Let all variables be defined as in the proof of Theorem 2 below. Under the assumptions of the lemma we then have

$$\begin{aligned} V_{nk_n}^2 &= \sum_{v=1}^{k_n} E\left[X_{n,v}^2 | \mathcal{F}_{n,v-1}\right] = \sum_{t=1}^{T} n^{-1} \sum_{i=1}^{n} \lambda_t' E\left[\psi_{it}\psi_{it}' | \mathcal{F}_{n,(t-1)n+i}\right] \lambda_t \\ &= \lambda' \bar{V}_n \lambda \xrightarrow{p} \lambda' V \lambda = \eta^2. \end{aligned} \tag{A.14}$$

We next show that

$$\text{for any } \varepsilon > 0, \qquad \sum_{v=1}^{k_n} E\left[X_{n,v}^2 \mathbf{1}\left(|X_{n,v}| > \varepsilon\right) \mid \mathcal{F}_{n,v-1}\right] \xrightarrow{p} 0. \tag{A.15}$$

To see this observe from the proof of Theorem 2 that under Assumption 1(a) the Condition (A.26) holds, i.e., $\sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta}\right] \to 0$ as $n \to \infty$. Hence

$$E\left\{\sum_{v=1}^{k_n} E\left[X_{n,v}^2 \mathbf{1}\left(|X_{n,v}| > \varepsilon\right) \mid \mathcal{F}_{n,v-1}\right]\right\} = \sum_{v=1}^{k_n} E\left[X_{n,v}^2 \mathbf{1}\left(|X_{n,v}| > \varepsilon\right)\right] \to 0$$

as $n \to \infty$ in light of inequality (A.29). The claim in (A.15) now follows from Chebychev's inequality. Furthermore, observe from the proof of Theorem 2 that under Assumption 1(a) also the Condition (A.28) holds, i.e., $\sup_n E\left[V_{nk_n}^{2+\delta}\right] < \infty$. Let $U_{nk_n}^2 = \sum_{v=1}^{k_n} X_{n,v}^2 = \lambda' \tilde{V}_n \lambda$, then observing that Condition (A.28) implies that V_{nk_n} is uniformly integrable it follows from Hall and Heyde (1980, Theorem 2.23) that

$$E\left[\left|V_{nk_n}^2 - U_{nk_n}^2\right|\right] \to 0. \tag{A.16}$$

By Chebychev's inequality this implies that $V_{nk_n}^2 - U_{nk_n}^2 \xrightarrow{p} 0$, which in turn implies that $\tilde{V}_n \xrightarrow{p} V$ as postulated in Assumption 1(c). ■

A.3 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 $C > 1$,

$$P\left(\eta^2 < C\right) = 1. \tag{A.17}$$

Define $X_{ni}^\dagger = X_{ni} \mathbf{1}\left\{\sum_{j=1}^{i-1} X_{nj}^2 \leq 2C\right\}$ with $X_{n1}^\dagger = X_{n1}$, and $S_{ni}^\dagger = \sum_{j=1}^{i} X_{nj}^\dagger$ 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 $E\left[S_{ni}^2\right] < \infty$, (iii) $E[S_{ni}|\mathcal{F}_{nj}] = S_{nj}$ 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 $E[X_{ni}|\mathcal{F}_{nj}] = 0$ for $0 \leq j < i$ and $1 \leq i \leq k_n$, since $E[X_{n1}|\mathcal{F}_{n0}] = 0$ by assumption, and for $2 \leq j < i$

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

We now establish that $\left\{S_{ni}^\dagger, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\right\}$ is also a zero mean, square integrable martingale array with, by construction, differences $X_{ni}^\dagger$. Since the random variables $X_{n1}, \ldots, X_{ni}$ are measurable w.r.t. $\mathcal{F}_{ni}$, clearly $S_{ni}^\dagger$ is measurable w.r.t. $\mathcal{F}_{ni}$. Also, since $\left|S_{ni}^\dagger\right| \leq |S_{ni}|$ clearly $E\left[S_{ni}^{\dagger 2}\right] \leq E\left[S_{ni}^2\right] < \infty$. Next observe that $E\left[X_{n1}^\dagger|\mathcal{F}_{n0}\right] = E[X_{n1}|\mathcal{F}_{n0}] = 0$ by assumption, and for $2 \leq j < i$

$$E\left[X_{ni}^\dagger|\mathcal{F}_{nj}\right] = E\left[E\left[X_{ni}^\dagger|\mathcal{F}_{ni-1}\right]|\mathcal{F}_{nj}\right] = 0.$$

By iterated expectations $E\left[X_{ni}^\dagger\right] = 0$ and thus $E\left[S_{ni}^\dagger\right] = 0$. Furthermore for $1 \leq j < i$, $E\left[S_{ni}^\dagger|\mathcal{F}_{nj}\right] = S_{nj}^\dagger$.This verifies that $\left\{S_{ni}^\dagger, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\right\}$ is indeed a zero mean, square integrable martingale array.

Next let $U_{nk_n}^2 = \sum_{i=1}^{k_n} X_{ni}^2$, then clearly $P(U_{nk_n}^2 > 2C) \to 0$ in light of (15). Consequently

$$P(X_{ni}^\dagger \neq X_{ni} \text{ for some } i \leq k_n) \leq P(U_{nk_n}^2 > 2C) \to 0, \tag{A.18}$$

which in turn implies $P(S_{nk_n}^\dagger \neq S_{nk_n}) \to 0$, and furthermore

$$E\left[\left|\zeta \exp(itS_{nk_n}^\dagger) - \zeta \exp(itS_{nk_n})\right|\right] \to 0$$

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 $S_{nk_n}^\dagger \xrightarrow{d} Z$ ($\mathcal{F}_0$-stably). Observe furthermore that in view of (A.18) the martingale differences $\{X_{ni}^\dagger\}$ satisfy that $\max_i \left|X_{ni}^\dagger\right| \xrightarrow{p} 0$ and $\sum_{i=1}^{k_n} X_{ni}^{\dagger 2} \xrightarrow{p} \eta^2$. Since $\left|X_{ni}^\dagger\right| \leq |X_{ni}|$ condition (16) implies furthermore that $E\left[\max_i X_{ni}^{\dagger 2}\right]$ is bounded in n.

We now show that $S_{nk_n}^\dagger \xrightarrow{d} Z$ ($\mathcal{F}_0$-stably). Let $U_{ni}^2 = \sum_{j=1}^{i} X_{nj}^2$ and $T_n^\dagger(t) = \prod_{j=1}^{k_n}\left(1 + itX_{nj}^\dagger\right)$ with

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

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

$$\begin{aligned} E\left[\left|T_n^\dagger(t)\right|^2\right] &= E\left[\prod_{j=1}^{k_n}\left(1 + t^2 X_{nj}^{\dagger 2}\right)\right] \leq E\left[\left\{\exp\left(t^2 \sum_{j=1}^{J_n - 1} X_{nj}^{\dagger 2}\right)\left(1 + t^2 X_{nJ_n}^{\dagger 2}\right)\right\}\right] \\ &\leq \left\{\exp(2Ct^2)\right\}\left(1 + t^2 E\left[X_{nJ_n}^{\dagger 2}\right]\right). \end{aligned}$$

Since $E\left[X_{nJ_n}^{\dagger 2}\right] \leq E\left[X_{nJ_n}^2\right]$ is uniformly bounded it follows from the above inequality that $E\left[\left|T_n^\dagger(t)\right|^2\right]$ is uniformly bounded in n.

Now define $I_n = \exp\left(itS_{nk_n}^\dagger\right)$ and $W_n = \exp\left(-\frac{1}{2}t^2 \sum_{i=1}^{k_n} X_{ni}^{\dagger 2} + \sum_{i=1}^{k_n} r\left(tX_{ni}^\dagger\right)\right)$ where $r(.)$ is implicitly defined by $e^{ix} = (1 + ix)\exp\left(-\frac{1}{2}x^2 + r(x)\right)$ as in Hall and Heyde (1980), p. 57. Then

$$I_n = T_n^\dagger(t)\exp\left(-\eta^2 t^2/2\right) + T_n^\dagger(t)(W_n - \exp\left(-\eta^2 t^2/2\right)). \tag{A.19}$$

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

$$E\left(I_n \zeta\right) \to E\left[\exp\left(-\eta^2 t^2/2\right)\zeta\right] \tag{A.20}$$

for any P-essentially bounded $\mathcal{F}_0$-measurable random variable ζ. Because $\mathcal{F}_0 \subset \mathcal{F}_{ni}$ it follows that

$\exp\left(-\eta^2t^2/2\right)\zeta$ is $\mathcal{F}_{ni}$-measurable for all n and $i \le k_n$. Hence,

$$\begin{aligned} E\left[T_n^{\dagger}(t)\exp\left(-\eta^2t^2/2\right)\zeta\right] &= E\left[\exp\left(-\eta^2t^2/2\right)\zeta\prod_j^{k_n}\left(1+itX_{nj}^{\dagger}\right)\right] \\ &= E\left\{E\left[\exp\left(-\eta^2t^2/2\right)\zeta\prod_j^{k_n}\left(1+itX_{nj}^{\dagger}\right)|\mathcal{F}_{nk_n-1}\right]\right\} \\ &= E\left\{\exp\left(-\eta^2t^2/2\right)\zeta\prod_j^{k_n-1}\left(1+itX_{nj}^{\dagger}\right)E\left[\left(1+itX_{nk_n}^{\dagger}\right)|\mathcal{F}_{nk_n-1}\right]\right\} \\ &= E\left\{\exp\left(-\eta^2t^2/2\right)\zeta\prod_j^{k_n-1}\left(1+itX_{nj}^{\dagger}\right)\right\} \\ &= \ldots \\ &= E\left\{\exp\left(-\eta^2t^2/2\right)\zeta E\left[\left(1+itX_{n1}^{\dagger}\right)|\mathcal{F}_{n0}\right]\right\} = E\left[\exp\left(-\eta^2t^2/2\right)\zeta\right]. \end{aligned}$$

Thus, in light of (A.19), for (A.20) to hold it suffices to show that

$$E\left[T_n^{\dagger}(t)\left(W_n - \exp\left(-\eta^2t^2/2\right)\right)\zeta\right] \to 0. \tag{A.21}$$

Let K be some constant such that $P(|\zeta| \le K) = 1$, then $E\left[\left|T_n^{\dagger}(t)\exp\left(-\eta^2t^2/2\right)\zeta\right|^2\right] \le K^2E\left[\left|T_n^{\dagger}(t)\right|^2\right]$ is uniformly bounded in n, since $E\left[\left|T_n^{\dagger}(t)\right|^2\right]$ 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.19) it follows furthermore that

$$E\left[\left|T_n^{\dagger}(W_n - \exp\left(-\eta^2t^2/2\right))\zeta\right|^2\right] \le 2E\left[|I_n\zeta|^2\right] + 2E\left[\left|T_n^{\dagger}(t)\exp\left(-\eta^2t^2/2\right)\zeta\right|^2\right]$$

is uniformly bounded in n, it follows that $T_n^{\dagger}(t)(W_n - \exp\left(-\eta^2t^2/2\right))\zeta$ is uniformly integrable. Having established uniform integrability, Condition (A.21) now follows since as shown by Hall and Heyde (1980, p. 58), $W_n - \exp\left(-\eta^2t^2/2\right) \xrightarrow{p} 0$ by using Conditions (14) and (15). Thus, it follows that $T_n^{\dagger}\left(W_n - \exp\left(-\eta^2t^2/2\right)\right)\zeta \xrightarrow{p} 0$. This completes the proof that $S_{nk_n}^{\dagger} \xrightarrow{d} Z$ ($\mathcal{F}_0$-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 $E\left[\zeta\exp(it\eta\xi)\right] = E\left[\zeta\exp(-\frac{1}{2}\eta^2t^2)\right]$ by iterated expectations, and thus $S_{nk_n} \xrightarrow{d} \eta\xi$ ($\mathcal{F}_0$-stably) in light of Proposition A.1(v). ∎

A.4 Proof of Central Limit Theorem for Panel Data

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 $\lambda'\psi_{(n)}$. In particular, we consider the representation $\lambda'\psi_{(n)} = \sum_{v=1}^{k_n} X_{n,v}$ with $k_n = Tn + 1$, defined by (22)-(23), and the corresponding information sets defined in (12). We recall the definitions $(t = 1, \dots, T, i = 1, \dots, n)$

$$\begin{aligned} X_{n,(t-1)n+i+1} &= n^{-1/2}\lambda_t'\psi_{it}, \\ \mathcal{F}_{n,(t-1)n+i} &= \mathcal{G}_{n,(t-1)n+i} \vee \mathcal{C} \end{aligned}$$

with $X_{n,1} = 0$. 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 λ_t is constant it is readily seen that $X_{n,v}$ is measurable w.r.t. to $\mathcal{F}_{n,v}$ by Assumption 1(b). Observing further that $\mathcal{F}_{n,(t-1)n+i} \subseteq \mathcal{B}_{n,i,t} \vee \mathcal{C}$ it follows from moment condition (8) of Assumption 2 that

$$\begin{aligned} E\left[X_{n,v}|\mathcal{F}_{n,v-1}\right] &= E\left[X_{n,(t-1)n+i+1}|\mathcal{F}_{n,(t-1)n+i}\right] \\ &= n^{-1/2}\lambda_t' E\left[\psi_{it}|\mathcal{F}_{n,(t-1)n+i}\right] = 0. \end{aligned} \tag{A.22}$$

Observe that $\|\lambda_t\| \le 1$ because $\|\lambda\| = 1$ and $\sup_t \|\lambda_t\|^2 \le \|\lambda\|^2$, and recall that $\mathcal{F}_{n,(t-1)n+i} \subseteq \mathcal{B}_{n,i,t} \vee \mathcal{C}$. Consider some γ with $0 \le \gamma \le \delta$, then

$$\begin{aligned} & E\left[\left|X_{n,v}\right|^{2+\gamma}|\mathcal{F}_{n,v-1}\right] = E\left[\left|X_{n,(t-1)n+i+1}\right|^{2+\gamma}|\mathcal{F}_{n,(t-1)n+i}\right] \\ = \; & \frac{1}{n^{1+\gamma/2}} E\left[\left|\lambda_t'\psi_{it}\right|^{2+\gamma}|\mathcal{F}_{n,(t-1)n+i}\right] \le \frac{1}{n^{1+\gamma/2}} \|\lambda_t\|^{2+\gamma} E\left[\|\psi_{it}\|^{2+\gamma}|\mathcal{F}_{n,(t-1)n+i}\right] \\ \le \; & \frac{1}{n^{1+\gamma/2}} \left\{E\left[\|\psi_{it}\|^{2+\delta}|\mathcal{F}_{n,(t-1)n+i}\right]\right\}^{\frac{2+\gamma}{2+\delta}} \\ \le \; & \frac{1}{n^{1+\gamma/2}} \left\{E\left[E\left[\|\psi_{it}\|^{2+\delta}|\mathcal{B}_{n,i,t} \vee \mathcal{C}\right]|\mathcal{F}_{n,(t-1)n+i}\right]\right\}^{\frac{2+\gamma}{2+\delta}} \le \frac{1}{n^{1+\gamma/2}} \left\{E\left[B_{it}|\mathcal{F}_{n,(t-1)n+i}\right]\right\}^{\frac{2+\gamma}{2+\delta}} \end{aligned} \tag{A.23}$$

using Lyapunov's inequality, iterated expectations and condition (5) postulated by Assumption 1(a). For $\gamma = 0$ this implies that

$$E\left[\left|X_{n,v}\right|^2|\mathcal{F}_{n,v-1}\right] \le \frac{1}{n}\left\{E\left[B_{it}|\mathcal{F}_{n,(t-1)n+i}\right]\right\}^{\frac{1}{1+\delta/2}}, \tag{A.24}$$

and for $\gamma = \delta$ we have

$$E\left[\left|X_{n,v}\right|^{2+\delta}|\mathcal{F}_{n,v-1}\right] \le \frac{1}{n^{1+\delta/2}} E\left[B_{it}|\mathcal{F}_{n,(t-1)n+i}\right]. \tag{A.25}$$

Let $V_{nk_n}^2 = \sum_{v=1}^{k_n} E\left[X_{n,v}^2|\mathcal{F}_{n,v-1}\right]$ denote the conditional variance. We next show that the following conditions

$$\sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta}\right] \to 0, \tag{A.26}$$

$$U^2_{nk_n} = \sum_{v=1}^{k_n} X^2_{n,v} \xrightarrow{p} \eta^2, \tag{A.27}$$

$$\sup_n E\left[V^{2+\delta}_{nk_n}\right] = \sup_n E\left[\sum_{v=1}^{k_n} E\left[X^2_{n,v}|\mathcal{F}_{n,v-1}\right]\right]^{1+\delta/2} < \infty. \tag{A.28}$$

are sufficient for Assumptions (14), (15), and (16) of Theorem 1. As discussed by Hall and Heyde (1980, p. 53) Condition (14) is equivalent to

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

Condition (14) is now seen to hold since

$$\sum_{v=1}^{k_n} E\left[X^2_{n,v}\mathbf{1}\left(|X_{n,v}| > \varepsilon\right)\right] = \sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta}\mathbf{1}\left(|X_{n,v}| > \varepsilon\right)/\,|X_{n,v}|^{\delta}\right] \leq \varepsilon^{-\delta}\sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta}\right] \to 0 \tag{A.29}$$

in light of condition (A.26). Condition (15) is the same as (A.27). Condition (16) is seen to hold since $E\left[U^2_{nk_n}\right] = E\left[V^2_{nk_n}\right]$ is uniformly bounded in light of Condition (A.28), using Lyapunov's inequality.

We next verify Conditions (A.26), (A.27) and (A.28). Utilizing (A.25) and Assumption 1(a) it then follows that

$$\sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta}\right] = \sum_{v=1}^{k_n} E\left\{E\left[|X_{n,v}|^{2+\delta}|\mathcal{F}_{n,v-1}\right]\right\} \leq \frac{nT+1}{n^{1+\delta/2}} \sup_{i,t} E\left[B_{it}\right] \to 0 \tag{A.30}$$

as $n \to \infty$, observing that the uniform integrability of B_{it} implies that $\sup_{i,t} E\left[B_{it}\right] < \infty$. This establishes condition (A.26). By Assumption 1(c) we have

$$\sum_{v=1}^{k_n} X^2_{n,v} = \sum_{t=1}^{T} n^{-1} \sum_{i=1}^{n} \lambda'_t \psi_{it} \psi'_{it} \lambda_t = \lambda' \tilde{V}_n \lambda \xrightarrow{p} \lambda' V \lambda = \eta^2. \tag{A.31}$$

This verifies (A.27).

Next observe that

$$\begin{aligned} E\left[V^{2+\delta}_{nk_n}\right] &= E\left[\left(\sum_{v=1}^{k_n} E\left[X^2_{n,v}|\mathcal{F}_{n,v-1}\right]\right)^{1+\delta/2}\right] \leq k_n^{\delta/2} E\left[\sum_{v=1}^{k_n}\left(E\left[X^2_{n,v}|\mathcal{F}_{n,v-1}\right]\right)^{1+\delta/2}\right] \\ &\leq k_n^{\delta/2} E\left[\sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta}|\mathcal{F}_{n,v-1}\right]\right] \end{aligned} \tag{A.32}$$

where the second line follows from inequality (1.4.3) in Bierens (1994) and the third from Lyapunov's inequality. Using (A.25) we have

$$E\left[V_{nk_n}^{2+\delta}\right] \leq \frac{k_n^{\delta/2}}{n^{1+\delta/2}} \sum_{t=1}^{T} \sum_{i=1}^{n} E\left[B_{it}\right] \leq \frac{(Tn+1)^{1+\delta/2}}{n^{1+\delta/2}} \sup_{i,t} E\left[B_{it}\right] < \infty$$

where the last inequality follows in light of Assumption 1(a). This establishes (A.28). Of course, in light of (A.24) and argumentation as above we also have

$$E\left[X_{n,v}^2\right] = E\left\{E\left[X_{n,v}^2|\mathcal{F}_{n,v-1}\right]\right\} \leq \frac{1}{n} E\left[\left\{E\left[B_{it}|\mathcal{F}_{n,(t-1)n+i}\right]\right\}^{\frac{1}{1+\delta/2}}\right] \leq \frac{1}{n}\left\{\sup_{i,t} E\left[B_{it}\right]\right\}^{\frac{1}{1+\delta/2}} < \infty,$$

which establishes that $X_{n,v}$ is square integrable.

Having verified all conditions of Theorem 1 it follows from that theorem that $\lambda'\psi_{(n)} \xrightarrow{d} v_\lambda^{1/2} Z_\lambda$ ($\mathcal{C}$-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. The claim in Part(a) of the theorem now follows from Proposition A.2(a).

To prove **Part (b)** of the theorem we show that the proof of Part (a) can be readily extended to $\check{\psi}_{it} = \psi_{it} - E\left[\psi_{it}|\mathcal{F}_{n,(t-1)n+i}\right]$ where now $X_{n,(t-1)n+i+1} = n^{-1/2}\lambda_t'\check{\psi}_{it}$. By construction $E\left[\check{\psi}_{it}|\mathcal{F}_{n,(t-1)n+i}\right] = 0$ and so clearly $E\left[X_{n,v}|\mathcal{F}_{n,v-1}\right] = 0$. Of course, given that ψ_{it} is $\mathcal{F}_{n,(t-1)n+i+1}$ measurable it follows that $\check{\psi}_{it}$ is also $\mathcal{F}_{n,(t-1)n+i+1}$ measurable. Next observe that

$$\begin{aligned}
E\left[\left\|\check{\psi}_{it}\right\|^{2+\delta}|\mathcal{F}_{n,v-1}\right] &\leq E\left[\left(\|\psi_{it}\| + \left\|E\left[\psi_{it}|\mathcal{F}_{n,(t-1)n+i}\right]\right\|\right)^{2+\delta}|\mathcal{F}_{n,(t-1)n+i}\right] \\
&\leq 2^{1+\delta} E\left[\|\psi_{it}\|^{2+\delta} + \left\|E\left[\psi_{it}|\mathcal{F}_{n,(t-1)n+i}\right]\right\|^{2+\delta}|\mathcal{F}_{n,(t-1)n+i}\right] \\
& \quad 2^{1+\delta} E\left[\|\psi_{it}\|^{2+\delta} + \left\|E\left[\psi_{it}|\mathcal{F}_{n,(t-1)n+i}\right]\right\|^{2+\delta}|\mathcal{F}_{n,(t-1)n+i}\right] \\
&\leq 2^{2+\delta} E\left[\|\psi_{it}\|^{2+\delta}|\mathcal{F}_{n,(t-1)n+i}\right]
\end{aligned}$$

Thus, by argumentation analogous as in (A.23) we get for $0 \leq \gamma \leq \delta$

$$E\left[\left|X_{n,v}\right|^{2+\gamma}|\mathcal{F}_{n,v-1}\right] \leq \frac{K}{n^{1+\gamma/2}}\left\{E\left[B_{it}|\mathcal{F}_{n,(t-1)n+i}\right]\right\}^{\frac{2+\gamma}{2+\delta}}$$

where K is a constant. From this we see that the inequalities (A.24) and (A.25) continue to hold with 1 replaced by K. Consequently the the remainder of the proof is seen to carry over, with $\check{V}$ in place of V, which establishes that under Assumptions 1 and 3,

$$\lambda'\left(\psi_{(n)} - \sqrt{n}b_n\right) \xrightarrow{d} \check{v}_\lambda^{1/2} Z_\lambda \text{ ($\mathcal{C}$-stably)} \tag{A.33}$$

where $\check{v}_\lambda = \lambda'\check{V}\lambda$. Note that when additionally Assumption 2 holds, $E\left[X_{n,(t-1)n+i+1}|\mathcal{F}_{n,(t-1)n+i}\right] = 0$ and thus $b_n = 0$, the result follows trivially from Part (a). Of course, (A.33) implies further that $\check{v}_\lambda^{1/2} Z_\lambda$ has the characteristic function $\phi_\lambda(s) = E\exp\{-\frac{1}{2}(\lambda'\check{V}\lambda)s^2\}$, $s \in \mathbb{R}$. By Proposition A.2 it follows from (A.33) that

$$\left(\psi_{(n)} - \sqrt{n}b_n\right) \xrightarrow{d} \check{V}^{1/2}\xi \ (\mathcal{C}\text{-stably}), \tag{A.34}$$

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

$$A\left(\psi_{(n)} - \sqrt{n}b_n\right) \xrightarrow{d} (A\check{V}A')^{1/2}\xi_* \ (\mathcal{C}\text{-stably}). \tag{A.35}$$

The claim in (18) holds observing that under Assumption 2 we have$\sqrt{n}b_n = 0$ and $\check{V} = V$. Obviously (A.35) also establishes the claim in (19) under Assumption 3(a). The claim that under Assumption 3(a) $A\psi_{(n)}$ diverges is obvious since under this assumption $b_n \xrightarrow{p} b$ and thus $A\psi_{(n)} = A\left(\psi_{(n)} - \sqrt{n}b_n\right) + \sqrt{n}Ab_n = O_p(n^{1/2})$, observing that the first term on the r.h.s. as well as Ab_n are $O_p(1)$. To verify the claim in (20) under Assumption 3(b) observe that by Proposition A.2(b)

$$\left(\psi_{(n)} - \sqrt{n}b_n, b, A, \vartheta\right) \xrightarrow{d} \left(\check{V}^{1/2}\xi, b, A, \vartheta\right) \ (\mathcal{C}\text{-stably})$$

for all real valued $\mathcal{C}$-measurable random variables ϑ. Since $\sqrt{n}b_n - b \xrightarrow{p} 0$ it follows furthermore that

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

which in turn implies in light of the continuous mapping theorem that

$$(A\psi_{(n)}, \vartheta) = \left(A(\psi_{(n)} - \sqrt{n}b_n) - A(\sqrt{n}b_n - b) + Ab, \vartheta\right) \xrightarrow{d} \left(A\check{V}^{1/2}\xi + Ab, \vartheta\right).$$

It now follows from Proposition A.1(iii) that

$$A\psi_{(n)} \xrightarrow{d} A\check{V}^{1/2}\xi + Ab \ (\mathcal{C}\text{-stably}).$$

The claim in (20) is now seen to hold by arguments analogous to those in the proof of Proposition A.2(b). Under Assumption 3(c) we have$\sqrt{n}b_n \xrightarrow{p} 0$, and thus in this case (20) holds with $b = 0$. ■

Proofs for Example 1. Using the definitions in the example it follows that $\sum_{v=1}^{k} X_{n,v}$ is measurable w.r.t. $\mathcal{G}_{n,k}$, $E\left[\sum_{v=1}^{k} X_{n,v}|\mathcal{G}_{n,k-1}\right] = \sum_{v=1}^{k-1} X_{n,v}$, and thus

$$E\left[\sum_{v=1}^{k} X_{n,v}\right]^2 \leq E\left[\sum_{v=1}^{2n+1} X_{n,v}^2\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[u_{i1}^2\right]\lambda_1^2 + \lambda_2^2 E\left[f^2\right]\frac{1}{n}\sum_{i=1}^{n} E\left[u_{i2}^2\right] = 1, \tag{A.36}$$

observing that $f = W_1(1) \sim N(0,1)$. Also,

$$\sum_{v=1}^{2n+1} X_{n,v}^2 = \lambda_1^2\frac{1}{n}\sum_{i=1}^{n} u_{i1}^2 + \lambda_2^2 f^2\frac{1}{n}\sum_{i=1}^{n} u_{i2}^2 \xrightarrow{p} \left(\lambda_1^2 + \lambda_2^2 f^2\right)$$

by the continuous mapping theorem and a law of large numbers for a triangular array of i.i.d. random variables. Furthermore, in light of (A.36)

$$E\left[\max_v X_{n,v}^2\right] \leq E\left[\sum_{v=1}^{2n+1} X_{n,v}^2\right] = 1 \text{ uniformly in } n.$$

Using Gaussianity we have $E\left[f^4\right] = 3$ such that

$$E\left[\sum_{v=1}^{2n+1} X_{n,v}^4\right] = \lambda_1^4 \frac{1}{n^2}\sum_{i=1}^{n} E\left[u_{i1}^4\right] + \lambda_2^4 E\left[f^4\right] \frac{1}{n^2}\sum_{i=1}^{n} E\left[u_{i2}^4\right] \to 0,$$

which implies that $\max_v |X_{nv}| \xrightarrow{p} 0$, recalling that in the proof of Theorem 2 we showed that condition (A.26) implies condition (14). Thus, $\{S_{nv}, \mathcal{G}_{nv}, 1 \leq k \leq 2n+1\}$ satisfies all conditions of Hall and Heyde (1980, Theorem 3.2) except for condition (3.21) which requires $\mathcal{G}_{n,v} \subseteq \mathcal{G}_{n+1,v}$.

For the discussion below Example 1 note that u_{i1} and f are jointly normal with distribution

$$\begin{bmatrix} u_{i1} \\ f \end{bmatrix} \sim N\left(0, \begin{bmatrix} 1 & 1/\sqrt{n} \\ 1/\sqrt{n} & 1 \end{bmatrix}\right).$$

This implies that $E\left[u_{i1}|f, u_{11}, ..., u_{i-1,1}\right] = n^{-1/2} f$ and thus $E\left[X_{n,i+1}|\mathcal{F}_{n,i}\right] = E\left[n^{-1/2} u_{i1}\lambda_1 | f, u_{11}, ..., u_{i-1,1}\right] = n^{-1} f \lambda_1$ for $i = 1, \ldots, n$. Observing further that $E\left[X_{n,n+i+1}|\mathcal{F}_{n,n+i}\right] = 0$ it follows that

$$\check{X}_{n,v} = \begin{cases} 0 & v = 1 \\ X_{n,v} - \frac{1}{n} f\lambda_1 & 1 < v \leq n+1 \\ X_{n,v} & n+1 < v \leq 2n+1 \end{cases}$$

is a martingale difference sequence with respect to the filtration $\mathcal{F}_{n,v}$, i.e. $E\left[\check{X}_{n,v}|\mathcal{F}_{n,v-1}\right] = 0$ for $1 \leq v \leq 2n+1$. Because

$$\sum_{v=1}^{n+1} \check{X}_{n,v} = \sum_{v=2}^{n+1}(X_{n,v} - \frac{1}{n} f\lambda_1) = n^{-1/2}\sum_{i=1}^{n} \varepsilon_{i1}\lambda_1 - f\lambda_1 = 0$$

it follows that

$$\sum_{v=1}^{2n+1} X_{n,v} = \sum_{v=1}^{2n+1} \check{X}_{n,v} + \sum_{v=2}^{n+1} E\left[X_{n,v}|\mathcal{F}_{n,v}\right] = \sum_{v=1}^{2n+1} \check{X}_{n,v} + f\lambda_1 = \sum_{v=n+2}^{2n+1} X_{n,i} + f\lambda_1.$$

Note that since $\sum_{v=1}^{n+1} \check{X}_{n,v}$ is zero the the joint limit in $\sum_{v=1}^{2n+1} \check{X}_{n,v}$ is degenerate in the sense that the component corresponding to λ_1 is zero. Yet, as demonstrated above, the term $\sum_{i=n+1}^{2n} \check{X}_{n,i} = \sum_{i=n+1}^{2n} X_{n,i}$ corresponding to λ_2 satisfies all conditions of Theorem 1. Thus it converges $\mathcal{C}$-stably to $(\lambda_2^2 f^2)^{1/2} Z$ with $Z \sim N(0,1)$ independent of f. The fact that f is measurable w.r.t $\mathcal{C}$ establishes the result. This example illustrates that Condition (13) is not always satisfied. However, in this particular example it is sufficient to concentrate on the non-degenerate part of $\sum_{v=1}^{2n+1} \check{X}_{n,v}$. ■

B Appendix B: Proofs for Section 3

Proof of Theorem 3. To prove **Part (a)** of the Theorem it is enough to show that for $\psi_{it} = c_{it}u_{it}$ as defined by (30) and (31) together with Assumptions 4 and 5 imply Assumptions 1 and 2. Recall that $\mathcal{B}_{n,i,t} = \sigma\left\{\left(x_{jt}^o, z_j, u_{j,t-1}^o, \mu_j\right)_{j=1}^n, u_{-i,t}\right\}$, $\mathcal{G}_{n,(t-1)n+i} = \sigma\left\{\left(x_{jt}^o, z_j, u_{j,t-1}^o, \mu_j\right)_{j=1}^n, (u_{jt})_{j=1}^{i-1}\right\}$ and $\mathcal{F}_{n,(t-1)n+i} = \mathcal{G}_{n,(t-1)n+i} \vee \mathcal{C}$, and observe that c_{it} is a function of $\left(x_{jt}^o, z_j\right)_{j=1}^n$ and thus measurable w.r.t. $\mathcal{G}_{n,(t-1)n+i}$. In light of Assumption 4(b) it then follows immediately that Assumption 1(b) holds, and correspondingly Assumptions 5 follows from Assumptions 2. Assumption 1(c) also follows immediately from Assumption 4(c).

To verify Assumption 1(a) let $\delta > 0$, $\rho > 1$, $\kappa > 1$, and K be as in Assumption 4. Furthermore, let τ be such that $1/\rho + 1/\kappa = 1/(1+\tau)$ and note that in light of the maintained assumptions $\tau > 0$. Next observe that by Assumption 4(a)

$$E\left[\|\psi_{it}\|^{2+\delta} |\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = \|c_{it}\|^{2+\delta} E\left[|u_{it}|^{2+\delta} |\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] \leq B_{it} \tag{B.1}$$

with $B_{it} = \phi_{it} \|c_{it}\|^{2+\delta}$. Applying the generalized Hölder's inequality and recalling that $\|\phi_{it}\|_\rho \leq K$ by Assumption 4(a) yields

$$\|B_{it}\|_{1+\tau} = \left\|\phi_{it} \|c_{it}\|^{2+\delta}\right\|_{1+\tau} \leq \|\phi_{it}\|_\rho \left\|\|c_{it}\|^{2+\delta}\right\|_\kappa \leq K \left\|\|c_{it}\|^{2+\delta}\right\|_\kappa. \tag{B.2}$$

Next observe that

$$\begin{aligned}
\left\|\|c_{it}\|^{2+\delta}\right\|_\kappa &= \left\{E\left[\left|\sum_{s=1}^{T^+} \|\pi_{st}h_{is}'\|^2\right|^{(1+\delta/2)\kappa}\right]\right\}^{1/\kappa} \leq \left\{(T^+)^{(1+\delta/2)\kappa} K_\pi^{(2+\delta)\kappa} \sum_{s=1}^{T^+} E\left[h_{is}h_{is}'\right]^{(1+\delta/2)\kappa}\right\}^{1/\kappa} \\
&\leq (T^+)^{1+\delta/2} K_\pi^{(2+\delta)} (T^+)^{1/\kappa} \sum_{s=1}^{T^+} \left\{E\left[h_{is}h_{is}'\right]^{(1+\delta/2)\kappa}\right\}^{1/\kappa},
\end{aligned} \tag{B.3}$$

using inequality (1.4.4) in Bierens (1994), and where K_π is a bound for the absolute elements of Π. Now let $b = (b_r)$ be some $R \times 1$ random vector with $\|b_r\|_{(2+\delta)\kappa} \leq K_b$, then

$$\begin{aligned}
E\left[\left(b'b\right)^{(1+\delta/2)\kappa}\right] &= R^{(1+\delta/2)\kappa} \sum_{r=1}^R E\left[|b_r|^{(2+\delta)\kappa}\right] \leq R^{(1+\delta/2)\kappa} \sum_{r=1}^R \left[\|b_r\|_{(2+\delta)\kappa}\right]^{(2+\delta)\kappa} \\
&\leq R^{(1+\delta/2)\kappa+1} K_b^{(2+\delta)\kappa}
\end{aligned} \tag{B.4}$$

using again inequality (1.4.4) in Bierens (1994). By Assumption 4(a) the $L_{(2+\delta)\kappa}$ norms of the element of h_{is} are uniformly bounded by some finite constant K. Observing further that the dimensions of h_{is} are bounded by $T(k_x + k_z)$ it follows from applying inequality (B.4) that $E[h_{is}h_{is}']^{(1+\delta/2)\kappa} \leq [T(k_x + k_z)]^{(1+\delta/2)\kappa+1} K^{(2+\delta)\kappa}$, and thus in light of (B.3):

$$\left\| \|c_{it}\|^{2+\delta} \right\|_\kappa \leq (T^+)^{2+\delta/2+1/\kappa} K_\pi^{(2+\delta)} [T(k_x + k_z)]^{(1+\delta/2)+1/\kappa} K^{(2+\delta)}. \tag{B.5}$$

Together with (B.2) this establishes that $\sup_{i,t} \|B_{it}\|_{1+\tau} < \infty$. Since this in turn implies the desired uniform integrability of B_{it} it follows that also Assumption 1(a) holds.

Having verified all conditions of Assumption 1 and observing that Assumptions 5 implies Assumption 2 it follows from Theorem 2(a) that

$$\psi_{(n)} \xrightarrow{d} V^{1/2}\xi \quad (\mathcal{C}\text{-stably}), \tag{B.6}$$

where $\xi \sim N(0, I_p)$, and ξ and $\mathcal{C}$ (and thus ξ and V) are independent. Recall that $M = [M_1, \ldots, M_T]$ with $M_t = \left[I_{p_t}, 0_{p_t \times (p_{T^+} - p_t)}\right]'$. Then observing that that MM' is diagonal, $M_t = M_{T^+} = I_{p_{T^+}}$ for $t \geq T^+$ it is readily seen that $\lambda_{\min}(MM') \geq 1$. Thus M has full row rank. Recalling that V is positive definite a.s. it follows that $\Phi = MVM'$ is also positive definite a.s. and thus by Proposition (A.2)(b)

$$m_{(n)} = M\psi_{(n)} \xrightarrow{d} MV^{1/2}\xi \quad (\mathcal{C}\text{-stably}). \tag{B.7}$$

From the definition of the GMM estimator and the model given in (27) we have

$$n^{1/2}(\tilde{\theta}_n - \theta_0) = \left(G_n'\tilde{\Xi}_n G_n\right)^{-1} G_n'\tilde{\Xi}_n m_{(n)}.$$

Observing further that by assumption Ξ, G, and V are $\mathcal{C}$-measurable it follows from (B.7) and Proposition A.2(b) that jointly

$$\left(\Xi, G, m_{(n)}, \vartheta\right) \xrightarrow{d} \left(\Xi, G, MV^{1/2}\xi, \vartheta\right).$$

for all real valued $\mathcal{C}$-measurable random variables ϑ. Since by assumptions $\tilde{\Xi}_n \xrightarrow{p} \Xi$ and $G_n \xrightarrow{p} G$ it follows furthermore that

$$\left(\tilde{\Xi}_n - \Xi, G_n - G, \Xi, G, m_{(n)}, \vartheta\right) \xrightarrow{d} \left(0, 0, \Xi, G, MV^{1/2}\xi, \vartheta\right). \tag{B.8}$$

Observing that $G'\Xi G$ is positive definite *a.s.* and employing the continuous mapping theorem yields

$$\left(n^{1/2}(\tilde{\theta}_n - \theta_0), \vartheta\right) \xrightarrow{d} \left((G'\Xi G)^{-1}G'\Xi MV^{1/2}\xi, \vartheta\right), \tag{B.9}$$

and thus by Proposition A.1(iii) $n^{1/2}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} (G'\Xi G)^{-1}G'\Xi MV^{1/2}\xi$ ($\mathcal{C}$-stably), which in turn implies that

$$n^{1/2}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} \Psi^{1/2}\xi_* \ (\mathcal{C}\text{-stably}) \tag{B.10}$$

with $\Psi = (G'\Xi G)^{-1}G'\Xi\Phi\Xi G(G'\Xi G)^{-1}$ and $\xi_* \sim N(0, I_d)$, and where ξ_* and $\mathcal{C}$ (and thus ξ_* and Ψ) are independent. The latter claim is easily verified by arguments analogous to those in the proof of Proposition A.2(b).

The claim that

$$n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}^{2}H_i'\pi_t\pi_t'H_i \xrightarrow{p} \Phi \tag{B.11}$$

follows from Assumption 4(c), observing that $\Phi = \sum_{t=1}^{T} M_t V_t M_t'$ and

$$n^{-1}\sum_{i=1}^{n}u_{it}^{2}H_i'\pi_t\pi_t'H_i = n^{-1}\sum_{i=1}^{n}\begin{bmatrix} u_{it}^2 c_{it}c_{it}' & 0'_{p_{T+}-p_t\times p_t} \\ 0_{p_{T+}-p_t\times p_t} & 0_{p_{T+}-p_t\times p_{T+}-p_t}\end{bmatrix} = M_t\tilde{V}_{t,n}M_t'.$$

Next assume that $E\left[u_{it}^2|\mathcal{F}_{n,(t-1)n+i}\right] = \sigma^2$. To see that in this case

$$\sigma^2 n^{-1}\sum_{i=1}^{n}H_i'\Pi\Pi'H_i = \sigma^2\sum_{t=1}^{T}n^{-1}\sum_{i=1}^{n}H_i'\pi_t\pi_t'H_i \xrightarrow{p} \Phi$$

observe that

$$\sigma^2 n^{-1}\sum_{i=1}^{n}H_i'\pi_t\pi_t'H_i = \sigma^2 n^{-1}\sum_{i=1}^{n}\begin{bmatrix} c_{it}c_{it}' & 0'_{p_{T+}-p_t\times p_t} \\ 0_{p_{T+}-p_t\times p_t} & 0_{p_{T+}-p_t\times p_{T+}-p_t}\end{bmatrix} \xrightarrow{p} M_tV_tM_t'.$$

since $\sigma^2 n^{-1}\sum_{i=1}^{n}c_{it}c_{it}' \xrightarrow{p} V_t$ in light of Assumption 4(c).

The proof of **Part (b)** follows from arguments analogous to those between (B.7) and (B.10), observing that B is $\mathcal{C}$-measurable.

To prove **Part (c)** observe that

$$\Phi_{(n)} = n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{it}u_{is}H_i'\pi_t\pi_s'H_i = n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}^2H_i'\pi_t\pi_t'H_i + o_p(1) = \Phi + o_p(1)$$

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 and utilizing (B.11). Since $\Phi = MVM'$ is positive definite a.s. as discussed in the proof of Part(a) it follows that $\Phi_{(n)}^{-1/2} \xrightarrow{p} \Phi^{-1/2}$. The claim follows in light of (B.7), and from arguments analogous to those between (B.7) and (B.10), observing that Φ is $\mathcal{C}$-measurable. ■

Proof of Theorem 4. Observe that

$$\begin{aligned} n^{1/2}(\tilde{\theta}_n - \theta_0 - \left(G_n'\tilde{\Xi}_nG_n\right)^{-1}G_n'\tilde{\Xi}_nb_n) &= \left(G_n'\tilde{\Xi}_nG_n\right)^{-1}G_n'\tilde{\Xi}_n\left(m_{(n)} - \sqrt{n}b_n\right) \\ &= \left(G_n'\tilde{\Xi}_nG_n\right)^{-1}G_n'\tilde{\Xi}_nM\check{\psi}_{(n)} \end{aligned}$$

where $\check{\psi}_{(n)}$ is defined in Assumption 6. For Part (a) we observe that by Theorem 2(b) $\check{\psi}_{(n)} \xrightarrow{d} \check{V}^{1/2}\xi$ ($\mathcal{C}$-stably) where ξ is independent of $\mathcal{C}$ and $\xi \sim N(0, I_p)$. Thus in light of and Proposition A.2(b)

$$\left(\tilde{\Xi}_n - \Xi, G_n - G, M, \Xi, G, \check{\psi}_{(n)}, \vartheta\right) \xrightarrow{d} \left(0, 0, M, \Xi, G, \check{V}^{1/2}\xi, \vartheta\right).$$

for all real valued $\mathcal{C}$-measurable random variables ϑ. The result then follows immediately from the continuous mapping theorem, and an application of Proposition A.1(iii). For Part (b) we observe that in light of Theorem 2(b) and Proposition A.2(b)

$$\left(\tilde{\Xi}_n - \Xi, G_n - G, \sqrt{n} b_n - b, \Xi, G, \check{\psi}_{(n)}, b, \vartheta\right) \xrightarrow{d} \left(0, 0, 0, \Xi, G, \check{V}^{1/2}\xi, b, \vartheta\right)$$

and the result again follows from the continuous mapping theorem and Proposition A.1(iii). Part (c) follows as a special case of Part (b) with $b = 0$. ■

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

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

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

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n} h_{it}' h_{is} u_{is}^{+} w_{it}^{+}(\tilde{\theta}_n - \theta_0) \\
= \;& n^{-1}\sum_{i=1}^{n} h_{it}' h_{is}\sum_{j=s}^{T}\pi_{sj}u_{ij}\sum_{l=t}^{T}\pi_{tl}w_{il}(\tilde{\theta}_n - \theta_0)
\end{aligned} \tag{B.12}$$

where

$$\left\| n^{-1}\sum_{i=1}^{n} h_{it}' h_{is}\sum_{j=s}^{T}\pi_{sj}u_{ij}\sum_{l=t}^{T}\pi_{tl}w_{il}(\tilde{\theta}_n - \theta_0)\right\| \le \left\|\tilde{\theta}_n - \theta_0\right\| n^{-1}\textstyle\sum_{i=1}^{n}\sum_{j=s}^{T}\sum_{l=t}^{T}\|\pi_{sj}\|\,\|\pi_{tl}\|\,\left\|h_{it}'h_{is}\right\|\,|u_{ij}|\,\|w_{il}\|$$

and

$$\begin{aligned}
E\left[\left\|h_{it}'h_{is}\right\|\,|u_{ij}|\,\|w_{il}\|\right] &\le E\left[\left\|h_{it}'h_{is}\right\|^2\right]^{1/2} E\left[|u_{ij}|^2\,\|w_{il}\|^2\right]^{1/2} \\
&\le E\left[\|h_{it}\|^4\right]^{1/4} E\left[\|h_{is}\|^4\right]^{1/4} E\left[|u_{ij}|^4\right]^{1/4} E\left[\|w_{il}\|^4\right]^{1/4}
\end{aligned}$$

by repeated application of the Cauchy-Schwarz inequality. By the boundedness of fourth moments all expectations are bounded and thus

$$n^{-1}\textstyle\sum_{i=1}^{n}\sum_{j=s}^{T}\sum_{l=t}^{T}\|\pi_{sj}\|\,\|\pi_{tl}\|\,\left\|h_{it}'h_{is}\right\|\,|u_{ij}|\,\|w_{il}\| = O_p(1).$$

Since by assumption $\left\|\tilde{\theta}_n - \theta_0\right\| = o_p(1)$ it follows that

$$n^{-1}\sum_{i=1}^{n} h_{it}' h_{is}\sum_{j=s}^{T}\pi_{sj}u_{ij}\sum_{l=t}^{T}\pi_{tl}w_{il}(\tilde{\theta}_n - \theta_0) = o_p(1).$$

The other terms appearing in B.12 can be treated in the same way. Therefore $\widetilde{\Phi}_{(n)} \xrightarrow{p} \Phi$ as claimed, and furthermore $\hat{\Psi}_n = \left(G_n' \widetilde{\Phi}_{(n)}^{-1} G_n\right)^{-1} \xrightarrow{p} \Psi = (G'\Phi^{-1}G)^{-1}$.

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

$$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \Psi^{1/2}\xi_*, \tag{B.13}$$

where ξ_* is independent of $\mathcal{C}$ (and hence of Ψ), $\xi_* \sim N(0, I_d)$. In light of (B.13), the consistency of $\hat{\Psi}_n$, and given that R has full row rank p^* it follows furthermore that under H_0

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

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

$$\left(R\hat{\Psi}R'\right)^{-1/2} n^{1/2}(R\hat{\theta}_n - r) \xrightarrow{d} \xi^* \tag{B.14}$$

where $\xi^* \sim N\left(0, I_{p^*}\right)$. Hence, in light of the continuous mapping theorem, T_n converges in distribution to a chi-square random variable with p^* degrees of freedom. The claim that $\hat{\Psi}_n^{-1/2}\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \xi^*$ is seen to hold as a special case of (B.14) with $R = I$ and $r = \theta_0$. ■

C Appendix C: Proofs for Section 4

Proof of Theorem 6. The proof follows the classical approach, see, e.g., Newey and McFadden (1994). Applying a first order Taylor approximation of $\partial L_n(\theta)/\partial\theta$ around θ_0 and employing the mean value theorem yields

$$0 = \sum_{t=1}^{T}\sum_{i=1}^{n} \psi_{it} + \sum_{i=1}^{n}\sum_{t=1}^{T} \frac{\partial\psi_{it}\left(\tilde{\theta}_n\right)}{\partial\theta'}\left(\hat{\theta}_n - \theta_0\right)$$

for some $\tilde{\theta}_n$ such that $\left\|\tilde{\theta}_n - \theta_0\right\| \leq \left\|\hat{\theta}_n - \theta_0\right\|$. (Strictly speaking the mean value theorem should be thought of as being applied component wise.) Next observe that for any $\eta > 0$,

$$\begin{aligned}& P\left(\left\|\frac{1}{n}\textstyle\sum_{i=1}^{n} \partial\psi_{it}\left(\tilde{\theta}_n\right)/\partial\theta' - B_t(\theta_0)\right\| > \eta\right) \qquad \text{(C.15)} \\ \leq\ & P\left(\left\|\frac{1}{n}\textstyle\sum_{i=1}^{n} \partial\psi_{it}\left(\tilde{\theta}_n\right)/\partial\theta' - B_t(\theta_0)\right\| > \eta, \left\|\tilde{\theta}_n - \theta_0\right\| \leq \delta\right) + P\left(\left\|\tilde{\theta}_n - \theta_0\right\| > \delta\right) \\ \leq\ & P\left(\sup_{\|\theta-\theta_0\|\leq\delta}\left\|\frac{1}{n}\textstyle\sum_{i=1}^{n} \partial\psi_{it}(\theta)/\partial\theta' - B_t(\theta)\right\| > \eta/2\right) \\ & + P\left(\left\|B_t(\tilde{\theta}_n) - B_t(\theta_0)\right\| > \eta/2\right) + P\left(\left\|\tilde{\theta}_n - \theta_0\right\| > \delta\right) \\ \to\ & 0\end{aligned}$$

as $n \to \infty$ in light of Assumption 8. Recalling that $B_t = B_t(\theta_0)$ is nonsingular it follows that

$$\left[n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\partial\psi_{it}\left(\tilde{\theta}_n\right)/\partial\theta'\right]^{-1} - B^{-1} = o_p(1)$$

where $B = \sum_{t=1}^{T} B_t$, and thus

$$n^{1/2}\left(\hat{\theta}_n - \theta_0\right) = \left[B^{-1} + o_p\left(1\right)\right]\left(n^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{n}\psi_{it}\right). \tag{C.16}$$

Assumption 7 maintains that Assumptions 1 and 2 are satisfied. It then follows from Theorem 2(a) that

$$\psi_{(n)} \xrightarrow{d} \operatorname{diag}\left(V_1^{1/2}, ..., V_T^{1/2}\right)\xi \;\; (\mathcal{C}\text{-stably})$$

where $\xi \sim N\left(0, I_p\right)$ with $p = dT$, and ξ and $\mathcal{C}$ (and thus ξ and V) are independent. Observing further that $n^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{n}\psi_{it} = A\psi_{(n)}$ with $A = [I_d, ..., I_d]$ it follows from Theorem 2(b) that

$$n^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{n}\psi_{it} \xrightarrow{d} \Omega^{1/2}\xi_* \;\; (\mathcal{C}\text{-stably}) \tag{C.17}$$

where $\Omega = \sum_{t=1}^{T} V_t$, $\xi_* \sim N\left(0, I_d\right)$, and ξ_* and $\mathcal{C}$ (and thus ξ_* and V) are independent. In light of (C.16) and (C.17), the continuous mapping theorem and Proposition A.2 it follows further that

$$\begin{aligned} n^{1/2}\left(\hat{\theta}_n - \theta_0\right) &\xrightarrow{d} \Psi^{1/2}\xi_*, \\ \Psi^{-1/2}n^{1/2}\left(\hat{\theta}_n - \theta_0\right) &\xrightarrow{d} \xi_* \end{aligned} \tag{C.18}$$

with $\Psi = B^{-1}\Omega B'^{-1}$.

Clearly (C.15) also holds with $\tilde{\theta}_n$ replaced by $\hat{\theta}_n$, which establishes that

$$\hat{B}_n = n^{-1}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}\partial\psi_{it}\left(\hat{\theta}_n\right)/\partial\theta' \xrightarrow{p} B.$$

By analogous arguments we also have

$$\hat{\Omega}_n = n^{-1}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}\psi_{it}\left(\hat{\theta}_n\right)\psi_{it}'\left(\hat{\theta}_n\right) \xrightarrow{p} \Omega.$$

This in turn establishes that $\hat{\Psi}_n = \hat{B}_n^{-1}\hat{\Omega}_n\hat{B}_n'^{-1} \xrightarrow{p} \Psi$. ∎

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] Aldous, D.J. and G.K. Eagleson (1978), On Mixing and Stability of Limit Theorems, *The Annals of Probability*, 6, 325-331.

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

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

[7] 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.

[8] Bai, J., 2009, Panel Data Model with Interactive Fixed Effects, *Econometrica 77, 1229-1279.*

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

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

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

[12] 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.

[13] 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.

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

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

[16] Bolthausen, E., 1982, On the Central Limit Theorem for Stationary Mixing Random Fields, *The Annals of Probability*, 10, 1047-1050.

[17] Billingsley, P., 1995, Probability and Measure, Third Edition, John Wiley & Sons, New York.

[18] Brock, W.A., and S. N. Durlauf, 2001, Discrete Choice with Social Interactions, *Review of Economic Studies*, 68, 235-260.

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

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

[21] Cox, D.R., 1975, Partial Likelihood, *Biometrika* 62, 269-276.

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

[23] Dedecker, J. and F. Merlevede, 2002, Necessary and Sufficient Conditions for the Conditional Central Limit Theorem, *The Annals of Probability*, 30, 1044-1081.

[24] Eagleson, G. K., 1975, Martingale Convergence to Mixtures of Infinitely Divisible Laws, *The Annals of Probability* 3, 557-562.

[25] Engle R.F., D.F. Hendry and J-F. Richard, 1983, Exogeneity, *Econometrica* 51*,* 277-304.

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

[27] 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.

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

[29] Jenish, N. and I. R. Prucha, 2009, Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields, *Journal of Econometrics* 150, 86-89.

[30] Jenish, N. and I. R. Prucha, 2012, On Spatial Processes and Asymptotic Inference under Near-Epoch Dependence, *Journal of Econometrics*, forthcoming.

[31] Kallenberg, O., 1997, *Foundations of Modern Probability*, Springer Verlag, New York.

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

[33] 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.

[34] Kelejian, H.H., and I.R. Prucha, 1985, Independent or Uncorrelated Disturbances in Linear Regressions, An Illustration of the Diffierence, *Economic Letters* 19, 35-38.

[35] 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.

[36] Kingman, J.F.C., 1978, Uses of Exchangeability, *The Annals of Probability* 6, 183-197.

[37] 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).

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

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

[40] McFadden, D. (1974), Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka (ed.), Frontiers in Econometrics (New York: Academic Press).

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

[42] Newey, W.K. and D. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, Chapter 36 ,Handbook of Econometrics, Volume IV, Edited by R.F. Engle and D.L. McFadden 0 1994 Elsevier Science B.V.

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

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

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

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

[47] 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.

[48] Réyni, A., 1963, On Stable Sequences of Events, *Sankhyā: The Indian Journal of Statistics, Series A*, 25, p. 293- 302.

[49] Rosenblatt, M., 1952, Remarks on a Multivariate Transformation, *The Annals of Mathematical Statistics*, 23, 470-472.

[50] 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.

[51] Schervish, M.J. , 1995, *Theory of Statistics,* Springer Verlag, New York.

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

[53] 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.

[54] Yu, J., R. de Jong and L.-F. Lee, 2012, Estimation for spatial dynamic panel data with fixed effects: The case of spatial cointegration, *Journal of Econometrics* 167, 16-37.

[55] Zellner, A. 1976, Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms, *Journal of the American Statistical Association,* 71, 400-405.